The Impact of Agglomeration Externalities on Product Innovation Output in Chinese Industrial Firms

Cindy Feng

Professor Charles Maxwell. Becker, Faculty Advisor

Professor Kent P. Kimbrough, Honors Seminar Instructor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2020

Acknowledgement

I would like to sincerely thank my thesis advisors, Professor Charles Becker and Professor Kent Kimbrough, for their extremely helpful suggestions and encouragement that accompanied me throughout the process of writing my thesis. My paper wouldn’t have achieved the current fulfillment without their instructive mentorship. My heartfelt gratitude goes to Professor Michelle Connolly, who offered me help for data and logistics problems I encountered during this process. Finally, I would like to thank Duke Economics Department for accommodating my need during the COVID-19 pandemic.

Abstract

Agglomeration externalities is defined as the economic benefits from concentrating firms, housing, and output. This study investigates the impact of agglomeration externalities of industrial firms on product innovation output in China. In the research, I specified the impact of agglomeration into three types: Marshallian or localization externalities, defined as the impact of collocating with same-industry firms; Urbanization economies, defined as the impact of collocating with different-industry firms, and Porter externalities, the impact of competing with same-industry firms as a result of localization. My result suggests endogenous spatial selection of firms account for most of the agglomeration impacts we observe. Despite so, urbanization economies is still impactful in boosting a firm's innovation performance, and should be taken into account as the government implements policies that boost firm performance.

JEL Codes: R3, R5, D24

Keywords: Agglomeration Theory, Urban Economics, Production, Emerging Markets

I. Introduction

Agglomeration externalities, defined as the benefits to entities from co-location of economic activities, has long been studied by a vast range of economic literature. It is motivated by a central question: why do cities exist in the first place? And how do cities contribute to the wealth of the area? The answer to this question is approached through several distinct aspects such as agglomeration of households, agglomeration of consumers and immigrants, or agglomeration of firms. Looking from the business aspect of this question, scholars are interested in whether business clusters help boost firm performance, increase firm survival rates, and contribute to the economic prosperity of the given area. Studies on these topics are often useful in offering policy implications for developing countries like China and India. Among developing countries, China gained its attention not only through its rapid economic development since the last century, but also through the inseparable link between the nation's economic growth, manufacturing edge, the presence of a large population and well-developed cities, and the emergence of agglomeration economies as a result of a series of widespread industrialization movements.

China has exhibited impressive industrialization and economic growth since the economic reform in 1978. During this process led by the paramount leader Deng Xiaoping, Chinese government attempted to mimic the historical sequence of Industrial Revolution in Europe through gradual and experimental approaches including subsidizing infrastructure buildup or exchanging manufactured goods instead of natural resources for machinery (Andors, 1979). The result is obvious: China has grown to become one of the world's largest manufacturing powerhouse, producing nearly 50 percent of the world's industrial goods including steel, cement, and vehicles (Wen, 2016). Along with this stunning progress is the increasing formation of industrial agglomeration, or industrial clusters, defined as the geographic concentration of industrial firms and activities in relatively small areas. Wen (2004) compared the concentration level of manufacturing firms between 1980, 1985, and 1995, discovering that industries have become more spatially clustered after the reform. In fact, as one of the tactics, central and local authorities in China established more than 100 economic zones in over 60 cities after 1995. Most of these zones are science parks and technological development areas that were built to boost firm performance (and therefore production and growth) through convenience of outsourcing, availability of resources, and access to industry knowledge (Zhang, 2014). Meanwhile, following the legalization of private sectors and private entrepreneurship in 1988, firms were granted the freedom to engage in location selection and optimization, in turn increasing the level of agglomeration. For example, Ge (2009) suggested that foreign trade and foreign investment are highly correlated with the significant increase in agglomeration from 1985 to 2005. Industries in contact with foreign businesses tend to cluster in areas that allowed them to easily access foreign markets. In short, during the economic reform, individual optimization behaviors, coupled with government policies, appeared to have substantially contributed to the existence of numerous agglomeration economies in China.

The question naturally lies in how government effort of establishing and encouraging industrial agglomeration has actually promoted firm performance and economic growth. Firm performance can be quantified through one commonly used metric – firm productivity. While extensive literature investigated the effect of spatial clusters on productivity among Chinese firms (Lin et al., 2011; Li & Gibson, 2014), only a few empirical studies investigated its impact on product innovation. Nevertheless, innovation is frequently considered as a key ingredient for firm success and economic growth, both in neoclassical theories (Solow, 1957; Fagerberg, 1994) and in empirical studies (Wong et al., 2005).

That said, this study contributes to existing research by investigating whether industrial agglomeration in China incentivizes and improves innovation behaviors among individual firms. The major data used for this study is a set of firm-level panel data for manufacturing industries – Annual Surveys of Industrial Firms (ASIF) from 1998 to 2007. The data is special in that it includes a direct measure of new product output instead of intermediate measures such as number of patents. Here new product output is defined, according to the National Bureau of Statistics of China, as either the output of completely new products introduced by new technology or new design, or the output of existing products of which their functions are extended or improved ("Description of ASIF", n.d.). In practice, this measure includes both government-certified new products, or uncertified new products invested by firms through their research and development processes. We should still be cautious when using this variable, since allowing for firms to report their own new products could potentially inflate the number if a firm reports a product with a new concept that however has little technological or structural improvement associated with it. Consider Diet Coke as an example. This could be counted as a new product since it is conceptually a breakthrough from the original product, and it does have a new "function" comparing to the original Coke, but we would find it less convincing to say that Diet Coke counts as a part of the firm's product innovation. Despite such pitfall, with this definition of new product output, we have gained exposure to a more workable and concrete measure than less direct measures such as patents.

The rest of the paper is structured as follows: Section II introduces the concept and theories of agglomeration economies, focusing on two types of agglomeration: localization and urbanization. Section III reviews key literature on agglomeration and firm performance in China. Section IV describes the theoretical framework and the measures for agglomeration. Section V illustrates the data and variables used for the study, and Section VI presents the empirical methodology. Finally, Section VII concludes the study by discussing policy implications and future concerns.

II. Background: What is Agglomeration?

This section introduces the terms and theories of agglomeration economies relevant to the rest of my research. Regional and urban economic theories of agglomeration concern the external economies a firm might benefit from by being located within the vicinity of other firms. The central idea that presumes the existence of such benefit is that a firm can absorb spillovers from nearby firms, whether the spillover is knowledge, access to suppliers and resources, or any other type of benefits that facilitate efficiency of firm activities. A primary distinction of agglomeration economies is made between localization and urbanization economies. Localization economies is defined as the benefits a firm obtains by co-locating near other firms within the same industry. Urbanization economies, on the other hand, is defined as the benefits a firm derives from co-locating near firms of different industries. Therefore, the distinction lies in whether the transmission of benefits occurs within or between industries.

