# Do Place-Based Policies Work? Micro-Level Evidence from China's Economic Zones Program

Yi Lu*, Jin Wang†, Lianming Zhu‡

This Version: July, 2015

**Abstract**

This paper examines the impact of a prominent place-based program in China – the Economic Zones program on economic activity of the targeted areas. To do so, we exploit two geo-coded comprehensive waves of Chinese firm censuses, which allow the construction of a panel dataset for areas before and after the zone establishment. By pairing areas across zone boundaries, we find that, first, the economic zones have a positive effect on the employment, output and capital of the targeted area. The program has also increased the number of firms located in the zones. Second, the extensive margin (firm births and deaths) plays a larger role in explaining the SEZ effect than the intensive margin (incumbents). Finally, the zones' effectiveness depends on program features. Across sectors, for firms in more capital-intensive sectors, the zones exhibit larger positive impacts on firm performance than those in more labor-intensive sectors. Meanwhile, location characteristics such as market potential and transportation accessibility are not critical factors in enhancing the program effects.

**Keyword**: Place-based Policy; Relocation; Heterogeneity; Market Potential
**JEL Classification**: H2, J2, R3, R5

*Yi Lu, Department of Economics, National University of Sinagapore, 1 Arts Link, Singapore, 117570. Email: ecsluyi@nus.edu.sg.

†Jin Wang, Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Email: sojinwang@ust.hk.

‡Lianming Zhu, Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan. Email: lmlianming@gmail.com. We thank Maitreesh Ghatak, Gerard Padró i Miquel, Albert Park, James Kung, Eric Verhoogen and the audience at various seminars for comments. All errors remain our own.

# 1 Introduction

Place-based programs—economic development policies aimed at fostering the economic growth of an area within their jurisdiction—have grown popular and been pursued by many governments around the world over the past several decades. By design, place-based policies can potentially influence the location of economic activity, as well as the wages, employment, and industry mix of targeted areas (Kline and Moretti, 2014a). Some economists are skeptical about the efficiency of the program (Glaeser and Gottlieb, 2008; Glaeser, Rosenthal, and Strange, 2010): workers and firms may move from other regions to the targeted areas and arbitrage away the benefits associated with the program without improving the welfare of local residents (Kline, 2010; Hanson and Rohlin, 2013). On the other hand, agglomeration economies are considered as an important rationale for policies that encourage new investment in a specific area (Kline and Moretti, 2014b; Combes and Gobillon, 2015).

While there is much research interest focusing on the programs in the United States and in Europe (see Neumark and Simpson, 2014, for a comprehensive review),[1] there have been few attempts to evaluate interventions in developing countries. Several questions loom especially large: who benefits and who loses from place-based programs? Which factors determine the effectiveness of such interventions? While developing countries usually suffer from poorly developed institutions and markets, would the assumptions and conceptual approaches from the place-based policy literature on the United States and Europe still hold for them? Very little progress has been made in addressing these issues largely due to the lack of longitudinal studies in developing countries, in particular, the research that traces a place-based program's effects on micro-level units such as firms and workers.

This paper presents a novel first step in this direction. Specifically, we document micro-level evidence on the incidence and effectiveness of place-based policies in developing countries, by evaluating China's Special Economic Zones (SEZs) program. As a prominent development strategy implemented worldwide, SEZs attempt to foster agglomeration economies by building clusters, increasing employment and attracting technologically advanced industrial facilities.[2]

China provides an ideal setting for exploring the causal effect of SEZs on regions and firms, which is of great policy relevance. In 1979, the first four SEZs were initiated by the

[1] Prominent examples include initiatives that target lagging areas, such as enterprise zones in the United States, and regional development aid within the European Union.

[2] The SEZs have been implemented by 135 countries (World Bank, 2008).

Chinese government as an experiment in the pursuit of pragmatic and innovative policies. After their early success, the horizon for SEZs has gradually expanded from the coastal areas to the central and western areas. This paper focuses on the wave between 2005 and 2008. In particular, in 2006, 663 provincial-level SEZs were established in China, among which 323 SEZs were granted in the coastal areas, 267 SEZs were established in the central areas, and 73 were granted in the western areas. Compared with the earlier waves, our sample is representative of spatial distributions and accounts for 42 percent of China's SEZs. Hence our estimates have large-scale implications.

Our analyses proceed in three stages. First, we examine the effect of SEZs on targeted areas' employment, output, capital, and the number of firms. Second, we further elucidate the mechanisms underlying the effects of SEZs. Specifically, our paper identifies the effects of the economic zone on two margins—the extensive margin (firm births, deaths, and relocation) and the intensive margin (the incumbents). Finally, we investigate the heterogeneous effects of the zones depending on program features and the characteristics of the targeted localities. We examine how zones influence firms in capital-intensive sectors and labor-intensive sectors differently, and how local determinants of agglomeration economies such as the zone's market potential and transport accessibility play a role in determining the effectiveness of the program.

A key innovation in our analysis is to construct a novel data set that merges a comprehensive data set of China's economic zones, which includes the establishment year, type and villages located within the boundary, with two geocoded Economic Censuses of Chinese firms in 2004 and 2008. The merged data set contains information on age, sector, address, investment, employment and output outcomes, and more importantly, the geographical proximity with respect to the zones, and dynamics at 3,143,445 firms. We then aggregate the individual firms to construct a panel data set by area and by year. The data series cover two periods, two years before the zone's establishment, and two years after the zone's establishment, allowing us to assess the effect of SEZs on the targeted area, and to provide novel evidence on how various margins contribute to the impacts. To the best of our knowledge, this is the first time that the outcomes of interest for SEZ areas can be precisely measured at a very disaggregated level, and comprehensive geocoded information on Chinese firms is compiled and analyzed in relation to SEZs.

The key challenge in identifying the causal effects of zone programs is the selection of appropriate control groups for economic zones (Neumark and Kolko, 2010). We start with a conventional difference-in-differences (DD) analysis at the village level, which is the most disaggregated geographic unit and smaller than a SEZ. We compare the changes of the performance among SEZ villages throughout the adoption of SEZ policies with the changes among non-SEZ villages during the same period. As an alternative approach, we follow Holmes (1998) and Neumark and Kolko (2010) in making use of detailed information on firm location and zone boundaries.[3] We exploit the discontinuity in treatment assignment that occurs at the zone boundary, a boundary discontinuity design (BD). We chose an area of 1,000 meters from the boundary of the SEZ and compare area performance on opposite sides of a zone boundary, presuming that observable and unobservable characteristics are likely to be very similar in the treated area that became an economic zone and the surrounding control area. To further address the endogeneity of artificially drawn boundaries (and hence treated and non-treated areas within this narrow range of the zone boundaries were still not balanced), we combine the DD and BD approaches, and conduct a BD-DD analysis. Specifically, we first obtain an estimate from the data without zone establishment, and then an estimate from the data with zone establishment. Assuming that confounding factors are fixed over time, we can then recover the true SEZ effect on the targeted areas from the difference in these two estimates. We conduct a series of analyses to further investigate the robustness of our findings, including experimenting with different bandwidth choices and two placebo tests to examine potential estimation biases due to the existence of unobservables or spillovers.

We present three classes of results. First, we find that the SEZ program had a significant and positive impact on targeted areas. Based on the BD-DD estimates, from 2004 to 2008, after two years of zones' establishment, the SEZ areas are 46.8 percent larger in employment, 54.5 percent larger in output, and 53.4 percent larger in capital than non-SEZ areas. Meanwhile, SEZs increase the number of firms by 23.5 percent.

Second, by decomposing the firms into three sub-samples, i.e., relocated firms, entrants and exiters, and continuing firms without zone status change in the period studied, we detect a sizable effect on major outcomes such as employment, capital, and output associated with firm births and deaths. We also find that incumbent firms in the zone show a significant improvement in performance. Relocation plays a minor role in the total SEZ effects. Overall, the results indicate that the program influences the targeted area through both extensive margins and intensive margins.

Third, the effects are potentially heterogeneous. We show that, across industries, the zones exhibit larger positive impacts on firms in capital-intensive industries than those in labor-intensive industries, consistent with the design of the SEZ program. However, across zones, we do not detect significantly larger effects for zones with higher market potential and infrastructure accessibility.

[3] Neumark and Kolko (2010) uses detailed GIS maps of California's enterprise zones to pick out a very narrow control ring (1,000 feet wide) around the zone.

The paper contributes to the literature that explores quasi-natural experiments to quantify the impact of place-based programs. A number of papers are close to our work. For instance, Criscuolo, Martin, Overman and Van Reenen (2012) investigates the causal effect of UK's Regional Selective Assistance (RSA) program on employment, investment, productivity and plant numbers (reflecting exit and entry). Givord, Rathelot, and Sillard (2013) examine the impact of the Zones Franches Urbaines (ZFUs) place-based tax-exemption program on business entry and exit rates, economic activity, employment, as well as on the financial strength of companies. Devereux, Griffith and Simpson (2007) as well as Briant, Lafourcade, and Schmutz (2015) uncover geographic factors that can account for the heterogeneity in the treatment effect: impacts of placed-based policies are more significant for locations with high market access.

We also see our paper as complementary to the literature that evaluates the aggregate effects of place-based policies in the presence of agglomeration externalities and infers the implications for productivity and welfare (Busso, Gregory, and Kline, 2013; Kline and Moretti, 2014b). In particular, the importance of firm dynamics in an urban economy, highlighted by Brinkman, Coen-Pirani, and Sieg (2015), is central to our decomposed analyses of SEZ effects attributable to entry, exit, incumbent and relocation.

Our paper relates to several recent papers, which explore China's establishment of SEZs and evaluate their impact on local economies. Wang (2013) and Alder, Shao, and Zilibotti (2013) examine the local (city-level) impact of SEZs on growth, capital formation and factor prices, while Cheng's (2014) analyses involve the estimation of the local (county-level) and aggregate impacts of SEZs. Extending the study to micro-domains, Schminke and Van Biesebroecke (2013) investigate the extensive margin effect of state-level zones on exporting performance.

Methodologically, our project builds on a broad strand of literature implementing the Geographic Regression Discontinuity (GRD) design (Black, 1999; Bayer, Ferreira, and McMillan, 2007; Dell, 2010; Keele and Titiunik, 2015). Lastly, our paper relates broadly to a set of studies examining the impact of taxation on firm-level outcomes such as location decision, entry and employment (Duranton, Gobillon and Overman, 2011; Brülhart, Jametti and Schmidheiny, 2012).

The remainder of the paper is organized as follows. Section 2 lays out the SEZ reform background. Section 3 describes the identification strategies and data in detail. Section 4 presents our baseline SEZ effect estimates and addresses various econometric concerns, followed by evidence on the mechanisms and on whether these effects are heterogeneous across industries and zones. Section 5 concludes.

## 2 Background

SEZs have been widely adopted place-based programs by Chinese governments. There are two main categories of SEZs: national-level and province-level economic zones. The former are more privileged and approved by the central government, while the latter are granted by provincial governments. Once formed, each zone has an administrative committee, which, on behalf of the respective government, takes the responsibility to direct and to administer the zone within the scope of its authority—such as project approval, local taxation, land management, finance, personnel, environmental protection, and public security.

Adopting special policies in an area within a jurisdiction, the goals of SEZs are to increase foreign direct investment, domestic investment, international trade, technological cooperation and innovation, and employment. To realize these goals, SEZs enjoy a certain degree of independence and authority to define preferential policies. In addition, SEZs work constantly to improve the infrastructure including the supply of utilities, telecommunications, transport, storage, and other basic installations and service facilities.[4]

The SEZ program includes a series of preferential policies, the most important of which are the following (Wang, 2013; Alder, Shao, and Zilibotti, 2013):

1. *Tax deduction and customs duty exemption.* Generous tax incentives—reduced corporate income tax rates 15%-24% as opposed to 33% firms normally pay—are available to foreign enterprises, technologically-advanced enterprises, and export-oriented enterprises. Customs duty exemption is given for equipment and machinery employed in the production of export product.

