Place-based Development: Evidence from Special Economic Zones in India*

Yeseul Hyun† Shree Ravi‡

August 2018

Abstract

Little is known about the aggregate and distributional effects of Special Economic Zones (SEZs) in developing countries. We investigate the influence of Indian SEZs by exploiting spatial variations in the timing of zonal operations. Using satellite and survey data, we establish that SEZs boosted economic activity within areas several times the size of the zones. The zones also drove a structural change in the local economy with resources shifting away from the informal sector and the formal sector growing in size and productivity. This growth, however, differently benefits workers at the higher end of the income and skill distributions.

*We thank Samuel Bazzi, Johannes Schmieder, Kehinde Ajayi and Dilip Mookherjee for their invaluable input, guidance and encouragement. We also thank Daniele Paserman, Martin Fiszbein, Aradhna Aggarwal, Michael Gechter, Amrit Amirapu, Nathaniel Young, Jonathan Hersh, Siyi Luo and conference and seminar participants at BU, Copenhagen Business School, University of Bordeaux and the Indian Institute of Foreign Trade for many useful comments. Representatives in the central government of India provided crucial assistance in data provision and verification. We gratefully acknowledge financial support from Department of Economics and the Institute for Economic Development at Boston University. Taeyeon Kim and Wenran Fan provided excellent research assistance.

†PhD Candidate, Boston University, yhyun@bu.edu

‡PhD Candidate, Boston University, sravi@bu.edu

1 Introduction

Special Economic Zones (SEZs) are a popular policy tool used by both developed and developing economies to boost growth in specific geographic areas and economic sectors. Governments typically direct infrastructure development and regulatory concessions at narrow regions in order to influence the location of large and productive firms within them. The resulting agglomeration of economic activity is expected to increase local competition, produce denser input and labor markets, and produce knowledge spillovers among firms and workers that drive long-term development. Empirical evidence of such benefits is, however, limited and focused on the study of such programs in developed economies.[1] The unique institutional and economic features of developing countries, such as the size of their informal sectors and tight credit markets, are certain to influence the success of their SEZ programs. There is hence a need to understand their zonal development initiatives separately.

In this paper, we provide one of the first empirical evaluations of the Indian SEZ policy to understand its aggregate and distributional impact on the Indian economy. India introduced the 2005 Special Economic Zones (SEZ) Act as one of the biggest pushes to industrial development in its history. Over ten years since 2005, it has annually invested roughly 0.5% of its GDP, totalling 62 billion US dollars. The size of its investment is substantial by both Indian and international standards.[2] As of 2016, 221 SEZs began operation across the nation, attracting firms through tax exemptions, infrastructural benefits and regulatory concessions, directly employing 1.4 million people and contributing to almost one-third of the annual national exports.[3]

However, the observed economic activity of SEZs is insufficient evidence of the policy's success. It is necessary to study the indirect effects of such spatial policies when evaluating them. One particular concern is that these observed benefits to regions receiving SEZs could be offset by losses elsewhere in the economy through resource relocations, which would produce little, if any, aggregate gains at considerable costs.[4]

Using the stages of approvals for SEZs as a source of quasi-experimental variation, our study not only reveals an increase in general economic activity in the immediate neighborhoods of SEZs but also confirms that regions several times the size of the zones benefit- up to the sub-national level of a district. Additionally,

[1] In the case of the US, while such policies seem to increase employment and wages (Busso, Gregory and Kline (2013); Empowerment Zones) the benefits are offset by losses elsewhere in the country (Kline and Moretti (2013); Tennessee Valley Authority program).

[2] Roughly half the amount was spent over a decade by the Indian government for financing infrastructural development and providing tax exemptions to all firms in two eastern Indian states (Shenoy, 2016). According to Kline and Moretti (2013) and Busso, Gregory and Kline (2013), the investments in the Tennessee Valley Authority program and the Empowerment Zones are around 20 billion and 3 billion US dollars, respectively.

[3] Statistics are sourced from the Ministry of Commerce website and Mukherjee and Bhardwaj (2016).

[4] The literature on such place-based development policies stresses the necessity to study the response of the wider private sector to the policies and measures their cost-effectiveness based on the programs' ability to form linkages with the local economy. See Duranton and Venables (2018) , Farole et al. (2011) and Aggarwal (2011) for a discussion.

we study how the SEZ policy interacts with the informal and formal sectors of the economy, a dual economy framework which is unique to low income countries. The most pertinent characteristics of informality in India are the absence of any form of regulation in production, omission from the tax base, and workers that are often unskilled and that do not receive social security benefits. While the Indian informal sector is characterized by low productivity, it is a major source of employment hiring about 80% of the total labor force.[5] Hence it is important to analyze the effects of SEZs in the dual economy framework for a more comprehensive understanding. Our analysis reveals that SEZs instigate a structural change of the economy, increasing firm size (in terms of production, employment and investment) and productivity in the formal sector while causing a shift of resources out of the informal sector and crowding out its production. Greater formality is considered more desirable in developing countries because it brings a larger part of the economy under beneficial government regulation (in matters such as worker safety and welfare) and broadens the tax base. Moreover, a shift of resources from the less productive informal sector would increase the overall productivity of the economy.[6] While our result show-cases the potential of such policies to bring about long-term development, we also find evidence suggestive of a rise in inequality in the short-term.

Our results are especially striking given the unique nature of Indian SEZs. The Indian government incentivized the participation of the private sector in zonal development and allowed SEZs to be of substantially smaller physical sizes than found elsewhere in the world. Small zones were doubted to be capable of producing significant additional economic activity. In addition, private sector participation in SEZ development was viewed as tax-free profit generation at the cost of issues such as misuse of land and inequities favoring large companies over small ones.[7]

The challenge to our identification strategy comes from the non-random nature of program location. This is a common concern for studies analyzing place-based development policies. Due to private sector participation in program location, however, we face the opposite concern of studies generally involving zones in developed economies, which usually target under-industrialized regions. In order to credibly isolate the effect of SEZs on the regional economy, we exploit a source of variation in the government-regulated approval process for SEZs. We consider only those regions surrounding SEZs that have reached the penultimate stage of approval before beginning operation. This stage, henceforth referred to as 'notification', ensures that both the regions and the developers possess qualities that make the project viable- the government signals its approval of the developer's detailed business plan, and the developer signals keenness in the region and his commitment to the plan by completing the purchase or rental of land. Forming our analysis based on

[5] Authors' calculations from the 2005 round of National Sample Survey (NSS) on Employment and Unemployment.

[6] Hsieh and Klenow (2009) estimate a 40% increase in overall productivity from reallocating resources to larger (formal) firms.

[7] See Aggarwal (2006) for a summary of the policy debate.

comparisons only among regions that were actively targeted by SEZ developers takes care of the first-order concern that the targeted regions may be different, for example in terms of potential for growth, than other parts of India.

While we base our analysis on the variation in the preparatory levels of SEZs, we adopt separate empirical frameworks for exploring different questions. Firstly, we trace the pattern of SEZ influence through time and space using granular satellite lights data as a proxy for economic activity. For this, we apply panel data analysis to a fourteen-year panel of all 1-square kilometer cells that are within 15 kilometers of any SEZ in our sample. This helps us establish that the beginning of operations inside an SEZ sets off an increase in economic activity not only within the SEZs, but also in the immediate neighborhoods around it. We find the effect to be moderately persistent across time and up to areas that are comparable in size to the administrative division of an Indian village. We also find that areas farther away are not hurt significantly (by a potential withdrawal of resources) thus recording net-positive effects up to areas spanning 1,200 square kilometers, one-fourth the size of a median district in our data-set.[8]

Our main results on the aggregate effects of SEZ spillovers are derived from a difference-in-differences framework using a rich set of variables on firms and workers drawn from formal and informal sector surveys. We compare outcomes between regions that have at least one operating SEZ (treated regions) and those with at least one SEZ that pass notification but no operating SEZ (control regions). In choosing control regions in this fashion, our strategy is in line with Busso, Gregory and Kline (2013) in not comparing areas to their geographical neighbours, limiting the concern of spillovers to the treated regions from control regions (such as movement of workers) that may mechanically bias our treatment effect. Additionally, we are able to separate the direct effects of SEZs (on firms and workers within SEZs) from the indirect effect on the local non-SEZ economy by studying subsets of firms and workers with varying possibilities of being located within SEZs. For robustness, we conduct a series of pre-trends analyses for the regions in our study to address the concern that there may be serious differences among areas of consideration in the timing of SEZ operations.

Our results indicate that the average labor productivity of formal sector manufacturing firms in treated districts increases differentially by 24% between 2005 and 2010. We also find evidence for within-industry expansion in formal production by 46%, employment by 18%, and investment in plant and machinery by 37% over the same time period. Along with productivity gains and an increase in the demand for labor, the wages in the treated formal sector experienced a differential increase of 14% over wages in the control district.

Our findings also suggest that the resulting agglomeration spillovers from SEZs structurally transform the economy away from informal lines of production towards greater formality. This is especially true in the

[8] The district is the main level of local government below the state in India.

case of informal manufacturing where we observe a halving of total production within industries of treated districts with total employment declining by 24% and labor productivity by 42%. One possible explanation for this trend is a selection effect driven by an increase in registrations among the most productive firms in the informal sector or those that previously stayed "under the radar" to escape taxes and other regulations. The presence of SEZs likely increased the demand for higher quantity and *quality* of local goods and services, thus motivating informal sector firms to register themselves to signal quality and expand customer base.[9] We also find evidence of a significant decrease in employment in firms at the *lower* end of the productivity spectrum, such as small household businesses. This suggests a reduction in "forced informality," which is usually a result of insufficient formal employment opportunities.