Localization economies, as specified above, refers to the benefits related to the concentration of same-industry firms within a certain area. It is primarily measured by the properties of a Marshallian vicinity[1], including population for labor market pooling, the availability of raw materials intermediate goods suppliers and final goods consumers, and the size of knowledge or technology spillovers (Marshall, 1890; Henderson, 2003). All such features, and hence the strength of localization economies, are expected to increase as the industry grows larger within the unit of area. In particular, the benefit from knowledge spillovers in a localization setting is referred to as Marshallian-Arrow-Romer (MAR) externalities (named after Marshall (1890), Arrow (1962), and Romer (1990)), the theory of which hypothesizes that information spillovers among same-industry firms promote firm-level successes as well as economic growth for the given region of study. An alternative measure for localization economies is the number of competing firms within the same industry (Porter, 1990). The underlying hypothesis is that local competition motivates growth by forcing innovation from firms that confront the threat of closing. Porter (1990) provided an example of the ceramics and gold jewelry industry in Italy as an industry where firms intensively compete for innovation and novel ideas. In reality, competition might not be only local, since firms such as Samsung and Apple compete in the cellphone market even though they are not in the same location.

Urbanization economies, defined as the benefit resulting from co-location of different-industry firms, is represented by the total population or total employment in a particular vicinity, city, region or any unit of area alike. The firm derives benefits mostly through savings from previous cost-inducing behaviors (e.g. transportation) that are now a part of the large-scale operations in the area. A more refined definition is proposed by Jacobs

[1] Marshallian vicinity: the model of localization economies proposed by Alfred Marshall in *Principles of Economics* (1890)

(1969) that emphasized the diversity of urban industry mix as an important component of urbanization externalities. A diverse industry mix in a particular area promotes exchange, imitation, and modification of products, activities and ideas, and so diversity itself is considered as a crucial source for knowledge spillovers.

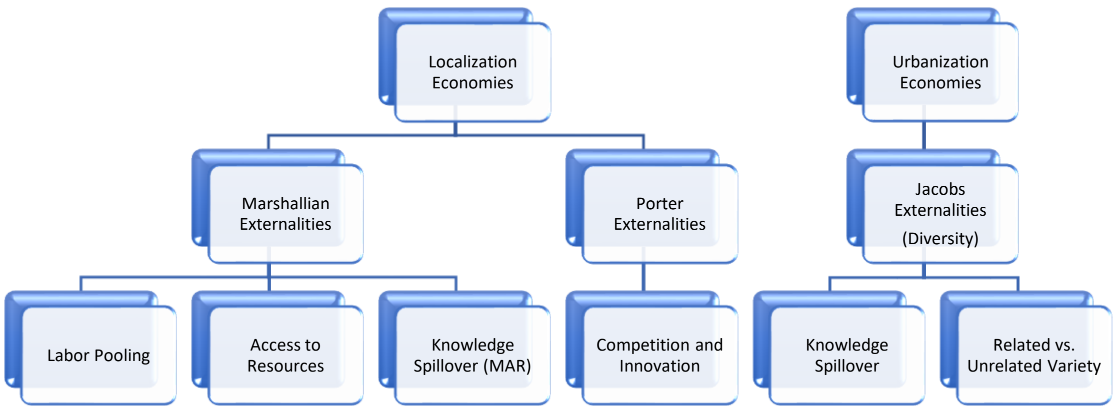

Figure [1] Hierarchical Structure of Agglomeration Theories

A further decomposition of Jacobs externalities involves the concept of related variety and unrelated variety (Boschma & Wenting 2007). Related variety takes into account the cognitive or technological relatedness between industries. The hypothesis is that a firm co-located near firms of different but similar industries absorbs inter-industry knowledge spillovers more quickly through better communication, but unrelated variety might also benefit a firm in a way similar to portfolio diversification strategy: the region can be more resistance to external shocks or unforeseen circumstances when it encompasses a great number of unrelated industries (Frenken et al., 2007), in which case the firm might stand a better chance to survive and seek for progress. Therefore, the net effect is ambiguous.

In figure 1, I plotted the hierarchical structure of agglomeration theories. One important aspect of my research is to find appropriate measures of agglomeration economies. The following section will discuss the key literature that constructed these valuable measures. Specifically, construction on six measures will be discussed. They are: localization economies, Porter externalities, urbanization economies (area size), Jacobs externalities (industrial diversity), related variety, and unrelated variety.

III. Literature Review

A plethora of past studies have proposed distinct ways to measure agglomeration. Martin et al. (2011) created useful measures of localization and urbanization economies. For a given firm, localization economies were approximated by the number of employees in other firms who are also working in the same industry and the same area. Since Marshallian externalities and the strength of localization economies are expected to increase with industry size, this measure comprehensively captures intra-industry spill-over effects. Indeed, this proxy can be improved by decomposing it into several proposed Marshallian externalities (such as labor pooling effect or access to suppliers and consumers), but constructing such decomposed measures normally requires input and output information. For example, Zheng & Zhao (2017) computed the difference between an industry's ideal input requirement and the city's actual employment composition across all other industries to approximate the extent to which the city provides suitable suppliers for the firm. While this measure is clearly more informative, the data for my study – Annual Surveys of Industrial Firms do not include any variables related to input values. Therefore, I intend to refer to the more doable strategies in Martin et al. (2011) to construct the index of localization economies.

In the same article, urbanization economies were similarly measured by the total number of employees of other industries in the given area. This is a very handy method that incorporates the effect of both city size and spill-over benefits gained from other industries. Therefore, the measures in Martin et al. (2011) are highly applicable to my research data and will be incorporated into my study.

While productivities have long been a target of research in China, but productivities related to innovation have been less explored than expected. Zhang (2015) is one exception in the field. In his article, he investigated agglomeration effect and product innovation using Annual Surveys of Industrial Firms from 1998 to 2007. Zhang (2015) grouped firms into "New Product Firms" (NPF), or firms that have strictly positive new product outputs in a given year, and "Non-NPFs", namely those who didn't. Although the author didn't explain the reason for doing so, the article successfully captured a firm's inclination or potential to engage in innovative behaviors. In other words, the author assumed that a firm's behavior in creating new products signalizes its intention to innovate and its potential to carry out innovation. In Zhang's model, he mentioned that spatial selection of firms could cause endogeneity issues, as firms that are new or less successful might choose to cluster in a business area, or it might be the reverse case. While Zhang didn't necessarily propose a solution to this reverse effect, I intend to minimize this concern by subsampling my data on a group that has little endogeneity naturally. This will be discussed further in Section VI.

In short, my contribution to the field of research will be twofold: I will investigate the effect of agglomeration on production innovation in China, which is a relatively new topic for Chinese studies on regional economics. My model also incorporates a more concrete measure of product innovation. As introduced previously, instead of number of patents, I will use “new product output”, defined as a completely new product or an extended/improved product developed through a firm’s research and investment. Finally, I will tackle the endogeneity concern on spatial selection of firms, which is an issue not addressed by past studies on the same topic, with a hope that doing so will motivate the development of better methodologies for studying this comparatively new theme.