2. *Discounted land-use fee.* Under Chinese law, all land is under state ownership. In the SEZs investors may lawfully obtain the rights for land development and business use. Preferential treatment—the duration of land use, the amount of the use fee, and the method of payment—is granted to investors further differentiated on the basis of types of businesses and uses. For example, the land use right is guaranteed for projects that have a total investment of at least US $10 million, or that are considered to be technologically advanced with a major influence on local economic development.[5]

3. *Special treatment in securing bank loans.* Foreign investment enterprises are encouraged to make use of domestic financial resources to finance their investment. The banks with which these firms hold accounts should put a priority on their loan applications.

[4] See China Provincial SEZ laws [various issues including Guangdong, Jiangsu, Anhui, Inner Mongolia, Shandong].

[5] Source: the government website of Zhejiang Province.

Figure 1 shows the geographic distribution of the establishment of SEZs in five waves in the past three decades: (1) the 1979-1983 wave; (2) the 1984-1991 wave; (3) the 1992-1999 wave; (4) the 2000-2004 wave; and (5) the 2005-2008 wave. In the first two waves, there were few SEZs established and they were mostly located in coastal regions and provincial capital cities. After Deng Xiaoping's famous Southern Tour in 1992, there was a huge surge of zone establishment (i.e., 93 state and 466 provincial SEZs), and a multi-level and diversified pattern of opening coastal areas and integrating them with river, border, and inland areas took shape in China. From 2000, aiming at reducing regional disparity, the first comprehensive regional development plan—the Western Development Strategy program—was launched; as a result, zone establishment was concentrated in inland cities. Recently, zone establishment has exhibited a more balanced development, i.e., between 2005 and 2008, 338 SEZs were granted in the coastal area, 269 in the central area, and 75 in the western areas. For detailed descriptions on these waves, see Appendix A.

[Insert Figure 1 here]

There are several types of SEZs (Alder, Shao, and Zilibotti, 2013), in which preferential policies have different focuses. For example, Economic & Technological Development Zones (ETDZs) are broadly defined zones with a wide spectrum of investors. Industrial Development Zones (IDZs) or High-Tech Industrial Development Zones (HIDZs) are formed to encourage investment in high-tech industries by offering incentives to enterprises/investors. Export Processing Zones (EPZs) and Bonded Zones (BZs) are intended for foreign trade, with the former eliminating or streamlining most customs procedures for business and the latter reducing tariff barriers. Table 1 shows the numbers of each zone type established in the five waves. State-level SEZs are more diverse, with EPZs and BZs being the dominant types in the recent waves. Provincial zones are either ETDZs or IDZs.

[Insert Table 1 here]

# 3 Estimation Strategy

## 3.1 Identification

To identify the effects of the SEZ program on the targeted area, we use three empirical strategies—specifically, a DD method, a BD design, and a BD-DD approach.

*DD Estimation.* In the DD estimation, we use village (the most disaggregated geographic unit in the data and smaller than the SEZ) as the unit of analyses, and compare the changes of the performance among SEZ villages before and after the adoption of SEZ policies with the changes among non-SEZ villages in the same county during the same period. Specifically, the DD estimation equation is

$$y_{vct} = \gamma SEZ_{vct} + \lambda_v + \lambda_{ct} + \varepsilon_{vct}, \tag{1}$$

where $y_{vct}$ represent outcome variables such as total output, labor, capital, and number of firms of village $v$ in county $c$ at year $t$; $SEZ_{vct}$ is an indicator that equals one if village $v$ adopts SEZ policies at year $t$ and zero otherwise; $\lambda_v$ is a village fixed effect, capturing time-invariant village-level characteristics such as geographic location; $\lambda_{ct}$ is a county-year fixed effect, capturing macro shocks common to all villages of county $c$ in year $t$, and $\varepsilon_{vct}$ is the error term. To control for potential heteroskadesticity and serial correlation, we cluster the standard errors at the village level (Bertrand, Duflo, and Mullainathan, 2004).

A prominent threat to our DD estimation is that SEZs were not selected randomly, implying that SEZ villages and non-SEZ villages could be systematically different ex ante. To construct more comparable control groups, we further restrict the units to be spatially closer, choosing the non-SEZ villages that reside in the same town as the SEZ villages. However, as our main analyses use two economic censuses (see the next section for more details), we are unable to analyze the pre-trends, and thus cannot verify the common trend assumption that treatment and control groups would follow the same time trend in the absence of the treatment. To circumvent this problem, we instead conduct a placebo test. Specifically, we randomly assign the SEZ policy adoption to villages, and construct a false SEZ status indicator, $SEZ_{vct}^{false}$. Given the random data generating process, $SEZ_{vct}^{false}$ is expected to cast zero effects; otherwise, it may indicate the misspecification of equation (1). We repeat the exercise 500 times to increase the power of the placebo test.

*BD Estimation.* As an alternative estimation approach, we use the BD framework based on the physical distance, an approach pioneered by Holmes (1998) and Black (1999) and widely applied in the literature (e.g., Bayer, Ferreira, and McMillan, 2007; Dell, 2010; Duranton, Gobillon, and Overman, 2011; Gibbons, Machin and Silva, 2013). Compared with the standard regression discontinuity design, our BD design involves a discontinuity threshold, which is a zone boundary that demarcates areas (Lee and Lemieux, 2010). Specifically, we restrict our analysis to a sample of areas within a short distance of the discontinuity—the zone boundary—with the identifying assumption being that other than zone policies (the treatment of interest), all geographical characteristics are continuous across the boundary. As a result, any discontinuity in **outcomes of interest** at a zone boundary is attributable to differences in the zone status. In the benchmark analysis, we consider a geographic distance of 1,000 meters (or 5/8 mile), and experiment with two different distances (i.e., 500 meters and 2,000 meters) in the robustness checks.

The BD estimation equation is

$$y_{az} = \gamma SEZ_{az} + \lambda_z + \varepsilon_{az}, \tag{2}$$

where $y_{az}$ concerns performance of area $a$ within 1,000 meters of the boundary of zone $z$; $SEZ_{az}$ is a dummy variable indicating whether area $a$ is located inside zone $z$ or not; $\lambda_z$ is a zone fixed effect, capturing the differences across zones; and $\varepsilon_{az}$ is the error term. To get conservative statistical inference, we cluster the standard errors at the zone level. Equation (2) is estimated using the year of economic census with zone establishment.

*BD-DD Estimation.* Despite their close distance to the zone boundary, areas inside and outside zones may still be different, particularly when zone boundaries were not randomly drawn. In other words, $\hat{\gamma}_{BD}$ becomes $\gamma + \eta$, where $\eta$ includes all the location differences (except for zone policies) across the zone boundary. To improve our identification further and address the concerns over the endogeneity of boundaries, we combine the DD and BD approaches, and conduct a BD-DD analysis. Specifically, we first estimate equation (2) using the year of data without zone establishment, and obtain $\hat{\gamma}_{BD,Control} = \eta_{Control}$. Similarly, we estimate the same equation using the year of data with zone establishment, and obtain $\hat{\gamma}_{BD,Treatment} = \gamma + \eta_{Treatment}$. Assuming that the underlying location characteristics are fixed over time except for the zone policies (i.e., $\eta_{Control} = \eta_{Treatment} = \eta$), we can then recover the true SEZ effect on the targeted areas from the BD-DD estimator $\hat{\gamma}_{BD-DD}$, i.e., $\hat{\gamma}_{BD-DD} = \hat{\gamma}_{BD,Treatment} - \hat{\gamma}_{BD,Control} = \gamma$.

The BD-DD estimation equation is

$$y_{azt} = \gamma SEZ_{azt} + \lambda_a + \lambda_{zt} + \varepsilon_{azt}, \tag{3}$$

where $y_{azt}$ measures performance of area $a$ within 1,000 meters of the boundary of zone $z$ at year $t$; $SEZ_{azt}$ is an indicator that equals one if area $a$ is inside zone $z$ with zone policies adopted in year $t$, and 0 otherwise; $\lambda_a$ is an area fixed effect, capturing all time-invariant area characteristics; $\lambda_{zt}$ is a zone (neighborhood)-year fixed effect, capturing unobserved shocks common to both sides of zone $z$ in year $t$. Including zone (neighborhood)-year fixed effects allows for flexible time trends across different zones. $\varepsilon_{azt}$ is the error term. To get conservative statistical inference, we cluster the standard errors at the zone level.

As checks on the identifying assumption underpinning our BD-DD analyses (i.e., underlying location characteristics are fixed over years except for the zone policies), we conduct two placebo tests. First, we focus on a sample of areas located outside the zones, and compare them within different distances of the zone boundary. Second, we compare areas located inside the zones but within different distances of the zone boundary. As areas have the same zone status in each of these two tests, the comparison among them within different distance ranges from the zone boundary should not show significant differences; otherwise, it indicates potential estimation bias due to the existence of unobservables or spillovers.

## 3.2 Data

*Census data.* The main data used in this study come from the first and second waves of the *Economic Census*, collected by the National Bureau of Statistics of China at the end of 2004 and 2008, respectively.[6] The advantage of Census data over the *Annual Survey of Industrial Firms* (ASIF), which is widely used in the literature (e.g., Hsieh and Klenow, 2009), is that the former is more comprehensive, covering all manufacturing firms in China, while the latter includes all state-owned enterprises (SOEs) and non-SOEs with annual sales of above five million RMB. In Table A1, we compare the differences between the Census and ASIF data in 2004 and 2008. The Census data, which represent the entire population of manufacturing firms, have clearly smaller and more dispersed sales, employment, and total assets than the ASIF data.

The Census data contain firms' full basic information, such as address, location code (i.e., 12-digit corresponding to village or community), industry affiliation, and ownership. We use information on firm address and location code to geographically locate a firm and identify whether it is in the zone or not (see *Coordinates data* and *Firm SEZ status* for details). The Census data that we have access to contain three firm-level financial and operational variables, that is, employment, output, and capital.

*Coordinates data.* In BD and BD-DD analyses, we aggregate outcomes from individual firms to areas that are close to the zone boundary. As a result, we need the precise geographical information (i.e., the coordinates) on firm location to determine firms' distances from the zone boundary. To this end, we search firms' addresses and obtain firms' geographic coordinates provided by Google's Geocoding API.[7] Specifically, for firms that report detailed Chinese addresses (i.e., an address with information on province, city, road name, numbering or building name), we use the following steps to obtain firm coordinates. We first enter

[6] The third wave was started in January 2014 and is still underway.

[7] We check for the robustness of the results using Baidu's Geocoding API services. Baidu is the Chinese version of Google, which provides Geocoding API services like Google, but has different coordinate system from Google.

the detailed Chinese address of a firm (for example, "157 Nandan Road, Xuhui District, Shanghai, China") in the Google Map, and then obtain a map where a red marker shows the specific location of the address (see Figure A1). The map allows us to infer whether the marked location is correct or not, that is, whether this is the map of the searched district (i.e., Xuhui District, Shanghai, China) and whether the searched road (i.e., Nandan Road) can be found on the map. Once the virtualized location is confirmed, we then collect the latitude and longitude of the address from the Google map. By repeating these steps for each firm in the Census data, we record coordinates for approximately 50.5 percent of firms with detailed addresses.

For the remaining 49.5 percent of the firms, we are unable to geocode their addresses mainly due to incomplete addresses, road name changes, and reporting errors.[8] To deal with this issue, we make use of a 12-digit location code reported by the firm and search instead the corresponding village or community where the firm resides.[9] An example of the search procedure is as follows. For a firm without detailed address (for example, "Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China"), we rely on the 12-digit location code reported by the firm (in this example, the 12-digit location code is "320585102202"). From the 12-digit location code, we can obtain which village or community that a firm belongs to (location code "320585102202" corresponds to "Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China"). Finally, we enter the name of the village or community and collect the latitude and longitude of the village or community from the Google map (see Figure A2).