Although the impact of the SEZ policy has been positive in terms of stimulating formalization in the economy, we find evidence suggestive of increasing inequality represented by a non-uniform effect on overall worker wages. While workers at the 90th percentile of the income distribution gain as much as 38% over the years in which their district was treated with SEZs, those at the lower end of the wage and educational distribution seem not to gain significantly. This suggests that workers are left out of the wage benefits due to their inability of being absorbed by the formal sector. It corroborates the current concern in both developed and developing countries about the problem of "skill gap" where the workforce is unable to fulfill the demand for skilled labor thus holding back further prospects of development.

Our work makes novel contributions to the study on place-based development policies. Firstly, we provide a general framework for evaluating place-based development policies using remote sensing data, which can be easily adapted for analyzing other such programs in India or other countries. Our framework is especially useful when policies are *not* implemented at an administrative level such as state, province or county[10]. Our cell-level analysis of policy impact provides a method for studying changes that are within these larger regions, such as industrial corridors, technology parks and other important spatially narrow infrastructure projects.

Secondly, we draw important insights from our focus on how zonal development programs affect the little-explored dual economic structure of the Indian economy. Earlier works studying SEZs and similar programs in India and China, such as Wang (2013), Alder, Shao and Zilibotti (2016), Chaurey (2016) and Shenoy (2016), do not touch upon the effect of such programs on the formal-informal sector dynamics. The closest work to ours, in this respect, is Magruder (2013) who finds that the change in minimum wage rule in Indonesia acted as a big-push mechanism leading to greater formalization.

Lastly, we contribute on a positive note to the mixed evidence on the role of firm agglomeration in boosting

[9] An upcoming work evaluates if increasing registrations and the quality channel are strong explanatory channels.

[10] Such policies allow for analytical methods such as spatial discontinuity design along administrative borders, as employed by Shenoy (2016) in his analysis of the New Industrial policy in the Indian state of Uttarakhand.

productivity and development of a region. Our analysis of the formal sector is comparable to the work of others in studying the impact of place-based programs in developed countries that do not have prominent informal sectors. Greenstone, Hornbeck and Moretti (2010) find positive productivity gains to firms located in the same county as "million dollar plants" in the United States while Kline and Moretti (2013) find that agglomeration gains from the Tennessee Valley Authority program are offset by losses elsewhere in the country. The evidence is also mixed in the case of programs in developing countries when only the impact on the formal economy is considered. Wang (2013) finds that municipalities receiving early waves of Chinese SEZs experience productivity gains while Chaurey (2016) does not find state-level productivity gains from firm agglomeration in the Indian state of Himachal Pradesh as a result of the New Industrial Policy.[11] Our findings of an increase in employment and wages in the formal sector of treated regions, both real and nominal, are in line with those of Kline and Moretti (2013), Busso, Gregory and Kline (2013), Wang (2013) and Chaurey (2016). Given the traditionally low level of labor mobility in India[12], real wage increases are also consistent with the prediction of Moretti (2010) that low labor mobility implies that any benefits from a shock to labor demand accrues to workers residing within a region.[13]

The paper proceeds in the following steps: Section 2 provides the reader with an overview of Indian SEZs. In Section 3, we study the pattern of spillovers caused by SEZs on the surrounding areas. Section 4 explores the aggregate effects of these spillovers and their implications for firms and workers. Section 5 summarizes our insights and highlights further areas of research.

2 The Indian SEZ Experience

The Government of India's SEZ policy was influenced by the success story of the Chinese SEZs. Impressed by the SEZs in Guangdong province in 2000, the Commerce Minister of India initiated changes in India's Export-Import policy, which converted existing Export Processing Zones (EPZs), which were industrial estates that produced export-oriented goods, to Special Economic Zones (SEZs).[14] SEZs were envisioned as comprehensive industrial townships with social facilities like housing blocks, schools and hospitals. The real growth in SEZ activity was kick-started by the SEZ Act of 2005, which officially proclaimed its objectives to be: (a) generation of additional economic activity, (b) promotion of exports of goods and services, (c)

[11] The movement of resources to treated regions from other regions is potentially responsible for muted gains at a spatially aggregated level. Glaeser and Gottlieb (2008) point out that this shift may still have an overall welfare impact on the aggregate economy if the elasticity of productivity to agglomeration is greater in places receiving the programs. Empirically, non-linearity of agglomeration effects is a challenge to establish and many papers, including this one, focus on attempting to document if there are positive net effects on the surrounding economy, up to a sub-national level of aggregation.

[12] Topalova (2010) finds that the landmark trade reform of 1991 that officially opened India to international trade had a surprisingly little impact on the already low inter-district migratory patterns.

[13] High mobility, on the other hand, would predict an in-migration of workers who would apply an upward pressure on land prices and cancel out the effect of any increase in nominal wages.

[14] EPZs do not form part of our sample.

promotion of investment from domestic and foreign sources, (d) creation of employment opportunities, and (e) development of infrastructure facilities.

The Indian SEZs differed in two key ways from SEZs and other place-based programs in the world, including the Chinese model. First, the minimum size requirement was much lower. This resulted in Indian SEZs being physically much smaller than municipality-sized SEZs in China and census tracts designated as Empowerment Zones in the United States. The size requirements were sector-specific; while Information Technology (IT) SEZs were allowed to be as small as 0.1 square kilometers, multi-product SEZs needed to be at least 10 square kilometers of area. The second distinguishing feature of Indian SEZs is that they were open to development by both the public and private sectors, resulting in 70% of the SEZs being either private or joint sector initiatives.[15] These features resulted in two main trends in zone location[16]: both public and private sector SEZs tended to locate in urbanized areas with already existing industrial clusters, or they clustered in belts to promote the development of a new industry within the state. Hence, despite the small size of an individual SEZ, the tendency to cluster increased the potential of agglomeration spillovers to impact regional productivity and economic growth. Studying this unique pattern of SEZ development could thus provide useful lessons to countries that find it economically and politically infeasible to develop large-sized SEZs.

Similar to other place-based development programs through the world, India provided largely fiscal incentive packages to the SEZ developers as well as to the firms locating within SEZs (henceforth referred to as units). Table A.1 in the Appendix provides an overview of these incentives. The most notable of these is the 100% tax exemption on profits for the first five years of operation, which converted to a 50% exemption in the next five after which the same rate was applied to any profit that was reinvested into SEZ activity.[17] Additionally, both developers and SEZ firms were exempted from paying the Minimum Alternate Tax (MAT), which is currently set at 18.5% of book profits in India.[18] The MAT is a compulsory tax levied on companies that make substantial profits but have low, or even zero, tax liability due to the host of deductions and exemptions available under the income tax law. According to a representative survey of SEZ developers and companies by Mukherjee and Bhardwaj (2016), 84% of the interviewed units declared that the tax exemptions formed the biggest motivating factor for them to begin production within SEZs.

Apart from tax benefits, both developers and companies wishing to locate within SEZs enjoyed an ease in administrative procedures through the "single window mechanism". Applications were reviewed jointly by both the Central and State governments through a single regulatory body, the Board of Approval (BoA),

[15] Authors' estimates.
[16] See Aggarwal (2011) for a detailed survey of SEZ developers on issues including zone location and development.
[17] The corporate tax rate is 35% in India.
[18] This exemption was reversed in 2011.

which was set up to facilitate a fast pace of clearances and resolution of bureaucratic red-tape typically surrounding the starting of a business venture. All of these incentives helped to create a relatively hassle-free environment for firms, wished to operate in a country not known for its ease of doing business[19].

Our empirical analysis hinges on the approval process for establishing SEZs. Applications to develop SEZs were submitted to the BoA, which met quarter-yearly and reviewed them based on the following criteria: the quality of the business plan, the plan for financing, land type targeted[20] and prior approvals of the state government.[21] While these factors were repeatedly highlighted during the decision making process, the relative importance of each, and whether this list is exhaustive, is unclear. If the application meets the requirements, the developing company is issued a formal approval. After this, it needs to revert to the BoA with documentation on land rental or purchase agreements as well as with any revisions to the development plan suggested by the BoA. At this stage, the body issues a notification for the SEZ. This is usually brought to the attention of the general public through news articles as well as notice boards erected at the site of the planned SEZ. Construction then commences and an SEZ is considered operational once the first unit starts production within it. Figure A.1 in the Appendix illustrates the approval process of the SEZs.

In our analysis, we only consider those SEZs that pass the penultimate stage, i.e. notification. The Ministry of Commerce and Industry is the primary source of information on Indian SEZs in the form of lists of notified and operational SEZs in India. We merge the lists to obtain information on the developer, date of notification, whether the zone has started operation, zone size, the industrial sector as well as the location of each SEZ (down to the village level, and occasionally to the street level). We then add the actual starting dates of operational SEZs, defined as the year in which the first unit within the SEZ becomes operational, which we source from newspaper articles, BoA meeting minutes and developers' websites. Our data-set, given the rich location details, can easily be analyzed at multiple levels of aggregation- at the neighborhood level with geo-coded location data, village, and district level. This is helpful in merging it with secondary data of different aggregation possibilities (as we do in this paper). Our sample includes all notified SEZs in the states of Andhra Pradesh, Gujarat, Maharashtra, Karnataka, Kerala, Uttar Pradesh and Tamil Nadu, which host more than 80% of total operational SEZs. At any point in time, we compare regions that host (or are about to host) operating or notified SEZs.

[19] In 2017, India ranked 130 out of the 190 countries considered in the World Bank's Ease of Doing Business Index, and has been consistently ranked below countries such as Iran, Nicaragua and Uganda.