IV. Theoretical Framework

1. Modelling Production Innovation Procedure

Consider a stochastic frontier production function:

$$Q_{it} = f(A_{it}, K_{it}, L_{it}, M_{it})\xi_{it} exp(V_{it}) \ (1)$$

where I_{it} represents the production output, K_{it}, L_{it} , and M_{it} represent the capital, labor, and material (raw or intermediate materials used) needed to produce the innovative products, and A_{it} represents the production capacity for firm *i* at time *t*. $\xi_{it} \in (0,1]$ models the level of production efficiency for the given firm. If $\xi_{it} = 1$, then the firm is at its optimal output given its technology embodied by the production function. When $\xi_{it} < 1$, the firm is not making the most of its input given its technology. Finally, $exp(V_{it})$ represents some random shock to the production function. If we assume that the production function is Cobb-Douglas, we would then obtain:

$$Q_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} M_{it}^{\gamma} \xi_{it} exp(V_{it}) \ (2)$$

By agglomeration theories mentioned above, A_{it} can be modeled as a function of urbanization, localization, and Porter externalities. This gives the following:

$$A_{it} = (LOC_{it}^{sz})^{\delta} (URB_{it}^{sz})^{\mu} (PORT_{it}^{sz})^{\eta} U_{it,production} \ (3)$$

Here $U_{it,production}$ denotes any other possible firm-specific heterogeneity that determines the production capacity for firm *i* from industry *s* and area *z* at time *t*.

Plugging (3) into (2), and log linearizing (2), we obtain the following:

$$q_{it} = u_{it,production} + \delta loc_{it}^{sz} + \mu urb_{it}^{sz} + \eta port_{it}^{sz} + \alpha k_{it} + \beta l_{it} + \gamma m_{it} + \ ln(\xi_{it}) + V_{it} \ (4)$$

where lower-case letters denote the log of upper-case variables in (3) and (2). Define v_{it}= V_{it} + $u_{it,production}$ + $u_{it,innovation}$, where the last term now accounts for firm heterogeneity in innovative production. Therefore, our final equation for innovation production is:

$$i_{it} = \delta' loc_{it}^{sz} + \mu' urb_{it}^{sz} + \eta' port_{it}^{sz} + \alpha' k'_{it} + \beta' l'_{it} + \gamma' m'_{it} + \ ln(\xi_{it}) + v_{it} \ (5)$$

Note that $k'_{it}, l'_{it}, and\ m'_{it}$ indicate the amount of inputs necessary to produce the given amount of innovative outputs. $\delta', \mu',\ \eta', \ldots, \gamma'$ are coefficients specific to the innovative production, different from those for the production function in (4). In reality, since the data doesn't allow us to differentiate between inputs for new product output and inputs for old product output, we can only let the total number of inputs enter our regression equation.

2. Modelling Agglomeration Economies

We now describe our measures of localization and urbanization economies:

$$loc_{it}{}^{sz} = \ln(employees_t{}^{sz} - \ employees_{it}{}^{sz} + 1) \ (5)$$

$$urb_{it}{}^{sz} = \ln(employees_t{}^{z} - \ employees_t{}^{sz} + 1) \quad (6)$$

Localization economies are measured by taking the difference between total employment in area *z* from industry *s* at time *t* and firm *i*'s total employment in area *z* from industry *s* at time *t*. Ultimately this gives the number of other employees working at the same industry and the same area. This number is strictly increasing in the strength of localization economies: the higher the index, the greater the number of other employees, and the stronger the localization economies. In the meantime, this difference will equal 0 when the given firm is a monopolist in its area and industry. That is, all employment of the industry at the area comes from this firm. To take into account the monopolist situation, we add the difference by 1 to avoid the unfavorable case when the index becomes ln(0). Therefore, localization index will be exactly equal to 0 when the firm is a monopolist and there is no localization economies.

Urbanization economies are measured by taking the difference between total employment in area z at time t and total employment in area z and industry s at time t. It measures the total number of employees working in other industries. The index is also strictly increasing in region size: the more urbanized the area, the higher the index. The proxy for urbanization will equal 0 when the area is perfectly specialized, that is, all employment comes from one industry in that region.

Porter externalities, or competition, is measured by taking the log of the reciprocal of the Herfindahl index, and therefore:

$$port_{it}^{sz} = ln\frac{1}{Herf_t^{sz}} \quad where \quad Herf_t^{sz} = \sum_{i \in S_t^{sz}} (\frac{employee_{it}^{sz}}{employee_t^{sz}})^2 \quad (7)$$

Here S_t^{sz} denotes the set of firms belonging to industry s and area z at time t. Herfindahl index at time t for industry s and area z is therefore calculated by the following procedure:

1. For each firm i in the industry s, area z, and time t, we obtain its industry share by dividing its employees over the total number of employees in industry s, city z, and at time t.

2. We square each firm's industry share.

3. We add the squared terms up.

Herfindahl index is originally proposed by economists Orris C. Herfindahl and Albert O. Hirschmanan. It is an effective measure of the concentration of market power, as the squared term exists in order to put larger weights on firms with larger industry share. If a company occupies a large share in the economy and therefore has greater market power, we would expect the index to be large. If each company represents a very small part of the economy, then we would expect the sum of small shares to be small as well. Linking back to the Porter externalities index, for firm *i* in industry *s*, city *z*, at time *t*, the index equals the log of the reciprocal of the local, industry-specific, and time-specific Herfindahl index, so if there is only one monopolistic firm in the given situation, the Herfindahl index will be 1 and the Porter index will be ln(1)=0, indicating that market concentration is high and competition is low; reversely, if Herfindahl index $\rightarrow 0$, then the internal term of the Porter index $\rightarrow \infty$, and therefore the monotonic transformation, i.e. log transformation, of the term will also approach infinity. In this case, market concentration is low and competition is high. This mathematical nature of the Porter index enables this measure to represent local competition in a clear-cut way.

V. Data and Variables

1. Basic Setup

The major data used for this study is the Annual Surveys of Industrial Firms (ASIF) 1998 – 2007 conducted by China's National Bureau of Statistics (NBS). The surveys include all state-owned or non-state-owned industrial firms with annual sales over 5 million RMB. According to a report on annual sales by Baidu[2], this cutoff likely excludes small enterprises but includes middle and large enterprises. The data set records the detailed address, street, city, province, and corresponding regional codes for each observation, and hence is extremely useful when constructing area-specific agglomeration indices. The data set also provides detailed firm-level statistics including but not limited to total output value, total

[2] See baijiahao.baidu.com/s?id=1657878955735267744&wfr=spider&for=pc.

employment, average wage, government subsidy, and industry codes. Note that NBS retracted the key variable of interest "new product output" after 2009, while government subsidy, a key variable that impacts firm performance and firm productivity, is also retracted after 2007, so surveys later than 2007 will not be used in this research.