In our analyses, we use all the data. However, to address possible measurement errors, we also conduct an analysis using only the sample of firms with detailed addresses (50.5 percent of the whole sample), and find similar results (see Table A2).

*Firm SEZ status.* In the Census data, firms do not directly report information about their SEZ status. To identify whether a firm is located inside an SEZ or not, we make use of the following data sources: (1) a comprehensive SEZ data set from the Ministry of Land and Resources of China, which contains rich information on SEZ boundaries. In particular, the boundary (four directions: south, north, west, and east) is featured by villages, communities, and sometimes roads. Based on such information, we use the maps to determine the vil-

[8]Incomplete address refers to an address that only has information on village, building, or street name, but with no number or building name.

[9]There are approximately 700,000 villages and communities in China. The habitable area of China is about 2.78 million square kilometers. On average, a village and community is about 4 square kilometers. In the Census data, the average number of firms in a village and community was 5.4 in 2004 and 6.7 in 2008. The statistics indicate the precision of using coordinates of a village and community when firms do not provide detailed address information.

lages or communities within the boundary; (2) various SEZs' official websites, which report detailed information of the villages and communities within the administrative boundary; (3) administrative division codes of the People's Republic of China at the village and community level in 2008, which are published by the National Bureau of Statistics and report for some economic zones the villages and communities under their administration; and (4) the Ministry of Civil Affairs website, which documents information on administrative divisions and codes at the village and community level.

For each zone, we therefore create a list of villages and communities that fall within its boundary. We then match the zone's village and community list with the Census data based on the villages and communities reported in the firms' addresses as well as the 12-digit location code (see Appendix B for a detailed discussion). Matched firms are identified to be located in the zones, and the rest are outside the zones. To verify our approach, we cross-check our matching by using the information that some firms report the SEZ names in their addresses.[10]

*Regression data.* For our analysis, we focus on SEZs established between 2004 and 2008. There were 682 SEZs granted during this period (specifically, 19 in 2005 and 663 in 2006), and there are substantial variations across the space. There were 338 SEZs granted establishment in the coastal area, followed by 269 in the central area and 75 in the western areas.[11] In terms of their types, the 682 SEZs can be classified into three groups: 19 state-level EPZs, 280 province-level ETDZs, and 383 province-level IDZs. In the analyses, we drop state-level zones from the sample to address the concern that they are not comparable with province-level zones.[12]

For the baseline DD empirical analysis, we aggregate individual firms to construct a panel data set by village and by year. Thus, for each village we can have two possible observations

[10] It could be argued that an SEZ boundary intersects a village and community, i.e., part of the village and community is within the zone boundary, while part of it is not designated as the zone area. This is not a concern in the case of China as for ease of administration: specifically, the local governments survey and appraise land, outline plans for future development based on the unit of village and community, and determine to designate the zone areas.

[11] The coastal area includes Liaoning, Beijing, Tianjin, Hebei, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, and Hainan. The central area includes Heilongjiang, Jilin, Inner Mongolia, Shanxi, Henan, Anhui, Hubei, Hunan, and Jiangxi. The western area includes Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang, Guizhou, Yunnan, Chongqing, Sichuan, and Tibet.

[12] In particular, in 2005, 18 state-level EPZs and one BZ were approved by the central government. First, the state-level zones have higher level administrative committees than provincial level SEZs and their committees enjoy more authority in managing the zones. Second, by design these EPZs and BZ mostly reside in the pre-established ETDZs—the overlapping problems. To take the Huizhou Export Processing Zone as an example, it is located inside the Guangdong Huizhou ETDZ, which was established in 1997. Therefore, the RD-DD identification strategies are not valid for this set of zones, as the pre-existing ETDZ confound the effect of the newly approved EPZs and BZs. See Wang (2013) for more details.

in 2004 and 2008, a year of data before and a year of data after the zone establishment to conduct our DD estimation. For regression within the same county, our sample consists of 60,782 villages in 600 counties: 4,072 SEZ villages and 56,710 non-SEZ villages that reside in the same county as the SEZ villages. For regression within the same town, our sample consists of 15,014 villages in 600 counties: 4,072 SEZ villages and 10,942 non-SEZ villages that reside in the same county as the SEZ villages.

For the purpose of our BD and BD-DD exercises, we need to calculate the distance of a firm from its neighboring SEZ boundary. While we have identified coordinates of each firm's location in *Coordinates data*, we do not have the accurate geocodes of each SEZ boundary, which prevents us from calculating the distance to the boundary directly.[13] We instead follow the approach used by Duranton, Gobillon, and Overman (2011) and determine the distance indirectly. Specifically, to determine whether a firm is located within 1,000 meters of the zone boundary, we search instead within a radius of 1,000 meters of the concerned firm.[14] As illustrated in Figure 2, for example, if a firm (firm A) is located outside the zone and there is another firm (firm B) located inside the zone within its 1,000-meter range, we designate the concerned firm as located within the 1,000 meter distance of the zone boundary; otherwise, it is not. Similarly, if a firm (firm C) is located inside the zone and there is another firm (firm D) located outside the zone within its 1,000 meter distance, we designate the concerned firm as located within the 1,000 meter distance of the zone boundary.

[Insert Figure 2 here]

By repeating these steps for all firms in the census data, we have a sample of 587 SEZs with 163,069 firms located within a 1,000 meter range of the zone boundaries in both 2004 and 2008: the 2008 sample contains 126,976 firms, approximately 42.8 percent of which are located inside SEZs; the corresponding numbers for the 2004 sample are 81,739 and 40.8 percent, respectively. We then aggregate those firms to construct a panel data set by area and by year. For each zone's 1,000 meter neighborhood, we have an area inside the zone and an area outside the zone, each has two observations in 2004 and 2008. Our regression sample for estimation consists of 587 areas.

*Summary statistics.* Table 2 shows summary statistics (the means and standard deviations on employment, output, capital and number of firms of the areas) for our main

[13] In particular, the most detailed Chinese GIS data are at the town level. The unavailability of village boundary GIS data renders an accurate geocoding of the zone boundaries impossible.

[14] On average, a village and community in China is about 4 square kilometers. By assuming a village and community is a circle, we calculate that the average radius of a village and community is about 1,000 meters. Therefore, in the benchmark analysis, we use a range of 1,000 meters from zone boundaries.

regression samples. Columns 1 to 4 present at the village level the key variables for the treatment and control group underlying our DD analyses. Specifically, the first two columns report SEZ villages and non-SEZ villages within the same county while columns 3 and 4 denote SEZ villages and non-SEZ villages within the same town. Columns 5 and 6 report, for the BD-DD exercise, the main variables for areas inside and outside zones within a 1000 meter of boundaries.

[Insert Table 2 here]

Panel a illustrates the areas' characteristics in 2008, i.e., two years after SEZs were established. The first and second columns show that SEZ villages on average have more employees, higher output, larger capital stock, and more firms than non-SEZ villages. As reported in column 3 and 4, the difference between the treatment and control group are still positive, however, at a decreased magnitude. This provides support for using non-SEZ villages spatially closer with SEZ villages as the comparison group. The last two columns show that when defining the control group to be along the boundary, the zone areas on average still had more employees, higher output, and larger capital stock than its neighboring 1000 meter ring, even though there were fewer firms in the zone areas.

Panel b reports the comparison between areas' characteristics in 2004, i.e., two years before SEZs were established in some locations. We find that there were significantly differences in areas' initial characteritics. For example, SEZ villages had more employment, output, capital and firms than non-SEZ villages. However, the differences in the outcome variables between SEZ areas and non-SEZ areas in 2004 are much smaller than those in 2008.

Overall, the aggregated area level data in Table 2 suggests that different areas are not identical to start with in terms of these measures. However, after SEZs were established in some localities, there are markedly a larger increase of economic activities in treatment areas than in control areas. In the next section we conduct rigorous analyses to shed light on first, whether these descriptive results are robust to controlling for other determinants of outcomes such as time-invariant and time varying differences between the areas; second, whether they should be interpreted as causal effects of SEZs. To do so, for each estimation method we present convincing evidence in support of the underlying identifying assumptions under which the coefficients of interest could be estimated.

# 4 Empirical Findings

## 4.1 Baseline Estimates

*DD Estimates.* Table 3 presents our DD estimates. In particular, the control group in column 1 consists of non-SEZ villages that reside in the same county (i.e., 6 digit code corresponding to county) as SEZ villages. Non-SEZ villages that reside in the same town (i.e., 9 digit code corresponding to town, which is an administrative unit below county) as SEZ villages are used as the control group in column 2. We examine four village-level outcomes—total employment, total output, total capital, and total number of firms. All these outcomes are measured in logarithm form, which makes it easy to interpret the magnitude of the effects. The magnitude of coefficients in column 1 (a broad DD) is similar to that in column 2 (when we restrict the units to be closer). Consistently, we find all the estimated coefficients of four outcomes are positive and statistically significant, suggesting that after zones' establishment, SEZ villages have higher employment, output and capital, as well as more firms than non-SEZ villages.

[Insert Table 3 here]

*BD and BD-DD Estimates.* Based on the regression sample of areas located within a 1,000 meter range of the zone boundary, Table 4 shows the coefficients of the impact of the SEZ program. Column 1 and 2 report BD and BD-DD estimates respectively. We find that for total employment, total output and total capital, all the estimated coefficients are positive and statistically significant (except for the BD estimate of total employment). All the BD-DD estimates are consistently larger and more significant than the corresponding BD estimates, pointing to the possibility of non-randomized zone boundaries. Meanwhile, we find that the BD estimate of the number of firms is negative and significant, while the BD-DD estimate is positive and significant. The results suggest that zones were established in places with initially a smaller number of firms but attracted more firms in two years.

[Insert Table 4 here]

*Economic Magnitude.* To gauge the economic magnitude of the SEZ effects, we rely on the BD-DD estimates.[15] We show that, after two years of the establishment of the zones, SEZs increase employment by 47.1 percent, output by 55.3 percent, capital by 54.7 percent,

[15] Note that DD estimates have smaller magnitude than BD-DD estimates. One possible explanation for this difference is the former calculates the average treatment effect while the latter estimates the local average treatment effect. Another possible explanation is that the two methods use different control groups—specifically, the DD analysis uses non-SEZ villages in the same county/town as the control group while the BD-DD analysis focuses on the areas located just opposite the zone boundary as the control group.

and the number of firms by 23.3 percent. These findings are largely consistent with those in the literature. For example, in studying the French Zones Franches Urbaines (ZFU), Givord, Rathelot, and Sillard (2013) find that the ZFU program increases the number of establishments via births and relocations. Criscuolo, Martin, Overman and Van Reenen (2012) find a large and statistically significant average effect of UK's RSA program on the treated for employment and investment.

## 4.2 Robustness Checks

In this subsection, we provide a battery of robustness checks on the aforementioned findings—specifically, a placebo test for the conventional DD analysis, sensitivity tests using different distance ranges from zone boundaries for the BD and BD-DD exercises, and two placebo tests for the BD-DD analysis.

*Randomly assigned SEZ Status in the DD analysis.* To check whether our DD estimation is misspecified, we randomly assign the villages in the 2008 data to be SEZ and non-SEZ units. Figure 3 shows the distribution of estimates from the 500 times of randomization. We find that for both control groups and four major outcomes, the distribution of the estimates from random assignments is centered around zero and statistically indifferent from zero. In addition, our benchmark estimates from Table 3, represented by vertical dash lines, clearly lie outside the estimates from the placebo tests. Taken together, these results imply that there are no substantial omitted variables in our DD specification.

[Insert Figure 3 here]

*Different distance ranges from zone boundaries.* In the BD and BD-DD benchmark analyses, we focus on a sample of areas located within a 1,000 meter distance of zone boundaries. To check whether our results are sensitive to the distance used, we conduct a standard check on the bandwidth selection in the BD analysis. In particular, we experiment with two alternative distances, specifically, within a 500 meter and a 2,000 meter distance of the boundaries. The estimation results are reported in columns 1-4 of Table 5. The magnitudes of the estimates from two alternative samples are similar to those from the benchmark sample, suggesting that our estimates are not biased due to the choice of a specific distance.