[20] The land should not only meet the minimum size requirements, but it should also be a contiguous area that is preferably waste land or unsuitable for double-crop cultivation.

[21] Information is derived from the BoA meeting minutes.

3 Patterns of SEZ Influence Across Space and Time

3.1 Data Overview

Our main data-set for analyzing the pattern of spillovers from SEZs is the Nightime Lights (henceforth referred to as NTL) time series. The raw data (National Geophysical Data Center, 2013) is obtained from the National Centers for Environmental Information. The resolution is 30 arc-seconds, which is equivalent to approximately 854 meters when measured around the center of India, and the data values range from 0 (background noise) to 63 (brightest). We clean the series from 2000 to 2013 by removing ephemeral events and gas flares to obtain persistent and stable lighting.

Nighttime luminosity has been extensively utilized as a proxy for economic growth and development (Henderson, Storeygard and Weil, 2012), degree of urbanization (Ma et al., 2012), degree of electrification (Min et al., 2013), population density (Sutton et al., 1997), etc.[22] The NTL has three main advantages- it is available annually at a high level of disaggregation and provides a neutral measure of a region's economic activity. In the Indian context, given that economic data is spatially coarse, available at infrequent intervals and may fail to capture the extent of economic activity in its vast unorganized sector, the usage of a neutral measure of economic activity enriches the overall examination and adds robustness to the findings derived from other data-sets that are available at higher levels of spatial aggregation. The annual availability further helps in confirming the effects of an SEZ that appear precisely in the year it becomes notified or operational, over neighbourhoods of varying proximity. Our reliance on nighttime lights data to capture general economic activity at the sub-national level is further bolstered by Bhandari and Roychowdhury (2011) who find that district-level GDP in India can be significantly explained by lights data and by Dugoua, Kennedy and Urpelainen (2018) who find that lights data has strong predictive power for village-level electrification in India. Furthermore, by being agnostic to the *nature* of economic activity captured by nighttime lights, we handle the criticism that sub-national comparisons of economic development using nighttime lights is problematic. Works such as Mellander et al. (2015) argue that the correlation between the nighttime lights and economic activity is weak for wages and strong for population and establishment density in the case of Sweden. In our case, even if nighttime lights are only correlated with population density, not economic growth, it still provides us with meaningful interpretation since we are also interested in spatial reallocation of human resources.

Using the NTL data and geo-coding the locations of all the SEZs in our data-set, we overlay a fine grid

[22] The frequency and disaggregated nature of the lights data allows for creative applications in studying a variety of interesting and previously unexplorable issues, such as regional political favoritism as in Hodler and Raschky (2014), and the effects of spatial distribution of ethnicities before colonization (Michalopoulos and Papaioannou, 2013).

layer over the map of India and use each cell as the unit of analysis. Each cell is defined as a square with the length of 0.01 decimal degree, which is approximately 1.025 kilometers at around the center of India. In order to see the indirect or spillover effects of SEZs on the cells, we restrict our attention to cells that are believed to be strictly outside of SEZs. For identifying cells outside the zones, we assume that SEZs are circularly shaped since the exact shape of SEZs are unknown. Using the area of the SEZ reported by its developer, we calculate the radius of the SEZ and draw a circle around the point. The circular shape assumption is a strong one and creates a concern that we might label area that is actually inside the SEZ as non-SEZ area. To avoid defining inside-SEZ cells as outside-SEZ ones, we take a conservative approach and increase the radii of SEZs by 10%.[23] We then record for each cell, the NTL reading over the years[24] and the distances between the centroid of the cell and the projected boundary of every SEZ in our database, thus linking SEZ-level information (such as notification and operation years) with cell-level information. We restrict our attention to cells that are at most 15 kilometers away from their closest SEZ, and the resulting number of observations (cells) is 62,386 per year.

We additionally use the Gridded Population of the World (GPW) data (Center for International Earth Science Information Network, 2016) series in order to measure the extent to which the effects recorded by our analysis of the NTL data is driven by population movements. This data-set, downloadable from the NASA Socioeconomic Data and Applications Center (SEDAC), is available at 5-year intervals from 2000. The lowest level of data resolution is the same as the NTL (roughly 1 kilometer-wide cells). The data for each cell is derived using the population listed by national and sub-national administrative units. In the case of India, this is the level of the sub-district. Each grid cell is assigned with values of population density per square kilometer according to a proportional allocation gridding algorithm, which allocates the same value to all cells within a sub-district.[25] We hence acknowledge that the data-set may have limitations for study at a granular level. However, since the cell-level population is at least not derived from its NTL reading, it is useful to analyze cell-level NTL per population. This provides us with an estimate for how important population movements are as an explanation for the effects produced by the NTL analysis.

Table 1 provides us with an overview of the SEZs in our sample. We have 251 notified SEZs in our sample, with 133 of them operational by 2014. The median size of an SEZ is about a third of a square kilometer, the size of around 32 soccer fields. The mean is much larger at 1.5 square kilometers due to the presence of a few exceptionally large SEZs such as that the Mundra SEZ in Gujarat, which spans 64 square kilometers. Since the IT and electronic sector SEZs make up 69% of the total and their minimum size requirements are small,

[23] See Figure 1 that depicts the fine grid of cells superimposed on circles that indicate SEZ locations as well as size.

[24] In cases there were two satellites collecting data for a cell, we take the average of the two readings.

[25] The allocation is based on an assumption that the population of a grid cell is the exclusive function of the land area within that pixel. Water area such as lakes, rivers, and ice-covered areas are excluded.

such size distribution is to be expected. It is also apparent from Panel A that SEZs are largely a private sector venture with 70% of the zones being developed by purely private or joint sector entities. The average year of notification of an SEZ is 2008 (Panel A) with operating SEZs being notified slightly earlier (Panel B). The difference between average year of notification and operation for SEZs in Panel B shows us that the developers take an average of two and half years to secure the necessary permits, complete substantial construction and attract their first tenants. Panel A of Table 1 also displays the average NTL and NTL per population in 3 kilometer-neighborhoods around SEZ boundaries. These neighborhoods experienced an increase in economic activity, which goes beyond the increase in population judging by the values of both variables after the initiation of the SEZ Act.

It is important to note that due to the NTL data being top-coded at 63, our analysis may be capturing the lower bound of the actual growth that took place. This is because SEZs tend to locate in relatively urban areas with already high values of NTL and may result in the right-censoring of our empirical results. We, therefore, anticipate spillover effects to be underestimated in our study. Figure A.3 indeed confirms that by the end of sample period, 2013, a non-negligible fraction of the data is top-coded implying that our estimates are conservative.[26]

3.2 Empirical Strategy

We utilize the fourteen-year panel data-set of cells, which is constructed as detailed in the earlier section. Our general approach is to compare cells based on their proximity to SEZs and measure the differential effects on them through time due to an SEZ's presence.

For clarity of interpretation, we focus on the time and distance dimensions in separate analyses. The time-dimension analysis uses an event-study framework and examines how a given area reacts to the event of a nearby SEZ beginning operation. The cells considered in the event-study are those that have exactly one SEZ within 3 kilometers that begin operation between 2006, the first year of SEZ operations, and 2013, the last year of observation of the NTL. In making this selection, we aim to (a) focus on the reaction of the immediate neighborhood and (b) reduce the number of SEZs that affect a cell.[27]

For such a cell i that is situated outside of SEZs in year t, we run the following specification:

$$log(light_{it}) = \alpha_i + \beta_t + \sum_{-13 \leq k \leq 7}^{k \neq -1} \gamma_k * D_{ikt} + \epsilon_{it}, \tag{1}$$

where the outcome variable is logged luminosity of cell i at time t, which is the logged transformation of

[26] See Section A.2 in the Appendix for a brief discussion.

[27] Since the treatment is at the SEZ level and the analysis at the cell level, it is possible that a cell is influenced by multiple SEZs in its vicinity, with overlapping notification and operation timelines, especially given the tendency of SEZs to cluster.

the cell's NTL value incremented by 1. A binary variable D_{ikt} takes the value of one if the SEZ within 3 kilometers away from cell i has been operating for k years in year t. Year 0 is the initial year of operation of an SEZ. We use cell and year fixed effects to add robustness to our findings. Additionally, we correct the standard errors for spatial autocorrelation following the specification of Conley (1999) up to a cutoff of 30 kilometers[28].

The estimates of interest are γ_k's, for $k \in [-13, 6]$. Each γ_k can be interpreted as the change in brightness (in log deviations) of a cell, k years since the operation of the nearby SEZ, relative to the year before its operation (γ_{-1} is normalized to 0). We expect γ_k to be positive and its magnitude to be increasing in k after the initial year of operation, indicating persistence in the effect of operating SEZs. Prior to operation, we should not expect any significant trend in γ_k since we do not expect regions to be affected by an SEZ even before it establishes its presence. We also use the same framework and have similar expectations from the coefficients when we study the event of notification of an SEZ.

