Place-Based Policies, Creation and Agglomeration Economies: Evidence from China's Economic Zone Program*

Yi Lu† Jin Wang‡ Lianming Zhu§

Tsinghua HKUST Osaka

This version: September, 2018

Abstract

Combining rich firm and administrative data, this paper examines the incidence and effectiveness of a prominent place-based policy in China: Special Economic Zones. Establishing zones is found to have had a positive effect on capital investment, employment, output, productivity and wages, and to have increased the number of firms in the designated areas. Net entry plays a larger role in generating those effects than incumbents. The special zone program's net benefits over three years are estimated to amount to about US$15.42 billion. Capital-intensive industries benefit more than labor-intensive ones from the zone programs.

Keywords: Place-based Policies; Agglomeration; Special Economic Zones; China

JEL Classification: H20, O25, R38

---

*We thank Prof. Matthew Shapiro, the Editor, and two anonymous reviewers for their careful evaluation of the paper and the substantial improvements that followed. We also thank Maitreesh Ghatak, Wen-Tai Hsu, Patrick Kline, James Kung, Pei Li, Gerard Padró i Miquel, Albert Park, Eric Verhoogen and the audience at various seminars for their comments. Wang acknowledges the financial support of a research grant from Hong Kong's Research Grants Council (no.16504715). Zhu acknowledges the financial support of research grants from JSPS KAKENHI (nos.16K17115, 18K12769 and 15H05728). All errors remain our own.

†School of Economics and Management, Tsinghua University, Beijing, 100084, China. Email: luyi@sem.tsinghua.edu.cn

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

§Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. Email: lianmingzhu@iser.osaka-u.ac.jp.

1 Introduction

Place-based programs—economic development policies aimed at fostering economic growth in a specific area within a larger 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 in the targeted area (Kline and Moretti, 2014a). Some economists are skeptical about the efficiency of such programs (Glaeser and Gottlieb, 2008; Glaeser, Rosenthal, and Strange, 2010). Firms may move from other regions to the targeted area and arbitrage away the benefits associated with the program without improving the welfare of local residents (Kline, 2010; Hanson and Rohlin, 2013). Still, agglomeration economies are considered an important rationale for policies that encourage new investment in a specific area (Kline and Moretti, 2014b; Combes and Gobillon, 2015).

Although there has been much research focused on such programs in the United States and 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? Do the economic gains substantially outweigh the costs? 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 policies in the United States and Europe still hold for them? Very little progress has been made in addressing these issues, largely because of a 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 study constitutes a novel step in that direction. Specifically, it documents micro-level evidence about the incidence and effectiveness of place-based policies in China's special economic zones (SEZs). SEZs are a prominent development strategy implemented worldwide (World Bank, 2008). They attempt to foster agglomeration economies by building clusters, increasing employment, and attracting technologically-advanced industrial facilities. China provides an ideal setting for exploring the effects of SEZs on regions and firms, which is of great policy relevance. In 1979 China launched its first four SEZs as an experiment in pragmatic and innovative policies. After their early success, China's horizon for SEZs has

[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.

gradually expanded from the coastal areas to the center and west. This study focuses on the wave of SEZs established between 2005 and 2008. In 2006, for example, 663 provincial-level SEZs were established in China, among which 323 were in coastal areas, 267 were in central areas, and 73 were in the west. That sample is more representative of the eventual spatial distribution than earlier waves, as it includes 42 percent of China's SEZs. Hence, estimates based on that sample have large-scale implications.

The analyses proceed in three stages. The first examines the local effects of an SEZ on the targeted area using a reduced-form approach. Welfare analysis is then applied to evaluate the SEZs' overall costs and benefits. Finally, the extent to which the effects of the zones depend on program features and the characteristics of the targeted localities is analyzed.

A key innovation has been the construction of a novel data set which merges comprehensive data on China's economic zones with two geocoded economic censuses in 2004 and 2008 covering all manufacturing firms. The zone data include the year in which each zone was established, the type of zone, and the villages located within its boundaries. The merged data set contains information on firms' ages, sectors, addresses, investment, employment, and output. That is supplemented, more importantly, with information on the firms' dynamics (entries, exits and continuing operations) and their geographic proximity to a zone. In all, 3,143,445 firms are covered. The individual firm data are aggregated to construct a panel data set by area and year. The data series cover two periods—two years before a zone's establishment and two years after—allowing an assessment of any effects of the SEZs on the targeted areas and providing interesting evidence about how various margins contribute to the impacts. Other important outcomes such as productivity and wages are analyzed using data from China's Annual Surveys of Industrial Firms (ASIFs) from 2004 to 2008. They cover all state-owned enterprises (SOEs) and non-SOEs with annual sales of more than five million yuan. This is the first time that the outcomes of interest for SEZs have been precisely measured on such a disaggregated level. It is also the first time that comprehensive geocoded information on Chinese firms has been compiled and analyzed in relation to SEZs, something which has not previously been possible in such studies. Only with very fine-grained geographic data can empirical analyses detect whether an SEZ has had a positive incremental effect on economic activity or simply displaced activity from an untreated area to an adjacent treated one.

The key challenge in identifying any causal effects of zone programs is selecting appropriate control groups, given the possible presence of spillovers. This study starts with a difference-in-differences (DD) analysis at the village level (the most disaggregated geographic unit and smaller than an SEZ) and then the county level. It compares the changes in performance among SEZ villages and counties with the changes among non-SEZ counterparts during the same period, conditional on a rich set of control variables. Beyond that, it investigates the robustness of the findings by checking parallel pre-trends between targeted and control areas using the ASIF data. As an alternative approach, the techniques of Neumark and Kolko (2010) and Briant, Lafourcade, and Schmutz (2015) are applied making use of the detailed information on firm location and zone boundaries. The discontinuity in treatment at the zone boundaries is exploited to combine a boundary discontinuity (BD) design with the DD setting (constituting a BD-DD analysis). Moreover, to examine any potential bias due to spillovers from nearby SEZ villages, we apply the unified estimation framework used by Miguel and Kremer (2004) and conduct concentric ring analyses following the lead of Kline and Moretti (2014b).

Moving beyond local effects, we compare the program's costs to the estimated magnitude of its impacts using a flexible estimation approach developed by Busso, Gregory, and Kline (2013) and Chaurey (2017). A back-of-the-envelope cost-benefit analysis is conducted. The distributional effects of the SEZ programs are considered, among which the main benefits include potential increases in firms' profits, workers' wages, and landlords' rental income. The corporate tax concessions that firms in SEZs typically enjoy are regarded as the main costs of the program.

The analyses yield three classes of results. First, the evidence shows that the SEZ program has had a significant and positive impact on the areas targeted. After two years, the SEZ areas have 58 percent more capital invested, 35 percent greater employment, and 49 percent larger output than the non-SEZ areas. The number of firms in the SEZs has increased by 29 percent. Productivity has increased by 1.5 percent on average within one year, and wage rates by 2.9 percent within two years, indicating agglomeration economies. There is relatively limited spillover in industrial activity between SEZs and the surrounding areas. The effects of SEZs mostly come from firm entering and exiting, with very limited effects from the previously existing firms.

Second, using the estimates from the back-of-the-envelope approach, the net present value of the benefits of the SEZ program during the 2006–2008 period is roughly US$22.33 billion. Comparing it with the total tax cost of US$6.91 billion, there are net benefits of US$15.42 billion from the zone program.

Finally, operating in a zone is most beneficial for firms in capital-intensive industries. Zones with better market potential or better access do not demonstrate significantly larger benefits. These findings are in line with the features of the SEZ programs, which typically subsidize capital investment.

This study 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. Givord, Rathelot, and Sillard (2013) examine the impact of *Zones Franches Urbaines* 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 Simpson (2007), like Briant, Lafourcade, and Schmutz (2015), uncover geographic factors that can account for heterogeneities in programs' effects, though in their study they find a significant impact of better market access. Along these lines, Rothenberg (2013) too emphasizes the role of transportation infrastructure in firm location choices, and hence the spatial distribution of economic activity. Chaurey (2017) has reported the only other study in a developing country (in India) that examines the impact of a location-based tax incentive scheme. That study finds that the program's heavy impact is driven by both the growth of existing firms and by the entry of new ones. There is no evidence of relocation or other spillovers between the treated and control areas. These Chinese findings are generally comparable with those of Chaurey, presumably due to the similar states of market development.

This study relates to a number of studies which have evaluated the aggregate effects of place-based policies in the presence of agglomeration externalities and inferred the implications for productivity and welfare, such as those of Busso, Gregory, and Kline (2013), and Kline and Moretti (2014b). But this study has been among the first to attempt a cost-benefit analysis of SEZs in the context of a developing economy. The significant net benefits estimated here could be closely linked to the institutional improvement brought by the SEZs in China.

This study also adds to industrial policy literature with special applications to Chinese SEZs. Rodrik (2008) has highlighted SEZs' utility as vehicles for China's integration into the global economy. Alder, Shao, and Zilibotti (2016) and Wang (2013) investigate the local (city-level) impact of SEZs on growth, capital formation, and factor prices, while Cheng (2014) estimates both the local (county-level) and aggregate impacts. The firm-level evidence developed in this study resonates particularly well with Wang's finding that the majority of the foreign direct investment attracted by the SEZs has been new activity rather than simply a reallocation from other non-SEZ areas (Wang 2013). Alder, Shao, and Zilibotti (2016) too detect no evidence of beggar-thy-neighbor effects on GDP. There have been a few other attempts to extend the studies of China's SEZs to micro-domains. Schminke and Van Biesebroecke (2013) investigate the extensive margin effect of national-level zones on firms' exporting performance. More recently, Zheng, Sun, Wu, and Matthew (2017) examine 110 national- and provincial-level industrial parks in eight major cities and any production and consumption spillovers they triggered.

Methodologically, this study relates to much previous work applying the geographic regression discontinuity designs (Holmes, 1998; 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; 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. Section 4 presents data in detail. Section 5 reports the baseline local SEZ effect estimates and addresses various econometric concerns, followed by evidence on the mechanisms in Section 6. Section 7 provides an analysis on the cost and benefit of this program. Section 8 investigates whether these effects are heterogeneous across industries, zones and firms of different size. The last section offers concluding remarks.

2 Background

In China, SEZs have been widely adopted as a key industrial policy aimed at increasing foreign and domestic investment, promoting international trade, and stimulating technological cooperation and innovation in a specific geographic area. China has two main categories of SEZs: national-level and province-level economic zones. The former have been approved by the central government and are more privileged, while the latter are provincial government initiatives. Geographically, national and provincial SEZs are mutually exclusive—a location cannot be both a provincial and a national SEZ at the same time. Each zone has an administration committee which administers the zone on behalf of the sponsoring government, handling project approval, local taxation, land management, finance, personnel, environmental protection, and security. Because SEZs are considered engines for economic growth, their success is linked to the political careers of regional government officials. That encourages their supervisors to strive for the best possible performance (Xu, 2011). SEZs enjoy a certain degree of independence and have authority to define, within limits, their own preferential policies. The most important preferential policies usually include the following (Wang, 2013; Alder, Shao, and Zilibotti, 2016):

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. Investors may lawfully obtain land development and business use rights through a contractual agreement, often after an auction or bidding process. To attract more industrial capital, SEZs set low land transfer fees (Wei and Zhao, 2009).[2] The duration of the agreement, the fees and the method of payment depend on the type of business. For example, in Guizhou province, for enterprises certified as high-tech the discounted rate could be only 25% to 35% of the regular fee. Export-oriented enterprises may receive a 10% to 20% discount on the normal fee. For infrastructure projects such as those improving transportation, telecommunications, water supply, energy supply or environmental protection the discount could be 20% to 30%.

3. *Special treatment in securing bank loans.* Foreign-invested enterprises are encouraged to make use of domestic finance for their investments. The banks prioritize their loan applications.