[Insert Table 5 here]

*Both outsize-zone or inside-zone areas in the BD-DD analysis.* As a further check on the identifying assumption for the BD-DD analysis, we perform two separate exercises. We start by comparing the performance of areas both located outside the zones: an area within a 1,000 meter distance of the boundaries versus an area within a 1,000 to 2,000 meter distance of the boundaries.[16] Given the fact that neither group enjoys zone policies, any substantial differences in their performances would indicate a misspecification of our BD-DD estimation. The estimation results are reported in column 5 of Table 5.

Similarly, we conduct another comparison of performance for areas both inside the zones: one within a 1,000 meter distance of the boundaries and the other within a 1,000 to 2,000 meter distance of the boundaries. The estimation results are reported in columns 6 of Table 5. Almost all estimates show no statistical significance, and the small magnitudes further suggest that the estimated effects from these two robustness exercises are close to zero. Combined with our benchmark estimates, we show that from 2,000 meters outside the zone boundaries to 2,000 meters inside the zone boundaries, the discontinuity in our outcomes is detected just at the zone boundaries. These results, therefore, provide support for the validity of our BD-DD identifying strategy.

Overall, we show through a wide range of placebo tests that our results are robust to several potential threats to our identifying assumption.

## 4.3 Mechanism

In the previous sections, we have established that SEZ areas have more employment, output, capital and number of firms than non-SEZ areas. China's SEZ program changes the capital and land costs and tax rates in some locations, which have profound influences on a firm's location choice and then its investment decisions. Specifically, facing the policy shocks, first, firms in the zone can respond along the intensive margin by varying decisions including inputs and outputs. Second, firms may respond along the extensive margin. That is, they can decide whether or not to enter the zone, to relocate to the zone, to exit the zone. The composition of active firms in the zone area is given by the sum of three components: (1) entrants: firms choosing to begin production in the zone; (2) relocated firms: firms that were operating in other locations and choose to relocate to the zone, surviving the shutdown shock; (3) incumbent firms: firms that were operating in the zone in the previous period, choose to continue production in the zone, surviving the shutdown shock. After the zones are established, what are the changes of the composition of firms located inside and outside the zones? To shed light on the underlying mechanisms, we decompose the SEZ effects into three parts: (1) new entrants and exiters, i.e., extensive margin effects from net entry; (2) firms switching from outside zones to inside zones or reverse, i.e., extensive margin effects

[16] For the second group, if a zone has a breadth less than 2,000 meters, we then use all the firms located more than 1,000 meters from the zone boundary within the zone.

from firm relocation; and (3) firms without changes of their zone status, i.e., intensive margin effects.[17]

BD-DD estimations are more rigorous in the identifying assumption than DD and BD estimations, which means the BD-DD estimations lead to more conservative and typically more credible inferences.[18] Our following analyses therefore use the BD-DD estimates. Specifically, our BD-DD estimator $\hat{\gamma}_{BD-DD}$ is

$$\begin{aligned}
\hat{\gamma}_{BD-DD} &= \frac{\partial \ln Y}{\partial SEZ} = \frac{\partial \ln \left[Y^{entry/exit} + Y^{inc} + Y^{re}\right]}{\partial SEZ} \\
&= \frac{Y^{entry/exit}}{Y}\frac{\partial \ln Y^{entry/exit}}{\partial SEZ} + \frac{Y^{inc}}{Y}\frac{\partial \ln Y^{inc}}{\partial SEZ} + \frac{Y^{re}}{Y}\frac{\partial \ln Y^{re}}{\partial SEZ} \\
&= \omega^{entry/exit}\hat{\gamma}^{entry/exit} + \omega^{inc}\hat{\gamma}^{inc} + \omega^{re}\hat{\gamma}^{re} \qquad (4)
\end{aligned}$$

where $Y$ represent area outcomes (i.e., total employment, total output, total capital, and total number of firms); $Y^{entry/exit}$, $Y^{inc}$, and $Y^{re}$ are the corresponding outcomes for the sample of entrants and exiters, incumbents, and relocated firms, respectively; $\hat{\gamma}^j \equiv \frac{\partial \ln Y^j}{\partial SEZ}$ are the BD-DD estimates from sample $j \in \{entry/exit, inc, re\}$; and $\omega^j \equiv \frac{Y^j}{Y}$ are the weights for sample $j$.

The decomposition analyses have demonstrated the necessity to distinguish the incumbents, relocated firms, and exiters and entrants. In other words, we need to trace three groups of firms from 2004 to 2008, respectively. To this end, we use the following steps. First, for firms that report unique IDs (i.e., their legal person codes) in the Census data, we use firm ID to match firms between the 2004 and 2008 Censuses. Second, for firms that have duplicate IDs, we use the firm name to link observations over time. Third, firms may receive a new ID as a result of restructuring, mergers, or privatization. For a firm for which no observation with the same ID can be matched over time, we use as much information as possible on firm name, location code, name of legal person representative, phone number and so on to find a match. Table A3 reports the number of new entrants (i.e., firms that exist in 2008, but do not exist in 2004), survivors (i.e., firms that exist in both 2004 and 2008), and exiters (i.e., firms that exist in 2004, but do not exist in 2008). Among 794,386 survivors, 92.7 percent of the firms are linked using firm ID, 4.7 percent using firm name, and 2.7 percent using other information. Finally, we classify survivors into relocated firms and incumbents based on their zone location changes over time—specifically, firms are designated as relocated firms if their coordinates changed from inside (outside) the zone in 2004

[17] For firms with location changes but without switching zone status (e.g., from outside to outside), we consider them as incumbents.

[18] We provide the decomposition using the DD and BD methods in Table A4, and find qualitatively similar results.

to outside (inside) the zone in 2008.

BD-DD estimates for three different samples are reported in Table A5, and the decomposition analyses are conducted in Table 6. We find that a majority of the SEZ effects come from the entry and exit margin—specifically, it accounts for 66.31 percent of the SEZ effect on total employment, 59.08 percent of the effect on total output, 61.38 percent of the effect on total capital, and 93.66 percent of the effect on the total number of firms. There is also a sizable effect at the intensive margin—that is, incumbent firms show significant improvement in their performance (i.e., becoming larger in terms of employment, output and capital), and this intensive margin accounts for 21.09–28.07 percent of the overall SEZ effects. Despite the large SEZ effects in the relocated firms sample, the contributions by this particular margin to the overall SEZ effects are similar to those by the intensive margin, presumably due to the small share of relocated firms in the total sample.

[Insert Table 6 here]

Overall, our decomposition indicates that the zones have a large and significant effect along both the extensive margin and intensive margin. We compare our findings to the literature and note that they are largely comparable to Criscuolo, Martin, Overman and Van Reenen (2012), which finds a large and statistically significant average effect of UK's RSA program on the treated for employment and investment with about half of the effects arising from incumbent firms growing (the intensive margin) and half due to greater net entry (the extensive margin). Our findings are also consistent with other works in the literature, however, the noteworthy difference is that we uncover more optimistic effects on incumbents. For example, Givord, Rathelot, and Sillard (2013) in the context of ZFU programs find that there is no evidence indicating an employment effect on existing businesses; employment growth mostly comes from the new businesses and firms relocated into the ZFUs. Meanwhile, reviewing the literature on U.S. enterprise zones and empowerment zones, Neumark and Simpson (2014) conclude that the evidence on generating employment is overall pessimistic.[19]

## 4.4 Heterogeneous Results

In this subsection, we investigate heterogeneous treatment effects by taking into account industry and zone characteristics. Because of the reduced capital costs, firms in capital-intensive sectors may be more likely to benefit from the zone program and thus shall exhibit larger effects. Firms produce goods and trade with various markets. The level of economic

[19] According to the review, only a few papers detect positive program effects, such as Busso et al. (2013), Freedman (2013), and Ham et al. (2011).

activity in a location is conditioned by that location's access to markets for its goods (Hanson, 2005). Productive amenities such as airports and **highways** also help reduce firms' trade and communication costs (Graham, Gibbons, and Martin, 2010; Combes and Gobillon, 2015). As a result, proximity to markets and infrastructure makes a zone more attractive for firms. We expect zones with larger market potential and transportation accessibility (local determinants of agglomeration) shall exhibit larger effects. Specifically, exploiting variations in capital-labor ratios at 4-digit industry level and transportation accessibility at the zone level, we compare the effects between capital-intensive and labor-intensive industries, the effects between spatially-integrated and spatially-isolated zones, and the effects between zones with high and low market potential.

*Capital-intensive vs. labor-intensive industries.* To investigate whether there are differential effects of SEZs between capital-intensive and labor-intensive industries, we divide industries into two categories, based on whether their capital-labor ratios in 2004 were above or below the sample median. The estimation results are reported in Table 7, with panel (a) for the capital-intensive industries, panel (b) for the labor-intensive industries, and panel (c) for the differences between these two groups.

[Insert Table 7 here]

We find that the SEZ effects are consistently stronger in capital-intensive industries than in labor-intensive industries. Specifically, in absolute terms, the employment effect of the SEZs in capital-intensive industries is larger than that in labor intensive industries by 10 percentage point, the output effect is larger by 20.9 percentage point, the capital effect is larger by 15.5 percentage point, and the number of firm effect is larger by 7.5 percentage point. These results are consistent with the features of the SEZ program—that is, the SEZ program largely provides subsidies on capital, which implies its effects should be magnified by the intensity of capital usage in production.

*Zones with high vs. low market potential.* To investigate whether there are differential SEZ effects across zones with different market potential, we first construct a market potential index for each zone in the spirit of Harris (1954), Rogers (1997). Specifically, we assume that the impact of trade and communication costs increase with the inverse of distance of zone $z$ to all prefecture-level cities within the province. The market potential $MP_z$ is defined as:

$$MP_z = \frac{\sum_{c \in PROV} GDP_c / dist_{zc}}{\sum_{c \in PROV} GDP_c},$$

where $PROV$ denotes province, $c$ denotes prefecture-level city, $GDP_c$ stands for city $c$'s

GDP, and $dist_{zc}$ is the distance between the zone's administrative committee $z$ and city $c$.[20] Following Briant el al. (2015), we divide the weighted summation of markets accessible from the zone by the total sum of market sizes in the province to mitigate the impact of large cities. Based on whether their market potential values in 2004 were above or below the sample median, we then classify the zones into two groups.

The estimation results are shown in Table 8, with panel (a) for zones with high market potential, panel (b) for zones with low market potential and panel (c) for the differences between these two groups.

[Insert Table 8 here]

We do not find statistically and economically significant differences in the SEZ effects between zones with high market potential and those with low market potential. These results imply that local determinants of agglomeration economies such as market potential do not play important roles in enhancing the program effectiveness in the context of China.

*Spatially-integrated vs. spatially-isolated zones.* To investigate whether there are differential SEZ effects across zones with different accessibility to transport infrastructure, we first construct an index for each zone, with a larger value indicating closer distance to an airport and highway density. In particular, as shown in Figure 4, we compute the distance of each zone's administrative committee to its nearest airport and rank the distance from largest to smallest ($rank\_airport$). Similarly, we rank the highway density of the city that the zone resides in. The zone's infrastructure accessibility index is constructed as $rank = (rank\_airport + rank\_highway)/2$, with a lower index indicating the zone is further away from transportation infrastructure. The full list of airports is compiled from China transportation yearbook 2005, while the data on highway density (mileage highways divided by land Area of the city) is from China regional statistical yearbook 2005.

