

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

Yi Lu<sup>†</sup>   Jin Wang<sup>‡</sup>   Lianming Zhu<sup>§</sup>

This Version: July, 2015

## Abstract

Exploiting two geo-coded firm censuses, this paper examines the impact of China's Special Economic Zones on economic activity in the targeted areas. We find that economic zones have a positive effect on employment, output and capital, and increase the number of firms. We also find that firm births and deaths play a larger role in explaining the zones' effects than incumbents and relocations. Finally, a zone's effectiveness depends on a program's features. Capital-intensive industries exhibit larger positive zone effects than labor-intensive ones. Location characteristics such as market potential and transportation accessibility are not critical factors in enhancing the program effects.

**Keyword:** Place-based Policies; Relocation; Market Potential; Special Economic Zones

**JEL Classification:** H2, J2, O2, R3

---

\*We thank Maitreesh Ghatak, Gerard Padró i Miquel, James Kung, Albert Park, Eric Verhoogen and the audience at various seminars for their comments. Wang acknowledges the financial support of a research grant from the Research Grants Council of Hong Kong (HKUST, No.16504715). All errors remain our own.

<sup>†</sup>Yi Lu, Department of Economics, National University of Singapore, 1 Arts Link, Singapore, 117570. Email: ecluyi@nus.edu.sg.

<sup>‡</sup>Jin Wang, Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Email: sojinwang@ust.hk.

<sup>§</sup>Lianming Zhu, Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan. Email: lmlianming@gmail.com.

# 1 Introduction

Place-based programs—economic development policies aimed at fostering the economic growth of an area within a 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 such programs (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),<sup>1</sup> 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? Since developing countries usually suffer from poorly-developed institutions and markets, would the assumptions and conceptual approaches of 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, 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, it sets out to document micro-level evidence on the incidence and effectiveness of place-based policies in developing countries by evaluating China’s Special Economic Zones (SEZs). As a prominent development strategy implemented worldwide, SEZs attempt to foster agglomeration economies by building clusters, increasing employment and attracting technologically-advanced industrial facilities.<sup>2</sup>

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

---

<sup>1</sup>Prominent examples include initiatives that target lagging areas such as enterprise zones in the United States and regional development aid within the European Union.

<sup>2</sup>SEZs have been implemented in 135 countries (World Bank, 2008).

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

Our analyses proceed in three stages. We examine the effect of an SEZ on the targeted area’s employment, output, capital, and the number of firms. We further elucidate the mechanisms underlying the observed effects. Specifically, our paper identifies the effects of the economic zones on two margins—the extensive margin (firm births, deaths, and relocation) and the intensive margin (the incumbents). Finally, we study 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, as well as 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 has been to construct a novel data set that merges a comprehensive data set of China’s economic zones, which includes the establishment year, the type of zones and the 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 to the zones, and the dynamics of 3,143,445 firms. We then aggregate these individual firms to construct a panel data set by area and by year. The data series cover two periods—two years before a zone’s establishment and two years after a zone’s establishment—allowing us to assess the effects of SEZs on the targeted areas 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 SEZs have been precisely measured on such a disaggregated level and over a whole universe of manufacturing firms. It is also the first time that comprehensive geocoded information on Chinese firms has been compiled and analyzed in relation to SEZs, something that has not been possible in previous work in this literature.

The key challenge in identifying the causal effects of zone programs is selecting appropriate control groups (Neumark and Kolko, 2010). This study starts 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 in performance among SEZ villages with the changes among non-SEZ villages during the same period. As an al-

ternative approach, we follow Holmes (1998) and Neumark and Kolko (2010) in making use of detailed information on firm location and zone boundaries.<sup>3</sup> We exploit the discontinuity in treatment at the zone boundary, a boundary discontinuity (BD) design. We choose an area 1,000 meters from the boundary of the SEZ to compare 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, we combine the DD and BD approaches in a BD-DD analysis. Specifically, we first obtain an estimate from the data without the zone and then another with the zone established. Assuming the confounding factors to be fixed over time, we isolate the SEZ's true effect on the targeted area from the difference in the two estimates. A series of analyses further investigate the robustness of the 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 has a significant and positive impact on the targeted areas. After two years, the SEZ areas have 47.1 percent greater employment, 55.3 percent larger output, and 54.7 percent larger capital than the non-SEZ areas. The number of firms has increased in the SEZs by 23.3 percent.

Decomposing the sample into relocated firms, entrants and exiters, and continuing firms without a zone status change in the period studied, our analysis reveals a sizable effect on employment, capital, and output associated with firm births and deaths. Incumbent firms show a significant improvement in performance. Relocation plays a modest role in the total SEZ effects. Overall, the results indicate that the program influences the targeted area through both extensive margins and intensive margins.

Finally, the zones exhibit larger positive impacts on firms in capital-intensive industries, while zones with higher market potential and infrastructure accessibility do not demonstrate significantly larger effects.

This paper fits into a large literature that explores quasi-natural experiments to quantify the impact of place-based programs. Criscuolo, Martin, Overman, and Van Reenen (2012) investigate the causal impact of the 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 Zones Franches Urbaines (ZFU) and their place-based tax-exemptions on business entry and exit rates, economic activity, employment, as well as on the financial strength of the companies. Devereux, Griffith, and

---

<sup>3</sup>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.

Simpson (2007), like Briant, Lafourcade, and Schmutz (2015), uncover geographic factors which can account for heterogeneities in the programs' effects—i.e., impacts of placed-based policies are more significant for locations with better market access.

We 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 their 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 also relates to several recent studies exploring China's SEZs to evaluate their impact on local economies. Alder, Shao, and Zilibotti (2013) and Wang (2013) examine the local (city-level) impact of SEZs on growth, capital formation and factor prices, while Cheng (2014) estimate both the local (county-level) and aggregate impacts. Extending such work to the micro-domains, Schminke and Van Biesebroecke (2013) investigate the extensive margin effect of state-level zones on exporting performance.

Methodologically, this study builds on much previous work applying the geographic regression discontinuity (GRD) design (Black, 1999; Bayer, Ferreira, and McMillan, 2007; Dell, 2010; Keele and Titiunik, 2015). It also relates broadly to a set of studies examining the impact of taxation on firm-level outcomes such as location decisions, entry and employment (Duranton, Gobillon, and Overman, 2011; Brühlhart, 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 in China. There are two main categories of SEZs: national-level and provincial-level economic zones. The former are approved by the central government and are more privileged, while the latter are granted by provincial governments. Each zone has an administration committee, which directs and administers the zone on behalf of the respective government handling project approval, local taxation, land management, finance, personnel, environmental protection and security.