Compared to the place-based programs in developed countries, China's SEZs have several distinctive features. First, China as a developing nation faces more governance and financing constraints than its more-advanced counterparts. Before the SEZ program, China's business environment was typified by weak institutions, including poor protection of private property rights, limited financial resources and weak infrastructure. None of this stimulated entrepreneurship. Small-scale regional SEZs were established aimed at policy experimentation and innovation. Within the zones, better institutions were provided aimed at reducing pre-existing distortions and improving economic efficiency (Alder, Shao, and Zilibotti, 2016). The early SEZs represented the Communist Party's commitment to the market economy and property rights protection, at least within those areas. Financial resources were directed to the targeted areas. The SEZs worked constantly to improve their utilities, telecommunications, transport, storage, and other basic installations and service facilities. At the same time, policies in the non-SEZ regions remained basically unchanged (Rodrik, 2008). That reduced resistance to the reform by substantially reducing the number of losers (Xu, 2011). The SEZs constituted a new reform path implementing transitional and heterodox institutional reforms. They managed to provide efficient incentives while maintaining the rents of those who were politically powerful (Qian, 2003). Taken together, those features of China's institutional landscape imply that the welfare consequences of place-based policy interventions

[2] Development zones were mostly created by city governments on rural land expropriated at below-market prices. Within the zone boundaries, municipalities have acquired large tracts of collectively-owned land following a formal requisitioning procedure. The administration committee of the SEZ then develops the now state-owned land by resettling the residents, paying compensation, destroying old construction, and installing new infrastructure. Plots developed in this way were eventually transferred to the zone's enterprises. See Wei and Zhao (2009).

may be different from those in a more developed setting such as the US or Europe where the policy environment is closer to a first-best world, and more delicate equity-efficiency trade-offs may be involved.

Figure 1 shows the geographic distribution of the SEZs established in five waves over 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 national 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. 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 for a few years concentrated in inland cities. More recently, zone establishment has been more balanced geographically. Between 2005 and 2008, 338 SEZs were established in coastal areas, 269 in the central area, and 75 in the west. For detailed descriptions of these waves, see Appendix A.

[Insert Figure 1 here]

There are several types of SEZs, in which the preferential policies have different focuses (Alder, Shao, and Zilibotti, 2016; Zeng, 2010). Economic and technological development zones (ETDZs) are broadly-defined zones with a wide spectrum of investors. High-tech industrial development zones (HIDZs) are intended to promote high-tech industrialization (such as software writing, integrated circuit and communications equipment manufacturing, biotechnology research, and so on) and to foster technological enterprise-based innovation. HIDZs and ETDZs have some similar functions, and the line between the two types of zone is blurred. Specialized industrial zones (SIZs) are cluster-type industrial parks aiming to develop particular industries, which should be consistent with local comparative advantages. Bonded zones were set up with three objectives: export processing, foreign trade, and logistics supported by bonded warehousing. Although they are physically inside China's borders, they function outside the country's customs regulations. Export processing zones (EPZs) are similar to bonded zones but are solely for export processing (to develop export-oriented industries).

Table 1 shows the number of each type of zone established in the five waves. National-level SEZs are more diverse, with EPZs being the dominant type in the recent waves. Provincial zones are usually ETDZs, HIDZs, or SIZs. In the earlier years SEZ status was not granted randomly. According to State Council documents, the central government authorized municipalities to establish SEZs based on a favorable geographic location, favorable industrial conditions and good human capital (Wang, 2013). Such site selection may have systematically biased the results because any positive effects could primarily reflect successful initial targeting of better-endowed areas which would be more responsive to treatment (Allcott, 2015). However, as the SEZ program was later expanded to other areas, it tended to be more representative of the eventual spatial distribution and less subject to such biases. This study therefore focuses on the latest granting wave. To further alleviate estimation biases arising from interactions between the old and new zones, areas covering zones from earlier waves are excluded from the analyses in constructing the comparison group (see Section 3 for details).

[Insert Table 1 here]

Even within a given city, economic zones were not randomly located—another dimension of the site selection bias. For example, zones tends to be located far from central business districts where farmland is more available and the opportunity cost is lower. That non-random siting of SEZs presents challenges in identifying their effects. How to choose a comparable control group? In this study, several estimation strategies are applied to address the identification issue (see Section 3 for details).

3 Estimation Strategy

The village is used as the unit of analysis in the baseline analysis, as it is the most disaggregated geographic unit in the data, and smaller than an SEZ. Difference-in-differences estimation compares one village's performance before and after the founding of an SEZ (the first difference) with the changes among non-SEZ counterparts during the same period (the second difference). Specifically, we use the following specification

$$Y_{vt} = \lambda_v + \gamma D_v \times Post_t + \lambda_{ct} + (\mathbf{X}_v \times \lambda_t)' \boldsymbol{\eta} + \varepsilon_{vt}, \tag{1}$$

where Y_{vt} is an outcome (the logarithm of capital, employment, output, number of firms, total factor productivity, the wage rate) of village v in year t; D_v denotes the program status with 1 if village v had an SEZ program and 0 otherwise; $Post_t$ is the indicator of a post-SEZ period; λ_v is a village fixed effects term capturing time-invariant village characteristics such as geographic location; λ_{ct} is a county-year fixed effects term capturing macro shocks common to all villages within the same county in year t; X_v is a vector of baseline village characteristics (to be discussed later); λ_t is a year fixed effect; and ε_{vt} is the error term. To control for potential heteroscedasticity and serial correlation, the standard errors are clustered at the village level. To check for any sensitivity to using the village as the unit of analysis, we conduct the DD estimations using counties. They comprise many villages and are bigger than an SEZ.

The unbiased estimation of γ hinges on two conditions. First, conditional on the controls (i.e., λ_v, λ_{ct}, $X_v \times \lambda_t$), that the regressor of interest $D_v \times Post_t$ is uncorrelated with the error term ε_{vt}—the conditional mean independence assumption. The second condition is that there is no spillover from the treatment villages to the control villages. In the next two subsections, we discuss several strategies to examine these two conditions in our research setting.

3.1 Conditional Mean Independence

Conditional mean independence condition is verified using three approaches.

Conditional on baseline characteristics. SEZ villages (or counties) are first compared with the non-SEZ areas in terms of a wide range of baseline characteristics. The covariates can shed light on how SEZ and non-SEZ areas differed at the inception of the SEZ program. All of the baseline characteristics are then included in the DD estimations to rule out the influence of the pre-program differences between the treatment and control groups. If the results with and without baseline controls are largely similar, that is taken as indicating that the DD estimates are not severely biased by incomparability between the treatment and control groups (Altonji, Elder, and Taber, 2005).

Pre-treatment parallel trends. A valid identification strategy requires that the treatment and control groups follow similar pre-program parallel trends. To confirm this, additional years of village and county outcome data were collected, and any temporal trends are analyzed comparing the treatment and control groups at both the village and county levels.

BD-DD estimation. Despite the various specifications used, the DD estimations could still have suffered from the non-random selection of the SEZ sites. As an alternative estimation approach, a boundary discontinuity framework is employed based on physical distance, an approach pioneered by Holmes (1998) and Black (1999) and widely applied in previous studies (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 in this case is an SEZ boundary (Lee and Lemieux, 2010). The premise of the BD framework is that close to the zone boundary, treatment and control areas should have very similar underlying characteristics except for the zone policies (the regressor of interest).[3] Any discontinuity in the outcomes at a zone boundary is assumed to be attributable to the zone's effects.

The traditional BD framework is imbedded in DD analyses and BD-DD analyses are conducted which compare areas close to the zone boundary before and after the zone's establishment. That provides a further control for any time-invariant differences in location characteristics among areas close to the zone boundary.

The BD-DD estimation equation is

$$Y_{azt} = \lambda_a + \gamma^z D_{az} \times Post_t + \lambda_{zt} + \varepsilon_{azt}, \tag{2}$$

where Y_{azt} measures performance in area a within 1,000 meters of the boundary of zone z in year t. D_{az} is an indicator variable set equal to one if area a is inside zone z with zone policies, and 0 otherwise. λ_a is an area fixed effect capturing all time-invariant area characteristics. λ_{zt} is a neighborhood-year fixed effect capturing unobserved shocks common to both sides of the zone z boundary in year t. Including neighborhood-year fixed effects allows for flexible time trends across different zones. ε_{azt} is again an error term. To ensure conservative statistical inference, the standard errors are clustered at the zone level.

3.2 Spillovers

The aforementioned framework does not allow for any spillover of the SEZ program from SEZ villages to non-SEZ villages, but such spillovers may in fact take place. For example, SEZ villages may attract industrial activity from non-SEZ villages (e.g., through relocation). That would constitute a negative spillover, but firms in non-SEZ villages might at the same time learn from their competitors in SEZ villages, an example of a positive spillover. Two techniques are applied to check on the magnitude of any spillovers and the associated biases in the DD estimates of the SEZs' effects.

Direct estimation of the spillover effect. To directly test for any spillover between villages, we apply the approach used by Miguel and Kremer (2004). Assume that spillovers between villages mostly take place within a broadly-defined region and are less important between such regions. On that assumption the data for aggregated geographic units, specifically counties, are used.[4] The analyses take advantage of the rich variations in the proportion

[3] Additionally, areas close to a zone boundary were required never to have been within any old zone (established before 2006). That rules out any concern that the estimates could have been complicated by interactions between old and new zones.

[4] We do not consider cross-county spillovers because subsequent ring analyses suggest very little program impact beyond 20km from a treated village, and the average distance across Chinese counties is about 1,200km.

of counties that are treated as well as the percentage of villages within a county that are treated. Specifically, we use the following augmented specification

$$Y_{vct} = \lambda_v + \gamma^v \cdot (D_{ct} \times D_v) + \sigma \cdot D_{ct} + \lambda_t + (\mathbf{X}_v \times \lambda_t)' \boldsymbol{\eta} + \varepsilon_{vct}, \tag{3}$$

where $D_{ct} = D_c \times Post_t$; and D_c denotes the treatment status as 1 if county c had one or more villages in an SEZ and 0 if it had none.

In this regression, σ captures the spillover effect on the untreated villages within a treatment county, whereas γ^v is the additional direct effect on a treated village. Their sum, $\gamma^v + \sigma$, is the overall treatment effect on a treated village. The equation (3) shows that if possible spillover σ is not accounted for, bias arises in the program effect estimates from the DD specification (1).

Concentric ring analysis. Another approach to estimating the magnitude of any spillovers and the importance of any biases is to exclude non-SEZ villages adjacent to the SEZ from the control group (Kline and Moretti 2014b; see also Neumark and Kolko, 2010 for a discussion of using adjacent areas as controls). Without sufficiently detailed information about village boundaries, that involves first calculating the distance between any two villages, and then excluding from the control group first the non-SEZ villages located within 2km of an SEZ village, then those within 4km, continuing step-wise out to 20km. This results in a series of estimates which show how sensitive the estimates of an SEZ's effects are to the use of the adjacent area in the control group, and how important any spillover effects are.

To further explore the robustness of the DD estimates to any spillover effects, an alternative framework used by Zheng, Sun, Wu, and Matthew (2017) is applied. It permits separately identifying the treatment effect and spillovers on a set of non-SEZ rings around the SEZ villages. Specifically, the specification is

$$\begin{aligned} Y_{vt} \;=\; & \lambda_v + \gamma D_v \times Post_t + \sum_{n=1}^{10} \sigma_n Ring(2(n-1), 2n)_v \times Post_t \\ & + \lambda_{ct} + (\mathbf{X}_v \times \lambda_t)' \boldsymbol{\eta} + \varepsilon_{vt}, \end{aligned} \tag{4}$$

where $Ring(2(n-1), 2n)_v$ are dummy variables indicating whether or not village v is located in the nth ring that is between $2(n-1)$ and $2n$ kilometers from its nearest SEZ village, $n = 1, 2, ..., 10$. In this regression, γ is the treatment effect on the SEZ village and σ_n is the spillover externality effect on the nearby nth ring.

4 Data

Firm data. The data used in this study came primarily from the economic censuses conducted by China's National Bureau of Statistics at the end of 2004 and 2008. The advantage of census data over the ASIF data often used in similar studies (e.g., Hsieh and Klenow, 2009) is that it is more comprehensive, covering all manufacturing firms in China, while the latter includes all SOEs and non-SOEs with annual sales of more than five million yuan. Table A1 in the appendix compares those 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 figures than the ASIF data.

The census data contain firms' full basic information, such as address, location code (a 12-digit code corresponding to a village or community), industry affiliation, and ownership. The address and the location code are used to locate a firm geographically and identify whether or not it is in a zone. The census data report employment, output, and capital for each firm.

Although the economic census achieves complete coverage of manufacturing firms, it has two shortcomings. First, the data were collected in two waves, only one of them in the pre-treatment period. That prevents comparing pre-treatment trends between the targeted and control areas. Also, the data set includes only three firm-level outcomes, from which the productivity and price impacts of SEZs cannot be directly inferred. To overcome these issues and provide comprehensive assessments of any SEZ effects, the analyses are augmented using the ASIF data to check for parallel pre-trends, investigate other important outcomes (such as changes in TFP and wages), and conduct cost-benefit analyses. In particular, the ASIF figures from 1998 (the first year to have the data) to 2008 are used for county-level parallel trend tests and those from 2004 (the first year to have detailed village administrative codes) to 2008 are used for village-level common-trend tests. The ASIF figures from 2004 to 2007/2008 are used to examine TFP and wages at both the village and county levels.[5] ASIF figures are also used to conduct the cost-benefit analyses at the county-level along the lines pioneered by Busso, Gregory, and Kline (2013) and by Chaurey (2017).