A few problems were tackled before using the panel data. In 2003, the old industry code classification system (GB/T 4754-1994) was replaced by a new classification system (GB/T 4754-2002). In order to retrieve a consistent classification over time, I converted industry codes into a common standard through the measure provided by Brandt et al. (2011). The end result is a 2-digit industry code for each firm. Aside from that, entries with missing data in key variables such as new product output or total employment were dropped. I also dropped entries with values that are likely erroneous, such as entries with negative values in new product output, total product output, capital, material, government subsidy, etc. Finally, to construct the panel data set, I matched firms throughout all years by their firm name, legal person, region code, industry code, phone number, etc. in a structured order using the method proposed by Brandt et al. (2011). The matched panel data set contains a sample of around 1.8 million entries in total, and around 180,000 entries for each year. Among all entries, no firm has other branches or subsidiaries, so there is no repeated observation per year.

As mentioned previously, we will use new product output as our dependent variable (denote here as Y). This variable represents the nominal value of the completely new products or the improved products introduced by firms through research and development. Unfortunately, due to the nature of the data we are unable to differentiate between the two types of new product output.

Next, we introduce our regressors, and they are categorized into firm characteristic variables, denoted as X, and agglomeration variables, denoted as Z. The following table introduces these variables and explains how each variable is practically computed.

Table [1.a]
Regressors

Z1: Urbanization Index (urb)	● log(the total number of employees in the county area - the total number of employees in the county area in the given industry +1); ● 1 unit increase corresponds to e-1 people increase in the difference
Z2: Localization Index (loc)	● log(the total number of employees in the county area in the industry - the total number of employees in the firm in the given industry and area+1); ● 1 unit increase corresponds to e-1 people increase in the difference
Z3: Porter Index (port)	● log(1/Herfindahl index) ● I used county as my scope for constructing the Herfindahl index.
X1: Frontier Production (ξ)	● Frontier production conceptually represents the technical inefficiency of a firm's production. It corresponds precisely to $ln(\xi_{it})$ in equation (7) and (8). Stata offers a command that predicts this value.
X2: Government Subsidy (sub)	● He & Zhu (2019) mentioned that firms heavily rely on government subsidy in China, so this variable is expected to play an important role in firm production. It corresponds to a part of A_{it}. ● In our data, research and development (R&D) subsidy, which is a determinant factor of innovation, is missing. Our hope is that government subsidy could control for a part of a firm's R&D subsidy, given that government subsidy does have R&D subsidy as one subcategory.
X3: Capital (K)	● Capital usage that corresponds to K_{it} in the model. ● Capital usage is measured by "paid-in capital" in the data.
X4: Material (M)	● Material usage that corresponds to M_{it} in the model. ● It is practically measured by "total industrial intermediate input" in the data.
X5: Labor (L)	● Labor usage that corresponds to L_{it} in the model. ● It is measured by "total number of employees averaged over the year" in the data. It is consistent with capital and labor as soon as we obtain the real values of the latter two through deflation.
X6: New product output of year *t-1* (Y[t-1])	● This will be an additional control added to the model, because we suspect that it is likely for firms with innovative production last year to have similar

	innovative production this year. So this helps account for firm heterogeneity in innovative production.
X7: Profit Dummy	● X7=1 if sales profit of the previous year>0, and X7=0 otherwise. This is based on the assumption that the firms would be more willing to invest in research of new product if their sales profit last year is positive.
X8, X9: Type of firms (foreign, joint)	● I set X6=1 if the company is a foreign affiliate, 0 otherwise. ● I then set X7=1 if the company is a joint firm, 0 otherwise. ● The reason for doing so is that Hu & Jefferson (2002) pointed out foreign affiliates might systematically enjoy better technology and productivity. To distinguish further, I included joint firm as a dummy variable as well.

Note that In the descriptive and the empirical analysis, all nominal values are deflated by the following GDP deflators with the base year being 2015.

Table [1.b]
GDP deflators 1998-2007,
base year 2015 (100)

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
57.79	57.05	58.23	59.42	59.78	61.34	65.60	68.16	70.84	76.33

Source: indexmundi.com

2. Multi-collinearity check and Principal Component Analysis (PCA)

Since agglomeration indices were all constructed from number of employees, I fear that multi-collinearity issue might exist or magnify the standard errors of the coefficient estimates in our empirical model. A pairwise correlation table shows that multi-collinearity is not a great concern among agglomeration variables. The correlations are all kept below 0.5. However, multi-collinearity among production factors is a concern, as all pairwise correlations are greater than 0.5. A potential solution is to remove one of the factors from the model, but here I used an alternative technique, which is principal component analysis.

Principal component analysis is a useful reduction technique that transforms linearly dependent columns into linearly independent ones. At the end we would obtain three components independent of each other. Although it is difficult to interpret what each of the components represents, we can refer to the correlation table to see how the components incorporated the information of the previous variables.

Table[2.a]
Multi-Collinearity Check

	Agglomeration			Production Factor			
	loc	urb	port		lnK	lnM	lnL
loc	1			lnK	1		
urb	0.44	1		lnM	0.55	1	
port	-0.11	-0.06	1	lnL	0.55	0.61	1

Table [2.b]
Components Correlation with Variables

	PC1	PC2	PC3
lnL	0.59	-0.36	-0.73
lnK	0.56	0.83	0.05
lnM	0.58	-0.44	0.69

From table [2.b] we can treat the first component (PC1) as a measure of size or scale, since it is correlated with the three inputs in a similar fashion. Given the positive correlation between PC2 and capital, and its negative correlation with the others, we can loosely treat PC2 as capital intensity, the amount of fixed or real capital in relation to other factors of production. Similarly, we can treat PC3 as material intensity. After PCA, the correlation between each pair of PC1, PC2, PC3 is 0, so we have successfully removed multi-collinearity.

3. Descriptive Statistics

Table 3 displays the basic summary statistics of our variables of interest. All nominal values are deflated by the GDP deflators provided by table [1.b].

The average number of employees for a firm is roughly 280 per year, indicating that most firms had a medium size. The standard deviation of 1367.76 indicates significant heterogeneity in firm size, and the maximum number of average employees is 194410. We can give a similar conclusion regarding firm size by looking at the total product output, where the mean output per year is 81315.39 RMB, and the standard deviation is 794804.7 RMB.

Table [3]
Summary Statistics for Variables

Variable	Observation	Mean	Standard Deviation	Minimum	Maximum
New Product	1,547,319	8314.02	280859.60	0	1.10×10^{8}
Total Product	1,797,103	81315.39	794804.70	0	1.86×10^{8}
New Product Density	1,521,662	0.031	0.14	0	1
Capital	1,799,550	33236.75	577139.40	0	3.05×10^{8}
Material	1,799,550	89881.38	856697.30	0	2.27×10^{8}
Average Worker	1,799,550	280.1277	1367.36	0	194410
Government Subsidy	1,799,550	243.1483	5440.46	0	1492742
Localization	1,799,550	5.74	3.30	0	12.66
Urbanization	1,799,550	10.64	1.48	0	14.62
Porter	1,792,599	1.34	1.16	0	5.20

Note: the unit for new product, total product, capital, material, and subsidy is (thousand RMB). The values are all deflated according to table [1].

