Appendix A

A: The procedure for identifying industrial clusters

Unlike EG index, in this paper we use continuous spatial measures to explore the geographic boundaries of a given clusters. Based on Zhu et al. (2019), we focus on identifying industries clusters using 2,035,909 Chinese manufacturing firms during the period 2003-2009. Our procedure to identify industrial clusters as follows.

(1) Identify the continuous geographic boundaries of a given industry

For the whole manufacturing industries, based on firms' address they registered, we recovered the geographic coordinates (longitude and latitude) of each manufacturing firm via the Google Maps JavaScript API. Then, we calculated the bilateral distance between every pair of firms and obtained the median distance d of manufacturing industries each year. Given a four-digit industry A comprising n_A firms, we constructed the industry-specific kernel density function of bilateral distances weighted by firms' employment as

$$\widehat{K_A}(d) = \frac{1}{h_A \sum_{i=1}^{n_A-1} \sum_{j=i+1}^{n_A} e_i e_j} \sum_{i=1}^{n_A-1} \sum_{j=i+1}^{n_A} e_i \cdot e_j \cdot f(\frac{d - d_{ij}}{h_A})$$

where d_{ij} is the Euclidean distance between firm i and firm j, d is the median distance of manufacturing industries as we mentioned before, e_i and e_j are the employment of firm i and firm j, respectively. h_A donates an industry-specific band width, while $f(\cdot)$ represents Gaussian distribution function. For each industries, we calculated its real kernel density at each kilometer and plotted real kernel density distribution.

Suppose firms are randomly distribution, the locations of manufacturing firms are all optional locations. For industry A with n_A firms, we randomly matched firm locations in all optional sample using a computer, and recalculated the simulated kernel density. Repeating this procedure 100 times and we obtained 100 simulated kernel density at each kilometer. For each kilometer, sorted kernel density values in ascending order and selected the upper bound $\overline{K_A}(d)$ at 99% values. Then, we used interpolation method to plot the simulated distribution of industry A, and obtained the geographic boundaries of a given industry by comparing the real kernel density and simulated distribution. Specifically, if the real kernel density is larger than the upper bound $\overline{K_A}(d')$ at d' kilometer, we consider d' as the geographic boundaries of industry A.

$$d_A = max\{0 \leq d \leq \overline{d}: \widehat{K_A}(d') > \overline{K_A}(d'), \forall d' \in (0, d)\}$$

(2) Compile inter-industry linkage dataset for each localized industry.

To capture the inter-industry linkages, we relied on three-digit industry connections to identify the input-output linkages of core industries. China's three-digit industry is often composed of multiple four-digit industries and there is a strong correlations between industries. We believe that three-digit industries are enough to evaluate vertical linkages. According to input-output consumptions from China's input-output table, we identified the most related upstream and downstream three-digit industries of each four-digit industries. Therefore, we compiled an inter-industry linkage dataset incorporated with core industry and its related industries.

(3) Identify firms within industrial clusters

For an industry A, we singled out firms within the core industries as the center of a cluster and calculated the employment density through an inter-industry linkage dataset within the range of geographic boundaries d_A, and denoted the maximal employment density (E_{A1}) as a critical value. Omitted all firms enclosed in E_{A1} and used the remaining firms to recalculate the employment density with a slightly smaller radius. Decreasing the radius until the remaining firms' maximal employment density reached the critical value, and we identified the second cluster E_{A2}. Repeated this procedure until we could not find another clusters whose employment density is higher than the critical value. Finally, we obtained all firms within clusters in industry A.

Appendix B: Tables and Figures

Table B1 Descriptive statistics of dependent variables (2003–2009)

Dependent variables	Overall		Before the establishment of development zones		After the establishment of development zones	
	Mean	SD	Mean	SD	Mean	SD
per_output	508.591	1105.227	349.812	747.560	586.940	1236.911
ROA	8.990	16.627	6.815	14.553	10.054	17.454
Growth	13.465	19.482	10.707	17.821	14.821	20.110
Employment	0.026	0.593	0.031	0.648	0.025	0.566
Export	0.064	0.230	0.079	0.258	0.057	0.213

Note: The unit of ROA, Growth, Employment and Export is percentage. The table was calculated by authors.