Firm SEZ status. The census data did not directly report information about each firm's SEZ status. To identify whether or not a firm is located within an SEZ, a comprehensive SEZ data set from the China's Ministry of Land and Resources is consulted. It defines SEZ boundaries in terms of villages, communities, and sometimes roads. Based on that information, maps are consulted to determine whether or not a village or community is

[5] In the 2008 ASIF data, no information on materials or value-added is available. That makes calculating TFP impossible for 2008, so the TFP analyses use data from 2004 to 2007.

within the boundary. The SEZs' official websites often report detailed information about the villages and communities within their administrative boundaries, and they too are consulted. The National Bureau of Statistics and the Ministry of Civil Affairs also report administrative divisions and codes at the village and community level on their websites. For some economic zones this includes information on 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, the firms' addresses as well as their 12-digit location codes are used. (See Appendix B for a detailed discussion.) To verify that approach, the results are checked by matching them against the SEZ names which some firms include in their addresses.[6]

Coordinates data. In the BD-DD analyses, the outcomes of individual firms are aggregated into areas close to the zone boundaries. This requires precise geographical information about firms' locations (specific coordinates) to determine each firm's distance from a zone boundary. The firms' addresses are used with Google's geocoding API to obtain their geographic coordinates.[7] 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, the firm's latitude and longitude are extracted from the Google map. This process allows determining the coordinates of approximately half of the firms.

To deal with incomplete addresses (those with only information on the village, building, or street name, but with no number or building name), road name changes and reporting errors, the remaining firms are searched for using their 12-digit location codes.[8] For example, a firm with the inexplicit address "Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China" also had a 12-digit location code of "320585102202" which corresponds to the more specific "Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China". The name of that village or community could then be used to collect the latitude and longitude information from Google maps (see Figure A2).

Village and county baseline characteristics. The village-level baseline characteristics include a village's distance from an airport and port, the capital-to-labor ratio and the number

[6] If an SEZ boundary bisected a village or community, only part of it would be in the zone. But this is not a concern in China where the local governments survey and appraise land and outline plans for future development based on village and community units.

[7] The robustness of these results is checked using Baidu's geocoding API service. Baidu is the Chinese version of Google. It provides a similar service, but uses a different coordinate system.

[8] There are approximately 700,000 villages and communities in China. The nation's habitable area is about 2.78 million square kilometers. On average, a village or community covers about 4 square kilometers. 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 village or community coordinates when firms do not provide a detailed address.

of firms in the village in 2004, all aggregated from the 2004 census data. The counties' baseline characteristics come from the provinces' 2004 statistical yearbooks and the China Population Census for 2000. Those sources report a rich set of variables including land area, total population, employment, GDP, the share of rural employment, export intensity, the ratio of government expenditure to revenue and much else, all in 2004. Cumulative GDP growth rate between 1998 and 2004 is also collected, along with the share of employment in agriculture and that in manufacturing, the mortality rate, and the share of eliminated illiteracy population, all as of 2000.[9] The airport and port distances, capital-labor ratios and number of firms in 2004 come from the 2004 census data.

Regression data. These analyses focus on SEZs established between 2004 and 2008. There were 682 SEZs established during that period (19 in 2005 and 663 in 2006), and there was substantial geographic variation. There were 338 SEZs established in the coastal area, 269 in the central area, and 75 in the west. Nineteen were national-level zones, most of which were EPZs. 615 were province-level ETDZs, 20 were province-level HIDZs, and 28 were province-level SIZs. National-level zones are excluded from the analyses because of the concern that they might not be fully comparable with province-level zones and because they are mostly EPZs within pre-established ETDZs—an overlapping problem.[10]

For the DD analyses, individual firms are aggregated to construct panel data sets by county (or village) and by year. Thus, each county (or village) has two observations in 2004 and 2008 in the DD estimation, a year of data before and a year of data after the zone's establishment. For the county-level regressions, the sample consists of 1,582 counties: 362 SEZ counties and 1,220 without an SEZ. For the village-level regressions, the analysis is restricted to SEZ villages and non-SEZ villages in the same county. The resulting sample comprises 59,949 villages in 580 counties: 3,963 SEZ villages and 55,986 non-SEZ villages.

The BD-DD analysis involves calculating each firm's distance from the nearest SEZ boundary. The coordinates of each firm's location have been established, but accurate geocodes for each SEZ's boundaries are not available, which prevent calculating the distance to the boundary directly.[11] Instead, the approach used by Duranton, Gobillon, and Overman

[9] It would have been ideal to have data on all of the control variables for 2004, but data availability requires augmenting the analysis with data on some covariates in 2000.

[10] In 2005, 19 national-level zones were approved by the central government, of which 18 were EPZs. Such national-level zones have higher-level administration committees than provincial-level SEZs, and their committees enjoy more authority in managing the zones. Those EPZs were mostly in pre-established ETDZs by design. To take the Huizhou Export Processing Zone as an example, it is located within the Guangdong Huizhou ETDZ, which was established in 1997. The DD and BD-DD identification techniques would not be applicable in that situation, as the pre-existence of the ETDZ confounds the effect of the newly approved EPZ. See Wang (2013) for more details.

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

(2011) is applied to determine the distance indirectly. To determine whether a firm is located within 1,000 meters of a zone boundary, a search within a radius of 1,000 meters of the firm is conducted,[12] as illustrated in Figure A3. If firm A is located outside a zone and 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 another firm (firm D) is found to be located outside the zone but within 1,000 meters of C, it is designated as located within 1,000 meters of the zone boundary.

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.[13] Those firms are then aggregated to construct a panel data set by area and by year. Each zone's 1,000 meter neighborhood comprises two areas—inside and outside the zone—and each has two observations in 2004 and 2008. The regression sample for estimation comprises 1,174 areas.

5 Empirical Findings

5.1 Comparing SEZ and non-SEZ areas

Table 2 lists the treatment and control group means for a variety of county and village characteristics in the initial year (all measured before the onset of the program), including a village's distance from an airport and port, capital-labor ratio and the number of firms. For counties it reports their total population, employment, GDP, the share of rural employment, export intensity, ratio of government expenditure to revenue, GDP growth rate, share of employment in agriculture and that in manufacturing, mortality rate, share of eliminated illiteracy population, land area, distance from an airport and port, capital-to-labor ratio and the number of firms.

[Insert Table 2 here]

SEZ areas tend to be closer to an airport and to have more manufacturing firms and capital than non-SEZ areas whether analyzed on a county or village basis. But Column 3 shows that on many dimensions there were significant differences between the SEZ and non-SEZ counties. The counties with an SEZ were on average more densely populated, economically better-developed and had greater fiscal capacity. They were also more accessible

[12] On average, a village or community in China covers about 4 square kilometers. Assuming villages and communities are circular allows calculating an average radius for a village or community of about 1,000 meters, which is why a range of 1,000 meters from a zone boundary is used in the analyses.

[13] The analysis is restricted to SEZs which had firms on each side of their boundary in both 2004 and 2008.

from a port. All of those baseline characteristics are included in the empirical analyses, interacted with year dummies in the DD estimations to control for the presence of the pre-program differences between the treatment and control groups.

5.2 Baseline Estimates

Table 3 presents the DD estimates using the village as the unit in the regressions. All of the regressions control for county-year fixed effects and village fixed effects.

[Insert Table 3 here]

Panels A and B report the results with and without controlling for baseline village characteristics (as illustrated in Table 2). Four outcomes reported in the two economic censuses are considered in columns 1–4: capital, employment, output, and the number of firms. The logarithms of those outcome data are presented to highlight the magnitude of the effects. The estimated coefficients of all four outcomes are consistently positive and statistically significant, suggesting that after the zones' establishment the SEZ villages gained investment, employed more labor, and produced more output than the non-SEZ villages. As would be expected, since they attract more firms. Meanwhile, panels A and B exhibit the same estimation patterns, with the former having slightly small magnitudes. These results suggest that the DD estimates may not have been entirely driven by the pre-program differences between SEZ and non-SEZ villages.

Given the limited accounting information collected in the economic censuses, the analyses of the efficiency impacts of the SEZ policies use the supplementary ASIF data from 2004 to 2008 (the years for which the 12-digit location codes were available).[14] Specifically, the ASIF data are applied in examining the relationship between SEZ status and total factor productivity and wage rates. TFP is estimated for each firm using the approach of Ackerberg, Caves, and Frazer (2015). They are then averaged for each village weighted using employment (see Appendix C for details of the firm productivity estimation). Firms' average wage rates are similarly aggregated to the village level, also weighted by employment. As is shown in columns 5 and 6, both estimates are both statistically and economically significant. These results suggest that after the establishment of the zones, the SEZ areas witness an increase in productivity. And firms in the zones pay higher wages than those outside.

To check the sensitivity of the results to the geographic unit used in the analysis, we report county-level DD estimates. The results are presented in Table 4. All of the regressions

[14]Table A2 in the Appendix shows consistent SEZ program effects on capital, employment, output and the number of firms using the ASIF data.

include year fixed effects and county fixed effects.

[Insert Table 4 here]

The estimated coefficients are consistently positive and statistically significant. They show consistent patterns in the two panels, suggesting that the DD estimates are not driven by the pre-program differences between SEZ and non-SEZ counties (i.e., baseline county characteristics as in Table 2). While the county-level DD estimates are smaller than the village-level ones, the differences reflect some muting due to the inclusion of non-SEZ villages in the SEZ counties. Without spillovers, the county-level DD estimator is simply the village-level version weighted by the outcome shares of the SEZ villages in the county.

To calculate the economic magnitude, we use the estimates from panel B of Table 3. Specifically, two years after the establishment of a zone, capital investment has increased by 57.9 percent on average, employment by 34.5 percent, output by 49.2 percent, and the number of firms by 29 percent. Those results compare well with those of previous studies. For example, Givord, Rathelot, and Sillard (2013) find that the French Zones Franches Urbaines program has 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 United Kingdom's employment and investment promotion program.

After one year of the SEZ program, productivity have improved by 1.5 percent on average. Wage rates have increased by an average of 2.9 percent after two years. Our findings are broadly consistent with other scholarly work on the agglomeration benefits of policies encouraging new business investment in a targeted area (Neumark and Simpson, 2014). In particular, Zheng, Sun, Wu, and Matthew (2017) analyze the impact of 43 state-level and 67 provincial-level industrial parks in China and find that they generate TFP increases and wage premiums within the targeted area.[15] The state-level parks had a larger effect on TFP than the provincial-level ones, but there was little difference in their effect on wages. That is consistent with the findings here. But our findings do contrast with those of Alder, Shao, and Zilibotti (2016) whose city-level analyses show that only state-level SEZs have a large and positive effect on local GDP per capita. They find that the effect of provincial-level SEZs is not significant. However, in comparing results from multiple studies, one should bear in mind that they differ greatly in the period studied, their primary samples (zones) and the data used.

[15] In that study the plants (including incumbent plants and new entrants) in the parks are 25.7% more productive after the introduction of the park. The wages in the parks are 12.7% higher after a park's creation.

5.3 Validity Checks

This subsection provides two tests on the DD identifying assumption as discussed in Section 3.1; that is, the check on the pre-trends, and the BD-DD estimation.

Check on the pre-trends. The ASIF data are used to analyze the pre-trends and verify the common trend assumption central to DD analysis. Specifically, the data from 2004 to 2008 are used for village-level DD analyses, and the regressor $D_v \times Post_t$ in the equation (1) is replaced by $D_v \times \lambda_t$. Ideally, the data series should be extended to earlier years to include a longer pre-treatment period, but the ASIF only in 2004 started to report the 12-digit location codes essential for pinning down the firms' villages. The regression results are reported in Table A3. Among all of the outcomes, the coefficients are small in the magnitude and mostly insignificant in 2005. But they become statistically significant and of greater magnitude from 2006. These results confirm that SEZ and non-SEZ villages had comparable trends prior to the granting event (before 2006) and that remarkable differences started to emerge right after the zones were established.

The analyses of the county-level pre-trends use data from 1998 (the first year in the ASIF data) to 2008. The regression results are displayed in Figure 2, in which 2005 (the year before the SEZs' establishment) is used as the reference year.

[Insert Figure 2 here]

For capital, employment and output, all the coefficients before 2006 are negative and largely insignificant, sharing similar small magnitude, but they become positive and gradually increase in magnitude after 2006. For the number of firms, TFP and wage rates, the coefficients show a similar pattern. These results also support the assumption of similar trends in the SEZ and non-SEZ counties before the SEZ program.