In order to better understand the meaning of the statistics for new product, I calculated new product density by dividing new product output value over total product output value. We therefore intuitively see that on average new product value occupied 3.1% of a firm's total product value per year, while the maximum share is 1, meaning that the firm only produced new products that year.

Finally, in terms of agglomeration externalities, we see that localization level is skewed to the right by comparing its max of 12.66 with its mean of 5.74. Specifically, for a given firm in a given county, there were on average around $e^{5.74}$ -1 (roughly 310) employees from other companies that were working in the same industry, and the maximum number of surrounding employees has peaked to $e^{12.66}$-1 (roughly 314895). This is an extremely high figure that might have belonged to some highly specialized industrial parks. In contrast, urbanization level is skewed to the left and exhibits lower standard deviation relative to its mean. This suggests that most firms were stably located in an urbanized setting. In the meantime, urbanization level was on average much higher than localization level, producing a mean of $e^{10.64}$ -1 (roughly 41771) surrounding employees from different industries in the county. The maximum number has reached to $e^{14.62}$ -1 (around 2235554), over seven-fold the max of surrounding specialized employees.

Figure [3.a] and [3.b] shows the histogram of localization level and urbanization level in both 1998 and 2007. Figure [3.a] reveals that most firms were located in areas with very little specialization. In 1998, a great number of firms had approximately 0 to 2 other industry-specific employees within their counties, while a small number of firms had on average about 1600 surrounding employees. In 2007, the number of firms in less specialized settings decreased, while the number of industry-specific surrounding employees for some of the other firms increased to about 1800. Turning to urbanization, the histogram of urbanization resembles a skewed normal distribution. The mode occurs at about 10 in 1998, indicating that most firms had roughly 22,025 surrounding employees from different industries. That number increased to about 162753 in 2007. In general, we can conclude that localization and urbanization level increased in 2007 as compared to 1998. This phenomenon could be explained by the increasing population in China over time (1274 million in 1998 to 1321 million in 2007[3]), as well as government's effort in building science parks and developing cities.

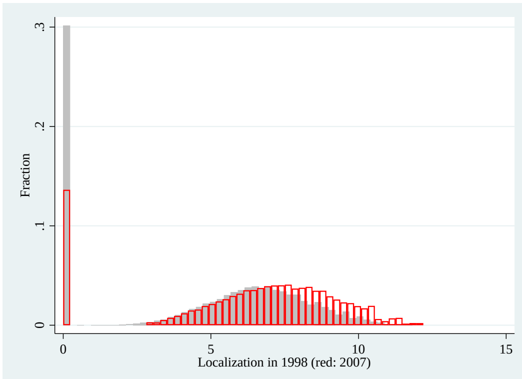

Figure [3.a] Histogram of Localization

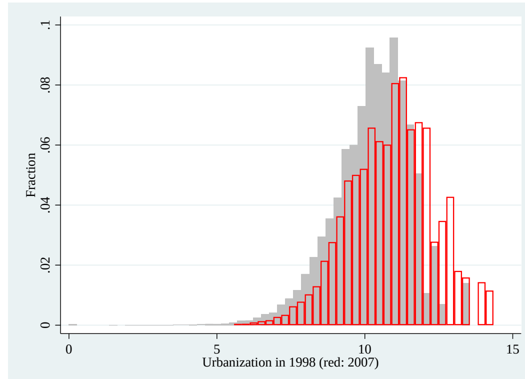

Figure [3.b] Histogram of Urbanization

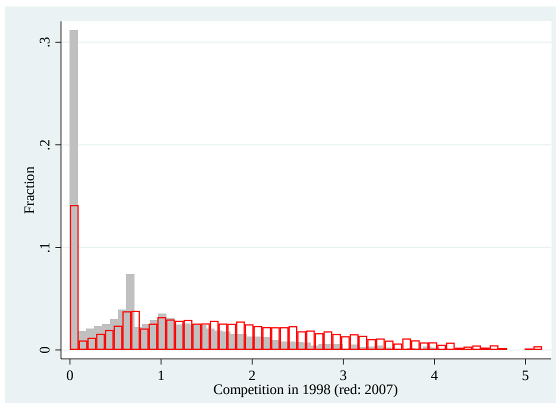

Figure [3.c] Histogram of Porter Index

The last column in table [3] shows that the Porter index is skewed to the right, indicating that most firms were located in cities with low competition from peers. From figure [3.c] we observe that most firms had very little competition in 1998, but the number of these firms decreased significantly in 2007. The trend in the histogram shifted right, implying that the peer effect increased over time. Counterintuitively, in table [2.a] we observed a neglectable negative correlation between localization and Porter. One possible

[3] Source: https://www.statista.com/statistics/263765/total-population-of-china/

cause could be the high industry shares among localized firms, which gave birth to a market with high specialization but low competition. Note that the max value of competition is ∞, which occurs when market concentration is low and the Herfindahl index is 0. In the data there are 26 entries that has a 0 Herfindahl index, and for the sake of the regression, I coded their corresponding Porter index as 1×10^{10} to capture this large competition effect.

4. A Brief Look at the Geographical Distribution

Figure [4.a] Geographical Distribution of Industrial Firms, 2005-2007

Figure [4.a] shows the geographical distribution of the total number of distinct industrial firms that existed throughout 2005-2007 at the level of province, autonomous region, and municipality. Among all 23 provinces, 5 autonomous regions and 4 municipalities, Guangdong had the highest amount of industrial firms (5252), followed by Jiangsu (2862), Shandong (2334), Heilongjiang (1712), Zhejiang (1644), Henan (1314), and Sichuan (1203). Most manufacturing firms clustered at the coastal area in the eastern part of China, with the exception of Sichuan and Heilongjiang. The western part of China had notably fewer industrial firms than the eastern part. Tibet had the least number of industrial firms (9), followed by Qinghai (48) and Shanghai (50).

Figure [4.b] Geographical Distribution of New Product Firms, 2005-2007

Figure [4.b] is the geographical distribution of the number of distinct industrial firms that introduced new products from 2005-2007. Henan had the largest number of new product firms (732), meaning that 55% of its industrial firms introduced new products over the span of the two years. The next is Sichuan, which had 626 new product firms, occupying 52% of its total industrial firms. The two provinces are closely followed by Guangdong (587, 11%), Beijing (445, 46%), Jiangsu (416, 14%), Zhejiang (365, 22%), and Shandong (362, 15%). Consistent with figure 4, the western part had very few new product firms, with Tibet ranking the lowest (1, 11%).

Finally, figure [5.a], figure [5.b], and figure [5.c] show the distribution of the level of localization, urbanization, and competition respectively. We first spot that provinces such as

Guangdong, Heilongjiang, and Chengdu, who had high numbers of industrial firms and new product firms, also enjoyed high levels of localization, urbanization, and competition. We also spot that the coastal areas in eastern China exhibited more agglomeration compared to those from the western part, while the same relationship is true for the number of new product firms. By cross checking the distributions observed from these maps, we spot a moderate positive connection between product innovation and agglomeration.