The goals of adopting special policies in an area within a jurisdiction are to increase foreign and domestic investment, promote international trade, stimulate technological co-

operation and innovation, and increase 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 their utilities, telecommunications, transport, storage, and other basic installations and service facilities.<sup>4</sup>

The most important preferential policies usually include the following (Wang, 2013; Alder, Shao, and Zilibotti, 2013):

1. *Tax deductions and customs duty exemptions.* Corporate income tax rates of 15%-24% as opposed to the 33% firms normally pay in China 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 exports.
2. *Discounted land-use fees.* Under Chinese law, all land is state-owned. In the SEZs investors may lawfully obtain land development and business use rights, the duration, the fees and the method of payment depending on the type of businesses. A total investment of at least US\$10 million, or which is considered to be technologically advanced with a major influence on local economic development may attract preferential treatment.<sup>5</sup>
3. *Special treatment in securing bank loans.* Foreign-invested enterprises are encouraged to make use of domestic financial resources to finance their investments. The banks put priority on their loan applications.

Figure 1 shows the geographic distribution of the SEZs established in five waves over 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 surge in zone establishment (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, China's first comprehensive regional development plan (the Western Development Strategy) was launched. As a result, zone establishment was then concentrated in inland cities. Recently, zone establishment has exhibited more balanced development. Between 2005 and 2008, 338 SEZs were granted in the coastal area, 269 in the central area, and 75 in the west. For detailed descriptions on

---

<sup>4</sup>See China provincial SEZ laws [various issues including Guangdong, Jiangsu, Anhui, Inner Mongolia, Shandong].

<sup>5</sup>Source: the government website of Zhejiang Province.

these waves, see Appendix A.

[Insert Figure 1 here]

There are several types of SEZs (Alder, Shao, and Zilibotti, 2013), in which the preferential policies have different focuses. Economic and 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 limited to such 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 type of zones 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 an SEZ program on its targeted area, we use three empirical strategies—specifically, a DD method, a BD design, and a BD-DD approach.

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

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

where  $y_{vct}$  is an outcome variable such as total output, labor, capital, or the 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 the villages in county  $c$  in year  $t$ .  $\varepsilon_{vct}$  is an error term. To control for potential heteroskedasticity and serial correlation, we cluster the standard errors at the village level (Bertrand, Duflo, and Mullainathan, 2004).

SEZs were not sited 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 contiguous, choosing the non-SEZ villages in the same towns as the SEZ villages. Since, however, the main analyses use data from 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 have followed the same time trends 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,  $SEZ_{vct}^{false}$  would be expected to have zero effects. Otherwise, it may indicate a 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 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, a 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 from 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 the zone status. In the benchmark analysis, we use a geographic distance of 1,000 meters (or 5/8 mile), but 500 meters and 2,000 meters are tested in the robustness checks.

The BD estimation equation is

$$y_{az} = \gamma SEZ_{az} + \lambda_z + \varepsilon_{az}, \quad (2)$$

where  $y_{az}$  concerns performance in 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.  $\varepsilon_{az}$  is an error term. To ensure conservative statistical inference, we cluster the standard errors at the zone level. Equation (2) is estimated using data from the economic census in the year after the zone was established.

*BD-DD Estimation.* Despite their proximity to the zone boundary, areas inside and



outside zones may differ, particularly if the zone boundaries have not been 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 the identification further and address the concerns about the endogeneity of boundaries, we combine the DD and BD approaches into a BD-DD analysis. We first estimate equation (2) using data from the year before the zone’s establishment, and obtain  $\hat{\gamma}_{BD,Control} = \eta_{Control}$ . We then estimate the same equation using data from the year after the zone’s 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.,  $\eta_{Control} = \eta_{Treatment} = \eta$ ), this estimates the SEZ’s true 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

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

where  $y_{azt}$  measures performance of area  $a$  within 1,000 meters of the boundary of zone  $z$  in 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 neighborhood-year fixed effect, capturing unobserved shocks common to both sides of zone  $z$  in year  $t$ . Including neighborhood-year fixed effects allows for flexible time trends across different zones.  $\varepsilon_{azt}$  is the error term. To ensure conservative statistical inference, we cluster the standard errors at the zone level.

To check the identifying assumption underpinning the BD-DD analyses (i.e., that the underlying location characteristics are fixed except for the zone policies), we conduct two placebo tests. The first focuses on a sample of areas located outside the zones and compare them within different distances of the zone boundary. Then, we compare areas within the zones but at different distances from the zone boundary. As the areas have the same zone status in each of these two tests, distance alone should not predict significant differences; otherwise, it would indicate 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, conducted by China’s National Bureau of Statistics at the end of 2004 and 2008.<sup>6</sup> The advantage of census data over the Annual Survey of Industrial Firms (ASIF) often used in the literature (e.g., Hsieh and Klenow, 2009) is that it is more

---

<sup>6</sup>The third wave was started in January 2014 and is still underway.

comprehensive, covering all manufacturing firms in China, while the latter includes only state-owned enterprises (SOEs) and non-SOEs with annual sales of more than five million yuan. Table A1 in the appendix compares these two data sources for 2004 and 2008. The census data, which represent the entire population of manufacturing firms, clearly show 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 (12-digit: corresponding to a village or community), industry affiliation, and ownership. We use address and location code to geographically locate a firm and identify whether it is in a zone or not (see the *Coordinates data* and *Firm SEZ status* sub-sections for details). The census data report employment, output, and capital for each firm.

*Coordinates data.* In the BD and BD-DD analyses, we aggregate outcomes of individual firms into areas close to the zone boundaries. This requires precise geographical information on firm locations (i.e., the coordinates) to determine the firms' distances from the zone boundaries. We search firms' addresses to obtain their geographic coordinates using Google's Geocoding API.<sup>7</sup> Each firm's detailed Chinese address (for example, "157 Nandan Road, Xuhui District, Shanghai, China") is first used with Google Maps to obtain a map with the specific location of the address indicated (see Figure A1). After confirming the correctness of the marked location, we extract the firm's latitude and longitude from the Google map. By this process, we determine coordinates for approximately 50.5 percent of the firms.

To deal with incomplete addresses,<sup>8</sup> road name changes, and reporting errors, we search the remaining firms using their 12-digit location codes.<sup>9</sup> So for a firm without detailed address (for example, "Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China"), we use the 12-digit location code (in this example the 12-digit location code is "320585102202"). The 12-digit code gives the village or community (location code "320585102202" corresponds to "Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China"). We then use the name of that village or community to collect the latitude and longitude of the village or community from Google maps (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

---

<sup>7</sup>The robustness of the results is checked by using Baidu's geocoding API service. Baidu is the Chinese version of Google. It provides a similar service, but has a different coordinate system.

<sup>8</sup>Incomplete address refers to an address that only has information on village, building, or street name, but with no number or building name.

<sup>9</sup>There are approximately 700,000 villages and communities in China. The habitable area of China is about 2.78 million square kilometer. On average, a village or community covers about 4 square kilometer. In the census data, the average number of firms in a village or community was 5.4 in 2004 and 6.7 in 2008. The statistics indicate the precision of using a village or community's coordinates when firms do not provide a detailed enough address.

percent of the whole sample), and find similar results (see Table A2).