BD-DD estimates. Table 5 shows the coefficients describing the impact of an SEZ program as estimated using the BD-DD framework with a sample of areas within 1,000 meters of a zone boundary. Consistent with the baseline estimates, the analyses show statistically and economically significant effects of the SEZ program. Moreover, the estimates are not sensitive to the inclusion of additional controls. The magnitudes of the coefficients are comparable to those from the baseline DD estimations, lending further support to the utility of this estimation framework.

[Insert Table 5 here]

5.4 Spillovers

One might be concerned that these DD estimates could be biased due to spillovers from SEZ villages to non-SEZ ones. The two sets of exercises laid out in Section 3.2 are used to address that possibility.

Direct estimation of the within-county spillover effect. Equation (3) in Section 3.2 allows the direct estimation of any spillover. The results are reported in Table 6. For employment, output and the number of firms, the spillovers are positive and significant, albeit small, suggesting a slight downward bias of the SEZ effects in the baseline estimations. There are no significant spillovers with respect to capital, productivity or wages. After correcting for the minor spillovers, sizable and positive effects of the SEZ program remain. Overall, although spillovers are considered an important concern in the place-based policy literature, they do not appear to have been empirically first-order in this Chinese setting.

[Insert Table 6 here]

Concentric ring analysis. The concentric ring approach elaborated in Section 3.2 allows a further check on spillovers in the village-level DD estimations. It generates a series of estimates $\hat{\gamma}_j$ (where $j \in \{1, 2, ..., 10\}$) corresponding to the exclusion of non-SEZ villages located within $2j$ kilometers of an SEZ village. Figure A4 in the Appendix plots the estimated coefficients $\hat{\gamma}_j$ with their 95 percent confidence intervals. Overall, using different sets of control villages yields positive and significant estimates and some evidence that is consistent with small positive spillovers from SEZ villages to nearby ones. Excluding adjacent non-SEZ villages does not substantially affect the estimates. Since spillover diminishes with distance from an SEZ village, these results suggest limited impact of spillovers.

That pattern is corroborated by the estimates of Equation (4). Figure A5 in the Appendix shows that the estimated treatment effects of being an SEZ village are quantitatively comparable to the benchmark estimates in Panel B of Table 3 and to the spillover estimates in Table 6. The estimated impact of spillover externality is negligible in terms of capital, employment, output and number of firms, though there are small spillover effects in terms of productivity and wages. The F-tests of the joint significance of the spillover effects show that they are statistically insignificant in all outcomes except for productivity.

Taken together, these exercises suggest that the baseline estimates of the SEZs' effects are not significantly biased by the relatively small spillovers from SEZ villages to non-SEZ villages nearby. In a closely related study, Zheng, Sun, Wu, and Matthew (2017) show that industrial parks generate net growth rather than a simple reshuffling of economic activity from the rest of the city. Our findings largely echo theirs but with notably less spillover.[16] In other related work, Chaurey (2017) examines the micro-level impact of a location-based tax incentive scheme in India. He reports finding large increases in employment, total output, fixed capital, and the number of firms, and finds no evidence of firms' relocating or of spillover of industrial activity between the treatment and control areas. His results are generally consistent with those of this study, probably because the two countries have similar levels of market development.

Using more aggregated city-level data, Wang (2013) has found that the majority of the direct investment attracted to China from overseas by the SEZs between 1978 and 2008 is new activity rather than simply relocation from non-SEZ areas. And Alder, Shao, and Zilibotti (2016) also find no evidence of beggar-thy-neighbor effects on GDP using data from 1988 to 2010. They report positive spillover from the state-level zones to nearby cities, which become stronger during the first 10 years, but no significant effect of provincial-level zones. Overall, these city-level findings to some extent confirm the limited negative spatial interactions detected in this study, along with the fairly small spillovers from provincial-level SEZs to nearby areas, at least in the short run.

6 Decomposition

These findings establish that SEZ areas had more invested capital, employment and output, attracted more firms, and showed increased productivity and higher wage rates than non-SEZ areas in the period studied. China's SEZ program significantly reduces the costs of capital and land and tax rates within the zone areas, which seems to have significantly influenced firms' location choices and investment decisions.

Firms could have responded to the SEZ policy initiatives by varying inputs and outputs (along the intensive margin) and by entering or exiting a zone (along the extensive margin). To further illuminate such decisions, the SEZ effects are decomposed into an extensive margin effect attributable to new entrants and exiters and an intensive margin effect generated by continuing firms.

6.1 Framework

The decomposition method developed by Foster, Haltiwanger, and Krizan (2001, 2008) is applied to the TFP and wage rates data. It separates the changes in the weighted average

[16] Zheng, Sun, Wu, and Matthew (2017) find that in an impact area extending 2km from an industrial park's boundary, the spillovers in terms of TFP and wages are 12% and 6%, respectively.

of firm-level productivity (or wages) into five margins. TFP will serve as an illustration. For each village v (the index is omitted), aggregate TFP in year t is

$$TFP_t = \sum_i \varphi_{it} tfp_{it},$$

where φ_{it} is the employment share of firm i in year t; and tfp_{it} is firm i's TFP in year t. The changes in aggregate TFP before and after the implementation of the SEZ program can then be decomposed as

$$\begin{aligned} \Delta TFP_t &= \underbrace{\sum_{i \in C} \varphi_{it-1} \Delta tfp_{it}}_{within} + \underbrace{\sum_{i \in C} \left(tfp_{it-1} - TFP_{t-1}\right) \Delta\varphi_{it}}_{between} + \underbrace{\sum_{i \in C} \Delta tfp_{it} \Delta\varphi_{it}}_{cross} \\ &\quad \underbrace{+ \sum_{i \in N} \varphi_{it} \left(tfp_{it} - TFP_{it-1}\right) - \sum_{i \in X} \varphi_{it-1} \left(tfp_{it-1} - TFP_{it-1}\right)}_{net\ entry}, \end{aligned} \tag{5}$$

where C, N, and X, represent the sets of continuing, entering, and exiting firms, respectively. The terms on the right side of equation (5) represent, in the order of their inclusion, the within-firm effect, the between-firm effect, the cross-effect, and the net entry effect.

The decomposition of outcomes like capital, employment and output is a bit different, as it tracks changes in the logarithm of the total summation. The technique embodies the spirit of Foster, Haltiwanger, and Krizan's (2001, 2008) method. The treatment of output will serve as an illustration. For each village v,

$$\ln S_t = \ln\left(\textstyle\sum_i s_{it}\right).$$

Hence, the decomposition becomes

$$\begin{aligned} \Delta \ln S_t &= \Delta \ln \left(\sum_i s_{it} \right) \\ &= \Delta \ln \left(\sum_{i \in C} s_{it} + \sum_{i \in N/X} s_{it} \right) = \Delta \ln \left(S_t^C + S_t^{N/X} \right) \\ &\simeq \underbrace{\varphi \Delta \ln S_t^C}_{within} + \underbrace{(1-\varphi) \Delta \ln S_t^{N/X}}_{net\ entry}, \end{aligned} \tag{6}$$

where $\varphi = \left(\frac{S_{t-1}^C}{S_{t-1}} + \frac{S_t^C}{S_t} \right)/2$. The two terms on the right side of equation (6) represent the within-firm effect and the net entry effect, respectively.

With these decompositions, each term (ΔY_v) can then be regressed on the SEZ program indicator along with the baseline controls, i.e.,

$$\Delta Y_v = \gamma^v D_v + (\lambda_{ct} - \lambda_{ct-1}) + \mathbf{X}_v' \boldsymbol{\eta} + \Delta\varepsilon_v, \tag{7}$$

which isolates how much each decomposition term contributes to the SEZ's total impact.

6.2 Results

The decomposition analyses require distinguishing the continuing firms, entrants and exiters. Each group is traced from 2004 to 2007/2008 using the firm IDs, names, addresses, and other information to carefully pin down the firm dynamics.[17] The capital, employment, and output decompositions use the 2004 and 2008 census data. For productivity decomposition the ASIF 2004 and 2007 data are exploited. For wage rate decomposition the ASIF data for 2004 and 2008 are used.

The decomposition results for capital, employment and output are presented in panel A of Table 7. Column 1 reports the total effects of the SEZ program. Columns 2 and 3 are the within-firm effects and the net entry effects, respectively. Most of the SEZ effects arise through firm births and deaths. They account for 80 percent of the changes in capital and employment, and 90 percent of the changes in output. The decomposition results for TFP and wages are reported in panel B, with column 1 for the total effects, and columns 2–5 for the within-firm effects, the between-firm effects, the cross-effects, and the net entry effects, respectively. Within-firm effects and between-firm effects are negligible. The net entry effects are statistically and economically significant, consistent with the patterns of capital, employment and output.

[Insert Table 7 here]

Overall, this decomposition indicates that the zones mostly promoted extensive margin effects. This result agrees with the findings of Givord, Rathelot, and Sillard (2013) who find no evidence of an employment effect on existing businesses. Employment growth in their sample comes mostly from new businesses and firms which relocated. Criscuolo, Martin,

[17] For firms which reported unique IDs (their legal person codes) in the census data, the tracing involves matching their firm ID in the 2004 and 2008 censuses. For firms with multiple 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 identified, as much information as possible on the firm's name, location code, the name of its legal representative person, phone number, and so on is used to find a match. A similar approach is applied in tracing firms in the ASIF data for 2004 and 2007/2008.

Overman, and Van Reenen (2012) do, though, find a large and statistically significant average effect of the United Kingdom's RSA program on employment and investment, with about half of the effects arising from incumbent firms growing (the intensive margin) and half caused by net entry (the extensive margin). However, in interpreting our results about the extensive margin, we caution that some firm births could be considered relocations if the SEZs attracted new-born firms from other regions. Some investors may simply have changed their location choices in establishing a new firm in response to an SEZ.

7 Cost-Benefit Analysis

The previous analyses document beneficial effects of the SEZ program, but the aggregate welfare implications of the program remain unclear, given its costs and possible redistribution of economic activity. We investigate that important issue using a flexible back-of-the-envelope cost-benefit estimation technique proposed by Busso, Gregory, and Kline (2013) and subsequently applied by Chaurey (2017). The SEZ program's main benefits have included increasing firms' profits, raising workers' wages, and generating rental income for landlords. It is shown that in this sample the increase in housing rents is negligible,[18] so the analysis focuses on wages and profits.

The effects of a zone on corporate profits ($\pi^{corporate}$) and wage bills (π^{wage}) are estimated using the county-level DD estimation. Table 8 reports the results.

[Insert Table 8 here]

Based on the estimated zone effect ($\hat{\gamma}^{corporate}$; $\hat{\gamma}^{wage}$), the counterfactual corporate profits ($\tilde{\pi}^{corporate}$) and wage bills ($\tilde{\pi}^{wage}$) are calculated as: $\tilde{\pi}^{corporate} = \pi^{corporate} / \left(1 + \hat{\gamma}^{corporate}\right)$ and $\tilde{\pi}^{wage} = \pi^{wage} / \left(1 + \hat{\gamma}^{wage}\right)$. Here, $\hat{\gamma}^{corporate}$ and $\hat{\gamma}^{wage}$ are the estimated zone effects on corporate profits and wage bills. The program's benefits can then be expressed as the total difference between the actual and counterfactual figures: $\pi^{corporate} - \tilde{\pi}^{corporate}$ for profits and $\pi^{wage} - \tilde{\pi}^{wage}$ for wages.

Table 9 shows that the actual corporate profits are 168.67 billion RMB in 2006, 223.76 billion RMB in 2007 and 253.29 billion RMB in 2008. The estimated benefits increases in corporate profits linked to the zone program are then 23.20 billion RMB in 2006, 30.77 billion RMB in 2007 and 34.83 billion RMB in 2008. Similarly, the actual wage bills are

[18] Data on housing costs from 2004 to 2008 are only available on the city level (one administrative level above a county), so the analysis of any SEZ effect on housing costs is conducted on the city level. Table A4 in the Appendix shows that there is no significant effect. That result is consistent with Wang's analyses of SEZs from 1978 to 2008 which find that any increase in housing rents is negligible (Wang, 2013).

155.23 billion RMB in 2006, 193.52 billion RMB in 2007 and 261.69 billion RMB in 2008. So the calculated benefits in terms of wage bills are 24.96 billion RMB, 31.11 billion RMB and 42.08 billion RMB. Adding the two categories of benefits and using a discount rate of 3%, the total gains from the zone program are roughly 178.62 billion RMB (or US$22.33 billion).