Figure [5.a] Localization

Figure [5.b] Urbanization

Figure [5.c] Porter

(all averaged over 2005 - 2007)

We might have recognized that the east and the west has distinct industrial structures. In *China's New Urbanization Path* by Jiquan Hu, the author addressed the large gap between the eastern part of China and the western part in terms of their manufacturing industries and regional economic growth. The eastern parts including Shanghai, Guangzhou, Zhejiang, etc, are coastal areas neighboring Japan, Korea, and other East Asian countries. Such geographical nature attracted a great number of manufacturing companies looking for efficient and profitable trades with other nations. Most eastern parts are also plain areas, making it easier to mobilize human or physical capital and attract talented workers. The western parts such as Tibet and Xinjiang, on the other hand, are mostly constituted by plateau and mountains, with very little concentration of residents and workers. This explains the low agglomeration level we observed. They are also not geographically friendly for people who are looking for work and trade, which is why most industrial firms won't choose to agglomerate there. But the western part has its own odds, which is the abundant amount of mineral and other natural resources, rendering it the largest hub for energy supply.

VI. Empirical Specification

1. Fixed Effects Regression Model

Referring back to table [1], I use a fixed effects regression that is specified as follows:

$$lnY_{it} = \widehat{\beta_{0t}} + \sum_{i=1}^{6} \widehat{\beta_{it}}\, lnX_{it} + \sum_{i=7}^{9} \widehat{\beta_{it}} X_{it} + \sum_{i=1}^{3} \widehat{\gamma_i}\, lnZ_{it} + T_i + F_i + \epsilon_{it} \quad (12)$$

where Y_{it} is the deflated new product output value, Z_{it} is the set of agglomeration indices, and X_{it} is the set of firm characteristic variables for firm *i* at year *t*. T_i is the time fixed effect, F_i is the firm fixed effect, and ϵ_{it} is the time-specific and firm-specific heterogeneity. Note that I lagged one year for the continuous variables in the vector *X* is because those values in *X*, such as capital, material, and labor, are reported as the end values of the given year, which will likely contribute to the production of the following year. Agglomeration variables do not suffer this constraint because the number of employees that were used to construct the indices were averaged over the year.

OLS estimates are reported in table [4]. The first column reports the estimates with only agglomeration indices and fixed effects included. In the second column, I further included the production factor controls – frontier production, capital, material, and labor. I included more controls - government subsidy, new product output in the previous year, and profit dummy - in the third column. Finally, I added the dummy variables indicating firm type into the forth column.

In all four columns the coefficients of localization economies are statistically significant. The relationship with the dependent variable is consistently positive. In column

(1) and (2), we observe that one percent increase in localization is associated with 0.0085% increase in new product output. The magnitude decreased in column (3) and (4) to 0.0065%. Urbanization also consistently exhibits a statistically significant positive effect with a larger magnitude throughout the four models: one percent increase in urbanization is expected to increase new product output by 0.23%. Porter externalities shows consistently a neglectable negative association with product innovation: one percentage increase in competition is expected to decrease new product output by $3.18 \sim 3.21 \times 10^{-6}$ %. In reality, porter externalites would have little practical economic significance on product innovation, while urbanization would be more relevant and important for new product outputs.

Other than the major agglomeration indices, frontier production, a measure of technical inefficiency, is negatively correlated with new product output with an effect of 0.12, implying that technical inefficiency could discount production output by 0.12%. We should be cautious when we interpret the coefficient of PC1 (scale), PC2 (capital intensity), PC3 (material intensity), and subsidy, because the coefficients are likely biased and inconsistent due to endogeneity: production factor inputs and government subsidy measured in deflated market values could still rely on the economic condition of the area the firm locates in, while the economic condition could also correlate with product innovation output through cost-benefit analysis that impacts research decision and production. Despite so, we still see a positive correlation between PC1, PC3, and new product output. Looking at lnY and profit dummy, we recognize that firms with positive sales profit and new product output in the previous year are more likely to have new product output in the given year. Finally, we didn’t spot any evidence showing innovation advantages for foreign firms, but we did find evidence that joint firms tend to produce 0.1% more than domestic firms.

Table [4]
OLS Estimate: Whole Sample

	(1) Agglomeration	(2) Production Factor	(3) Research Decision	(4) Firm Type
localization	0.0085***	0.0080***	0.0065***	0.0065***
	[0.0018]	[0.0019]	[0.0019]	[0.0019]
urbanization	0.23***	0.24***	0.23***	0.23***
	[0.0068]	[0.0070]	[0.0069]	[0.0069]
porter	-3.18×10^{-6}***	-3.21×10^{-6}***	-3.21×10^{-6}***	-3.21×10^{-6}***
	[3.56×10^{-7}]	[4.02×10^{-7}]	[3.98×10^{-7}]	[3.98×10^{-7}]
$\ln\xi$[n-1]		-0.12***	-0.12***	-0.12***
		[0.0035]	[0.0035]	[0.0035]
PC1[n-1]		0.040***	0.037***	0.037***
		[0.0022]	[0.0022]	[0.0022]
PC2[n-1]		-0.028***	-0.024***	-0.024***
		[0.0038]	[0.038]	[0.0038]
PC3[n-1]		0.016***	0.0093**	0.0094**
		[0.0043]	[0.0044]	[0.0043]
lnsub[n-1]			-0.0094***	-0.0094
			[0.0011]	[0.0011]
lnY[n-1]			0.083***	0.083***
			[0.0014]	[0.0014]
Profit dummy			0.042***	0.042***
			[0.0053]	[0.0053]
Foreign dummy				0.038
				[0.033]
Joint dummy				0.10***
				[0.031]
Time fixed effect	Yes	Yes	Yes	Yes
Firm fixed effect	Yes	Yes	Yes	Yes
Number obs.	1,440,762	1,388,692	1,388,502	1,388,502
R^2	0.0014	0.14	0.17	0.17

Note: standard errors are reported in the bracket. *, **, and *** indicate significance at level 0.1, 0.05, and 0.01.

Table [5.a] groups the sample through the threshold of 25%, 50%, and 75% percentile of average number of total employees, a variable that could be treated as a measure of firm size. We can steadily see an increasing positive correlation of urbanization and new product, from 0.085% to 0.29%, as we go from the lowest to the highest percentile group. This might suggest that urbanization has a larger impact on product innovation incentivization and improvement for larger firms than for smaller firms. We also see an interesting trend on the coefficient of the Porter index: the effect is extremely small and negative for the lowest percentile group, but for the other groups it is positive in an increasing manner with firm size. This might explain why we obtained a neglectable positive correlation when we averaged over all groups. Finally, localization has a statistically significant negative impact on new product output for medium sized firms, but we have found no evidence on how it impacts the smallest and the largest percentile group. The heterogeneity in the two polarized group might explain why we observed a small positive coefficient in table [4].