For the distance-dimension analysis, we classify the SEZs at any point in time into three main age groups: $period0$ denotes years before the SEZ is notified, $period1$ covers the post-notification and pre-operation years, and $period2$, the years after operation. For a cell i that is not located in any SEZ, an SEZ in the x-th distance ring ($x - 1$ to x kilometers away from the cell) exerts an effect that is dependent on whether the SEZ is in period 0, 1 or 2 of its lifetime. In order to study the varying effect of an SEZ across distances, we conduct the following analysis for a particular value of x:

$$log(light_{it}) = \alpha_i + \beta_t + \gamma_x * period1_{ixt} + \delta_x * period2_{ixt} + \sum_{d=x+1}^{15} \sum_{\theta=0}^{2} \lambda_d^{\theta} period\theta_{idt} + \epsilon_{it}, \quad (2)$$

The main outcome variable is logged luminosity of cell i in year t and we run the above specification for each integer value of x from 0 to 15. For the x_{th} distance ring analysis, we consider all cells that have their *closest* SEZ in this ring. $period\theta_{ixt}$ is the number of $period\theta$ SEZs that are in the x_{th} ring away from cell i in year t, for θ values 0, 1 and 2.[29] The average difference in the effects of period 2 and period 0 SEZs (the base group), denoted by δ_x, captures the additional effect felt on a cell due to the operation of an SEZ in the x_{th} ring. The average change to the NTL of a cell due to the notification of an SEZ in the x_{th} ring is similarly denoted by γ_x. δ_x can be thought of as a long-run effect of an SEZ on a cell, and γ_x, the short-run effect. Both measures are useful in developing an understanding of the changes that the local economy experiences, although the latter effect is of greater economic interest. The double summation term indicates that we control for all farther away SEZs affecting the cell (but not within the x_{th} ring), in whatever period of life

[28] The results are robust when the cutoff is alternatively set to 10, 20, 50, and 100 kilometers.

[29] The empirical results are qualitatively unchanged when the terms become dummy variables of whether or not there is at least one SEZ in each period.

they may be.[30] Just as in Specification 1, the standard errors are corrected for spatial autocorrelation.

We expect δ_x to be decreasing in x, i.e. cells closer to SEZs experience the most positive effects while cells farther away experience less positive or even negative effects. The latter would be the case if there was a movement of resources away from farther areas to areas closer to SEZs. This is a reasonable expectation given that the zones bring in new firms that attract workers as well as other firms to the region due to an increase in opportunities for work. We also expect that the effects of operating SEZs are greater in magnitude than the effects of notified SEZs at all distances from the SEZs ($\delta_x \geq \gamma_x$).

Using the specification above, we are also able to test if changes to the neighborhoods of SEZs in terms of NTL is driven by population movements, a channel we can test at the granular level with the main outcome variable being logged lights per population. Since the population data is available at 5 year intervals, the period of life of an SEZ is updated every 5 years. We expect that a significant portion of the expected increase in NTL at neighborhoods close to SEZs will be driven by an increase in population in that neighborhood. The magnitudes should, however, be taken as less reliable than in the analysis using only NTL data due to the way in which population data is constructed for India, as discussed in Section 3.1.

3.3 Results

Inside SEZs: Event Study of Activity within SEZs

An event-study of areas inside SEZs confirms the relevance of Specification 1. Figure 2a reports the percent change in the cells' NTL (derived from the γ_k's), which shows a clear and persistent increase in activity after the beginning of SEZ operation. The figure, however, also indicates an upward trend in economic growth prior to the SEZs' operation. One possible explanation for this trend is the preparatory activity undertaken in and around SEZs after notification (in terms of building, road construction, setting up water distribution networks, etc.). As discussed in Section 3.1, the data shows that it took 2 to 3 years on average after their notification for the operating SEZs in our sample to begin operation. Figure 2b, which studies the event of notification, confirms that the pre-operation growth in Figure 2a may indeed be driven by post-notification activities.

SEZ Spillovers over Time

We then restrict our attention to areas strictly outside, but within 3 kilometers from any SEZ to study the spillover effect of SEZs across time. Figure 2c displays a similar trend as the within-SEZ analysis. When an SEZ starts operating, the immediate neighborhood experiences a significant increase in NTL, which is lower

[30] Figure A.2 contains a pictorial representation of Specification 2 for greater clarity.

than the increase within SEZs as expected. It is persistent in magnitude but loses significance over time. We suspect that this is due to the small number of SEZs older than 4 years old by 2013, considering the mean year of initial operation is 2009 according to Table 1.

We still observe the upward trend in NTL prior to operation. Since the area of analysis is physically outside zones, the increase in post-notification activity can be a result of both construction activities that may extend outside zones such as building external connecting roads as well as the surrounding economy preparing for the impending shock to local demand for goods and services. Examples of the latter would be the construction of hostels and residential properties to host potential out-of-area SEZ workers. This may still not explain the upward-sloping trend prior to even the notification of SEZs as Figure 2d shows.[31] However, we can still argue that regions around notified SEZs experience a noticeable increase in the slope of NTL, i.e. growth rates, upon notification.

Figures A.4a and A.4b display results from an alternate specification in which we modify the event to be the year of the *earliest* operating/notified SEZ within 3 kilometers. This specification is more flexible in the selection of cells than the current one and allows for the presence of multiple SEZs within the 3 kilometer radius. For example, these cells could be representative of more urban areas, which attract a greater number of SEZs. This does not, however, produce significantly different patterns, adding to the robustness of our finding.[32]

SEZ Spillovers across Space

Figures 3a and 3b, plot γ_x's and δ_x's in Specification 2, illustrating the spatial extent of SEZ spillovers in the long- and short-run, respectively. The resulting trends are in line with our expectations. In the long- and short-run scenarios, we observe a sharp increase in NTL in neighborhoods of operational SEZs and notified yet not operating SEZs (relative to the average level in the pre-notification period), respectively. The positive impact especially continues to be significant for areas with the closest SEZ within 4 kilometers in the long run.

Interestingly, from Figure 3a, we do not find evidence of a zero-sum relocation of resources due to the SEZ policy, at least at the level of aggregation considered in this paper- at areas up to 15 kilometers away from SEZs.[33] While areas at a distance greater than 5 kilometers from SEZs seem to return to their normal growth pattern, there is no strong evidence of farther regions being negatively affected by a withdrawal of resources. This area is equivalent to quarter the size of a median district in our data-set. We cannot,

[31] The pattern is preserved when we control for time trends in addition to year fixed effects.

[32] We also experiment with increasing the radius of the neighborhood to 5 kilometers to get similar results, which are not included in the data appendix and are available on request.

[33] This holds true when we check patterns over a wider area of 20 kilometers' radius (results of which are presented in Figure A.5) that spans 1,200 square kilometers.

however, rule out zero-sum effects over wider regions with this framework.

One concern about Specification 2 is that it considers cells that have their closest SEZ in the x_{th} distance ring and thus looks at different sets of cells for the analysis across distances, i.e. cells that do have SEZs in the x_{th} distance ring are not considered for the x_{th} distance ring analysis if they have at least one SEZ that is *nearer* than x kilometers.

We carry out an alternate, less restrictive specification by considering all cells that have at least one SEZ in the x_{th} distance ring, and controlling for *both* nearer and farther away SEZs when studying the effects of SEZs at a particular distance ring x around a cell.[34] Results from this alternate specification is displayed in Figures A.6a and A.6b. While the pattern of influence is similar to the ones produced with the original specification, there is stronger evidence that regions farther away than 5 kilometers do not get hurt and in fact experience positive and significant effects due to operational SEZs.

In Figures 3c and 3d, we adopt Specification 2 to show the effect of having an SEZ nearby by the distance to the closest SEZ on NTL per population while acknowledging the drawback of the population data detailed in Section 3.1. The population movements do seem to drive some, but not all, of the increase in NTL especially in the long-run scenario. This suggests that there are other channels at play, especially in the case of effects produced by operating SEZs.[35]

The chief takeaway from the analysis in this section is that we find evidence of Indian SEZs producing persistent and positive spillovers on the local economy as measured by nighttime lights. While they boost economic activity in areas up to 4 kilometers from zones (comparable to the size of a typical Indian village), farther away areas do not seem to be significantly hurt. SEZs, therefore, can be viewed as generators of net positive effects on areas several times the actual size of the zones and even comparable to the size of an Indian district. An important caveat to our findings is that we are unable to determine if there are net positive effects at the national level. It is possible that activity shifts from elsewhere in the country to areas around SEZs, producing little if any net benefits at the aggregate level. Our results speak only to regional positive effects. This finding, however, motivates a deeper look into the forces behind the spillovers generated in the local economy. In addition, it also lend credibility to the results derived in the following section which use more traditionally available data-sets on firms and workers in the Indian economy that are available at the spatially aggregated level of the village or district.

[34] See Section A.5 in Appendix.

[35] Acknowledging the limitation of the GPW data, we run a separate analysis examining the influence on SEZs on sub-district level population movements by aggregating the NTL and GPW data at the sub-district level. A detailed explanation and the corresponding result is presented in Section A.7 in Appendix. In short, it is unlikely that there are any population movement across sub-districts that can be attributed to the introduction fo SEZs.

4 Aggregate Effects of SEZs on Firms and Workers

We have shown in the previous section that the introduction of SEZs into a region promotes general economic activity. The effect appears positive up to a level of geographical aggregation, a district, which is of interest to political and administrative authorities. A district is the main level of local governance below the state and is divided further into sub-districts that consist of villages and towns. In our sample, there are 68 districts, each of which is, on average, divided into 19 sub-districts. Each sub-district is further divided into an average of 9 villages. In order to understand the general equilibrium effects of SEZ activity, we now examine individual and firm behavior in response to the introduction of SEZs.[36] We employ multiple data sets at different levels of administrative units- the village, sub-district and district.