[Insert Figure 4 here]

We then divide zones into two groups, based on whether their accessibility indices in 2004 were above or below the sample median. The estimation results are reported in Table 9, with panel (a) for spatially-integrated zones, panel (b) for spatially-isolated zones, and panel (c) for the differences between these two groups. We also do not find statistically and economically significant differences in the SEZ effects across zones with different infrastructure

[20] Note that Harris (1954) and Rogers (1997) use city as a regression unit, and the market potential of a city is the weighted average GDP from other cities. In our setting, economic zones are smaller units than cities. We therefore also include the city where economic zones reside into the calculation of market potential.

accessibility

[Insert Table 9 here]

Taken together, we find that capital-intensive sectors benefit more from the zone programs compared with labor-intensive sectors, but the effects of SEZs are quite similar across zones with different market potential and different infrastructure accessibility. Our results resonate with previous work that put emphasis on the characteristics of the industry in analyzing heterogenous effects of place-based policies (Combes and Gobillon, 2015), but are contrast with findings in the literature on the role of regional characteristics, for example, the study of the French ZFU program (Briant, Lafourcade, and Schmutz, 2015). These findings suggest that the complementary role of regional and industry characteristics to place-based development programs may hinge on the research context.

# 5 Conclusion

In this paper, we exploit a natural experiment in the establishment of China's economic zones. Using firms as the unit of analysis, we elucidate the mechanisms underlying the zone effects and unfold the determinants of program effectiveness. By focusing on a prominent place-based policy in China, we address several important questions: whether the zone works; for whom it works; and for designing more efficient programs, what works and where it works (Neumark and Simpson, 2014).

Unpacking the distributional effects of the zone program, we show that the program have demonstrated a large effect on the targeted areas along extensive margins, especially entry and exit, while relocations play a minor role in the program effect. The incumbent firms have experienced a sizable improvement in their performance (intensive margins). The findings may help to diffuse the more general pessimism toward zone programs in developing countries.

Another important finding is that the zone program's effectiveness depends crucially on the design of the policies. China's economic zone program offers various subsidies for capital investment. The resulting zone effects are significantly larger for firms in capital-intensive sectors than firms in labor-intensive sectors. Meanwhile, location characteristics such as market potential and transportation accessibility seem not to be critical factors in enhancing the program effects. Overall, our analysis serves as a reminder that, to make an effective policy, one has to pay close attention to the circumstances of the heterogeneity of agents influenced.

We, however, caution our study as a first step towards understanding the micro-foundations of place-based policies in developing countries. Much remains to be done. While our paper focuses on evaluating short-term effects of the zones (two years after zones' establishment), further efforts can be extended to investigating whether similar results can be found in the long run. Our empirical analysis of micro-level impacts of zones has not really engaged with aggregate productivity and welfare implications. To make further progress on this issue would require developing a general equilibrium model. It would be also interesting, in particular, to uncover the links from local institutions (political, economic, and social) to the effects of zones (Becker, Egger, and Ehrlich, 2013).[21] Such analyses will undoubtedly be of great benefit in addressing how SEZ policy interventions should be implemented in specific contexts.

[21] Becker, Egger, and Ehrlich (2013) investigate the heterogeneity across regions of EU member states in their ability, e.g., to utilize transfers from the European Commission. Only those regions with sufficient human capital and good-enough institutions are able to turn transfers into faster per capita income growth and per-capita investment.

# References


[1] Alder, Simon, Lin Shao, and Fabrizio Zilibotti. 2013. "Economic Reforms and Industrial Policy in a Panel of Chinese Cities." CEPR Discussion Paper 9748.

[2] Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588-638.

[3] Becker, Sascha O., Peter H. Egger, and Maximilian von Ehrlich. 2013. "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects." *American Economic Journal: Economic Policy*, 5(4): 29-77.

[4] Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economic*s, 114(2): 577-599.

[5] Briant, Anthony, Miren Lafourcade, and Benoit Schmutz. 2015. "Can Tax Breaks Beat Geography? Lessons from the French Enterprise Zone Experience." *American Economic Journal: Economic Policy*, 7(2): 88–124.

[6] Brinkman, Jeff, Coen-Pirani Daniele, and Holger Sieg. 2015. "Firm Dynamics in an Urban Economy." *International Economic Review*, forthcoming.

[7] Brülhart, Marius, Mario Jametti, and Kurt Schmidheiny. 2012. "Do Agglomeration Economies Reduce the Sensitivity of Firm Location to Tax Differentials?" *Economic Journal*, 122(563): 1069-1093.

[8] Busso, Matias, Jesse Gregory, and Patrick Kline. 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review*, 103(2): 897-947.

[9] Cheng, Yiwen. 2014. "Place-Based Policies in a Development Context – Evidence from China." UC Berkeley, mimeo.

[10] Combes, Pierre-Philippe and Laurent Gobillon. 2015 The Empirics of Agglomeration Economies. In Handbook of Urban and Regional Economics, vol. 5, G. Duranton, V. Henderson and W. Strange (eds.), Elsevier-North Holland, Amsterdam.

[11] Criscuolo, Chiara, Ralf Martin, Henry Overman, and John Van Reenen. 2012. "The Causal Effects of an Industrial Policy." NBER Working Papers 17842.

[12] Dell, M. 2010. "The Persistent Effects of Peru's Mining Mita." Econometrica, 78(6):1863-1903.

[13] Devereux, Michael P., Rachel Griffith, and Helen Simpson. 2007. "Firm Location Decisions, Regional Grants and Agglomeration Externalities." Journal of Public Economics, 91(3–4): 413-435.

[14] Duranton, Gilles, Laurent Gobillon, and Henry G. Overman. 2011. "Assessing the Effects of Local Taxation using Microgeographic Data." Economic Journal, 121(555): 1017-1046.

[15] Freedman, Matthew. 2013. "Targeted Business Incentives and Local Labor Markets." Journal of Human Resources, 48(2): 311-344.

[16] Gaubert, Cecile. 2014. "Firm Sorting and Agglomeration." UC Berkeley, working paper.

[17] Gibbons, S., S. Machin and O. Silva. 2013. "Valuing School Quality Using Boundary Discontinuities." Journal of Urban Economics, 75(May): 15-28.

[18] Givord, Pauline, Roland Rathelot, and Patrick Sillard. 2013. "Place-Based Tax Exemptions and Displacement Effects: An Evaluation of the Zones Franches Urbaines Program." Regional Science and Urban Economics, 43(1): 151-163.

[19] Graham, Daniel J., Stephen Gibbons, and Ralf Martin. 2010. "The Spatial Decay of Agglomeration Economies: Estimates for Use in Transport Appraisal." Report for Department of Transport.

[20] Glaeser, Edward L., and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policies." Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, 39(1): 155-253.

[21] Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange. 2010. "Urban Economics and Entrepreneurship." In: Edward L. Glaeser, Stuart S. Rosenthal and William C. Strange (eds.), Cities and Entrepreneurship: Journal of Urban Economics, 67(1): 1-14.

[22] Ham, John C., Charles Swenson, Ayse Imrohoroglu, and Heonjae Song. 2011. "Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Communities." Journal of Public Economics, 95(7–8): 779-797.

[23] Hanson, Andrew, and Shawn M. Rohlin. 2013. "Do Spatially Targeted Redevelopment Programs Spill-Over?" Regional Science and Urban Economics, 43(1): 86-100.

[24] Hanson, Gordon H. 2005. "Market Potential, Increasing Returns, and Geographic Concentration." Journal of International Economics, v67(1), 1-24.

[25] Harris, Chauncy D. 1954. "The Market as a Factor in the Localization of Industry in the United States." Annals of the Association of American Geographers, 44(4): 315-348.

[26] Holmes, Thomas J.. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." Journal of Political Economy, 106 (4), 667-705.

[27] Hsieh, Chang-Tai, and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." Quarterly Journal of Economics, 124(4): 1403-1448.

[28] Keele, Luke J., and Rocio Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." Political Analysis, 23(1): 127-155.

[29] Kline, Patrick. 2010. "Place Based Policies, Heterogeneity, and Agglomeration." American Economic Review, 100(2): 383-387.

[30] Kline, Patrick, and Enrico Moretti. 2014a. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Policies." Annual Review of Economics, 6: 629-662.

[31] Kline, Patrick, and Enrico Moretti. 2014b. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." Quarterly Journal of Economics, 129(1): 275-331.

[32] Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." Journal of Economic Literature, 48(2): 281-355.

[33] Lucas Jr., Robert E., and Esteban Rossi-Hansberg. 2002. "On the Internal Structure of Cities." Econometrica, 70(4): 1445-1476.

[34] 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.

[35] Neumark, David, and Helen Simpson. 2014. "Place-Based Policies." NBER Working Papers 20049.

[36] Rogers, Cynthia L. 1997. "Job Search and Unemployment Duration: Implications for the Spatial Mismatch Hypothesis." Journal of Urban Economics, 42(1): 109-132.

[37] Schminke, Annette, and Johannes Van Biesebroeck. 2013. "Using export market performance to evaluate regional preferential policies in China." Review of World Economics, 149(2), pp 343-367.

[38] Wang, Jin. 2013. "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities." *Journal of Development Economics*, 101: 133-147.

[39] World Bank. 2008. Special Economic Zones: Performance, Lessons Learned, and Implications for Zone Development.

# Appendix

## Appendix A: Five Waves of Economic Zones

The zone granting waves are as follows (Figure 1):

1979-1983: In the late 1970s, China's State Council approved small-scale SEZ experiments in four remote southern cities: Shenzhen, Zhuhai, and Shantou in Guangdong Province, as well as Xiamen in Fujian Province. China started with virtually zero foreign direct investment and almost negligible foreign trade before 1978, so those zones were considered a test base for liberalization of trade, tax, and other policies nationwide.

1984-1991: Supported by the initial achievements of the first group of SEZs, the central government expanded the SEZ experiment in 1984. Fourteen other coastal cities were opened to foreign investment. From 1985 to 1988, the central government included even more coastal municipalities in the SEZ experiment. In 1990, the Pudong New Zone in Shanghai joined the experiment along with other cities in the Yangtze River valley. An important pattern of this economic zone granting wave is that cities with better geographical location, industrial conditions, and human capital were selected. Forty-six state-level development zones and 20 province-level development zones were established from 1984 to 1991.

1992-1999: Following Deng Xiaoping's famous Southern Tour in 1992, which sent a strong signal for continuous reform and economic liberalization, the State Council of China opened a number of border cities and all the capital cities of the inland provinces and autonomous regions. This period had witnessed a huge surge of development zone establishment. Ninety-three state-level development zones and 466 province-level development zones were created within municipalities to provide better infrastructure and achieve agglomeration of economic activities. As a result, a multi-level and diversified pattern of opening coastal areas and integrating them with river, border, and inland areas took shape in China.

2000-2004: From 2000, aiming at reducing regional disparity, the State Council launched the Western Development Strategy, China's first comprehensive regional development plan to boost the economies of western provinces. The success of coastal development zones sheds light on the effectiveness of the program in attracting investment and boosting employment. As a result, more development zones were granted by the central authority and the provincial governments in inland cities. China's entry into the WTO in 2001 led to an increasing number of state-level Export Processing Zones and Bonded Zones. In total, 64 state-level development zones and 197 province-level development zones were established between 2000 and 2004.

2005-2008: From 2005, 682 SEZs were established. In terms of the geographical distribution, 338 SEZs were granted in coastal areas, followed by 269 in central areas, and 75 in western areas. In terms of granting authority, there are 19 state-level development zones and 663 province-level development zones.