Table [5.a]
OLS Estimate: Firm Size

	(1) <25%	(2) 25%~50%	(3) 50%~75%	(4) >=75%
loc	0.0024	-0.013***	-0.016***	-0.0027
	[0.0034]	[0.0042]	[0.0047]	[0.0052]
urb	0.085***	0.11***	0.15***	0.29***
	[0.012]	[0.012]	[0.013]	[0.017]
port	-1.08×10^{-6}	0.091***	0.11***	0.089***
	[3.65×10^{-7}]	[0.013]	[0.016]	[0.021]
All controls	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Firm fixed effect	Yes	Yes	Yes	Yes
Number obs.	309,457	371,320	351,089	356,636
R^2	0.13	0.12	0.15	0.2

Note: standard errors are reported in the bracket. *, **, and *** indicate significance at level 0.1, 0.05, and 0.01.

Table [5.b] groups the sample by geographic locations. As we recognized in figure [4] and [5], the west and the east are significantly different in their economic structures, which is why I sorted out the west as a subsample in the following regression. For the eastern part, I also subsampled it by the “Northern Industrial Base (IB)” and the “Southern Industrial Base” (Baidu; n.d.). The major two industrial bases in the north are Beijing-Tianjing-Tangshan industrial base and Southern Liaozhong (including Shenyang, Dalian, Liaoyang, etc.)

industrial base. These two bases mostly specialize in heavy industries such as iron, machinery, soil, and chemicals. On the other hand, the major two industrial bases in the south are Shanghai-Nanjing-Hangzhou industrial base and the Pearl River Delta (including Guangzhou, Shenzhen, Zhuhai, etc.) industrial base. These southern two bases mostly specialize in light industries such as clothing, food, toy, electronic goods, etc. The difference in specialization might result in different relationship with agglomeration variables.

Table [5.b]
OLS Estimate: Geographic Specification

	(1) West	(2) Northern IB	(3) Southern IB
loc	-0.021	-0.0031	0.0066**
	[0.014]	[0.0034]	[0.0027]
urb	-0.0051	0.067***	0.3***
	[0.02]	[0.015]	[0.009]
port	-1.32×10^{-6}	-3.84×10^{-6}***	-4.14×10^{-7}***
	[1.43×10^{-6}]	[8.12×10^{-7}]	[0.016]
All controls	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes
Firm fixed effect	Yes	Yes	Yes
Number obs.	9,086	371,320	854,451
R^2	0.0033	0.12	0.14

Note: standard errors are reported in the bracket. *, **, and *** indicate significance at level 0.1, 0.05, and 0.01.

From column (1) in table [5.b] we find no evidence for the existence of agglomeration externalities in the west. All agglomeration variables are statistically insignificant. In column (2) we find evidence of a positive correlation between urbanization and new product output among firms in the northern industrial base. For firms in the southern industrial base, we discover that both localization and urbanization positively impact new product output, and the magnitude for urbanization is larger than that in the northern IB. Competition effects in both IBs are neglectable and of little practical significance. In conclusion, the regression result with geographic specification suggests that southern IB might be the most benefitted recipient of agglomeration externalities.

2. Robustness Check

One major issue that could bias the estimate is spatial selection of firms. For example, if firms with better productivity selected to locate in agglomerated areas in the first place, then our estimates would be inconsistent, as firm productivity would now reversely affect the agglomeration indices assigned to the firms.

Alfred Marshall proposed that geographic features play an important role in a firm's spatial selection decision. For example, manufacturing companies would likely choose to locate in areas with less urbanization in order to dispose of waste and avoid polluting the inner cities. Other industries would likely locate in areas with available natural resources that would help lower transaction and transportation cost for needed materials. In an economy with international trades, manufacturing firms might also locate in coastal areas to reduce the export cost for their goods.

In the meantime, state-owned firms in China traditionally have less freedom to engage in spatial selection, as they are established by local government of each of the cities and counties, hence we would expect agglomeration indices to be orthogonal with heterogeneity among state-owned firms. Since we observed from table [5.a] that firms with a size under the 25 percentile tend to exhibit distinct effects, especially for competition, compared to firms with a size over the 25 percentile, table [6.a] and [6.b] shows the regression result for state-owned firms with a size under and over the 25 percentile threshold respectively. I also included the interaction terms of firm size and agglomeration indices to take into account the idiosyncratic agglomeration effects we observed among firms with different sizes.

Table [6.a]
OLS Estimate: State-owned Firms
(>=25% in firm size)

	(1) West	(2) Northern IB	(3) Southern IB
loc	0.032	-0.0089	0.029
	[0.0034]	[0.0088]	[0.019]
urb	-0.045*	0.10***	0.0068
	[0.012]	[0.03]	[0.036]
port	0.080	-0.21***	-0.74
	[0.21]	[0.079]	[0.11]
loc*L	-0.000043	7.21×10^{-6}***	-1.51×10^{-6}
	[0.0000313]	[2.62×10^{-6}]	[7.10×10^{-6}]
urb*L	0.000024	-3.15×10^{-7}	8.13×10^{-6}**
	[0.000020]	[2.95×10^{-6}]	[3.85×10^{-6}]
port*L	0.00018	0.00021***	0.00021*
	[0.00042]	[0.000073]	[0.00012]
All controls	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes
Firm fixed effect	Yes	Yes	Yes
Number obs.	2,036	28,880	21,308
R^2	0.02	0.44	0.43

Note: standard errors are reported in the bracket. *, **, and *** indicate significance at level 0.1, 0.05, and 0.01.

In table [6.a], we recognize that localization is no longer statistically significant for any geographic specification for medium to large-size firms. We also notice that the magnitude of urbanization has decreased intensively: urbanization likely decreases new product output of firms in the western part by 0.045% and increases firms in northern IBs by 0.1%. We haven't found any evidence to conclude that urbanization plays a part in promoting product innovation in Southern IBs. The statistically significant interaction term for urbanization implies that large firms benefit from urbanization more. Consider the max number of firm size, which is 194410 people, we recognize that urbanization can increase innovation production of the largest-size firm by 1.5% if we assume that the main urbanization effect is 0. Similarly, in northern IBs, the firm with the largest size could benefit from an increase of roughly 1.4% in new product output by 1% increase in localization. A similar reasoning applies to the porter interaction term: the largest firm (either in Northern IBs or in Southern IBs) potentially could receive a 40% increase in product innovation output through increase in competition. Finally, in table [6.b], we find it hard to conclude any evidence about agglomeration effects for small-size firms.