Table B2 Descriptive statistics

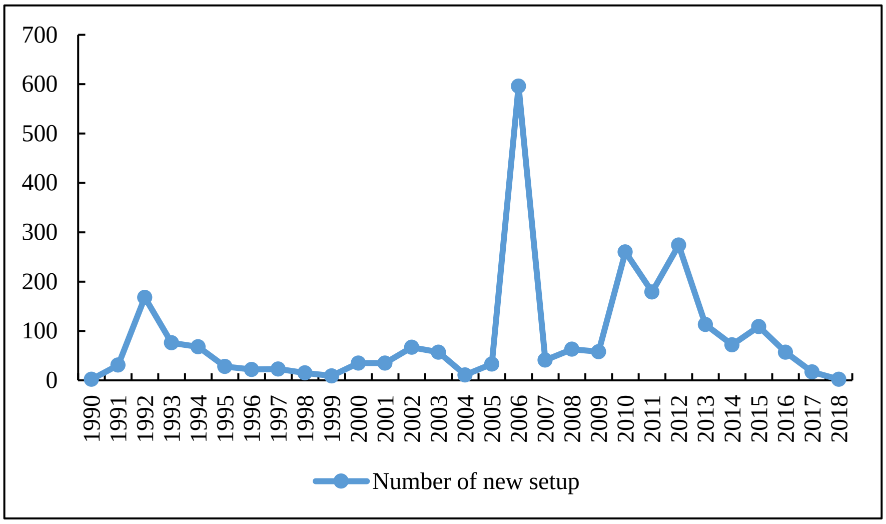
Variable	Obs	Mean	Std. Dev.	Min	Max
DZ*Policy	943163	0.069	0.253	0.000	1.000
DZ_scale*Policy	943163	0.086	0.342	0.000	2.754
DZ_num*Policy	943163	0.103	0.437	0.000	6.000
DZ_main*Policy	943163	0.069	0.253	0.000	1.000
Age	943163	1.953	0.771	0.000	4.111
Large	943163	0.057	0.232	0.000	1.000
HHI	943163	0.215	0.236	0.004	1.000
PGDP	912230	10.451	0.863	6.826	12.743
FDI	911857	11.284	1.616	3.091	13.916
Density	938054	0.558	0.681	0.0001	13.333

Note: The unit of Density is thousand per/km^2, Age, PGDP, and FDI are in logarithmic form. The table was calculated by authors.

Table B3 Effects of development zones on firms' performance: DID estimation (Core industries)

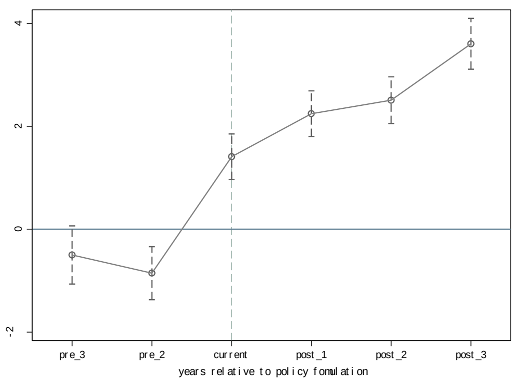
Dependent variable	Per_output	ROA	Growth	Employment	Export
	(1)	(2)	(3)	(4)	(5)
*DZ*Policy*	73.17***	3.167***	3.591***	0.036***	0.008***
	(8.103)	(0.269)	(0.911)	(0.014)	(0.002)
Age	127.5***	4.030***	-40.93***	-0.017	0.034***
	(10.59)	(0.218)	(1.327)	(0.018)	(0.002)
Large	341.2***	0.879***	9.123***	0.055***	0.053***
	(15.81)	(0.283)	(0.903)	(0.017)	(0.004)
HHI	-29.26***	-1.436***	-1.010	0.0512***	0.080***
	(10.88)	(0.251)	(1.022)	(0.019)	(0.005)
PGDP	341.2***	8.201***	7.626***	0.170***	0.025***
	(28.38)	(0.506)	(2.166)	(0.034)	(0.006)
FDI	17.05***	1.030***	2.411***	0.018***	0.005***
	(3.963)	(0.105)	(0.397)	(0.006)	(0.001)
Density	7.940**	0.593***	0.027	-0.003	0.009***
	(3.248)	(0.126)	(0.358)	(0.017)	(0.002)
Intercept	-3617.5***	-102.3***	10.23	-1.708***	-0.264***
	(293.0)	(5.703)	(23.41)	(0.367)	(0.072)
Industry FE	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
Observations	329591	317822	246298	249417	321040
R^2	0.758	0.661	0.468	0.357	0.879

Note: Estimations are based on the China Industry Survey dataset between 2003 and 2009 using DID specification with firm, industry, region and year fixed effects. Robust standard errors are reported in parentheses. Statistics significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. The table was calculated by authors.