4.1 Key Variables

At the village level, we analyze firm and worker numbers using the Economic Census (EC). The strength of the EC data is that it is a complete enumeration of all enterprises in India (except those engaged in crop plantation and cultivation) and is identifiable at the village level. Through firm-level information on employee size, industry and ownership type, it provides us with an overview of the distribution of activity across industrial sectors, both manufacturing and services, in every village or town. The EC data covers information on 28 million firms located in the districts of our interest in each of the two rounds available, 2005 and 2012.[37]

At the district level, we make use of more detailed firm characteristics such as production and wages to shed light on the mechanism behind SEZ effects. For studying movements within the formal manufacturing sector[38], we use the Annual Survey of Industries (ASI) dataset. The ASI dataset is an annual survey of firms in manufacturing that are considered formal, i.e. those registered under the Factories Act[39]. It is a complete enumeration of firms above a 100 in worker strength with an annual survey of a repeated random cross-section of smaller firms. This data contains more information than the EC, which allows for a deeper analysis of firm-level variables such as size (employment, asset base and production), new firm formation and costs of production including wages and rents. The data set covers around 30,000 firms in the districts of our interest annually from 2000 to 2009. We also make use of a similarly rich set of firm-level information provided by the informal sector counterpart to the ASI, the National Sample Survey (NSS) Unorganized Manufacturing and Services quinquennial survey data, which covers firms in the unregistered sector of the

[36] By 'firm', we imply establishment or factory-level information.
[37] Results pertaining to village- and sub-district level analyses can be found in the Appendix.
[38] Nationally representative surveys on formal service firms have not been conducted so far in India.
[39] Registration under the Factories Act is required for firms above 10 workers if the unit uses power, and above 20, if not. This is also the standard definition of formality adopted by researchers on the Indian economy.

Indian economy. Each survey round in our study (2000, 2005 and 2010) contains information on around 35,000 firms in the districts of our interest.

We complement the firm analysis with worker-level information which allows us to analyze worker wage effects within districts taking into consideration individual characteristics such as education level attained and household demographics. This information is derived from the NSS Employment and Unemployment Surveys. The data is a repeated cross-section of a nationally representative sample of workers from across all industrial activities. Information on firm type and industry, wages, household characteristics, education and consumption is provided. The data set covers around 104,000 workers in every round considered (2000, 2005 and 2010) among the districts in our study. While the worker-level survey data does not have explicit indicators for whether the worker is employed in the formal or the informal sector, we make the distinction using the 10- worker rule of the Factory Act in order to analyze the effects of SEZs separately on the informal and the formal work-force.

Table A.2 in the Appendix provides an overview of the regions and population studied in 2005. The regions in our study were on average much denser than the all-India average of 382 people per square kilometer. More than 90% of the working population in a district received no education above secondary level. Average firm size was small with 96% of the firms employing below 10 workers. This indicates the highly skewed firm size distribution and the vast size of the informal sector.[40] The extent of informality in the economy is also apparent from indicators such as the proportions of firms that hire no workers at all (40%), operate without power (34%), do not have external financing options (96%) and rely on informal sources of finance (around 40%). We also find evidence of the low level of productivity in the informal sector with wages in the formal sector manufacturing being on average almost 10 times that of those in the informal sector.

4.2 Empirical Strategy

In this section we adopt a difference-in-differences framework which chiefly requires that both treated and control regions follow common trends prior to the initiation of the SEZ policy. This is a challenge since the locational choices of such place-based policies are not random. In the case of developed economies, zones are usually located in under-industrialized regions. In the Indian context, however, because SEZ development was mainly a private sector initiative, the program targeted regions with a greater degree of urbanization, human capital quality and profitability. A simple comparison of areas with SEZs and those without would therefore be unsatisfactory as it will likely violate the common pre-trends assumption and bias our difference-in-differences estimate of effects upwards.

Our solution is to adopt a similar strategy to the one commonly used in the literature on place-based

[40] The figure is comparable to estimates in Amirapu and Gechter (2014).

development policies. We use the the approval process of SEZs as a source of quasi-experimental variation to compare regions that have at least one SEZ that has passed notification but none that is operating in them (the control group) with those that have at least one operating in them (the treated group). The exact set of regions that fall into treatment and control categories differ according to the frequency and level of aggregation allowed by the multiple data-sets used.[41] In the case of Busso, Gregory and Kline (2013) and Kline and Moretti (2013), control regions were formed out of rejected areas. There is an added strength in our case such that the control areas were never disqualified by administrative authorities and were indeed expected to have operating SEZs in the near future.

A point immediately in favor of our identification strategy is that it takes care of the first-order concern that areas attracting SEZs, both notified and operational, may be fundamentally different from other areas in terms of worker and industrial composition and growth potential. Table 3 provides some evidence for this when we compare across treated, control and other districts in the states in our sample. Both treated and control districts are almost ten times as dense as other districts in the states, consistent with the trend of SEZs locating in relatively urban areas. Treated and control districts also seem to have substantially different worker compositions than the rest of the state, especially with respect to the proportion of workers employed in agriculture and manufacturing. The rest of the state seems to predominantly depend on agriculture for its livelihood with 64% in related professions compared to only 7% in treated or control districts. The average monthly income of workers and their education level in the districts compared against each other in our analysis are also clearly higher with the proportion employed in informal household businesses being lower by around 10 percentage points. This table tells that results from our proposed comparison of treatment and control districts will be more credible than a simple comparison of SEZ and non-SEZ districts.

The second-order concern comes from the fact that treated regions seem to be disproportionately among those with earlier notified SEZs, as shown in the bottom panel of Table A.3. This could imply that the order of timing in application and notification is correlated with unobservables relevant for the outcomes studied. However, anecdotal evidence points to developers being spatially restricted in their choice of location. Locations were not chosen solely based on profitability but also giving weight to the own-state bias of SEZ developers, both public and private. State governments always started SEZs within their states, and private sector developers usually chose locations within the state or district in which they are headquartered.[42] It is then quite likely that the control areas did not form part of the choice set of developers in treated regions. This could be more due to their out-of-state location rather than growth potential. Moreover, we find that 60 percent of the control districts were targeted *earlier* by SEZ developers that did not manage to reach the

[41] Refer to Table A.3 in Appendix for an overview.
[42] Evidence is gathered from interviews with SEZ developers in Tamil Nadu.

stage of notification. We derive this information from the BoA meeting minutes which discuss the decisions made on all the SEZs that ever applied for approvals. These areas attracted initial interest around the same time as treated areas according to the bottom rows of Table A.3. We find that SEZs in both control and treated districts received early formal approvals within the first year of the SEZ Act. The average difference between the two groups receiving a formal approval was only about 8 months.

Given that treated and control regions received developer interest within a comparable time frame, one is also naturally concerned about the delays to notification and eventual operation of SEZs in control regions, and whether this was correlated with the potential of the regions and the selective success of the SEZ policy. If the set of selection criteria for regions were completely known, as was in the case of Busso, Gregory and Kline (2013) in their evaluation of Empowerment Zones, we could use techniques such as propensity score weighting to increase the comparability between treated and control regions. In our case, however, the exact set of conditions used by the BoA to deem an SEZ notification-worthy, is unknown. What we know from the BoA's meeting minutes is that these included other criteria than economic indicators for the region- such as the ability of the developer to possess the land, obtain environmental clearances and propose a viable development plan. The difficulties in land acquisition and clearing administrative steps including environmental clearances, which were prerequisites for notification, are also well documented by Mukherjee and Bhardwaj (2016) and Aggarwal (2006). Arguably these measures are related to state-level factors such as the lack of land banks, or delays in administrative clearing that can be believed to be time-constant at least within the short-frame of our analysis of the SEZ policy. There is also the possibility that treated regions differ in the number of 'capable' developers than control regions given that they have earlier operating SEZs; we address these issues with the addition of region fixed effects to our framework which would take into account non-time varying differences in business and administrative potential among regions.

We derive additional evidence of common trends among treated and control regions by utilizing NTL data to compare time trends among cells in 5 kilometer-neighborhoods of SEZs that eventually became operational versus those that never became operational (by 2013, the last observation year for NTL). Figures 4a and 4b show that both kinds of regions did not experience significantly different pre-trends (as captured by trends in NTL) before notification, regardless of whether the SEZs in them ever began operation. Since the areas covered by a 5 kilometer radius is roughly twice the size of an average village in our sample, the analysis confirms common trends among relative large portions of land. We also find evidence for 'developer seriousness' in control regions from the trend in Figure 4b where areas around notified SEZs seemed to experience an increase in economic activity in the initial couple of years after notification judging from the spike in economic activity before the region returned to its normal growth path. This supports our belief that the developers of notified SEZs were committed to the region and to the project.

We also take advantage of the ASI data to analyze pre-trends in district totals of production, assets used and employment in formal manufacturing sectors prior to the SEZ Act. Figure 5 shows that there are no discernible differences in trends among treated and control districts, except for a slightly higher trend in employment in treated districts. We also carry out the following falsification test using the same information at the firm-level, to see if operating and notified SEZs produced effects on formal firm activity even before their introduction in districts:

$$log(y_{fidt}) = \alpha_0 + \alpha_i + \beta_t + \gamma * No.EventuallyOperatingSEZs_d + \delta * No.EventuallyNotifiedSEZs_d + \varepsilon_{fidt} \quad (1)$$

The outcome variable y_{fidt} takes the logged values of variables related to an average firm f in 2-digit industry i in district d at time t: such as production, investment, employment, wages and average productivity. The main regressors are the number of eventually operating and notified SEZs that the district receives after 2005. Since the analysis is over the time period between 2000 and 2005 (before the announcement of the SEZ Act.), the corresponding coefficients,γ and δ , should not show any significance if there had not been any differential growth pattern before the introduction of the SEZ Act. The results can be found in Table 2 and confirm that there were not any significant differences in growth patterns in the pre-periods.