## Appendix B: Identifying Each 12-Digit Location Code within a Zone's Boundaries

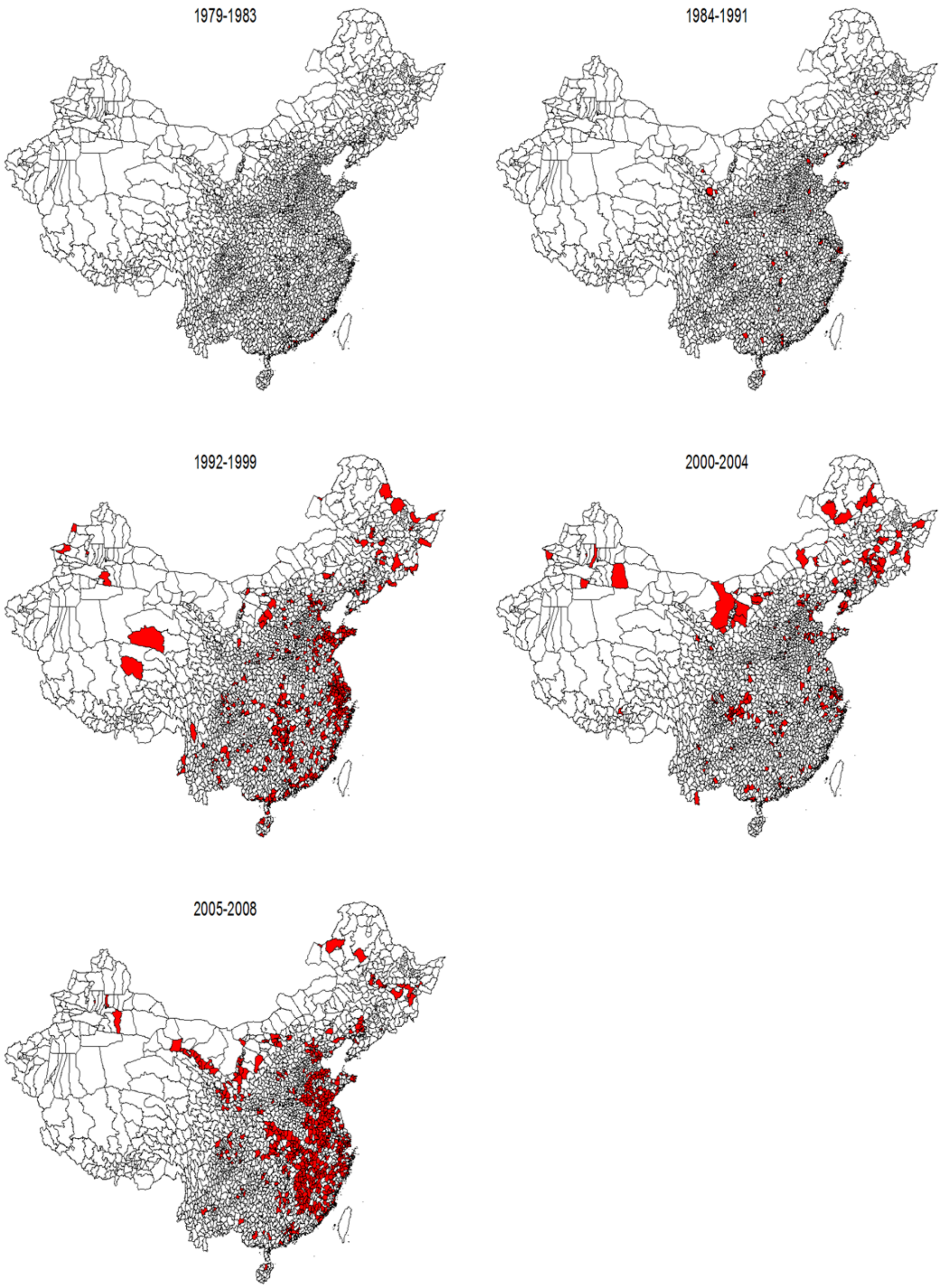
We rely on the firm's administrative location code to identify whether a firm is located inside the SEZ or not. We summarize three cases that we use to find the administrative location code within the zone's boundaries. For each case, we pick a zone to illustrate how we obtain the administrative location code.

1. A zone with an independent zone admin code. For example, Anhui Nanling Industrial Zone (zone code: S347063) has an independent 12-digit administrative location code: 340223100400 (Anhui Nanling Industrial Zone Community).

2. A zone is equivalent to a town/street, i.e., all villages/communities under the town/street will be within the zone's boundaries. For example, Shandong Fei County Industrial Zone (zone code: S377099) administrates Tanxin Town (administrative location code: 371325105). The 9-digit town code is enough to pin down the zone area.

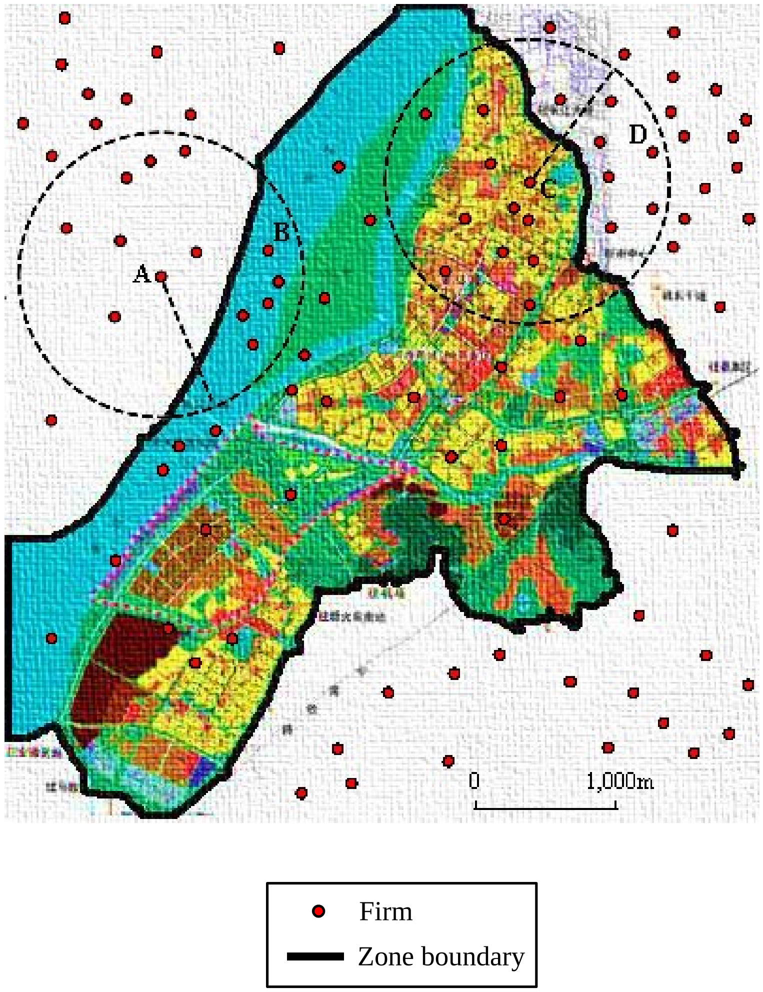
3. A zone resides in several villages and communities. For example, Hubei Yunmeng Economic Development Zone (zone code: S427040) administrates the following eight villages and one community: Xinli Village (administrative location code: 420923100201), Heping Village (420923100202), Qianhu Village (420923100203), Hebian Village (420923100204), Zhanqiao Village (420923100205), Quhu Village (420923103220), Zhaoxu Village (420923103223), Sihe Village (420923104209), and Qunli Community (420923100007).

Figure 1 Special Economic Zone Granting Waves



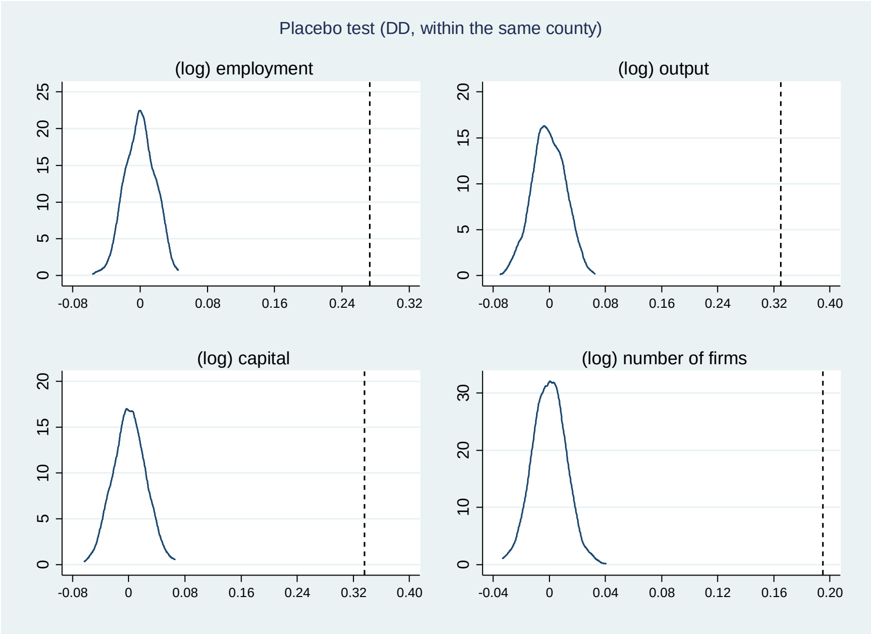

*Notes:* There are five granting waves of SEZs: first, 1979-1983; second, 1984-1991; third, 1992-1999; fourth, 2000-2004; and fifth, 2005-2008. In each wave, counties where SEZs were newly established are denoted by regions shaped in red.

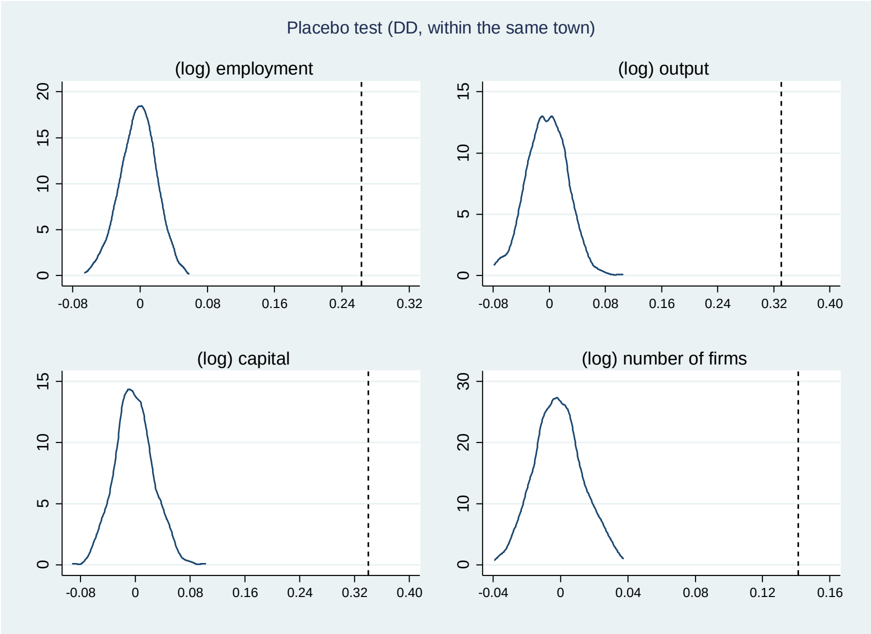
Figure 2 Firms near the Zone Boundary




*Notes*: We describe how we determine whether a firm is located within 1,000 meters of the zone boundary. As illustrated in the figure, if a firm (firm A) is located outside the zone and within its 1,000-meter range, there is another firm (firm B) located inside the zone, the concerned firm (firm A) is designated as located within the 1,000 meter distance of the zone boundary. If a firm (firm C) is located inside the zone and within its 1,000 meter distance, there is another firm (firm D) located outside the zone, the concerned firm (firm C) is designated as located within the 1,000 meter distance of the zone boundary.

Figure 3 Placebo Test




*Notes*: We randomly assign the SEZ policy adoption to villages, and construct a false SEZ status indicator. We conduct the estimation using Equation (1) and repeat the exercise 500 times to increase the power of the placebo test. The figure shows the distribution of estimates from the 500 times of randomization.