[Insert Table 9 here]

The corporate tax forgone can be estimated as the difference between the counterfactual taxes and actual taxes paid. The counter-factual taxes can be estimated as counterfactual corporate profits ($\tilde{\pi}^{corporate}$) $\times$ the statutory tax rate (33 percent in 2006 and 2007; 25 percent in 2008). As Table 9 shows, the estimated corporate tax breaks are 15.87 billion RMB in 2006, 23.30 billion RMB in 2007 and 18.36 billion RMB in 2008. Using a discount rate of 3% produces a total cost of 55.30 billion RMB (US$6.91 billion). Comparing the costs and the benefits yields an estimated net benefit of 123.32 billion RMB (or US$15.42 billion) from the zone program.

8 Heterogeneous Effects

Operating in an SEZ could have different impact on firms with different characteristics, operating in different zones and different industries. Industry variations in capital-labor ratios, zones' different access to transportation and different firm sizes allow examining possible differences in the impact. All of these tests are conducted with the village–level data using the same set of controls as in the benchmark village-level DD analysis.

Capital-intensive vs. labor-intensive industries. Because of the capital cost reductions available, firms in capital-intensive sectors may have been more likely to benefit from a zone program and to have derived greater benefits. To investigate that possibility, the industries are categorized based on whether their average capital-labor ratios in 2004 were above or below the sample median. The results estimating the differential impact on the two groups are reported in Panel A of Table 10 (see Table A5 in the Appendix for the estimates from the two subsamples).

[Insert Table 10 here]

Most of the SEZ effects are indeed consistently stronger in the capital-intensive industries. In absolute terms, after the implementation of an SEZ there is on average 10.6 percent more capital investment and 10.9 percent larger output among the capital-intensive industries compared to the labor-intensive ones. Those results are in line with the features designed into the SEZ programs, which typically subsidize capital investment.

Good vs. poor infrastructure. Most firms trade in multiple markets. The level of economic activity in a location depends in part on that location's access to markets for its goods (Hanson, 2005). Good airports and highways help reduce firms' trade and communication costs, so proximity to markets and good infrastructure should make a zone more attractive (Graham, Gibbons, and Martin, 2010; Combes and Gobillon, 2015). To investigate whether there are such differential effects, an infrastructure index is constructed for each SEZ county. It includes the county's distance from the nearest airport and port, the highway density of the city in which the county resides, and the county's market potential. For each SEZ county, its distance from the nearest airport and port is measured, and those distances are ranked from greatest to least. That yielded the sub-indices ($rank\_airport$ and $rank\_port$). The county's highway density is treated similarly to obtain the sub-index ($rank\_highway$).[19]

A county's market potential is then quantified using a technique in the spirit of Harris (1954) and of Rogers (1997). The impact of trade and communication costs is assumed to increase with the inverse of a county's distance from all the other counties within the same province. The market potential MP_s of SEZ county s is therefore defined as

$$MP_s = \frac{\sum_{c \in PROV} GDP_c / dist_{cs}}{\sum_{c \in PROV} GDP_c},$$

where $PROV$ denotes a province, c denotes a county, GDP_c stands for county c's gross domestic product (GDP), and $dist_{cs}$ is the distance between SEZ county s's administrative headquarters and county c.[20] Following Briant, Lafourcade, and Schmutz (2015), the weighted sum of the markets accessible from an SEZ county is divided by the total size of all of the markets in the province to mitigate the impact of large counties. The SEZ counties' market potentials are then ranked from lowest to highest, resulting in the fourth sub-index ($rank\_mp$).

A zone's infrastructure index is then the average of the ranks associated with the four dimensions: $rank = (rank\_airport + rank\_port + rank\_highway + rank\_mp)/4$, so a larger index indicates a zone with better infrastructure. The SEZ counties are then divided into two groups based on whether their infrastructure index in 2004 was above or below the

[19] The list of airports is compiled from China's 2005 Transportation Yearbook, while the data on highway density (miles of highways divided by the land area of the city) are from China's 2005 Regional Statistical Yearbook.

[20] Note that both Harris (1954) and Rogers (1997) use a city as the regression unit, and their market potential for a city is the weighted average of the GDPs of the other cities. In China, economic zones are smaller units than counties. The county where an economic zone resides is therefore also included in the calculation of market potential.

sample median.

The estimation results for the difference in policy impact on the two groups are reported in Panel B of Table 10 (see Table A6 for additional details). No statistically or economically significant differences in the SEZ effects are found between the zones with good and poor infrastructure. That implies that accessibility and the surrounding region's market potential are not critical factors in determining a zone's impact.

Firms of large vs. small size. To investigate whether firm size is important in determining a zone's beneficial effects, the firms are sorted into two groups based on whether their total sales in 2004 are above or below the sample median. They are then aggregated to the village level for estimation. The differential impact between large firms and smaller ones is reported in Panel C of Table 10 (see Table A7 for additional details). The SEZs tends to attract larger firms. That too echoes the program's design, in that larger firms are more likely to make large capital investments. But there are no statistically significant differences between small and large firms in terms of the other outcomes.

Taken together, these results indicate that capital-intensive firms benefit more from the zone program than labor-intensive ones, but the effects of an SEZ are quite similar regardless of an SEZ's accessibility and for firms of different size. That resonates with the findings of previous work which has emphasized the characteristics of the industry in analyzing the effects of place-based policies (Criscuolo, Martin, Overman, and Van Reenen, 2012; Combes and Gobillon, 2015). But these findings contrast with previous findings on the role of regional characteristics, for example those of Briant, Lafourcade, and Schmutz (2015). They do though compare well with the work by Alder, Shao, and Zilibotti (2016) who find in their study of Chinese SEZs that market access had no significant relationship with a city's GDP. One possible explanation is that the later SEZs established in the wave of 2006 were less subject to selection compared to the earlier waves. The characteristics of their locations may not therefore have differed as much. Overall, these findings suggest that the complementary roles of regional and industry characteristics in place-based development programs may hinge on the specific context.

9 Conclusions

This study exploits a natural experiment involving the establishment of China's economic zones which targeted firms rather than individuals. By focusing on a prominent place-based policy in China, the study has addressed whether or not zones work, for whom, and also what works and where (Neumark and Simpson, 2014). It does so by constructing a data set with geo-coded information about firms with relatively fine granularity. The findings constitute the first compelling evidence about the local economic effects of zones, their benefits and costs, and some determinants of program effectiveness, at least in China. Given the large number of developing countries implementing similar zone programs, the findings have important implications for policy (World Bank, 2008) and the design of more effective SEZs.

China's zone programs have demonstrated a large effect on the targeted areas in terms of extensive margins, especially via entries and exits. Existing firms have experienced limited improvement in their performance. There have also been productivity benefits and price impacts arising from locating in an SEZ, which indicate the presence of agglomeration economies. There are relatively limited spillovers in industrial activity between SEZs and non-SEZ areas.

In monetary terms, the program is estimated to have brought a net benefit of US$15.42 billion within three years of its implementation. These findings may help to dispel the general pessimism about zone programs in developing countries.

Another important finding is that a zone's effectiveness depends crucially on the design of its policies. China's economic zones offer various subsidies for capital investment, and operating in a zone is significantly more beneficial for capital-intensive firms than for more labor-intensive ones. Zones with better market potential or better access do not demonstrate significantly larger benefits. That finding serves as a reminder that formulating effective policy requires paying close attention to the circumstances of the agents to be influenced.

This study has been a first step toward understanding the micro-foundations of place-based policies in developing countries. Much remains to be done. This study evaluates only short-term effects (two years after the zones' establishment) due to data limitations. Further efforts should more precisely investigate the long-term impacts of the zones with a structural approach and better data. It would be interesting, in particular, to uncover any links between local political, economic, and social institutions and the effects of zones (Becker, Egger, and von Ehrlich, 2013).[21] Such analyses would undoubtedly be of great benefit in defining how SEZ policy interventions should best be implemented in specific contexts.

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

References

[1] Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

[2] Alder, Simon, Lin Shao, and Fabrizio Zilibotti. 2016. "Economic Reforms and Industrial Policy in a Panel of Chinese Cities." *Journal of Economic Growth*, 21: 305–349.

[3] Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics*, 130(3): 1117–1165.

[4] Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–184.

[5] Au, C. and J.V. Henderson. 2006a. "Are Chinese Cities Too Small?" *Review of Economic Studies*, 73(3): 549–576.

[6] Au, C. and J.V. Henderson. 2006b. "How Migration Restrictions Limit Agglomeration and Productivity in China." *Journal of Development Economics*, 80(2): 350–388.

[7] 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.

[8] 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.

[9] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.

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

[11] 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.

[12] 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.

[13] 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.

[14] Chaurey, Ritam. 2017. "Location-Based Tax Incentives: Evidence From India." *Journal of Public Economics*, 156: 101–120.

[15] Cheng, Yiwen. 2014. "Place-Based Policies in a Development Context: Evidence from China." University of California, Berkeley, working paper.

[16] 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.

[17] 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.

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

[19] 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.

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

[21] Foster, Lucia, John Haltiwanger and Chad Syverson. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis*, ed. Charles R. Hulten, Edwin R. Dean and Michael J. Harper, 303–372, University of Chicago Press, Chicago IL.

[22] Foster, Lucia, John Haltiwanger and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review,* 98(1): 394–425.

[23] Gibbons, Stephen, Stephen Machinb, and Olmo Silva. 2013. "Valuing School Quality Using Boundary Discontinuities." *Journal of Urban Economics*, 75(May): 15–28.

[24] 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.

[25] Glaeser, Edward L., and Joshua D. Gottlieb. 2008. "The Economics of Place-making Policies." *Brookings Papers on Economic Activity*, 39(1): 155–253.

[26] Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange. 2010. "Urban Economics and Entrepreneurship." *Journal of Urban Economics*, 67(1): 1–14.

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

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

[29] Hanson, Gordon H. 2005. "Market Potential, Increasing Returns, and Geographic Concentration." *Journal of International Economics*, 67(1): 1–24.

[30] 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.

[31] 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.

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

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

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

[35] 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(1): 629–662.

[36] 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.

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

[38] Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72: 159–217.

[39] 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.

[40] Neumark, David, and Helen Simpson. 2014. "Place-Based Policies." National Bureau of Economic Research (NBER) Discussion Paper 20049.

[41] Qian, Yingyi. 2003. "How Reform Worked in China." In *In Search of Prosperity: Analytic Narratives on Economic Growth*, Dani Rodrik (ed.), 297–333. Princeton University Press, Princeton NJ.

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

[43] Rodrik, Dani. 2008. "Second-Best Institutions." *American Economic Review*, 98(2): 100-104.

[44] Rothenberg, Alex. 2013. "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways." University of California, Berkeley, working paper.

[45] 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.

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

[47] Wei, Yaping, and Min Zhao. 2009. "Urban Spill Over vs. Local Urban Sprawl: Entangling Land-use Regulations in the Urban Growth of China's Megacities." *Land Use Policy*, 26(4): 1031–1045.

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

[49] Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature*, 49(4): 1076–1151.

[50] Zeng, Douglas Zhihua. 2010. "Building Engines for Growth and Competitiveness in China: Experience with Special Economic Zones & Industrial Clusters." World Bank, Washington DC.

[51] Zheng, Siqi, Weizeng Sun, Jianfeng Wu, and Kahn E. Matthew. 2017. "The Birth of Edge Cities in China: Measuring the Spillover Effects of Industrial Parks." *Journal of Urban Economics*, 100: 80–103.

Appendices

A Five Waves of Economic Zone Formation

The waves of zone establishment shown in Figure 1 are as follows.

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 in 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 national-level development zones and 20 province-level development zones were established from 1984 to 1991.

1992–1999: After Deng Xiaoping's famous southern tour in 1992 the State Council opened several 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 national-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 the coastal development zones demonstrated their effectiveness in attracting investment and boosting employment. As a result, more development zones were granted by the central authorities and the provincial governments in inland cities. China's entry into the World Trade Organization in 2001 led to an increasing number of national-level export processing zones and bonded zones. In total, 64 national-level development zones and 197 province-level development zones were established between 2000 and 2004.

2005–2008: From 2005, an 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 national-level development zones and 663 province-level development zones were formed.

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

Each firm's administrative location code is used to locate it as either within an SEZ or not. These three cases summarize the process.

1. Some SEZs have their own administrative codes. For example, Nanling Industrial Zone (zone code: S347063) in Anhui has an independent 12-digit administrative location code: 340223100400 (Anhui Nanling Industrial Zone Community).

2. Some zones are co-terminous with a town or a Chinese administrative area termed a street. All villages or communities under the town or street will then be within the zone's boundaries. For example, Fei County Industrial Zone (zone code: S377099) in Shandong encompasses all of Tanxin town (administrative location code: 371325105). The 9-digit town code is enough to pin down its zone status.