Given the supporting evidence for parallel trends among districts with notified SEZs and those with operational SEZs before the initiation of the SEZ policy, we first utilize the annual data on the formal manufacturing sector in a generalized difference-in-differences framework to study the year-on-year effects of additional SEZs as they become notified or operational within a district. We only consider districts that have been notified with at least one SEZ before 2010. The treatment variables are now the stock of operating SEZs in district d at time t and the stock of notified but not yet operating SEZs in the same district at time t:

$$y_{fidt} = \alpha_0 + \alpha_1 * No.OperatingSEZs_{dt} + \alpha_2 * No.NotifiedSEZs_{dt} + \beta_i + \gamma_d + \delta_t + \varepsilon_{fidt} \quad (2)$$

y_{fidt} is the outcome variable of an average formal manufacturing firm f in industry i and district d at time t. These include logged values of production, average labour productivity (defined as total production per worker), wages, employment and value of plant and machinery. $No.OperatingSEZs_{dt}$ refers to the total number of operational SEZs in district d at time t and $No.NotifiedSEZs_{dt}$ refers to the number of notified but not yet operational SEZs in district d at time t. District, 2-digit industry and year fixed effects are included with the standard errors being clustered at the district level. In the above specification, we assume that every SEZ, operational and notified, has a uniformly additive effect on the outcomes of a firm in a district-industry cell.

We expect positive effects on firm-level measures such as average labor productivity, employment and production due to the presence of an additional operating SEZ, i.e. α_1 to be positive. This would provide proof of the push that SEZs give to local demand and of the productivity spillovers that they are capable of generating. The specification also allows us to evaluate the changes that are brought about by notified SEZs in the region (α_2) which could reveal the mechanism behind the increase in light activity following SEZ notification shown in the previous section.

In order to analyze the formal sector data along with the less frequently available informal sector data, we use the following framework:

$$y_{irt} = \alpha_0 + \alpha_1 T_r + \alpha_2 T_r AFTER_t + \beta_i + \gamma_r + \delta_t + \varepsilon_{irt} \tag{3}$$

y_{irt} is the average firm-or worker-level outcome variable (such as logged worker wages or firm size) in industry i in region r in year t. T_r is the treatment indicator, which takes the value of 1 for regions that were treated with at least one operational SEZ before the post-treatment period as stated in Table A.3 in Appendix. The value is 0 for regions that have at least one SEZ notified before this time but none operational yet. $AFTER_t$ is the time indicator, which takes the value 1 for the post-treatment period, and 0 otherwise. Region, industry[43] and year fixed effects are included with standard errors cluster-robust at the level of the region.

α_2 is the coefficient of interest, which describes the change in an outcome such as average employment of a firm located in the treated region with respect to the control region due to the presence of at least one operational SEZ. Depending on our analysis of the formal or the informal sector, our expectations differ about the effect that SEZs are bound to have on firms. For formal sector firms, we expect positive productivity spillovers that encourage production, and boost investment and employment. We also expect an increase in wages paid by the formal firm due to the increase in productivity as well as greater demand for labor. For firms in the informal sector, we expect a priori that the increase in labor demand and wages in the formal sector may lead to a reduction in sustenance-level self-employment with workers moving to formal firms that are expanding, paying more and offering greater job security. The increase in demand for local goods by SEZs could also motivate more productive firms in the informal sector to pay the cost of being regulated and gain from the increased profitability of being formal. This would result in a reduction in the overall size of the informal sector in terms of employment, assets and production.

[43] In an alternate specification, we consider region-industry fixed effects.

4.3 Results

Effects on Formal Firms in the non-SEZ economy

To study the changes in an average firm in formal manufacturing, we use a 9-year district-industry panel and follow Specification 2. Column 2 of Table 4 shows us that every additional operating SEZ results in a 2.2% increase in an average formal firm's production, accompanied by a 1.5% increase in asset usage and 1% increase in employment. Labour productivity, both average and marginal (i.e. wages paid), experience a significant increase of 1.8% and 1.2% respectively. Assuming a constant returns to scale Cobb-Douglas production function, this would point to an increase in total factor productivity between 0.7 % and 1.2% depending on the value of output elasticity of capital.

Note that the presence of an additional notified SEZ (the values of which are presented in Table A.4) also seems to positively impact firm investment and the wage level. This is consistent with our findings on the increases in NTL upon SEZ notification, which we hypothesize could be due to an increase in demand for activities related to the development of an SEZ as well as the anticipation effect of a bigger customer base for firms in the future. This would induce a greater demand for labour and capital, reflected in the increase in wage and investment.

In order to distinguish between direct effects on firms beginning production within SEZs and spillovers on firms outside the zones, we explored the effects of SEZs on firms in different employment size bins- size 1: (0,10), 2: [10,20), 3: [20,100), and 4: [100, .). We do this because all firms within SEZs must necessarily belong to the registered sector. So we may just be capturing the effect of their presence in column 2 of Table 4. Since we are mainly interested in spillovers, the size-wise analysis helps us check if effects are just concentrated among larger firms, which are more likely to be firms within SEZs than the smaller ones. From the rest of Table 4, we see that this is not the case. Small firms with employment under 20 also seem to benefit from significant increases in production and investment in districts treated with an additional operating SEZ. One plausible reason we do not notice an increase in employment among size 1 and 2 firms could be because previously unregistered informal firms join these bins, finding it profitable to register themselves in order to establish supply linkages with SEZ firms and workers. This would pull down the average size of formal firms in the under-20 category because the switchers are likely to be small (due to the 10-worker rule for registration under the Factories Act). Firms, regardless of size, experience increases in average labor productivity and pay higher wages as they become more productive and increase their demand for labor.

We also find evidence of every additional operational SEZ instigating a 1% increase in the proportion of new firm formation across size categories, implying that firm formation is not only restricted to large, newly operating SEZ firms. We additionally resolve the concern that the positive spillovers we observe in Table 4 come only from these newly formed firms rather than old, already existing firms in the districts. Table 5 splits the samples into firms that are older than 4 years (i.e. older than the announcement of the SEZ Act of 2005) and those that are not. Every additional operating SEZ has a strong effect on new firms, especially with respect to production and productivity. Put differently, the presence of an additional operating SEZ stimulates the formation of new firms that are 18.6% more productive and pay 5% higher wages. Interestingly, the effects of SEZs on old firms are also positive and significant, albeit lower in magnitude. Old firms, existing before the SEZ policy came into force, also experience an increase of 0.4% in productivity and wages, expand in employment by 0.9% and show signs of production and investment expansion of 1.3% with every additional SEZ operating in their district. The stronger effects on new firms is reasonable considering that new firms are probably direct results of SEZs coming into districts and have the flexibility to adopt the latest technology or best practices right upon formation and do not face the inertia that old firms may face in changing or upgrading production methods to increase competitiveness.

Effects on Informal Firms in the non-SEZ Economy

Our analysis of the informal sector highlights the opposite effects SEZs have on the formal and informal portions of the economy. Columns 1 and 3 of Table 6 exhibit the values of coefficient α_2 from Specification 3 with district, industry and year fixed effects, and Columns 2 and 4 instead employ district-industry fixed effects and year effects. The presence of at least one operational SEZ in a district has made the average informal manufacturer experience a halving of value-added and total production and a decrease in asset usage by 32% compared to a firm located in a district without an operational SEZ. The firms also shrink in size with respect to employment by about 20%. Since the average number of workers in these firms is 3.9, this would imply the exit of 0.78 workers from a firm on average. Labor productivity, as measured by gross value added or production per worker, and average wages paid are also negatively impacted with almost a halving of wages paid to an average worker in the treated district relative to the control.

While the presence of an SEZ seems discouraging to unregistered manufacturing firm activity, Columns 3 and 4 of Table 6 show that the negative influence does not extend to the unregistered service sector. Here we observe an almost equal and opposite trend, especially in firm-level investment and employment. This is in line with the expectations of big push models such as the one formulated by Magruder (2013), which expect most of the formalization to happen in the tradeable and industrializable sectors such as manufacturing. In the case of manufacturing, the tradability factor results in firms facing greater pressure of losing business to those outside the region if they did not cope with productivity increases of competitors or demand for higher quality products. Also, since manufacturing is generally industrializable (i.e. production is scalable at an industrial level), informal manufacturing may be crowded out when its formal sector counterpart receives a big push in productivity and demand. On the other hand, since services are not often industrializable and tradeable, increases in local demand has to be satiated by local service firms, both informal and formal. This justifies the expectation that the informal service sector does not face crowding out as in the case of informal manufacturing due to SEZ presence.

Aggregate Effects on the Dual Economy

We analyze the informal and formal sectors within the same empirical framework by considering two pre-treatment years (2000 and 2005) and one post-treatment year (2010) for the formal sector analysis instead of the annual data. We then use the totals within a district-industry cell of production, investment and employment in formal and informal sectors as outcomes. The results derived from Specification 3 are shown in Table 7.

While the formal sector in any 2-digit industry in a treated district experiences a boost of 46% in production, 37% in investment and 18% in employment, the informal sector of the same district-industry group experiences opposite effects of a larger magnitude in all the three parameters. The gains to labour productivity in the formal sector is also accompanied by an even greater loss in the informal sector. Just as in the firm-level results, we observe that total activity within informal services expands significantly unlike in informal manufacturing.

Given a productivity distribution within the informal sector, its decline as observed in Table 6 and the second panel of Table 7 could be driven by the movement of firms at the both ends of the distribution. The most productive firms are likely to switch out of the unregistered sector in order to gain from the increase in demand for goods generated by SEZs, which would explain the decrease in average productivity and total output in the informal sector. Alternatively, they could be shutting down if they fail to cope with the increased competition from the expanding formal sector. Informal sector crowding out could also happen due to less productive firms shutting down because the workers no longer have to resort to subsistence activities ("forced informality") with more job opportunities in services or formal manufacturing.