*Firm SEZ status.* The census data do not directly report information about each firm’s SEZ status. To identify whether a firm is located inside an SEZ or not, we make use of the following data sources. A comprehensive SEZ data set from the Ministry of Land and Resources of China defines SEZs’ boundaries in terms of villages, communities, and sometimes roads. Based on that information, we use the maps to determine whether or not a village or community lies within the boundary. The SEZs’ official websites, report detailed information about the villages and communities within their administrative boundaries. The National Bureau of Statistics and the Ministry of Civil Affairs website also reported administrative divisions and codes at the village and community level, including information on for some economic zones the villages and communities under their administration.

A list of villages and communities within each zone is thus created. Matching the list with the census data, we use the firms’ addresses as well as the 12-digit location codes (see Appendix B for a detailed discussion) to identify the firms located in each zone, and those outside any zone. To verify this approach, we cross-check the results by matching against the SEZ names some firms include in their addresses.<sup>10</sup>

*Regression data.* The analysis focuses on SEZs established between 2004 and 2008. There were 682 SEZs established during that period (specifically, 19 in 2005 and 663 in 2006), and there were substantial geographic variations. 338 SEZs were established in the coastal area, 269 in the central area and 75 in the western area.<sup>11</sup> Nineteen were state-level EPZs, 280 were province-level ETDZs, and 383 were province-level IDZs. In the analyses, we exclude state-level zones because of the concern that they might not be comparable with province-level zones.<sup>12</sup>

For the baseline DD analysis, we aggregate individual firms to construct a panel data set by village and by year. Thus, each village could have two observations in 2004 and 2008, a

---

<sup>10</sup>It could be argued that where an SEZ boundary bisects a village or community only part of it is in the zone. This is not a concern in China where the local governments survey and appraise land, outline plans for future development based on village and community units to designate the zone areas.

<sup>11</sup>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.

<sup>12</sup>In 2005, 18 state-level EPZs and one BZ were approved by the central government. Such state-level zones have higher-level administration committees than provincial-level SEZs and their committees enjoy more authority in managing the zones. By design these EPZs and BZ mostly reside in pre-established ETDZs—an overlapping problem. To take the Huizhou Export Processing Zone as an example, it is located inside the Guangdong Huizhou ETDZ, which was established in 1997. The BD-DD identification strategies are not valid for this set of zones, as the pre-existing ETDZ confounds the effect of the newly approved EPZs and BZs. See Wang (2013) for more details.

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

The BD and BD-DD exercises involve calculating each firm’s distance from the nearest SEZ boundary. Coordinates of each firm’s location in Coordinates data are known, but accurate geocodes of each SEZ boundary are not, which prevents calculating the distance to the boundary directly.<sup>13</sup> We instead apply the approach used by Duranton, Gobillon, and Overman (2011) to determine the distance indirectly. To determine whether a firm is located within 1,000 meters of a zone boundary, we search within a radius of 1,000 meters of the firm,<sup>14</sup> as illustrated in Figure 2. If firm A locates outside the zone and is found to be within 1,000 meters of firm B inside the zone, A is designated as being within 1,000 meters of the zone boundary; otherwise, it is not. Similarly, if firm C is located inside a zone and there is another firm (firm D) located outside the zone but within 1,000 meters, C is designated as located within 1,000 meters of the zone boundary.

[Insert Figure 2 here]

Repeating these steps for each firm in the census data yields a sample of 587 SEZs with 163,069 firms located within 1,000 meters of their boundaries: the 2008 sample contains 126,976 firms, approximately 43 percent of which are located inside an SEZ; the corresponding numbers for the 2004 sample are 81,739 and 41 percent.<sup>15</sup> We then aggregate those firms to construct a panel data set by area and by year. Each zone’s 1,000 meter neighborhood has two areas inside and outside the zone, each has two observations in 2004 and 2008. The regression sample for estimation consists of 587 areas.

*Summary statistics.* Table 2 shows summary statistics (the means and standard deviations of the employment, output, capital and number of firms in the areas) for the main regression samples. Columns 1 to 4 present the village level data. 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 summarize

<sup>13</sup>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.

<sup>14</sup>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.

<sup>15</sup>We restrict to SEZs that in each year 2004 and 2008, both sides of the zone boundary have firms.

the data for the BD-DD exercise.

[Insert Table 2 here]

Panel a illustrates the areas' characteristics in 2008, i.e., two years after the SEZs were established. The first and second columns show that SEZ villages on average have more workers employed, greater output, larger capital stock, and more firms than non-SEZ villages. As reported in columns 3 and 4, the difference between the treatment and control group is of a decreased magnitude, though still positive. The last two columns show that when the control group is defined to be along the boundary, the zone areas on average still had more employees, more output, and larger capital stock than the neighboring 1,000 meter ring, even though there were fewer firms in the zone areas.

Panel b compares the areas' characteristics in 2004, two years before SEZs were established in some locations. There were significant differences in areas' initial characteristics. 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 suggest that the different areas were not identical to start with in terms of these measures. However, after SEZs were established in some localities there was a markedly larger increase in economic activities in the treatment areas than in the 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 the DD estimates. The control group in column 1 consists of non-SEZ villages in the same county (6 digit code corresponding to a county) as SEZ villages. Non-SEZ villages in the same town (9 digit code) as SEZ villages are used as the control group in column 2. Four village-level outcomes are considered: employment, output, capital, and number of firms. The logarithm form of these outcomes are presented to highlight the magnitude of the effects. The magnitudes of the coefficients in column 1

(a broad DD) are similar to those in column 2 (when we restrict the units to be closer). All the estimated coefficients of the four outcomes are consistently positive and statistically significant, suggesting that after the zones' establishment, the SEZ villages have gained employment, output and capital, as well as more firms than the non-SEZ villages.

[Insert Table 3 here]

*BD and BD-DD estimates.* Using the sample of areas within 1,000 meters of a zone boundary, Table 4 shows the coefficients describing the impact of the SEZ program. Columns 1 and 2 report the BD and BD-DD estimates respectively. All the estimated coefficients for employment, total output and capital are positive and statistically significant except for the BD estimate of total employment. All the BD-DD estimates are consistently larger than the corresponding BD estimates, pointing to the possibility of non-random zone boundaries. The BD-estimated coefficient of the number of firms term is negative and significant, while that estimate using BD-DD is positive and significant. These results suggest that zones were established in places with a smaller number of firms but attracted more firms in the two years.

[Insert Table 4 here]

*Economic impact.* To gauge the economic significance of the SEZ effects, we use the BD-DD estimates.<sup>16</sup> Two years after the establishment of the zones, the SEZs' employment has increased by 47.1 percent, output by 55.3 percent, capital by 54.7 percent, and the number of firms by 23.3 percent. These findings are largely consistent with those of previous studies. For example, Givord, Rathelot, and Sillard (2013) find that the French Zones Franches Urbaines program have significant effects on both business creation and employment. Criscuolo, Martin, Overman, and Van Reenen (2012) also find a large and statistically significant average effect of the UK's employment and investment promotion program.