3. Some zones take in several villages or communities. For example, Yunmeng Economic Development Zone (zone code: S427040) in Hubei administers eight villages and one community: Xinli Village, Heping Village, Qianhu Village, Hebian Village, Zhanqiao Village, Quhu Village, Zhaoxu Village, Sihe Village, and Qunli Community. An enterprise in any of them will be within the zone.

C Estimation of Firm TFP

Consider the following Cobb-Douglas production function (in logarithmic form):

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \epsilon_{it}, \tag{A1}$$

where y_{it} is the logarithm of firm output, l_{it}, k_{it}, and m_{it} are the logarithms of the employment, capital, and materials inputs. ω_{it} is the firm's productivity, and ϵ_{it} takes in measurement error and any unanticipated shocks to output.

Obtaining consistent production function estimates $\boldsymbol{\beta} = (\beta_l, \beta_k, \beta_m)$ requires controlling for unobserved productivity shocks potentially leading to simultaneity and selection biases. A control function based on a static input demand function is used as a proxy for the unobserved productivity.

The control function approach initiated by Olley and Pakes (1996) and extended by Levinsohn and Petrin (2003) is applied. The following material demand function is used as a proxy for the unobserved productivity:

$$m_{it} = m_t\left(\omega_{it}, l_{it}, k_{it}\right). \tag{A2}$$

Inverting (A2) yields the control function for productivity:

$$\omega_{it} = h_t\left(l_{it}, k_{it}, m_{it}\right).$$

In the first stage, unanticipated shocks and measurement errors (ϵ_{it}) are purged by estimating the following equation:

$$y_{it} = \phi_t\left(l_{it}, k_{it}, m_{it}\right) + \epsilon_{it}, \tag{A3}$$

That yields a predicted output ($\hat{\phi}_{it}$).

(A1) and (A3) from the first stage estimation can then be used to express productivity:

$$\omega_{it}\left(\boldsymbol{\beta}\right) = \hat{\phi}_{it} - \beta_l l_{it} - \beta_k k_{it} - \beta_m m_{it}. \tag{A4}$$

To estimate the production function coefficients $\boldsymbol{\beta}$, the technique of Ackerberg, Caves, and Frazer (2015) is applied and moments are formed based on innovation in the productivity shock ξ_{it} in law of motion for productivity:

$$\omega_{it} = g\left(\omega_{it-1}\right) + \xi_{it}.$$

Using (A4), $\omega_{it}\left(\boldsymbol{\beta}\right)$ is non-parametrically regressed against $g\left(\omega_{it-1}\right)$ to obtain the innovation term $\xi_{it}\left(\boldsymbol{\beta}\right) = \omega_{it}\left(\boldsymbol{\beta}\right) - E\left(\omega_{it}\left(\boldsymbol{\beta}\right)|\omega_{it-1}\left(\boldsymbol{\beta}\right)\right)$.

The moment conditions used to estimate the production function coefficients are then:

$$E\left(\xi_{it}\left(\boldsymbol{\beta}\right)\mathbf{Y}_{it}\right) = 0,$$

where $\mathbf{Y}_{it}$ contains lagged labor and materials, and current capital.[22]

Once the production function coefficients $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_l, \hat{\beta}_k, \hat{\beta}_m\right)$ have been estimated, a firm's

[22] Following the lead of other scholars, labor and materials are treated as flexible inputs and their lagged values are used to construct the moments. Capital is considered a dynamic input with adjustment costs, so its current value is used in forming the moments.

total factor productivity can be computed as:

$$\hat{\omega}_{it} = \hat{\phi}_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_m m_{it}.$$

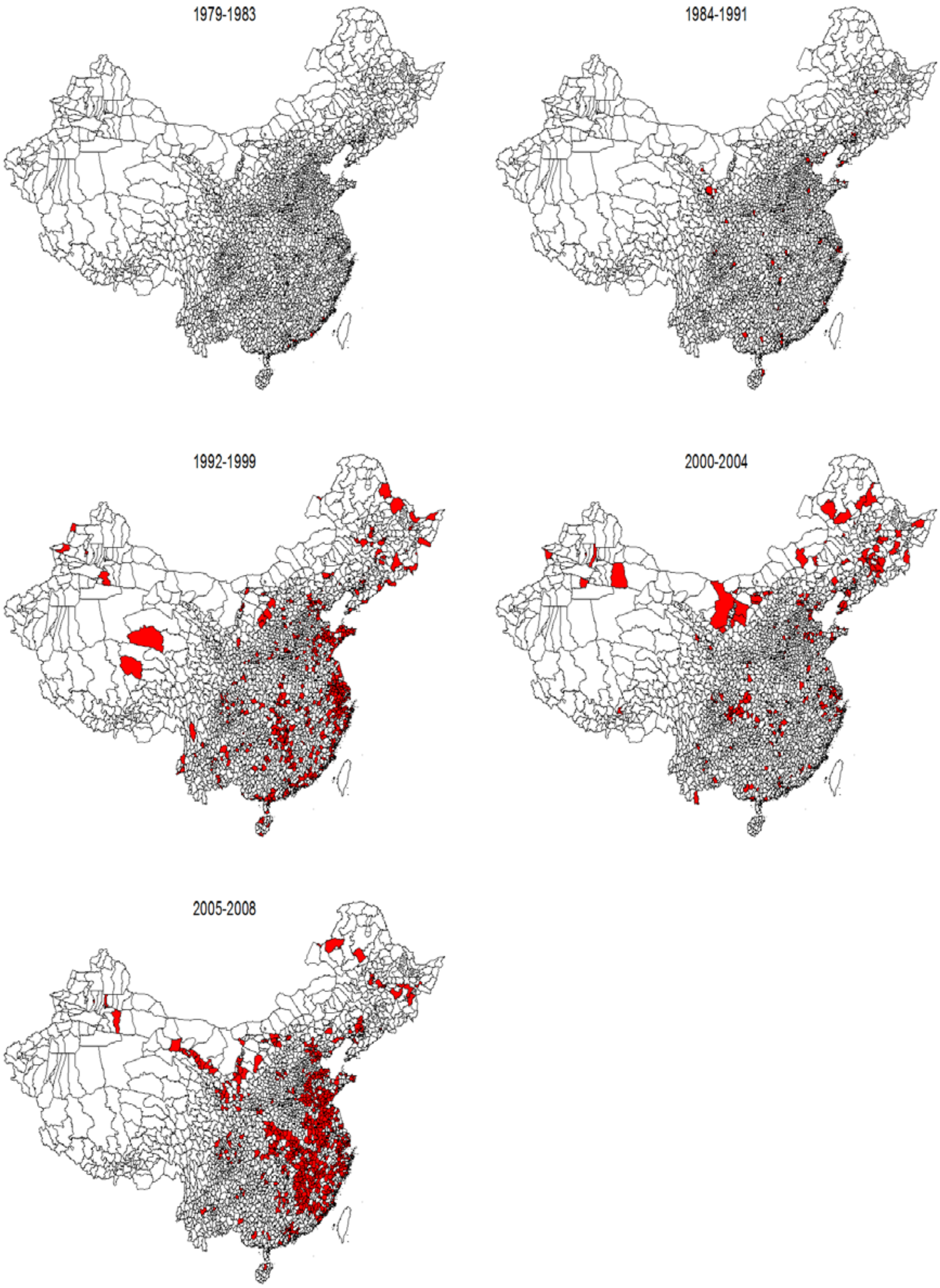
References

[1] Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

[2] Levinsohn, James A., and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–341.

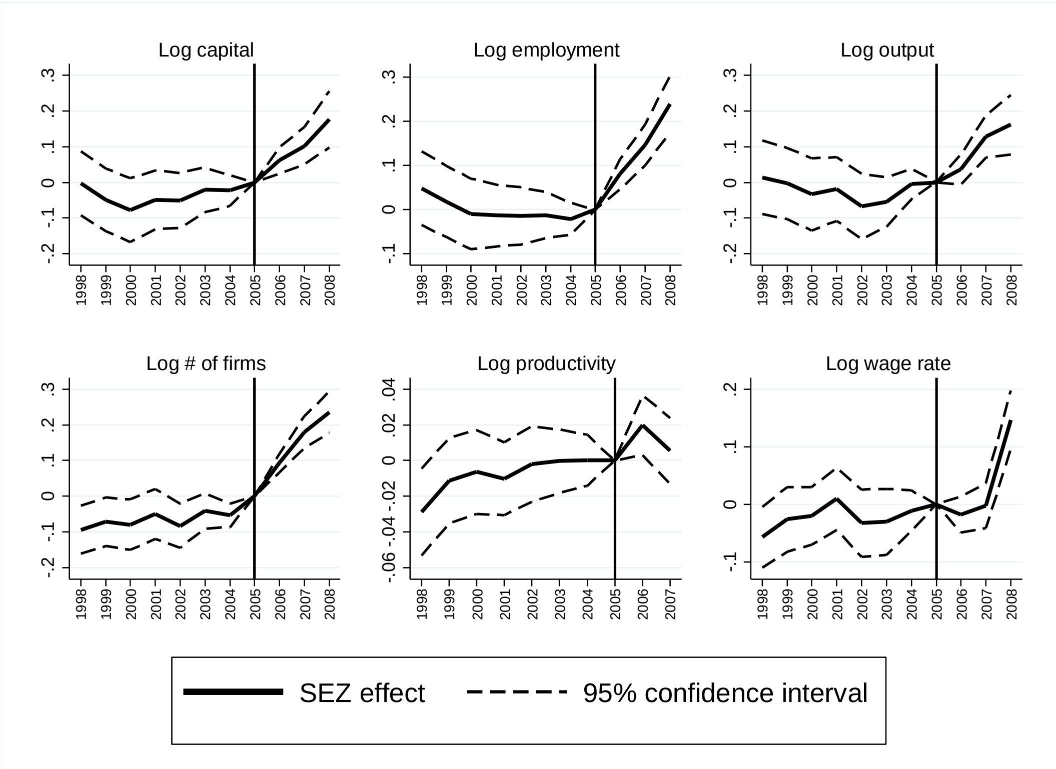
[3] Olley, Stephen G., and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

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 by the shaded areas.

Figure 2 Test for Parallel Trends at the County-Level Analysis



Notes: Any temporal trends are analyzed comparing the treatment and control groups at the county level. ASIF data are used for the analysis. The plots connected by the solid line indicate changes in outcomes compared to 2005 (the period immediately before the SEZ treatment) conditional on baseline characteristics, county and year fixed effects. The dashed lines indicate the 95% confidence intervals where standard errors are clustered at the county level.

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				
National-level Economic Zones, of which:		46	93	64	19
By Type					
1. Economic and Technological Development Zones		12	20	17	
2. High-tech and Industrial Development Zones		26	27		
3. Export Processing Zones			1	39	18
4. Bonded Zones		4	11		
5. Border Economic Cooperation Zones			14		
6. Other		4	20	8	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		18	430	169	615
2. High-tech and Industrial Development Zones		2	29	14	20
3. Specialized Industrial Zones			7	14	28
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, national-level development zones, and province-level economic zones are distinguished. The comprehensive SEZs are the zones established in Shenzhen, Shantou, Zhuhai, and Xiamen. National-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

	(1) SEZ	(2) Non-SEZ	(3) Difference
Panel A. Village level			
Log distance from an airport (kilometers)	3.766	3.877	−0.111
	(0.013)	(0.003)	(0.013)
Log distance from a port (kilometers)	4.847	4.918	−0.072
	(0.022)	(0.005)	(0.021)
Log capital-labor ratio in 2004	1.729	1.158	0.571
	(0.020)	(0.005)	(0.018)
Log number of firms in 2004	4.630	4.167	0.463
	(0.015)	(0.005)	(0.018)
Panel B. County level			
Log population in 2004	3.912	3.317	0.595
	(0.032)	(0.025)	(0.050)
Log employment in 2004	0.883	0.291	0.592
	(0.034)	(0.023)	(0.047)
Log GDP in 2004	3.529	2.693	0.836
	(0.037)	(0.029)	(0.057)
Share of rural employment in 2004	10.436	10.619	−0.183
	(0.279)	(0.173)	(0.352)
Export intensity in 2004	0.013	0.006	0.007
	(0.002)	(0.001)	(0.002)
Ratio of government expenditure to revenue in 2004	2.861	5.865	−3.004
	(0.062)	(0.182)	(0.337)
GDP growth rate from 1998–2004	0.542	0.583	−0.041
	(0.018)	(0.009)	(0.019)
Share of employment in agriculture in 2000	0.754	0.790	−0.037
	(0.007)	(0.004)	(0.008)
Share of employment in manufacturing in 2000	0.142	0.134	0.008
	(0.004)	(0.002)	(0.005)
Mortality rate in 2000	0.062	0.065	−0.003
	(0.0005)	(0.0004)	(0.0008)
Share of eliminated illiteracy population in 2000	0.021	0.022	−0.001
	(0.001)	(0.001)	(0.001)
Log land area (10 thousand square kilometers)	−1.746	−1.419	−0.328
	(0.032)	(0.029)	(0.056)
Log distance from an airport (kilometers)	4.320	4.534	−0.214
	(0.032)	(0.020)	(0.041)
Log distance from a port (kilometers)	5.713	6.354	−0.641
	(0.049)	(0.026)	(0.055)
Log capital-labor ratio in 2004	4.704	4.767	−0.063
	(0.029)	(0.019)	0.039
Log number of firms in 2004	5.593	4.244	1.348
	(0.048)	(0.040)	(0.077)

Note: This table reports the summary statistics of the treatment and control samples (SEZ and non-SEZ). Panel A shows initial village-level characteristics, and Panel B reports county-level characteristics in the initial year. Columns 1 and 2 show means and standard deviations in parentheses. Column 3 reports the unconditional difference between the treatment and control group.