On the bottom panel of Table 7, we provide evidence of movements of resources, in terms of workers and firms, between the formal and informal sectors of treated district economies. Using the NSS worker-level surveys, we observe an 8.2% decrease in the proportion of workers employed in small household manufacturing businesses that are generally less productive, providing some support for a reduction in forced informality. Correspondingly, the worker-level survey also indicates that the worker composition in the formal sector, as proxied by the 10-worker rule, rises by 7.9% in treated districts over control districts. The last two columns in the bottom panel use the Economic Census to analyze changes in firm counts in districts- both in the total number of firms as well as the number of informal firms, proxied by the ten-worker rule. A 21.1% decrease in the number of informal firms in treated over control districts accompanied by no significant change in the total number of firms also corroborates the story of increased formalization brought about by SEZs in the local economy. An upcoming paper, Ravi (2018), weighs the relative importance of two channels- informal firm deaths versus increased registrations into the formal sector- in explaining this tend of SEZ-driven economic formalization.

Effects on the Overall Wage Distribution

Table 8 uses worker-level survey data and reveals the results of quantile regressions on worker wages in the 90th, 75th, 50th, 25th and 10th percentile of the income distribution. While the average wage level in a district treated with at least one SEZ differentially increases by 17.9% relative to that of the control district, this differential wage increase is clearly not uniform across the distribution of workers: workers in the upper end of the distribution seem to gain the most in treated districts relative to control districts, with the 90th percentile wage earners experiencing the maximum differential wage increase of around 42%. There is no significant increase in wages among workers in the lower end of the wage distribution in treated relative to control districts, with the 10th percentile workers in treated districts even appearing to experience a relative decline in wages, albeit statistically insignificant.

Table 9, which also uses worker-survey data, shows that this pattern is driven by increases in formal sector wages in both manufacturing and services as well as increases in the returns to higher education. Workers who have above secondary school level of education (junior college and above) experience a wage increase of 66% between 2005 and 2010. Workers with lower education levels do not seem to gain significantly in terms of wages. This is in line with the fact that SEZs directly increase wages in the formal sector and not in the informal sector and the general tendency of education and formality of occupation being positively correlated.

Interestingly, we also observe a slight but significant decrease in the proportion of people in treated districts that are educated below the primary level, the results of which are in Table 10. This could be interpreted as a result of in-migration of more educated workers in order to take advantage of the higher paying labor market in treated districts. While data limitations prevent us from directly testing movements of workers across the education or skill spectrum, there is some support for a general increase in population density in areas treated with operational SEZs from the analysis of NTL per population in Section 3. Given the short period of analysis, it is less likely that the effect could be due to the local population being driven to invest in higher education.

5 Concluding Remarks

We show that SEZs did not only benefit firms locating within them but also produced local economic spillovers which reflected at the aggregate level of a district. We then delve deeper to show evidence for positive productivity spillovers and firm expansion in the formal manufacturing sector, as well as for crowding out of the informal manufacturing sector. We thus present evidence of SEZs driving a structural change of the economy towards greater formalization.

The results are especially striking because of the nature of the Indian SEZs which are smaller and privatized compared to those in other countries. However, the non-uniform gains in wages among workers serves as a caution about low-skilled workers potentially losing out when spatial development policies are implemented. These lessons could be useful for implementing such programs in other developing countries that share similar political and economic realities.

While we do not carry out explicit cost-benefit calculations, our work prompts questions about the cost effectiveness of the SEZ policies. Cost-benefit calculations of such programs in the past show mixed results. While Busso, Gregory and Kline (2013) estimate net moderate benefits to the development of Empowerment Zones, Chaurey (2016) casts a doubt in the case of the New Industrial Policy Scheme that the increase in reported profits could come from either a true increase in production or simply more truthful reporting.[44] In our case, a cost-benefit analysis is trickier to carry out because of difficulties in calculating the cost of foregone tax revenues. Tax holidays are directed at a much smaller subset of firms and not to all firms in a particular state or census tract. In this case, weighted survey data on income cannot be used in conjunction with the tax rate to calculate a tax bill, as is done in Shenoy (2016). The focus of this paper is hence on an evaluation of the the benefits side, leaving the estimation of the cost-effectiveness of this venture as work for future papers.

Another interesting future area of research is to compare benefits across the different types of zonal development programs launched by the Indian government in recent years. The recently launched National Investment and Manufacturing Zones (NIMZs) differ from the SEZs in some important ways, being mainly state-led initiatives, not offering complete tax holidays and also being centers for domestic as well as export-oriented production. These differences could lead to different outcomes, and it would be of great policy interest to compare the effects of the different zonal development styles and draw conclusions on the optimal design for the Indian context.

[44] The paper also does not take into account the influence on informal sector workers, who could be losers in this policy as shown in our research.

References

Aggarwal, Aradhna. 2006. "Special Economic Zones: Revisiting the Policy Debate." *Economic and Political Weekly*, 41(43/44).

Aggarwal, Aradhna. 2011. "Promoting Agglomeration Economies and Industrial Clustering Through SEZs: Evidence from India." *Journal of International Commerce, Economics and Policy*, 2(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.

Amirapu, Amrit, and Michael Gechter. 2014. "Indian Labor Regulations and the Cost of Corruption: Evidence from the Firm Size Distribution." *Working Paper.*

Bhandari, Laveesh, and Koel Roychowdhury. 2011. "Night Lights and Economic Activity in India: A Study using DMSP-OLS Night Time Images." *Proceedings of the Asia-Pacific advanced network*, 32: 218–236.

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.

Center for International Earth Science Information Network. 2016. "Gridded Population of the World, Version 4 (GPWv4): Population Density." NASA Socioeconomic Data and Applications Center. (accessed November 16, 2016).

Chaurey, Ritam. 2016. "Location-based Tax Incentives: Evidence from India." *Journal of Public Economics*, 101–120.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of econometrics*, 92(1): 1–45.

Dugoua, Eugenie, Ryan Kennedy, and Johannes Urpelainen. 2018. "Satellite Data for the Social Sciences: Measuring Rural Electrification with Night-time Lights." *International Journal of Remote Sensing*, 39(9): 2690–2701.

Duranton, Gilles, and Anthony J. Venables. 2018. "Place-based Policies for Development." *World Bank Policy Research Working Paper; no. WPS 8410.*

Farole, Thomas, et al. 2011. *Special Economic Zones: What have we Learned?* World Bank, Poverty Reduction and Economic Management Network.

Glaeser, Edward L, and Joshua D Gottlieb. 2008. "The Economics of Place-Making Policies." National Bureau of Economic Research.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Opernings ." *Journal of Political Economy*, 118(3): 536–598.

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer space." *The American Economic Review*, 102(2): 994–1028.

Hodler, Roland, and Paul A Raschky. 2014. "Regional Favoritism." *The Quarterly Journal of Economics*, 129(2): 995–1033.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocatiion and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4).

Kline, Patrick, and Enrico Moretti. 2013. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics*, 129(1): 275–331.

Magruder, Jeremy R. 2013. "Can Minimum Wages Cause a Big Push? Evidence from Indonesia." *Journal of Development Economics*, 100(1): 48–162.

Ma, Ting, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan. 2012. "Quantitative Estimation of Urbanization Dynamics using Time Series of DMSP/OLS Nighttime Light Data: A Comparative Case Study from China's Cities." *Remote Sensing of Environment*, 124: 99–107.

Mellander, Charlotta, José Lobo, Kevin Stolarick, and Zara Matheson. 2015. "Night-time Light Data: A Good Proxy Measure for Economic Activity?" *PloS one*, 10(10): e0139779.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-Colonial Ethnic Iinstitutions and Contemporary African Development." *Econometrica*, 81(1): 113–152.

Min, Brian, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. 2013. "Detection of Rural Electrification in Africa using DMSP-OLS Night Lights Imagery." *International journal of remote sensing*, 34(22): 8118–8141.

Moretti, Enrico. 2010. "Local Labor Markets." *Handbook of Labor Economics*, 4b.

Mukherjee, Arpita, and Bhavook Bhardwaj. 2016. "Imposition of MAT on SEZs: Concerns and the Way Forward."

National Geophysical Data Center. 2013. "Version 4 DMSP-OLS Nighttime Lights Time Series." National Oceanic and Atmospheric Administration. (accessed April 30, 2016).