## 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 alternative distance from zone boundaries for the BD and BD-DD exercises, and two placebo tests for the BD-DD analysis.

---

<sup>16</sup>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.

*Randomly-assigned SEZ status in the DD analysis.* To check whether the DD estimation could have been mis-specified, we randomly assign the villages in the 2008 data to be SEZ or non-SEZ. Figure 3 shows the distribution of estimates from 500 of such randomizations. For both control groups and the four major outcomes, the distribution of the estimates from random assignments is centered around zero and not significantly different from zero. In addition, the benchmark estimates from Table 3, represented by vertical dashed lines, clearly lie outside the estimates from the placebo tests. Taken together, these results imply that there are no significant omitted variables in the DD specification.

[Insert Figure 3 here]

*Distance from a zone boundary.* The BD and BD-DD benchmark analyses focus on areas within 1,000 meters of a zone boundary. To check whether the results are sensitive to the distance used, two alternative distances are tested, specifically, 500 and 2,000 meters. The estimation results are reported in columns 1–2 of Table 5. The magnitudes of the estimates from two alternative samples are similar to those from the benchmark sample, suggesting that the estimates are not biased due to the choice of a specific distance.

[Insert Table 5 here]

*Both outside and inside 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 compare the performance of areas outside the zones within 1,000 meters of the boundaries with that of areas 1,000 to 2,000 meters from the boundaries.<sup>17</sup> Since neither group enjoys zone policies, any substantial difference in their performance would indicate a mis-specification in the BD-DD estimation. The estimation results are reported in column 3 of Table 5.

Similarly, we conduct another comparison of performance for areas inside the zones, one within 1,000 meters of a boundary and the other 1,000 to 2,000 meters from the boundary. The estimation results are reported in column 4 of Table 5. Almost all of these 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 the benchmark estimates, the tests show that from 2,000 meters outside the zone boundaries to 2,000 meters inside, the discontinuity in outcomes is detected just at the zone boundaries. These results therefore provide support for the validity of the BD-DD identifying strategy.

Overall, this wide range of placebo tests shows that our results are robust to several

---

<sup>17</sup>For the second group, if a zone has a breadth less than 2,000 meters, all the firms located more than 1,000 meters from the boundary within the zone are used.

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 has profound influence on firms' location choices and investment decisions. When facing the policy shocks, firms in a zone can respond along the intensive margin by varying inputs and outputs. They can also respond along the extensive margin by deciding whether or not to enter the zone, to relocate to the zone, to exit the zone. The composition of active firms in a zone area is given by the sum of firms choosing to begin production there, incumbent firms which choose to continue production in the zone, and firms operating in other locations which choose to relocate to the zone, despite the disruption involved.

After the zones are established, what are the changes in the composition of the sets of firms located inside and outside the zones? And how the composition change explain our findings on the SEZ effects? To shed light on the underlying mechanisms, we decompose the SEZ effects into: (1) new entrants and exiters; (2) firms which do not change their zone status; and (3) firms switching from outside to inside a zone or the reverse.

BD-DD estimations are more rigorous in their identifying assumptions than DD and BD estimations, which means the BD-DD estimations lead to more conservative and typically more credible inferences.<sup>18</sup> The following analyses therefore use the BD-DD estimates. Specifically, the BD-DD estimator  $\hat{\gamma}_{BD-DD}$  is

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

where  $Y$  represents the area outcomes (i.e., employment, output, capital, and 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  $j \in \{entry/exit, inc, re\}$ .  $\omega^j \equiv \frac{Y^j}{Y}$  are the weights for sample  $j$ .

The decomposition analyses have demonstrated the necessity of distinguishing the incumbents, relocated firms, and exiters and entrants. In other words, three groups of firms

---

<sup>18</sup>Table A3 presents the decomposition using the DD and BD methods with qualitatively similar results.



from 2004 to 2008 must be traced. This involves the following steps. First, for firms which report unique IDs (their legal person codes) in the census data, firm ID is used to match them between the 2004 and 2008 censuses. For firms with duplicate IDs, the firm name is used to link observations over time. Firms may receive a new ID as a result of restructuring, merger, or privatization. For a firm for which no observation with the same ID could be matched over time, we use as much information as possible on the firm’s name, location code, the name of the legal representative person, phone number and so on to find a match. Table A4 reports the number of new entrants (i.e., firms that exist in 2008, but did not exist in 2004), survivors (i.e., firms which existed in both 2004 and 2008), and exiters (i.e., firms that existed in 2004 but not in 2008). Among the 794,386 survivors, 92.7 percent are linked using a firm ID, 4.7 percent using a firm name, and 2.7 percent using other information. Finally, we classify survivors into relocated firms and incumbents based on any zone changes. Firms are designated as relocated if their coordinates changed from inside (outside) a zone in 2004 to outside (inside) in 2008.

BD-DD estimates for the three sub-samples are reported in Table A5, and the decomposition results are reported in Table 6. Most of the SEZ effects comes from firm births and deaths—specifically, it accounts for 66.31 percent of the SEZ effect on employment, 59.08 percent of the effect on output, 61.38 percent of the effect on capital, and 93.66 percent of the effect on the number of firms. Incumbent firms also show significant improvement in their performance (in terms of employment, output and capital), accounting for 21.09–28.07 percent of the overall SEZ effects. Despite the large SEZ effects in the relocated firms sample, their contributions to the overall SEZ effects is similar to that by the incumbents, presumably due to the small share of relocated firms in the total sample.

[Insert Table 6 here]

Overall, this decomposition indicates that the zones have a large and significant effect along both the extensive margin and intensive margin. This companies well with the findings of Criscuolo, Martin, Overman, and Van Reenen (2012), which find a large and statistically significant average effect of the UK’s RSA program on employment and investment, with about half of the effects arising from incumbent firms growing (the intensive margin) and half due to net entry (the extensive margin). Our findings are however more optimistic than those of Givord, Rathelot, and Sillard (2013) with respect to incumbents. They find no evidence indicating an employment effect on existing businesses; employment growth mostly comes from the new businesses and firms which relocated. 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.<sup>19</sup>

#### 4.4 Heterogeneous Impacts

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 exhibit larger effects. Firms produce goods and trade with various markets. The level of economic activity in a location depends in part on 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. Zones with better market potential and transportation (local determinants of agglomeration) would therefore be expected to exhibit larger effects. By exploiting variations in capital-labor ratios at the 4-digit industry level and transportation at the zone level, we seek differences in the effects between capital-intensive and labor-intensive industries, between zones with high and low market potential, and between spatially-integrated and isolated zones.