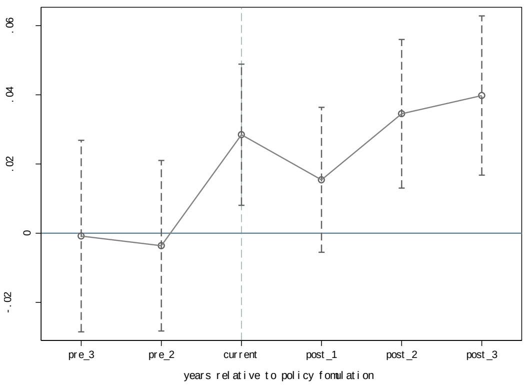


Data source: The Catalogue of China's Development Zones (2018 version), the figure was made by authors.

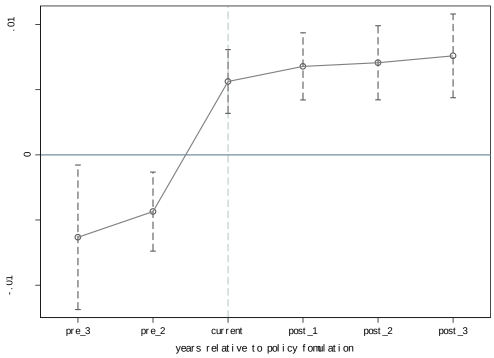
Figure B1 Number of new setup development zones during the period 1984-2018



(a) Dependent variable=*Growth*



(b) **Dependent variable**= *Employment*



(c) Dependent variable= *Export*

Note: these figures were made by authors.

Figure B2: Parallel trend: the coefficients of yearly interaction terms pre-period and post-period.

Appendix C: Group Classification of Heterogeneity Analysis

Different industries group: Based on the different factor intensity in China' manufacturing industries, we divide them into three categories according to the convention. Specifically, labour-intensive industries contain agricultural food processing, food manufacturing and textile, etc; capital-intensive industries consist of chemical products manufacturing, ferrous metal smelting and rolling processing, etc; technology-intensive industries include pharmaceutical manufacturing, communications equipment, electronic equipment manufacturing, etc.

Eastern group: Eastern group contains 11 coastal provinces and municipalities, including Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan.

Yangtze River Delta city cluster: According to city planning of Yangtze River Delta (2016), Yangtze River Delta city cluster contains 26 cities, including 9 cities in Jiangsu province (Nanjing, Wuxi, Changzhou, Suzhou, Nantong, Yancheng, Yangzhou, Zhenjiang and Taizhou), 9 cities in Zhejiang province (Hangzhou, Ningbo, Jiaxing, Huzhou, Shaoxing, Jinhua, Zhoushan and Taizhou), 8 cities in Anhui province (Hefei, Wuhu, Ma 'anshan, Tongling, Anqing, Chuzhou, Chizhou, Xuancheng) and Shanghai.

Central Plains city cluster: The Central Plains city cluster consists of 30 cities with stronger industrial bases in central China, which comprises 18 cities in Henan province (Zhengzhou, Kaifeng, Luoyang, Nanyang, Anyang, Shangqiu, Xinxiang, Pingdingshan, Xinyang, Zhumadian, Xuchang, Jiaozuo, Zhoukou, Hebi, Puyang, Luo river, Sanmenxia and Jiyuan), 3 cities in Shanxi province (Changzhi, Jincheng and Yuncheng), 2 cities in Hebei province (Xingtai, Handan), 2 cities in Shandong province (Liaocheng, Heze) and 5 cities in Anhui (Huaibei, Bengbu, Suzhou, Fuyang, Bozhou).

REFERENCES

City planning of Yangtze River Delta (2016). Retrieved from https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/201606/t20160603_962187.html?code=&state=123, Jun.2016.

City planning of Central Plains city cluster (2016). Retrieved from http://www.gov.cn/xinwen/2017-01/05/content_5156816.htm, Dec.2016.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4), 1077-1106. https://doi.org/10.1111/0034-6527.00362

Ellison, G., & Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5), 889-927. https://doi.org/10.1086/262098

Zhu, X., Liu, Y., He, M., Luo, D., & Wu. Y. (2019). Entrepreneurship and industrial clusters: Evidence from China industrial census. *Small Business Economics*, 52(3):595-616. https://doi.org/10.1007/s11187-017-9974-3