Table [6.b]
OLS Estimate: State-owned Firms
(<25% in firm size)

	(1) West	(2) Northern IB	(3) Southern IB
loc	0.036	0.0041	-0.0016
	[0.024]	[0.0088]	[0.011]
urb	0.0018	0.0055	-0.0026
	[0.012]	[0.021]	[0.029]
port	5.81×10^{-8}	1.78×10^{-7}	-3.47×10^{-9}
	[1.01×10^{-7}]	[1.2×10^{-7}]	[6.24×10^{-8}]
loc*L	-0.0016	-0.00026	0.00037
	[0.0012]	[0.00048]	[0.00064]
urb*L	0.00010	0.00050	-0.000073
	[0.00022]	[0.00018]	[0.00027]
port*L	0.0012	0.0012	0.00033
	[0.0025]	[0.0019]	[0.0017]
All controls	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes
Firm fixed effect	Yes	Yes	Yes
Number obs.	12,52	11,961	6,463
R^2	0.002	0.02	0.02

Note: standard errors are reported in the bracket. *, **, and *** indicate significance at level 0.1, 0.05, and 0.01.

3. Discussion

Our result shows that spatial selection accounts for most of the effects that agglomeration indices captured, and this is especially prominent for localization. Indeed, aside from the localized benefits such as natural resources and distance to ports, Alfred Marshall proposed some other benefits, such as knowledge spill-over, that would occur after a firm chooses its location. However, recent research, such as Zhang (2014) and Tavassoli & Jienwatcharamongkhol (2016), failed to find evidence that shows how localization benefits individual firms after they entered the location. It might be the case that the opportunity costs, such as regulations in science parks, outweighed the benefit brought by labor pooling, transportation, or knowledge spill-over. We did notice that urbanization externalities are more impactful in boosting a firm's innovation performance. This is also consistent with past research results from Martin et al. (2011) and Zhang (2014). The result might imply that industrial diversity and inter-industry knowledge spill-over are more helpful for firms than localized benefits. It is a reasonable speculation since industrial firms might depend on technologically related firms from other industries to develop its new technology and new design, in a similar way how interdisciplinary studies incorporate knowledge from different fields to achieve new discoveries. Industrial firms can also trade outputs to and obtain necessary inputs from different sectors more easily in an area with a larger size or population base.

In the meantime, we also discover that larger firms are more likely to benefit from agglomeration. This could be due to their market power, as firms with larger size tend to attract more workers, synthesize resources faster, and have better connection, communication, or intellectual interaction with other firms. Although it would be illuminating if we could discover methods to help smaller firms perform better, we couldn't find evidence that small firms benefit from agglomeration and spatial clustering. We also recognize that firms in the west tend not to benefit from any agglomeration in terms of product innovation output. This could be due to their low agglomeration level that is too small to make a positive impact. We could call for policies that promote formation of spatial clusters, but the key question would be whether it is necessary to change the economic structure of the west. In general, we recognize that localized clusters such as science parks might have not delivered their goals of boosting firm performance, as well as innovation, as expected.

VII. Conclusion

In this study, I investigated the effect of agglomeration activities on product innovation in order to reflect the industrialization and economic growth of China. I followed the theoretical framework proposed by Martin et al. (2011), and constructed the corresponding agglomeration variables including localization economies, urbanization economies, and competition. I employed a fixed effects regression with firm-specific controls (productivity, firm size, etc.) and agglomeration variables. I then conducted a robustness check by assuming exogeneity of treatment variables among state-owned firms. I discussed the latter regression results with policy implications.

My study implies that policymakers could promote innovation by encouraging firms to locate in industrially diversified cities and areas. In the meantime, government designing science parks should also take into consideration the industrial composition across the park: introducing a diversified pool of enterprises will be more effective than constructing a specialized setting.

Finally, future research should investigate agglomeration externalities in a more specific manner. Given the data constraint, this paper cannot construct measures such as labor pooling effect, related variety, unrelated variety, etc. Future research can also improve the methodology by constructing a structural equation model for this topic, taking into account a firm's decision in research and development, the probabilistic trial of success in innovation, and the process of realizing innovation into actual production.

Reference:

Andors, S. (1979). China's Industrialization in Historical Perspective. In: China's Road to Development (Second Edition). United Kingdom: Pergamon.

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2011). Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics*, *97*: 339–351.

Boschma, R. A., & Wenting, R. (2007). The spatial evolution of the British automobile industry: Does location matter? Industrial and Corporate Change, *16*: 213–38.

Carlino, G., & Kerr, W. (2015). Agglomeration and Innovation. *Handbook of Regional and Urban Economics*, *5*: 349-404, DOI: https://doi.org/10.1016/B978-0-444-59517-1.00006-4.

Fagerberg, J. (1994). Technology and International Differences in Growth Rates. Journal of Economic Literature, 32: 1147–75.

Frenken, K., OORT, F., & Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, *41*(5): 685–697.

Ge, Y. (2009) Globalization and Industry Agglomeration in China. *World Development*, *37*(3): 550-559. DOI: 10.1016/j.worlddev.2008.07.005.

Gordon, I., McCann, P. (2005). Innovation, agglomeration, and regional development, *Journal of Economic Geography*, *5*(5): 523-543. DOI: https://doi.org/10.1093/jeg/lbh072.

He C., & Zhu, S. (2019). How Do Agglomeration Economies Contribute to Firm Survival in China?. In: Evolutionary Economic Geography in China. Economic Geography. Singapore: Springer.

Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial Development in Cities. *Journal of Political Economy*, *103*(5): 1067-1090.

Hu, A., & Jefferson, G. (2002). FDI impact and spillovers: Evidence from China's electronics and textile industries. *The World Economy*, *25*: 1063-1076.

Jacobs, J. (1969). The economy of cities. New York, MA: Random House.

Li, C., & Gibson, J. (2014). Agglomeration Economies in China: Locations and Effects. *Working Papers in Economics,* University of Waikato.

Lin, H., Li, H., & Yang, C. (2011). Agglomeration and Productivity: Firm-level Evidence from China's Textile Industry. *China Economic Review*, *22*: 313-329. DOI: 10.1016/j.chieco.2011.03.003.

Marshall, A. (1920). Principles of economics; an introductory volume. London: Macmillan.

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. Journal of Urban Economics, 69: 182-195.

National Bureau of Statistics. (n.d.). Variable Description of Annual Surveys of Industrial Firms. p.3, n.d.

Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, *March/April*: 73-91.

Solow, R. (1957). Technical Change and the Aggregate Production Function. *Review of Economics and Statistics*, *39*: 312–20.

Wen, M. (2004). Relocation and agglomeration of Chinese industry. *Journal of Development Economics*, 73(1): 329-347. DOI: 10.1016/j.jdeveco.2003.04.001.

Wen, Y. (2016). China's Rapid Rise: From Backward Agrarian Society to Industrial Powerhouse in Just 35 Years. *Regional Economist*.

Wong, P., Ho, Y., & Autio, E. (2005). Entrepreneurship, Innovation and Economic Growth: Evidence from GEM data. *Small Business Economics*, *24*(3): 335-350. DOI: 10.1007/s11187-005-2000-1.

Zhang, H. (2014). Agglomeration and product Innovation in China.

Zheng, L., Zhao, Z. (2017). What drives spatial clusters of entrepreneurship in China? Evidence from economic census data. China Economic Review, 46: 229-248. DOI: 10.1016/j.chieco.2017.09.014.