*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]

The SEZ effects are consistently stronger in the capital-intensive industries. In absolute terms, the employment effect of SEZ residence on capital-intensive industries is larger than that in labor-intensive industries by 10 percent, the output effect is 20.9 percent greater, the capital effect is 15.5 percent larger, and the number of firm effect is larger by 7.5 percent. These results are consistent with the features of the SEZ program—the SEZs provide subsidies for capital investment, implying 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

---

<sup>19</sup>According to their review, only a few papers such as those of Busso, Gregory, and Kline (2013), Freedman (2013), and Ham, Swenson, Imrohoroglu, and Song (2011) report positive program effects.

SEZ effects in zones with different market potential, we construct a market potential index for each zone in the spirit of Harris (1954) and Rogers (1997). Specifically, the impact of trade and communication costs is assumed to increase with the inverse of a zone  $z$ 's distance from all the 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 a province,  $c$  denotes a prefecture-level city,  $GDP_c$  stands for city  $c$ 's GDP, and  $dist_{zc}$  is the distance between the zone  $z$ 's administrative committee and city  $c$ .<sup>20</sup> Following Briant, Lafourcade, and Schmutz (2015)'s lead, the weighted sum of the markets accessible from the zone is divided by the total size of all markets in the province to mitigate the impact of large cities. Then zones are then grouped based on whether their market potential values in 2004 were above or below the sample median.

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]

No statistically or economically significant differences in the SEZ effects are found between the zones with high and low market potential. These results imply that local determinants of agglomeration economies such as market potential do not play important roles in enhancing SEZ effectiveness in China.

*Spatially-integrated vs. spatially-isolated zones.* To investigate whether there are differential SEZ effects among zones with different accessibility, an index is first constructed for each zone, with a larger value indicating greater proximity to an airport and highways. As shown in Figure 4, each zone's distance to the nearest airport is computed and these are ranked from greatest to least ( $rank\_airport$ ). The highway density of each zone's city is ranked similarly. The zone's infrastructure accessibility index is then  $rank = (rank\_airport + rank\_highway)/2$ , with a lower index indicating poorer accessibility. The full list of airports is compiled from 2005 Transportation Yearbook, while the data on highway density (highways mileage divided by the land area of the city) is from China's 2005

---

<sup>20</sup>Note that Harris (1954) and Rogers (1997) use city as a regression unit, and the market potential of a city is the weighted average of the GDP from other cities. In China, economic zones are smaller units than cities. The city where an economic zone resides is therefore also included in the calculation of market potential.

Regional Statistical Yearbook.

[Insert Figure 4 here]

The zones are then divided into two groups based on whether their accessibility index in 2004 was above or below the sample median. The estimation results are reported in Table 9, with panel (a) for more accessible zones, panel (b) for spatially-isolated zones, and panel (c) for the differences between these two groups. Again, no statistically or economically significant differences are evident.

[Insert Table 9 here]

Taken together, we find that capital-intensive firms benefit more from zone programs than labor-intensive ones, but the effects of SEZs are quite similar across zones with different market potential and different infrastructure accessibility. These results resonate with those of previous work which emphasize the characteristics of the industry in analyzing the effects of place-based policies (Crisciolo, Martin, Overman, and Van Reenen, 2012; Combes and Gobillon, 2015), but they contrast with previous findings on the role of regional characteristics, for example, those of Briant, Lafourcade, and Schmutz (2015). These findings suggest that the complementary roles of regional and industry characteristics in place-based development programs may hinge on the specific context.

## 5 Conclusion

This paper exploits a natural experiment in the establishment of China's economic zones. Using firms as the unit of analysis has helped elucidate the mechanisms underlying the observed zone effects and reveal the determinants of program effectiveness. By focusing on a prominent place-based policy in China, it has addressed several important questions: whether zones work; for whom; and for designing more efficient programs, what work and where they work (Neumark and Simpson, 2014).

The zone programs have demonstrated a large effect on the targeted areas along extensive margins, especially via entries and exits. Relocations play only a minor role. Incumbent firms experience a sizable improvement in their performance. These findings may help to diffuse the general pessimism about zone programs in developing countries.

Another important finding is that a 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 for labor-intensive firms. Location characteristics such as market potential and

transportation accessibility seem, however, not to be of critical importance. Overall, this analysis serves as a reminder that to make an effective policy one has to pay close attention to the circumstances of the agents influenced.

This study has been a first step towards understanding the micro-foundations of place-based policies in developing countries. Much remains to be done. This study evaluates short-term effects (two years after the zones' establishment) due to data limitation, but further efforts should be extended to investigate zones' long term impacts. This empirical analysis of micro-level impacts 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).<sup>21</sup> Such analyses will undoubtedly be of great benefit in addressing how SEZ policy interventions should be implemented in specific contexts.

---

<sup>21</sup>Becker, Egger, and Ehrlich (2013) investigate the heterogeneity among EU member states in terms of their ability 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.” Center for Economic and Policy Research (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 Economics*, 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, Daniele Coen-Pirani, and Sieg Holger. 2015. “Firm Dynamics in an Urban Economy.” *International Economic Review*, forthcoming.
- [7] Brühlhart, 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.” University of California, 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.” National Bureau of Economic Research (NBER) Discussion Paper 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.” University of California, Berkeley, working paper.
- [17] Gibbons, Stephen, Stephen Machin, and Olmo 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.” London: Department for Transport.
- [20] Glaeser, Edward L., and Joshua D. Gottlieb. 2008. “The Economics of Place-Making Policies.” *Brookings Papers on Economic Activity*, 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*, 67(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." National Bureau of Economic Research (NBER) Discussion Paper 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): 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 Zone Formation

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 in Xiamen of Fujian Province. China started with virtually no foreign direct investment and almost negligible foreign trade before 1978, so those zones were considered a test base for the 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 locations, 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: Deng Xiaoping's famous Southern Tour in 1992 sent a strong signal for continuous reform and economic liberalization, and the State Council of China subsequently opened a number of border cities and all the capital cities of the inland provinces and autonomous regions. This period witnessed a huge surge in the establishment of development zones. 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 activity. 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 its western provinces. The success of coastal development zones demonstrated 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, additional 682 SEZs were established. In terms of their geographical distribution, 338 were in coastal areas, 269 in central areas, and 75 in western areas. In