Table 3 The SEZ Effects: Village-Level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Without controlling for covariates						
SEZ*post2006	0.334	0.271	0.330	0.192	0.007	0.026
	(0.032)	(0.026)	(0.035)	(0.021)	(0.007)	(0.013)
Panel B. Controlling for covariates						
SEZ*post2006	0.579	0.345	0.492	0.290	0.015	0.029
	(0.034)	(0.028)	(0.038)	(0.023)	(0.007)	(0.013)
Village FEs	Yes	Yes	Yes	Yes	Yes	Yes
County-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	580	580	580	580	405	406
Observations	119,898	119,898	119,898	119,898	43,830	56,600

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. In columns 1-6, the dependent variable is the natural log of the measure of capital, employment, output, number of firms, productivity and wage rate, respectively. Covariates include village-level characteristics listed in Panel A of Table 2, interacted with the year dummy. The standard errors are reported in parentheses, clustered by county. All regressions control for village fixed effects and county-year fixed effects.

Table 4 The SEZ Effects: County-Level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Without controlling for covariates						
SEZ*post2006	0.153	0.121	0.188	0.076	0.023	0.063
	(0.035)	(0.024)	(0.038)	(0.023)	(0.007)	(0.014)
Panel B. Controlling for covariates						
SEZ*post2006	0.132	0.155	0.172	0.115	0.015	0.047
	(0.039)	(0.028)	(0.046)	(0.024)	(0.008)	(0.015)
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	1,582	1,582	1,582	1,582	1,478	1,502
Observations	3,164	3,164	3,164	3,164	5,868	7,434

Note: All observations are at the county-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. In columns 1-6, the dependent variable is the natural log of the measure of capital, employment, output, number of firms, productivity and wage rate, respectively. Covariates include county-level characteristics listed in Panel B of Table 2, interacted with the year dummy. The standard errors are reported in parentheses, clustered by county. All regressions control for county and year fixed effects.

Table 5 The SEZ Effects: BD-DD Estimations

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Without controlling for covariates						
SEZ*post2006	0.547	0.471	0.553	0.233	0.016	0.038
	(0.066)	(0.049)	(0.069)	(0.038)	(0.009)	(0.014)
Panel B. Controlling for covariates						
SEZ*post2006	0.633	0.495	0.602	0.266	0.016	0.039
	(0.071)	(0.051)	(0.070)	(0.042)	(0.009)	(0.014)
Area FEs	Yes	Yes	Yes	Yes	Yes	Yes
Neighborhood-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,348	2,348	2,348	2,348	3,714	4,710

Note: Observations are at the area-year level within 1,000 meters of a zone boundary. In columns 1-6, the dependent variable is the natural log of the measure of capital, employment, output, number of firms, productivity and wage rate, respectively. In columns 1-4, Census 2004 and 2008 are used. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. Area fixed effects and neighborhood-year fixed effects are included in the specification. Covariates include area-level characteristics which are averaged from the villages in the area, interacted with the year dummy. The standard errors are clustered at the zone level, reported in parentheses.

Table 6 Spillover Effect

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Spillover effect	0.032	0.073	0.119	0.052	−0.001	0.025
	(0.046)	(0.034)	(0.054)	(0.023)	(0.005)	(0.020)
Additional direct effect	0.663	0.353	0.547	0.375	0.017	0.029
	(0.062)	(0.049)	(0.065)	(0.040)	(0.010)	(0.023)
Covariates*year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Village FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	1,779	1,779	1,779	1,779	1,577	1,587
Observations	142,202	142,202	142,202	142,202	82,716	109,088

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 with full sample of 103,263 villages in 1,779 counties including both SEZ counties and non-SEZ ones are used for analysis. In columns 5, ASIF data from 2004 to 2007 with full sample of 32,149 villages in 1,577 counties are used. In column 6, ASIF data from 2004 to 2008 with full sample of 33,826 villages in 1,587 counties are used. In columns 1-6, the dependent variable is the natural log of the measure of capital, employment, output, number of firms, productivity and wage rate, respectively. Covariates include village-level characteristics listed in Panel A of Table 2. The standard errors are reported in parentheses, clustered by county. All regressions control for village fixed effects and year fixed effects.

Table 7 Decomposition

	(1)	(2)	(3)	(4)	(5)
Panel A.	Total growth	Within	Net entry		
Log capital	0.579	0.129	0.485		
	(0.034)	(0.015)	(0.038)		
Log employment	0.345	0.010	0.304		
	(0.028)	(0.011)	(0.030)		
Log output	0.492	0.098	0.402		
	(0.038)	(0.018)	(0.040)		
Panel B.	Total growth	Within	Between	Cross	Net entry
Log productivity	0.015	0.001	0.001	−0.013	0.026
	(0.011)	(0.007)	(0.003)	(0.004)	(0.011)
Log wage rate	0.048	0.004	0.002	−0.061	0.104
	(0.022)	(0.025)	(0.005)	(0.016)	(0.019)

Note: In Panel A, Census data in 2004 and 2008 aggregated at the village level are used for analysis. In Panel B, ASIF data in 2004 and 2007 aggregated at the village level are used for analysis on log productivity; ASIF data in 2004 and 2008 aggregated at the village level are used for analysis on log wage rate. The SEZ effects are decomposed into: (1) new entrants and exiters, or the extensive margin effect; and (2) continuing firms, or the intensive margin effect. SEZ effects are estimated using equation (12). Covariates listed in Panel A, Table 2 are included in all specifications. The standard errors are clustered at the county level, reported in parentheses.

Table 8 SEZ Effects: Corporate Profits and Wage Bills

	(1)	(2)
Dependent variable	Log corporate profits	Log wage bills
SEZ*post2006	0.159	0.192
	(0.066)	(0.029)
Covariates*year dummies	Yes	Yes
County FEs	Yes	Yes
Year FEs	Yes	Yes
Number of clusters	1,428	1,502
Observations	6,363	7,440

Note: All observations are at the county-year level. ASIF data from 2004 to 2008 are aggregated for the analysis. Covariates include initial county-level characteristics listed in Panel B, Table 2. The standard errors are clustered at the county level, reported in parentheses. All regressions control for county and year fixed effects.

Table 9 Cost and Benefit Analysis: A Back-of-the-Envelope Approach

	(1)	(2)	(3)	(4)
Panel A. Benefits				
Year	Actual value (billion RMB)	SEZ effect	Counterfactual value (billion RMB)	Benefits (billion RMB)
(1) Corporate profits				
2006	168.67	0.1595	145.47	23.20
2007	223.76	0.1595	192.99	30.77
2008	253.29	0.1595	218.45	34.83
(2) Wage bills				
2006	155.23	0.1916	130.27	24.96
2007	193.52	0.1916	162.41	31.11
2008	261.69	0.1916	219.62	42.08
Panel B. Costs (corporate tax)				
Year	Counterfactual corporate profits (billion RMB)	Statutory tax rate	Actual taxes paid (billion RMB)	Costs (billion RMB)
2006	145.47	33%	32.14	15.87
2007	192.99	33%	40.38	23.30
2008	218.45	25%	36.25	18.36

Note: The calculations on benefits and costs of the SEZ program using a back-of-the-envelope approach are shown in the Table. See Section 7 for details.

Table 10 Heterogeneous Effects of the SEZ Program

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Capital versus Labor-Intensive Industries						
Differential Treatment Impact	0.106	0.067	0.109	0.025	0.010	0.002
	(0.060)	(0.049)	(0.066)	(0.034)	(0.014)	(0.026)
Panel B. SEZ Counties with Good versus Poor Infrastructure						
Differential Treatment Impact	−0.045	0.003	−0.012	−0.068	−0.017	0.021
	(0.072)	(0.058)	(0.084)	(0.042)	(0.014)	(0.026)
Panel C. Firms with Large versus Small Size						
Differential Treatment Impact	0.065	0.055	0.029	0.049	0.004	−0.006
	(0.047)	(0.036)	(0.053)	(0.019)	(0.013)	(0.026)

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. Panel A, B and C reports the differences in the SEZ effects between capital-intensive and labor-intensive industries, between SEZs with good and poor infrastructure, and between large and small firms. Capital and labor intensity are defined at the 4-digit level based on a capital-labor ratio above or below the median value in 2004. SEZ counties with good (poor) infrastructure index are those with infrastructure indices above (below) the median in 2004: a larger index indicates better infrastructure. Firms with large (small) size are those with sales above (below) the median in 2004.

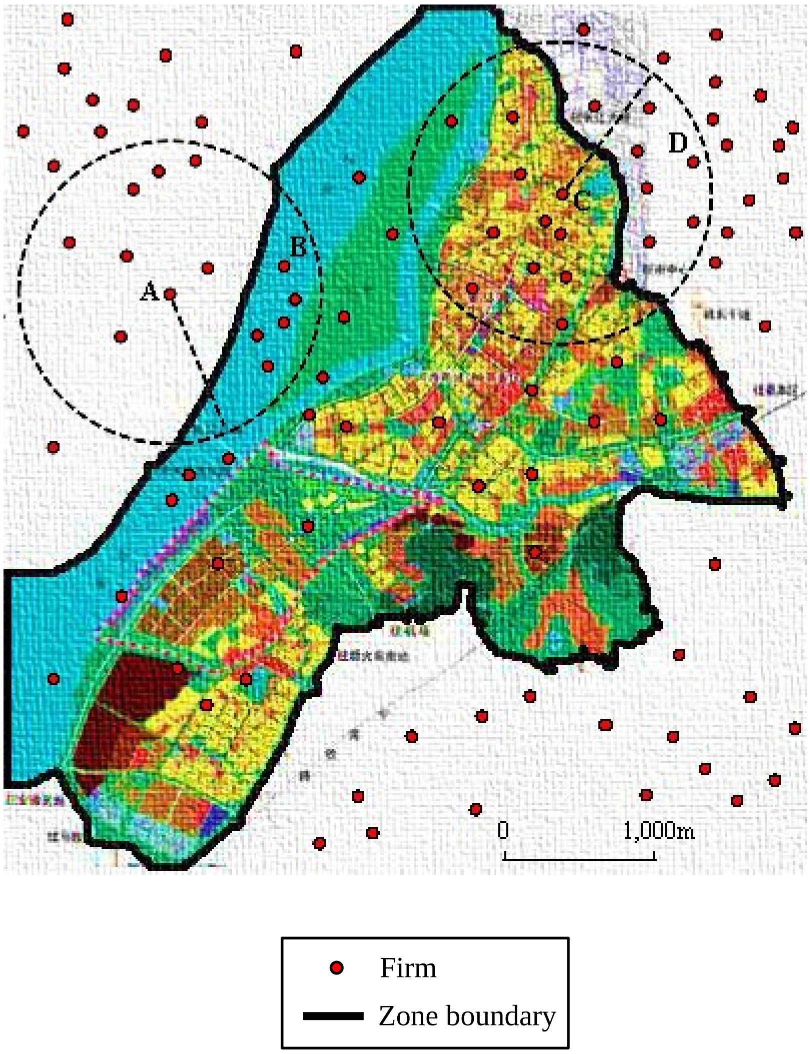
Figure A1 Searching for a Detailed Address with Google Maps

Notes: To obtain coordinates for firms that report detailed Chinese addresses, the address (for example, "157 Nandan Road, Xuhui District, Shanghai, China") is searched 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 are read from the map.