Figure 4 Geography: An Example of Tianjin Wuqing Economic Zone


*Notes*: To serve as an example, we illustrate how to measure a zone –Tianjin Wuqing ETDZ–'s geographic conditions: the distance of the zone's administrative committee to its nearest airport (Tianjin Binhai International Airport).

Figure A1 Searching for A Detailed Address on Google Maps


*Notes:* The example illustrates how we obtain coordinates for firms that report detailed Chinese addresses. We first enter the detailed Chinese address of a firm (for example, “157 Nandan Road, Xuhui District, Shanghai, China”) in the Google Map, and then obtain a map where a red marker shows the specific location of the address. Once the virtualized location is confirmed, we then collect the latitude and longitude of the address from the Google map.

Figure A2 Searching for Villages and Communities on Google Maps


*Notes:* The example illustrates how we obtain the coordinates for villages and communities. We first enter the name of the village or community, followed by the name of the town, city, and province that the village or community belongs to (for example, "Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China") in the Google Map. The specific location of the village or community is then denoted by a red marker. Once the virtualized location is confirmed, we then collect the latitude and longitude of the village or community from the Google map.

Table 1 SEZ Program Granting Waves

| Granting Period | 1979-1983 | 1984-1991 | 1992-1999 | 2000-2004 | 2005-2008 |
|---|---|---|---|---|---|
| Number of Zones Newly established | 4 | 66 | 559 | 261 | 682 |
| # Comprehensive SEZs | 4 | | | | |
| # State-level Economic Zones in which: | | 46 | 93 | 64 | 19 |
| By Types | | | | | |
| 1. Economic and Technological Development Zone | | 12 | 23 | 17 | |
| 2. High-tech and Industrial Development Zone | | 26 | 27 | | |
| 3. Export Processing Zone | | | 1 | 39 | 18 |
| 4. Bonded Zones | | 4 | 11 | 6 | 1 |
| 5. Border Economic Cooperation Zones | | | 15 | 1 | |
| 6. Other | | 4 | 16 | 1 | |
| By Regions | | | | | |
| 1. Coastal Region | | 36 | 60 | 39 | 15 |
| 2. Central Region | | 6 | 18 | 12 | 2 |
| 3. Western Region | | 4 | 15 | 13 | 2 |
| # Province-level Economic Zones in which: | | 20 | 466 | 197 | 663 |
| By Types | | | | | |
| 1. Economic and Technological Development Zone | | 16 | 401 | 112 | 279 |
| 2. Industrial Development Zone | | 4 | 65 | 85 | 384 |
| in which: High-tech and Industrial Development Zone | | 3 | 29 | 14 | 19 |
| By Regions | | | | | |
| 1. Coastal Region | | 7 | 277 | 76 | 323 |
| 2. Central Region | | 7 | 138 | 71 | 267 |
| 3. Western Region | | 6 | 51 | 50 | 73 |

*Note* : Five granting waves of Economic Zones are 1979-1983, 1984-1991, 1992-1999, 2000-2004 and 2005-2008. For each period, the number of development zones newly established is provided, in which comprehensive SEZs, state-level development zones and province-level economic zones are distinguished. Comprehensive SEZs refer to four economic zones established in Shenzhen, Xiamen, Zhuhai and Xiamen. State-level development zones are granted by the central government. Province-level development zones are granted by the provincial governments. Coastal region includes Liaoning, Beijing, Tianjin, Hebei, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, and Hainan. Central region includes Heilongjiang, Jilin, Inner Mongolia, Shanxi, Henan, Anhui, Hubei, Hunan, and Jiangxi. Western region includes Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang, Guizhou, Yunnan, Chongqing, Sichuan, and Tibet.

Table 2 Summary Statistics

| | Within the same county | | Within the same town | | Within 1,000 meters | |
|---|---|---|---|---|---|---|
| | Inside the zone | Outside the zone | Inside the zone | Outside the zone | Inside the zone | Outside the zone |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | | Panel (a). Year 2008 | | | |
| (log) employment | 6.02 | 4.76 | 6.02 | 5.11 | 8.11 | 8.03 |
| | (1.63) | (1.63) | (1.66) | (1.66) | (1.46) | (1.46) |
| (log) output | 11.53 | 9.95 | 11.53 | 10.35 | 13.91 | 13.65 |
| | (2.11) | (2.11) | (2.13) | (2.13) | (1.89) | (1.89) |
| (log) capital | 11.24 | 9.49 | 11.24 | 9.98 | 13.67 | 13.37 |
| | (2.13) | (2.13) | (2.14) | (2.14) | (1.88) | (1.88) |
| (log) number of firms | 2.19 | 1.39 | 2.19 | 1.71 | 3.86 | 4.21 |
| | (1.14) | (1.14) | (1.20) | (1.20) | (1.21) | (1.21) |
| | | | Panel (b). Year 2004 | | | |
| (log) employment | 5.59 | 4.61 | 5.59 | 4.95 | 7.47 | 7.86 |
| | (1.65) | (1.65) | (1.67) | (1.67) | (1.45) | (1.45) |
| (log) output | 10.36 | 9.09 | 10.36 | 9.50 | 12.42 | 12.71 |
| | (2.13) | (2.13) | (2.17) | (2.17) | (1.89) | (1.89) |
| (log) capital | 10.23 | 8.77 | 10.23 | 9.31 | 12.43 | 12.67 |
| | (2.19) | (2.19) | (2.20) | (2.20) | (1.94) | (1.94) |
| (log) number of firms | 1.73 | 1.16 | 1.73 | 1.42 | 3.28 | 3.87 |
| | (1.07) | (1.07) | (1.11) | (1.11) | (1.17) | (1.17) |

*Note* : Means and standard errors on employment, output, capital, and number of firms are given in parentheses. All variables are defined at the area-year level. Panel a and b report the areas' characteristics in 2004 and 2008 respectively. Columns 1-2 report SEZ villages and non-SEZ villages within the same county. Columns 3-4 illustrate SEZ villages and non-SEZ villages within the same town. Columns 5-6 report the main variables for areas inside and outside zones within a 1,000 meter of boundaries.

Table 3 The SEZ Effects: DD Estimation

| | Within the same county | Within the same town |
|---|---|---|
| | (1) | (2) |
| Panel (a). Dependent variable: (log) employment | | |
| *InsideZone* | 0.273*** | 0.263*** |
| | (0.020) | (0.025) |
| Observations | 121,564 | 30,028 |
| Panel (b). Dependent variable: (log) output | | |
| *InsideZone* | 0.330*** | 0.331*** |
| | (0.026) | (0.032) |
| Observations | 121,564 | 30,028 |
| Panel (c). Dependent variable: (log) capital | | |
| *InsideZone* | 0.336*** | 0.340*** |
| | (0.025) | (0.031) |
| Observations | 121,564 | 30,028 |
| Panel (d). Dependent variable: (log) number of firms | | |
| *InsideZone* | 0.195*** | 0.141*** |
| | (0.015) | (0.018) |
| Observations | 121,564 | 30,028 |

*Note:* Observations are at the village-year level. *InsideZone* is an indicator variable for whether the village is inside the zone or not. Panels a-d report the estimation results on the natural log of the measure of employment, output, capital, and number of firms respectively. In column 1, non-SEZ villages within the same county as the SEZ villages are used as the control group. In column 2, non-SEZ villages within the same town are used as the control group for SEZ villages. Village fixed effects and county-year fixed effects are included in all specifications. Standard errors in parentheses. Standard errors are clustered at the village level. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table 4 The SEZ Effects: BD and BD-DD Estimation

| Within 1,000 meters | BD | BD-DD |
|---|---|---|
| | (1) | (2) |
| Panel (a). Dependent variable: (log) employment | | |
| *InsideZone* | 0.084 | 0.471*** |
| | (0.098) | (0.040) |
| Observations | 1,174 | 2,348 |
| Panel (b). Dependent variable: (log) output | | |
| *InsideZone* | 0.261** | 0.553*** |
| | (0.123) | (0.056) |
| Observations | 1,174 | 2,348 |
| Panel (c). Dependent variable: (log) capital | | |
| *InsideZone* | 0.307** | 0.547*** |
| | (0.122) | (0.054) |
| Observations | 1,174 | 2,348 |
| Panel (d). Dependent variable: (log) number of firms | | |
| *InsideZone* | −0.350*** | 0.233*** |
| | (0.078) | (0.031) |
| Observations | 1,174 | 2,348 |

*Note:* Observations are at the area-year level within a 1,000 meter distance of the zone boundaries. *InsideZone* is an indicator variable for whether the area is inside the zone or not. Panels a-d report the estimation results on the natural log of the measure of employment, output, capital, and number of firms respectively. In column 1, the data of 2008 is used for the BD analysis. Zone (neighborhood) fixed effects are included in the specification. In column 2, the data of both 2004 and 2008 are used for the BD-DD exercise. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table 5 Robustness Checks

| | I. Alternative distance ranges | | II. 0-1000 meters vs. 1000-2000 meters | |
|---|---|---|---|---|
| | Within 2000 meters | Within 500 meters | Outside the zone | Inside the zone |
| | (1) | (2) | (3) | (4) |
| Panel (a). Dependent variable: (log) employment | | | | |
| *Dummy* | 0.494*** | 0.412*** | −0.023 | −0.080 |
| | (0.036) | (0.044) | (0.038) | (0.056) |
| Observations | 2,428 | 2,256 | 2,188 | 1,384 |
| Panel (b). Dependent variable: (log) output | | | | |
| *Dummy* | 0.567*** | 0.447*** | −0.073 | −0.082 |
| | (0.051) | (0.062) | (0.049) | (0.074) |
| Observations | 2,428 | 2,256 | 2,188 | 1,384 |
| Panel (c). Dependent variable: (log) capital | | | | |
| *Dummy* | 0.584*** | 0.423*** | −0.025 | −0.159** |
| | (0.050) | (0.060) | (0.051) | (0.079) |
| Observations | 2,428 | 2,256 | 2,188 | 1,384 |
| Panel (d). Dependent variable: (log) number of firms | | | | |
| *Dummy* | 0.267*** | 0.203*** | −0.010 | −0.060 |
| | (0.028) | (0.035) | (0.028) | (0.044) |
| Observations | 2,428 | 2,256 | 2,188 | 1,384 |

*Note:* Observations are at the area-year level. In Columns 1-2, the regression samples consist of areas within a 2,000 meter and 500 meter distance of zone boundaries that are reported in the column heading. *Dummy* is an indicator variable for whether the area is inside the zone or not. In Column 3, the regression sample include areas both outside the zones: areas within a 1,000 meter distance of the boundaries versus those within a 1,000 to 2,000 meter distance of the boundaries. *Dummy* is an indicator variable for whether the area is within a 1,000 meter range or not. In column 4, the sample is composed of areas both inside the zone: areas within a 1,000 meter distance of the boundaries and areas within a 1,000 to 2,000 meter distance of the boundaries. *Dummy* is an indicator variable for whether the area is within a 1,000 meter range or not. Panels a-d report the estimation results on the natural log of the measure of employment, output, capital, and number of firms respectively. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table 6 Mechanism: A Decomposition Analysis

| Dependent variable: | (log) employment | (log) output | (log) capital | (log) number of firms |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Entrants and Exiters | 0.281 | 0.283 | 0.305 | 0.204 |
| Incumbents | 0.094 | 0.134 | 0.114 | 0.000 |
| Relocaters | 0.078 | 0.114 | 0.111 | 0.028 |

*Note:* The zone's total effect has been decomposed into three parts: the effect due to entrants and exiters, incumbents, and relocaters respectively. In Columns 1-4, the dependent variables are the natural log of the measure of employment, output, capital, and number of firms that are reported in the column heading. The BD-DD estimate of the SEZ effect based on the full sample is a weighted average of the BD-DD estimates on each subsample. The weights used are the share of employment, output, capital, and firm numbers of the corresponding subsample in the full sample.