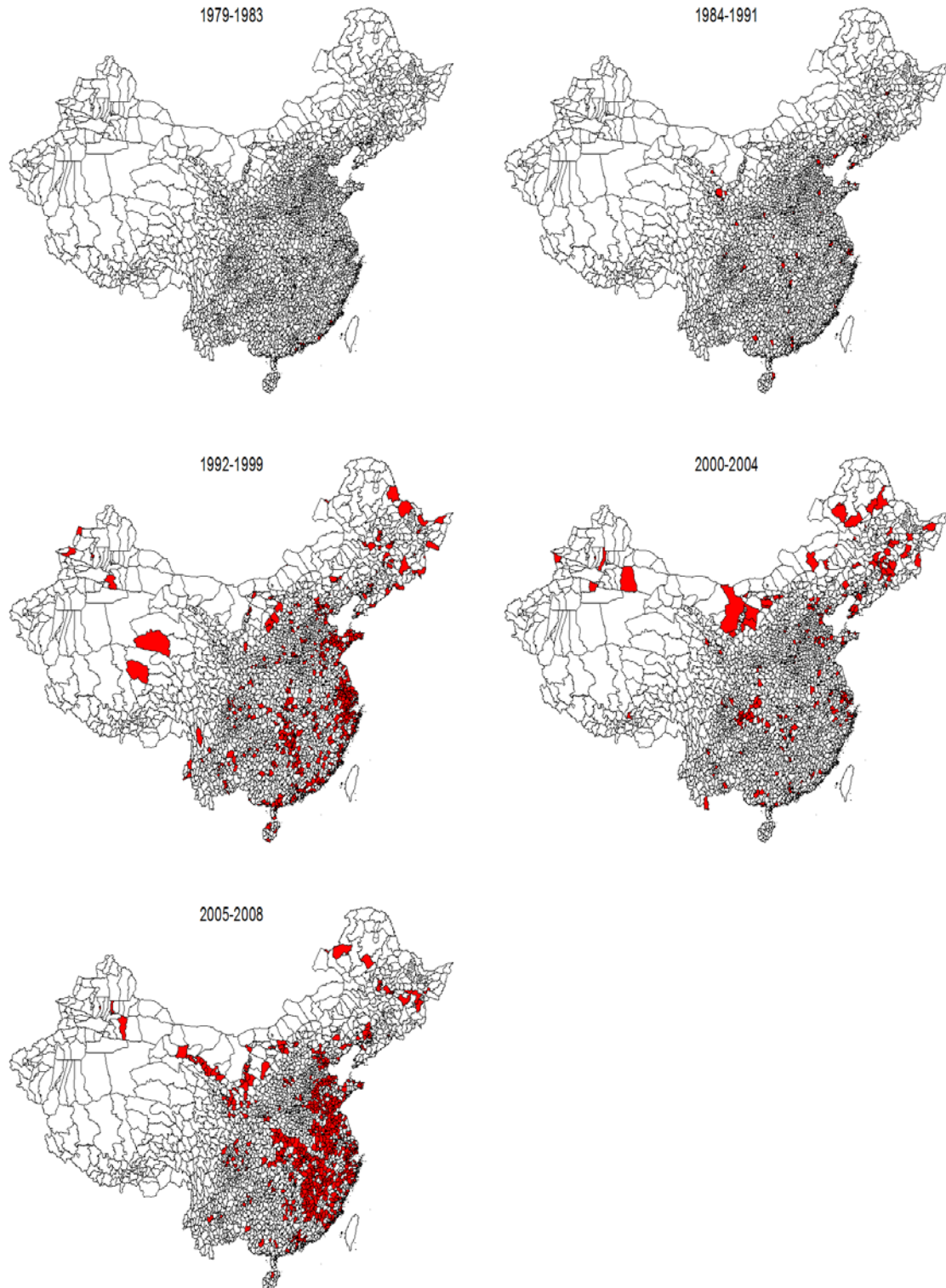
terms of granting authority, 19 state-level development zones and 663 province-level development zones were formed.

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

The administrative location code of each firm is used to locate it inside an SEZ or not. These three cases summarize the process.

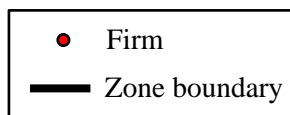
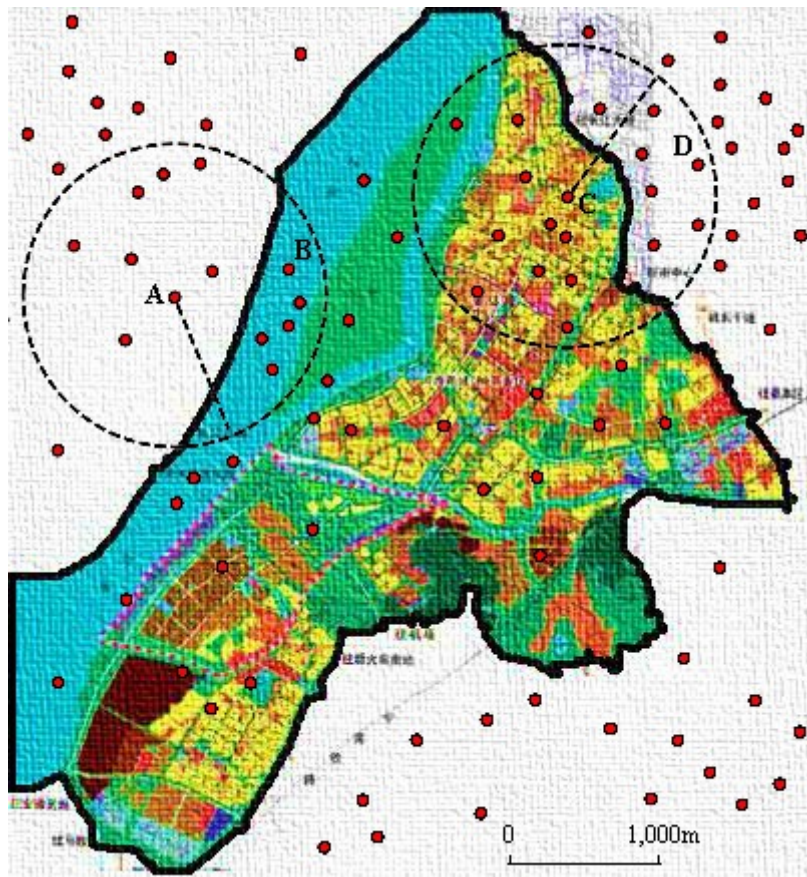
1. An SEZ with its own administrative code. For example, Anhui Nanling Industrial Zone (zone code: S347063) has an independent 12-digit administrative location code: 340223100400 (Anhui Nanling Industrial Zone Community).
2. A zone is equivalent to a town or street and all villages or communities under the town or street will be within the zone's boundaries. For example, Shandong Fei County Industrial Zone (zone code: S377099) encompasses all of Tanxin town (administrative location code: 371325105). The 9-digit town code is enough to pin down the zone area.
3. A zone takes 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 Zones by Wave