Figure A2 Searching for Villages and Communities with Google Maps

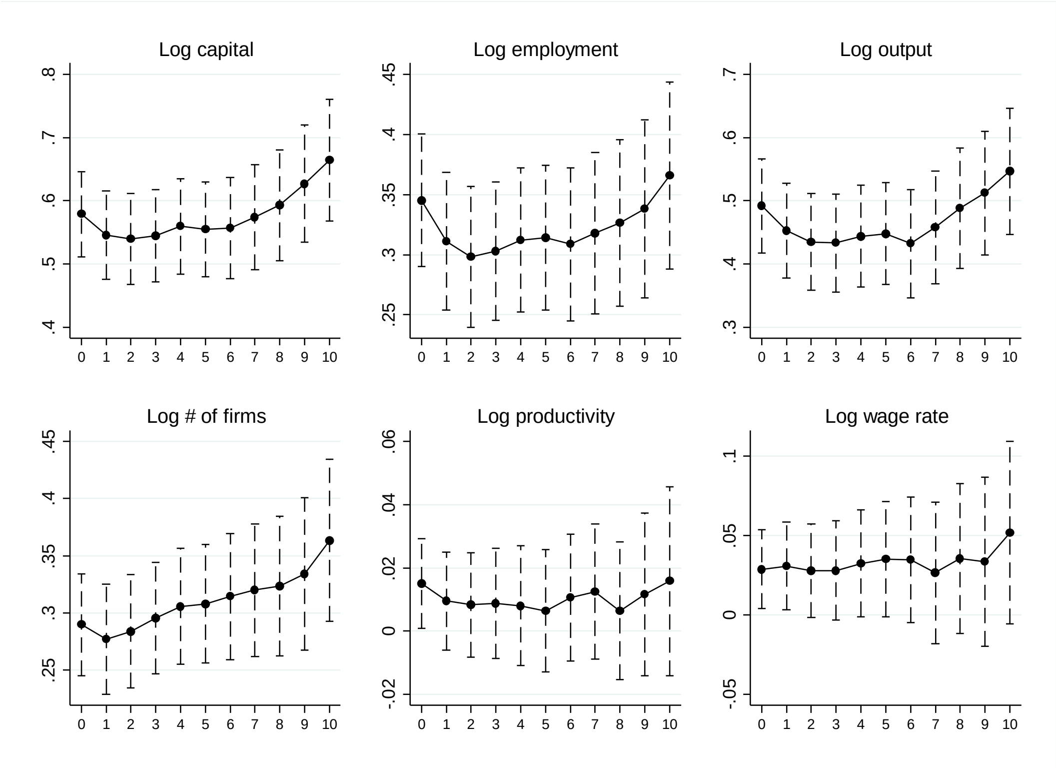
Notes: To obtain the coordinates for villages and communities, the name of the village or community is entered followed by the name of the town, city, and province to which the village or community belongs (for example, "Liunan Village, Liuhe Town, Taicang City, Suzhou, Jiangsu Province, China"). The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

Figure A3 Firms Near a Zone Boundary



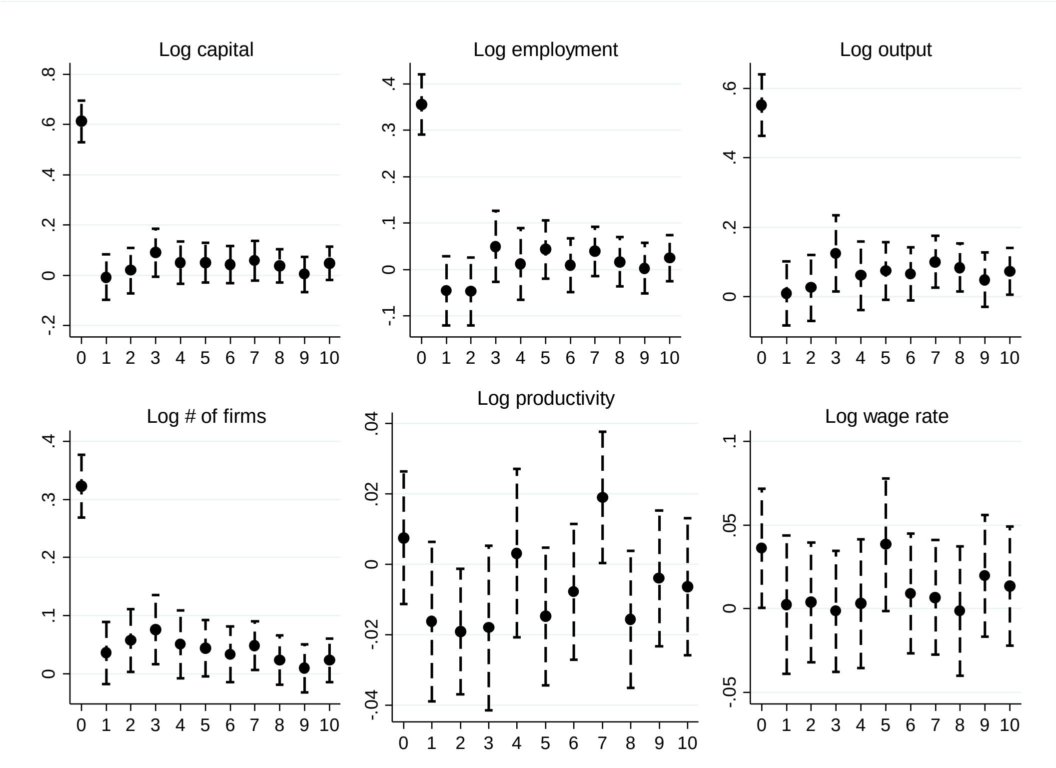
Notes: If a firm (firm A) is located outside the zone and within 1,000 meters, and there is another firm (firm B) located inside the zone, firm A is designated as located within 1,000 meters of the zone boundary. If a firm (firm C) is located inside the zone and within 1,000 meters, and there is another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

Figure A4 Ring Analysis



Notes: In this analysis, we step-wisely exclude from the control group, the non-SEZ villages within the 2km of the SEZ villages, those within the 4km of the SEZ villages, continuing to the exclusion of those within the 20km of SEZ villages. The benchmark estimate is plotted in the horizontal line labeled as 0 without exclusion of any control SEZ villages. The estimates labeled from 1, 2, 3,…, 10 indicate exclusion of SEZ villages within 2, 4, 6,…, 20 kilometer distance of the SEZ villages.

Figure A5 Ring Analysis: An Alternative Model Specification



Notes: In this analysis, we provide an alternative spillover estimates using equation (4). The treatment effect on the SEZ village is plotted in the horizontal line labeled as 0. The estimates labeled from 1, 2, 3,…, 10 indicate the spillover externality effect on the nearby 1st, 2nd, 3rd,…10th ring between 0 and 2, 2 and 4, 4 and 6, …, 18 and 20 kilometers from its nearest SEZ village.

Table A1 Comparison between the 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								
Capital	1,321,026	14091	120	13800	256,999	62,093	2,731	89,541
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
Panel (b). Year 2008								
Capital	1,788,227	20558	303	20265	382,842	80,805	3,575	107,615
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

Note: p10 and p90 denote the 10th and 90th percentiles. Sources: Economic Census and Anual Survey of Industrial Firms for 2004 and 2008.

Table A2 Village-Level Analysis (ASIF Data from 2004-2008)

	(1)	(2)	(3)	(4)
Dependent variable	Log capital	Log employment	Log output	Log # of firms
SEZ*post2006	0.294	0.223	0.261	0.195
	(0.035)	(0.031)	(0.038)	(0.021)
Covariates*year dummies	Yes	Yes	Yes	Yes
Village FEs	Yes	Yes	Yes	Yes
County-year FEs	Yes	Yes	Yes	Yes
Number of clusters	406	406	406	406
Observations	56,772	56,772	56,772	56,772

Note: All observations are at the village-year level. ASIF data from 2004 to 2008 are used. Covariates include village-level characteristics listed in Panel A, Table 2. The standard errors are clustered at the county level. All regressions control for village fixed effects and county-year fixed effects.

Table A3 Testing for Pretrends: Village-Level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
SEZ*year2005	0.011	0.055	0.057	0.025	−0.001	0.016
	(0.027)	(0.027)	(0.033)	(0.017)	(0.007)	(0.014)
SEZ*year2006	0.17	0.15	0.184	0.107	0.015	0.023
	(0.037)	(0.035)	(0.041)	(0.023)	(0.009)	(0.017)
SEZ*year2007	0.296	0.239	0.294	0.196	0.014	0.039
	(0.045)	(0.039)	(0.048)	(0.027)	(0.010)	(0.018)
SEZ*year2008	0.489	0.415	0.442	0.367		0.054
	(0.065)	(0.058)	(0.067)	(0.037)		(0.021)
Covariates*year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Village FEs	Yes	Yes	Yes	Yes	Yes	Yes
County-year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,772	56,772	56,772	56,772	43,830	56,600

Note: In columns 1-4 and 6, ASIF data from 2004 to 2008 are used for the analysis. In column 5, ASIF data from 2004 to 2007 are used. Covariates include village-level characteristics listed in Panel A of Table 2. The standard errors are reported in parentheses, clustered by county. All regressions control for village fixed effects and county-year fixed effects.

Table A4 The SEZ Effects on Housing Prices

	(1)	(2)
Dependent variable	Log house prices	Log house prices
SEZ*post2006	0.015	0.020
	(0.026)	(0.026)
City FEs	Yes	Yes
Year FEs	Yes	Yes
Covariates*year dummies	No	Yes
Number of clusters	284	284
Observations	1,362	1,362

Note: All observations are at the city-year level. The dependent variable is the natural log of the housing prices. If a city has any county under its administration granted with SEZs, the treatment indicator equals one since 2006. Covariates include average county-level characteristics (as listed in Panel B of Table 2) for those under a city. The standard errors are reported in parentheses, clustered by city. All regressions control for city and year fixed effects.

Table A5 Heterogeneous Effects by Industrial Capital-Labor Ratio

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Capital-Intensive Industries						
SEZ*post2006	0.604	0.375	0.547	0.296	0.013	0.027
	(0.044)	(0.036)	(0.053)	(0.027)	(0.009)	(0.016)
Number of clusters	560	560	560	560	392	395
Observations	58,784	58,784	58,784	58,784	23,421	31,274
Panel B. Labor-Intensive Industries						
SEZ*post2006	0.498	0.308	0.438	0.272	0.003	0.025
	(0.035)	(0.027)	(0.035)	(0.022)	(0.009)	(0.016)
Number of clusters	549	549	549	549	398	400
Observations	98,540	98,540	98,540	98,540	30,615	39,696

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. Capital and labor intensity are defined at the 4-digit level based on a capital-labor ratio above or below the median value in 2004. Standard errors are in parentheses. The standard errors are clustered at the county level.

Table A6 Heterogeneous Effects by Infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. SEZ Counties with Good Infrastructure						
SEZ*post2006	0.569	0.35	0.491	0.271	0.009	0.037
	(0.044)	(0.037)	(0.049)	(0.029)	(0.009)	(0.017)
Number of clusters	288	288	288	288	240	241
Observations	78,698	78,698	78,698	78,698	25,206	32,382
Panel B. SEZ Counties with Poor Infrastructure						
SEZ*post2006	0.614	0.347	0.502	0.339	0.025	0.016
	(0.053)	(0.041)	(0.056)	(0.034)	(0.012)	(0.019)
Number of clusters	292	292	292	292	165	165
Observations	41,200	41,200	41,200	41,200	18,624	24,218

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. SEZ counties with good (poor) infrastructure index are those with infrastructure indices above (below) the median in 2004: a larger index indicates better infrastructure. Standard errors are in parentheses. The standard errors are clustered at the county level.

Table A7 Heterogeneous Effects by Firm Size

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log capital	Log employment	Log output	Log # of firms	Log productivity	Log wage rate
Panel A. Firms with Large Size						
SEZ*post2006	0.362	0.212	0.342	0.153	−0.006	0.011
	(0.030)	(0.024)	(0.035)	(0.014)	(0.008)	(0.017)
Number of clusters	525	525	525	525	394	398
Observations	71,022	71,022	71,022	71,022	21,300	27,953
Panel B. Firms with Small Size						
SEZ*post2006	0.297	0.157	0.313	0.103	−0.010	0.017
	(0.036)	(0.028)	(0.043)	(0.017)	(0.008)	(0.016)
Number of clusters	560	560	560	560	385	394
Observations	66,154	66,154	66,154	66,154	24,357	30,613

Note: All observations are at the village-year level. In columns 1-4, Census data 2004 and 2008 are used for analysis. In column 5, ASIF data from 2004 to 2007 are used. In column 6, ASIF data from 2004 to 2008 are used. Firms with large (small) size are those with sales above (below) the median in 2004. Standard errors are in parentheses. In panels A and B, the standard errors are clustered at the county level.