Table 7 Heterogeneous Effects by Industrial Capital-Labor Ratios

| Dependent variable: | (log) employment | (log) output | (log) capital | (log) number of firms |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Panel (a). Capital-Intensive Industries | | | | |
| *InsideZone* | 0.524*** | 0.667*** | 0.638*** | 0.291*** |
| | (0.051) | (0.073) | (0.071) | (0.033) |
| Observations | 2,124 | 2,124 | 2,124 | 2,124 |
| Number of industries | 242 | 242 | 242 | 242 |
| Panel (b). Labor-Intensive Industries | | | | |
| *InsideZone* | 0.424*** | 0.458*** | 0.483*** | 0.216*** |
| | (0.041) | (0.053) | (0.051) | (0.032) |
| Observations | 2,260 | 2,260 | 2,260 | 2,260 |
| Number of industries | 243 | 243 | 243 | 243 |
| Panel (c). Difference | | | | |
| *InsideZone* | 0.100 | 0.209** | 0.155* | 0.075 |
| | (0.065) | (0.090) | (0.088) | (0.046) |

*Note:* Observations are at the area-year level. In panels a and b, capital (labor) intensive industries correspond to the industries at the 4-digit level whose capital-labor ratio are above (below) the median value in 2004. Panel c reports the difference of the SEZ policy impact on a capital intensive industry and a labor intensive industry. *InsideZone* is an indicator variable for whether the area is inside the zone or not. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. In panel c, standard errors are computed using the delta method. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table 8 Heterogeneous Effects by the Market Potential of Zones

| Dependent variable: | (log) employment | (log) output | (log) capital | (log) number of firms |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Panel (a). Zones with High Market Potential | | | | |
| *InsideZone* | 0.490*** | 0.599*** | 0.577*** | 0.230*** |
| | (0.053) | (0.074) | (0.073) | (0.042) |
| Observations | 1,192 | 1,192 | 1,192 | 1,192 |
| Number of zones | 309 | 309 | 309 | 309 |
| Panel (b). Zones with Low Market Potential | | | | |
| *InsideZone* | 0.451*** | 0.506*** | 0.517*** | 0.237*** |
| | (0.059) | (0.084) | (0.079) | (0.046) |
| Observations | 1,156 | 1,156 | 1,156 | 1,156 |
| Number of zones | 310 | 310 | 310 | 310 |
| Panel (c). Difference | | | | |
| *InsideZone* | 0.039 | 0.093 | 0.059 | −0.008 |
| | (0.080) | (0.112) | (0.107) | (0.062) |

*Note:* Observations are at the area-year level. In panels a and b, zones with high (low) market potential correspond to zones whose market potential indices are above (below) the median in 2004: a larger index stands for higher market potential. Panel c reports the difference of the SEZ policy impact on zones with high market potential and those with low market potential. *InsideZone* is an indicator variable for whether the area is within the zone or not. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. In Panel c, standard errors are computed using the delta method. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table 9 Heterogeneous Effects by Zones' Infrasturcture Accessibility

| Dependent variable: | (log) employment | (log) output | (log) capital | (log) number of firms |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Panel (a). Spatially-integrated Zones | | | | |
| *InsideZone* | 0.490*** | 0.545*** | 0.517*** | 0.252*** |
| | (0.050) | (0.069) | (0.068) | (0.040) |
| Observations | 1,196 | 1,196 | 1,196 | 1,196 |
| Number of zones | 311 | 311 | 311 | 311 |
| Panel (b). Spatially-isolated Zones | | | | |
| *InsideZone* | 0.451*** | 0.561*** | 0.579*** | 0.214*** |
| | (0.062) | (0.089) | (0.084) | (0.047) |
| Observations | 1,152 | 1,152 | 1,152 | 1,152 |
| Number of zones | 308 | 308 | 308 | 308 |
| Panel (c). Difference | | | | |
| *InsideZone* | 0.039 | −0.015 | −0.063 | 0.038 |
| | (0.080) | (0.113) | (0.108) | (0.062) |

*Note:* Observations are at the area-year level. In panels a and b, spatially-integrated (spatially-isolated) zones corresponds to the zones whose infrastructure accessibility index are above (below) the median: a higher index indicates that a zone is more accessible to transportation infrastructures such as airports and highways. Panel c reports the difference of the SEZ policy impact on spatially-integrated zones and spatially-isolated zones. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. In Panel c, standard errors are computed using the delta method. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table A1 Comparison between Census and ASIF data

| | *Economic Census* | | | | *Annual Survey of Industrial Firms* | | | |
|---|---|---|---|---|---|---|---|---|
| | Obs. | Mean | p10 | p90 | Obs. | Mean | p10 | p90 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel (a). Year 2004 | | | | | | | | |
| Employment | 1,321,026 | 64 | 4 | 118 | 256,999 | 221 | 25 | 426 |
| Output | 1,321,026 | 14692 | 21 | 14607 | 256,999 | 68,451 | 5,424 | 95,593 |
| Capital | 1,321,026 | 14091 | 120 | 13800 | 256,999 | 62,093 | 2,731 | 89,541 |
| Capital-labor ratio | 1,319,227 | 207 | 9 | 295 | 255,558 | 293 | 36 | 537 |
| Output-labor ratio | 1,319,227 | 144 | 3 | 302 | 255,558 | 327 | 54 | 658 |
| Panel (b). Year 2008 | | | | | | | | |
| Employment | 1,822,419 | 58 | 4 | 100 | 382,838 | 194 | 25 | 350 |
| Output | 1,738,045 | 27578 | 360 | 30765 | 383,779 | 111,150 | 6,340 | 160,229 |
| Capital | 1,788,227 | 20558 | 303 | 20265 | 382,842 | 80,805 | 3,575 | 107,615 |
| Capital-labor ratio | 1,787,873 | 316 | 23 | 477 | 382,272 | 374 | 54 | 736 |
| Output-labor ratio | 1,737,794 | 299 | 33 | 538 | 382,165 | 577 | 102 | 1,190 |

*Note:* Table A1 reports number of observations, mean, 10th and 90th percentile of firm employment, output, capital, capital-labor ratio, and output-labor ratio based on *Economic Census* and *Anual Survey of Industrial Firms* for 2004 and 2008.

Table A2 The SEZ Effects: BD-DD Estimation (Aggregating Firms with Detailed Addresses)

| | |
|---|---|
| Panel (a). Dependent variable: (log) employment | |
| *InsideZone* | 0.601*** |
| | (0.058) |
| Observations | 1,828 |
| Panel (b). Dependent variable: (log) output | |
| *InsideZone* | 0.719*** |
| | (0.080) |
| Observations | 1,828 |
| Panel (c). Dependent variable: (log) capital | |
| *InsideZone* | 0.732*** |
| | (0.075) |
| Observations | 1,828 |
| Panel (d). Dependent variable: (log) number of firms | |
| *InsideZone* | 0.375*** |
| | (0.043) |
| Observations | 1,828 |

*Note:* Observations are at the area-year level within a 1,000 meter distance of the zone boundaries. To address possible measurement errors, in this exercise the panel data set by area and by year is constructed by aggregating firms with detailed addresses (50.5 percent of the whole sample). *InsideZone* is an indicator variable for whether the area is inside the zone or not. Panels a-d report the estimation results on the natural log of the measure of employment, output, capital, and number of firms respectively. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. ***, ** and * denote significance at the 1, 5 and 10% level respectively.

Table A3 Firms in 2004 and 2008

| | All firms | Entrants | Survivors | | | | Exiters |
|---|---|---|---|---|---|---|---|
| | | | Total | Linked using | | | |
| | | | | Firm ID | Firm name | Other information | |
| Census 2004 | 1,321,026 | – | 794,386 | 736,228 | 37,030 | 21,128 | 526,640 |
| Census 2008 | 1,822,419 | 1,028,033 | | | | | – |

*Note:* "All firms" denotes the full sample. "Entrants" refers to firms that exist in 2008, but do not exist in 2004; "Survivors" refers to firms that exist in both 2004 and 2008; "Exiters" refers to firms that exist in 2004, but do not exist in 2008. The following steps are used in identifying "Entrants", "Survivors" and "Exiters". First, for firms that report unique IDs (i.e., their legal person codes) in the Census data, firm ID is used to match firms in the 2004 census with those in the 2008 census. Second, for firms that have duplicate IDs, firm name is used to link observations over time. Third, firms may receive a new ID because of restructuring, mergers, or privatization. For a firm for which no observation with the same ID can be matched over time, information such as firm name, location code, name of legal person representative, phone number and so on is used in the matching process.

Table A4 Mechanism: A Decomposition Analysis Based on DD-Estimates

| Dependent variable: | (log) employment | (log) output | (log) capital | (log) number of firms |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Panel (a). Within the Same County | | | | |
| Entrants and Exiters | 0.148 | 0.158 | 0.176 | 0.116 |
| Incumbents | 0.038 | 0.067 | 0.055 | 0.000 |
| Relocaters | 0.030 | 0.033 | 0.032 | 0.010 |
| Panel (b). Within the Same Town | | | | |
| Entrants and Exiters | 0.132 | 0.146 | 0.168 | 0.093 |
| Incumbents | 0.041 | 0.067 | 0.056 | 0.000 |
| Relocaters | 0.044 | 0.051 | 0.051 | 0.012 |

*Note:* The zone's total effect has been decomposed into three parts: the effect due to entrants and exiters, incumbents, and relocaters respectively. Panel a reports the decomposition results in which SEZ villages and non-SEZ villages are within the same county. Panel b reports the decomposition results in which SEZ villages and non-SEZ villages are within the same town. In Columns 1-4, the dependent variables are the natural log of the measure of employment, output, capital, and number of firms that are reported in the column heading. The DD estimate of the SEZ effect based on the full sample is a weighted average of the DD estimates on each subsample. The weights used are the share of employment, output, capital, and firm numbers of the corresponding subsample in the full sample.

Table A5 Mechanism: BD-DD Estimation on Three Subsamples

| Within 1,000 meters | Incumbents | Entrants and Exiters | Relocaters |
|---|---|---|---|
| | (1) | (2) | (3) |
| Panel (a). Dependent variable: (log) employment | | | |
| *InsideZone* | 0.167*** | 0.782*** | 1.012*** |
| | (0.020) | (0.065) | (0.183) |
| Observations | 2,248 | 1,968 | 748 |
| Panel (b). Dependent variable: (log) output | | | |
| *InsideZone* | 0.225*** | 0.897*** | 1.309*** |
| | (0.037) | (0.093) | (0.242) |
| Observations | 2,248 | 1,968 | 748 |
| Panel (c). Dependent variable: (log) capital | | | |
| *InsideZone* | 0.194*** | 0.923*** | 1.319*** |
| | (0.027) | (0.093) | (0.236) |
| Observations | 2,248 | 1,968 | 748 |
| Panel (d). Dependent variable: (log) number of firms | | | |
| *InsideZone* | 0 | 0.407*** | 0.470*** |
| | – | (0.044) | (0.097) |
| Observations | 2,248 | 1,968 | 748 |

*Note:* Observations are at the area-year level within a 1,000 meter distance of the zone boundaries. The three panel data sets are constructed by aggregating subsamples "Incumbents", "Extrants and Exiters" and "Relocaters" by area and by year respectively. "Incumbents" refer to continuing firms with no zone status changes between 2004 and 2008. "Entrants and Exiters" refer to firms that exist in 2004 and exit in 2008 (Exiters), and firms that do not exist in 2004 and enter in 2008 (Entrants). "Relocaters" refer to continuing firms that change their zone status, i.e., firms that relocate from outside the zone to inside the zone, and firms that relocate from inside the zone to outside the zone. *InsideZone* is an indicator variable for whether the area is within the zone or not. Panels a-d report the estimation results on the natural log of the measure of employment, output, capital, and number of firms respectively. Area fixed effects and zone (neighborhood)-year fixed effects are included. Standard errors in parentheses. Standard errors are clustered at the zone level. ***, ** and * denote significance at the 1, 5 and 10% level respectively.