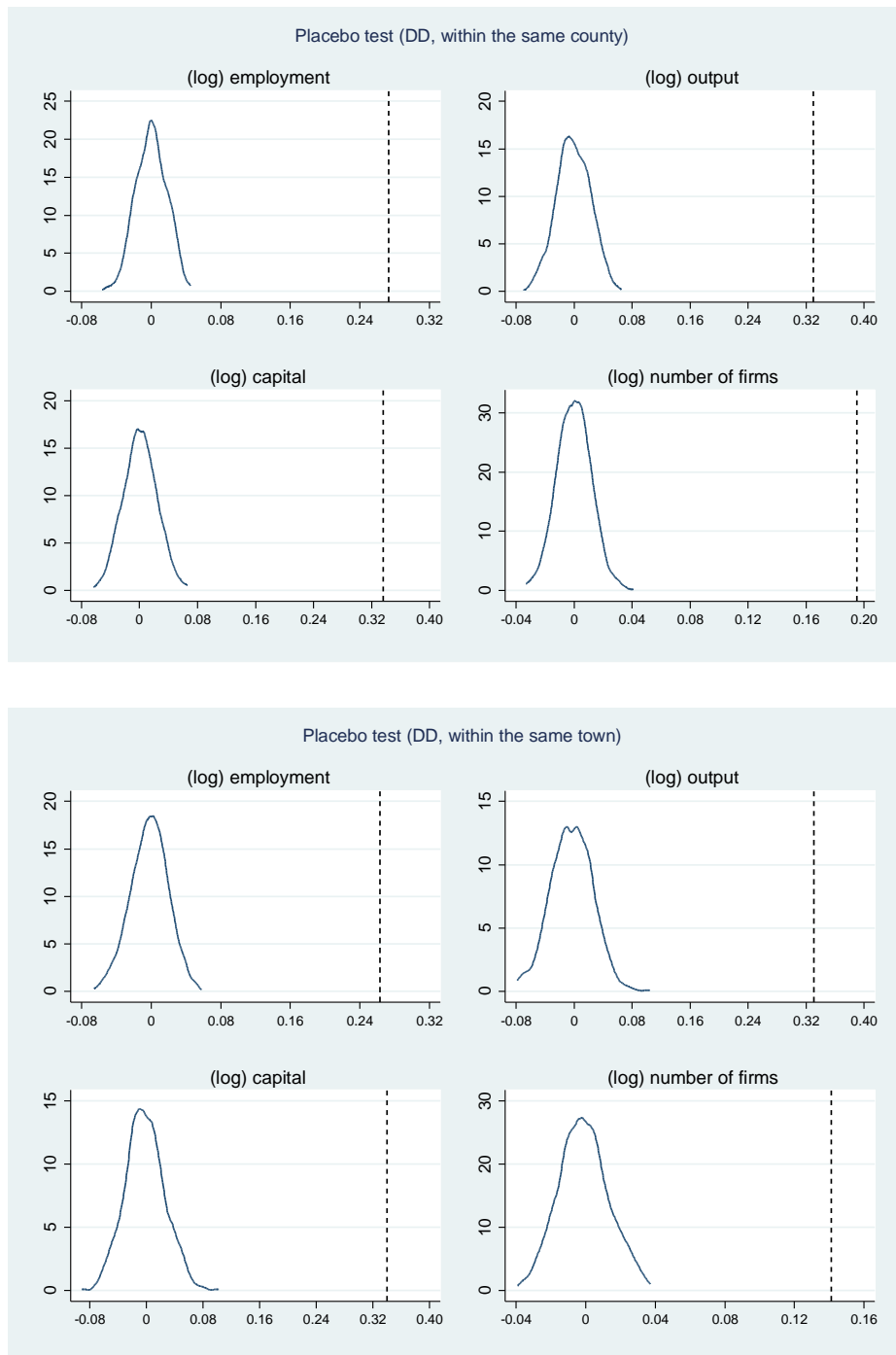
*Notes:* There have been five granting waves of SEZs: 1979-1983; 1984-1991; 1992-1999; 2000-2004; and 2005-2008. In each wave, counties where SEZs were newly established are indicated in color.

Figure 2 Firms near a Zone Boundary



*Notes:* If a firm (firm A) is located outside the zone and within 1,000 metres, there is another firm (firm B) located inside the zone, firm A is designated as located within 1,000 metres of the zone boundary. If a firm (firm C) is located inside the zone and within 1,000 meters, there is another firm (firm D) located outside the zone, firm C is designated as located within the 1,000 meters of the zone boundary.

Figure 3 Placebo Test



*Notes:* SEZ status is randomly assigned to villages to construct a false SEZ status indicator. This is repeated 500 times and the false data are fitted using Equation (1). The figure shows the distribution of estimates from the 500 estimations. The vertical dashed lines present the results from Table 3.

Figure 4 Geography: An Example using Tianjin’s Wuqing Economic Zone



*Notes:* An illustration shows how to measure the distance from a zone’s administrative committee to the nearest airport.

Table 1 SEZ Wave Breakdown

Wave	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, of which:		46	93	64	19
By Type					
1. Economic and Technological Development Zones		12	23	17	
2. High-tech and Industrial Development Zones		26	27		
3. Export Processing Zones			1	39	18
4. Bonded Zones		4	11	6	1
5. Border Economic Cooperation Zones			15	1	
6. Other		4	16	1	
By Region					
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, of which:		20	466	197	663
By Type					
1. Economic and Technological Development Zones		16	401	112	279
2. Industrial Development Zones		4	65	85	384
of which: High-tech and Industrial Development Zones		3	29	14	19
By Region					
1. Coastal Region		7	277	76	323
2. Central Region		7	138	71	267
3. Western Region		6	51	50	73

*Note* : For each period the number of development zones newly established is provided, and comprehensive SEZs, state-level development zones and province-level economic zones are distinguished. The comprehensive SEZs are the zones established in Shenzhen, Shantou, Zhuhai, and Xiamen. State-level development zones are granted by the central government. Province-level development zones are granted by the provincial governments. The coastal region includes Liaoning, Beijing, Tianjin, Hebei, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, and Hainan. The central region includes Heilongjiang, Jilin, Inner Mongolia, Shanxi, Henan, Anhui, Hubei, Hunan, and Jiangxi. The 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 (1.63)	4.76 (1.63)	6.02 (1.66)	5.11 (1.66)	8.11 (1.46)	8.03 (1.46)
(log) output	11.53 (2.11)	9.95 (2.11)	11.53 (2.13)	10.35 (2.13)	13.91 (1.89)	13.65 (1.89)
(log) capital	11.24 (2.13)	9.49 (2.13)	11.24 (2.14)	9.98 (2.14)	13.67 (1.88)	13.37 (1.88)
(log) number of firms	2.19 (1.14)	1.39 (1.14)	2.19 (1.20)	1.71 (1.20)	3.86 (1.21)	4.21 (1.21)
Panel (b). Year 2004						
(log) employment	5.59 (1.65)	4.61 (1.65)	5.59 (1.67)	4.95 (1.67)	7.47 (1.45)	7.86 (1.45)
(log) output	10.36 (2.13)	9.09 (2.13)	10.36 (2.17)	9.50 (2.17)	12.42 (1.89)	12.71 (1.89)
(log) capital	10.23 (2.19)	8.77 (2.19)	10.23 (2.20)	9.31 (2.20)	12.43 (1.94)	12.67 (1.94)
(log) number of firms	1.73 (1.07)	1.16 (1.07)	1.73 (1.11)	1.42 (1.11)	3.28 (1.17)	3.87 (1.17)

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

Table 3 The SEZ Effects: DD Estimation

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

*Note:* Observations are at the village-year level. *InsideZone* is an indicator variable for whether a village is inside a zone or not. Panels a to d report the estimation results using the natural logarithm of the employment, output, capital, and number of firms data. 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 are in parentheses. The 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 Estimations

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

*Note:* Observations are at the area-year level within 1,000 meters of a zone boundary. *InsideZone* is an indicator variable for whether the area is inside the zone or not. Panels a to d report the estimation results using the natural logarithm of employment, output, capital, and number of firms respectively. In column 1, the 2008 data is used for the BD analysis. Neighborhood fixed effects are included in the specification. In column 2, both 2004 and 2008 data are used for the BD-DD exercise. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. The 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. 0-1,000 meters vs. 1,000-2,000 meters	
	Within 2,000 meters	Within 500 meters	Outside the zone	Inside the zone
	(1)	(2)	(3)	(4)
Panel (a). Dependent variable: (log) employment				
<i>Dummy</i>	0.494*** (0.036)	0.412*** (0.044)	-0.023 (0.038)	-0.080 (0.056)
Observations	2,428	2,256	2,188	1,384
Panel (b). Dependent variable: (log) output				
<i>Dummy</i>	0.567*** (0.051)	0.447*** (0.062)	-0.073 (0.049)	-0.082 (0.074)
Observations	2,428	2,256	2,188	1,384
Panel (c). Dependent variable: (log) capital				
<i>Dummy</i>	0.584*** (0.050)	0.423*** (0.060)	-0.025 (0.051)	-0.159** (0.079)
Observations	2,428	2,256	2,188	1,384
Panel (d). Dependent variable: (log) number of firms				
<i>Dummy</i>	0.267*** (0.028)	0.203*** (0.035)	-0.010 (0.028)	-0.060 (0.044)
Observations	2,428	2,256	2,188	1,384

Note: Observations are at the area-year level. In columns 1 and 2, the regression samples consist of areas within 2,000 meters and 500 meters of a zone boundary as reported in the column headings. *Dummy* is an indicator variable for whether the area is inside a zone or not. In column 3, the regression sample include areas outside the zones: areas within 1,000 meters versus those within 1,000 to 2,000 meters. *Dummy* is an indicator variable for whether the area is within 1,000 meters or not. In column 4 the sample is composed of areas inside the zone. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. The 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 zones' total effect has been decomposed into the effects due to entrants and exiters, incumbents and relocaters. The BD-DD estimate of the SEZ effect based on the full sample is a weighted average of the BD-DD estimates for 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 Ratio

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

*Note:* Observations are at the area-year level. Capital and labor intensity are defined at the 4-digit level based on a capital-to-labor ratio above or below the median value in 2004. Panel c reports the difference attributable to SEZ residence. *InsideZone* is an indicator variable for whether the area is inside or outside a zone. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. In panels a and b, the standard errors are clustered at the zone level. In panel c, the standard errors are computed using the delta method. \*\*\*, \*\* and \* denote significance at the 1, 5 and 10% level respectively.

Table 8 Heterogeneous Effects by Market Potential

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

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

Table 9 Heterogeneous Effects of Accessibility

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

*Note:* Observations are at the area-year level. In panels a and b, spatially-integrated (spatially-isolated) zones are those with infrastructure accessibility index above (below) the median: a higher index indicates greater accessibility. Panel c reports the difference in SEZ policy impact on the spatially-integrated and spatially-isolated zones. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. In panels a and b, the standard errors are clustered at the zone level. In panel c, the standard errors are computed using the delta method. \*\*\*, \*\* and \* denote significance at the 1, 5 and 10% level respectively.



Figure A1 Searching for a Detailed Address with Google Maps



*Notes:* To obtain coordinates for firms which report detailed Chinese addresses, the address (for example, “157 Nandan Road, Xuhui District, Shanghai, China”) is searched for in Google Maps to get a red marker showing the specific location of the address. Once that location is confirmed, the latitude and longitude of the address is read from the map.

Figure A2 Searching for Villages and Communities with Google Maps



*Notes:* To obtain the coordinates for villages and communities the name of the village or community, is entered followed by the name of the town, city, and province to which the village or community belongs (for example, “Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China”). The specific location is then denoted by a red marker. Once that location has been confirmed, the latitude and longitude of the village or community is read from the map.

Table A1 Comparison between the Census and ASIF data

	<i>Economic Census</i>				<i>Annual Survey of Industrial Firms</i>			
	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:* p10 and p90 denote the 10th and 90th percentiles. Sources: Economic Census and Annual Survey of Industrial Firms for 2004 and 2008.

Table A2 BD-DD Estimation (Aggregating Firms with Detailed Addresses)

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

Note: Observations are at the area-year level within 1,000 meters of the zone boundaries. To address possible measurement errors, 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 a zone or not. Panels a to d report the estimation results using the natural logs of the measures of employment, output, capital, and number of firms. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. The standard errors are clustered at the zone level. \*\*\*, \*\* and \* denote significance at the 1, 5 and 10% level respectively.

Table A3 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 zones' total effect has been decomposed into the effect due to entrants and exiters, incumbents and relocaters. 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 where SEZ villages and non-SEZ villages are within the same town. The DD estimate of the SEZ effect based on the full sample is weighted by the DD estimates for each subsample. The weights used are the share of employment, output, capital, and firm numbers of the corresponding subsample in the full sample.

Table A4 Firms in 2004 and 2008

	All firms	Entrants	Total	Survivors			Exiters
				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 which exist in 2008, but not in 2004. "Survivors" refers to firms present in both 2004 and 2008. "Exiters" refers to firms which existed in 2004, but not in 2008. For firms with 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. For firms with duplicate IDs, firm name is used. Firms may receive a new ID because of restructuring, merger, or privatization. For a firm for which no observation with the same ID could be matched over time, information such as the firm name, location code, the name of legal person representative, phone number and so on is used in the matching process.

Table A5 Mechanism: BD-DD Estimations with Three Subsamples

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

Note: Observations are at the area-year level within 1,000 meters of the zone boundaries. The three panel data sets are constructed by aggregating the "Incumbents", "Entrants and Exitors" and "Relocaters" subsamples by area and by year. "Incumbents" refers to continuing firms with no zone status changes between 2004 and 2008. "Entrants and Exitors" refers to firms which existed in 2004 but not in 2008 (Exitors), and firms which did not exist in 2004 but did in 2008 (Entrants). "Relocaters" refers to continuing firms which changed their zone status from outside to inside the zone, or vice versa. *InsideZone* is an indicator variable for whether the area is within a zone or not. Area fixed effects and neighborhood-year fixed effects are included. Standard errors are in parentheses. The standard errors are clustered at the zone level. \*\*\*, \*\* and \* denote significance at the 1, 5 and 10% level respectively.