Ravi, Shree. 2018. "Understanding the Effect of Indian Special Economic Zones on the Informal Economy." *Working Paper.*

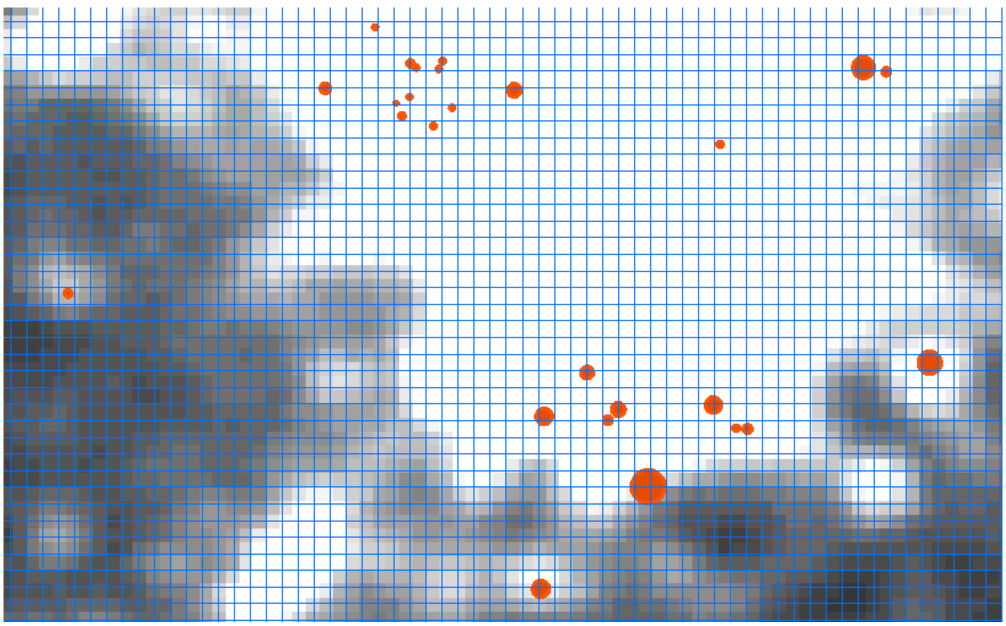
Shenoy, Ajay. 2016. "Regional Development through Place-based Policies: Evidence from a Spatial Discontinuity." *Working Paper.*

Sutton, Paul, Dar Roberts, Chris Elvidge, and Henk Meij. 1997. "A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States." *Photogrammetric Engineering and Remote Sensing*, 63(11): 1303–1313.

Topalova, Petia. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *IMF Working Paper.*

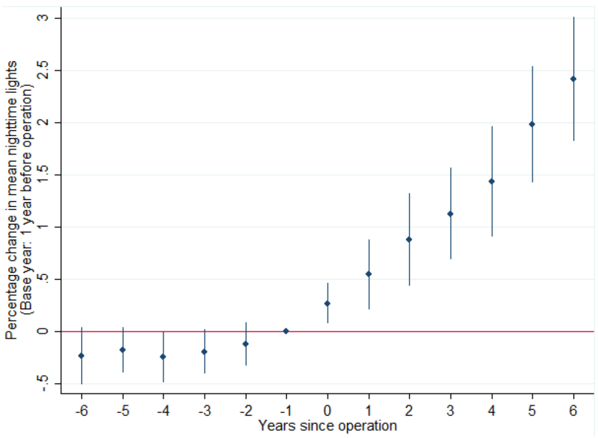
Wang, Jin. 2013. "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities." *Journal of Development Economics*, 101: 133–147.

Figure 1: Construction of the Cell-level Panel

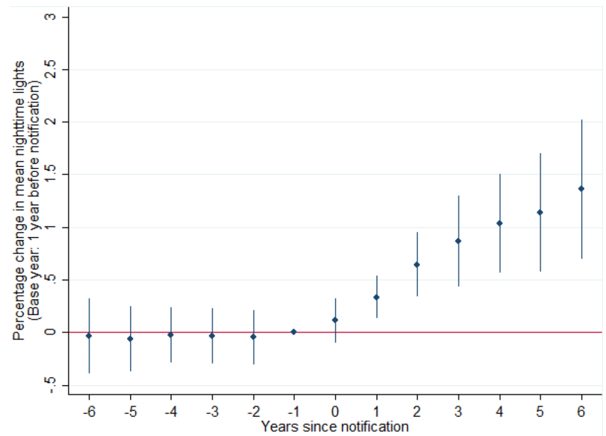


Note: Figure shows how the cell-level panel is constructed. A grid of 1 square-kilometer cells is overlaid on the map of India and cells that are strictly outside geo-coded SEZ locations, which are depicted by the circles, are considered for the analysis of spillover effects.

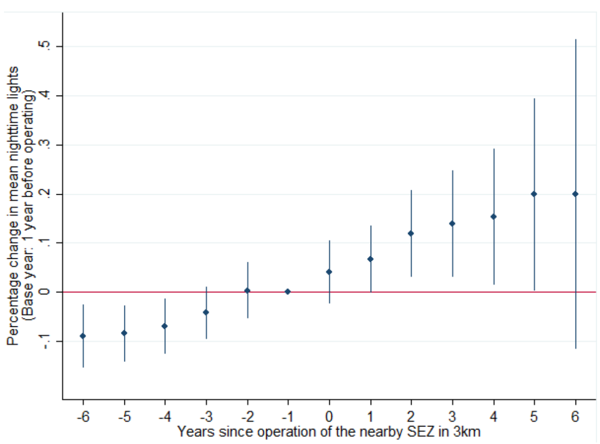
Figure 2: Event Study of SEZs and SEZ Neighborhoods



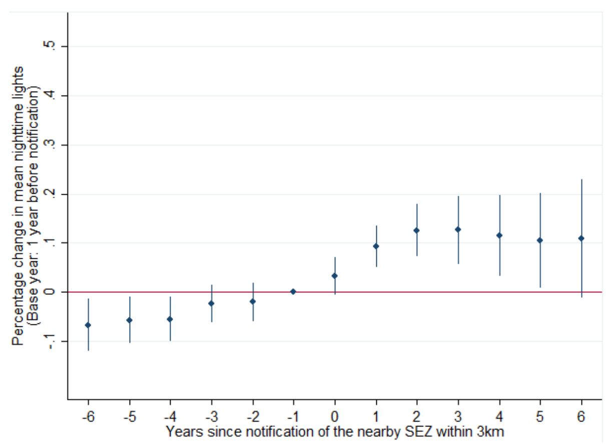
(a) Event: Operation, Inside SEZs



(b) Event: Notification, Inside SEZs



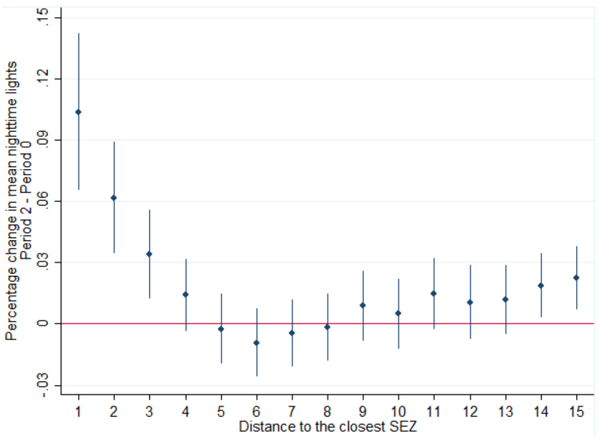
(c) Event: Operation, SEZ Neighborhood



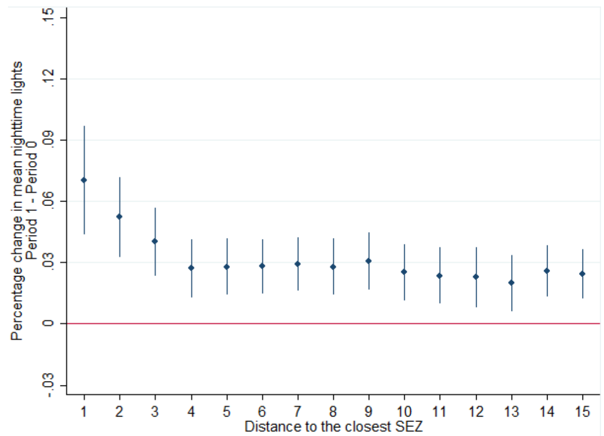
(d) Event: Notification, SEZ Neighborhood

Note: Figures plot 0.01*percentage changes in NTL backed out from $\gamma_k' s$ in Specification 1. Figures (a) and (b) are obtained from cells within SEZs and (c) and (d) are results from cells within 3 kilometers away from SEZs. The base year is the year before operation for Figures (a) and (c) and the year before notification for Figures (b) and (d). Cell and year fixed effects are included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

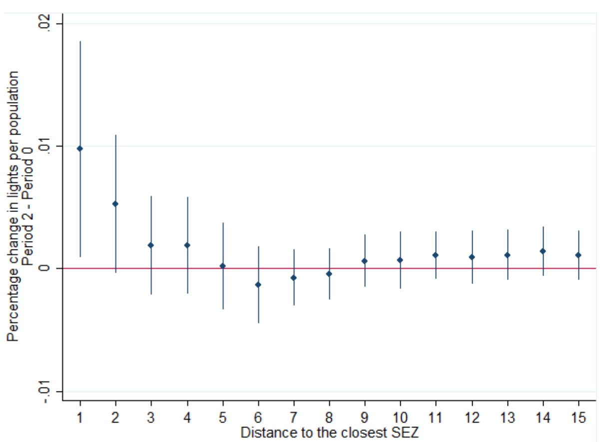
Figure 3: Effect of SEZs across Distances



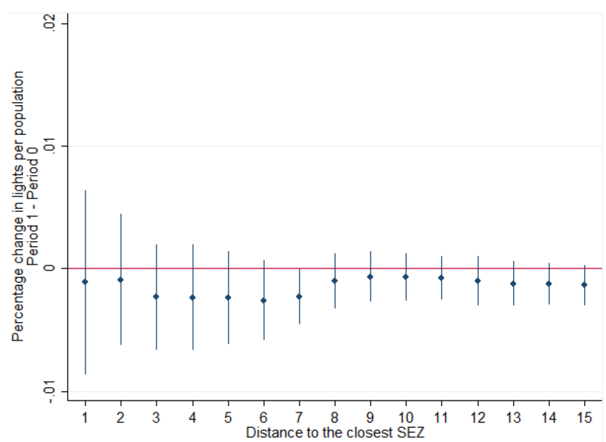
(a) Long Run Effect: NTL



(b) Short Run Effect: NTL



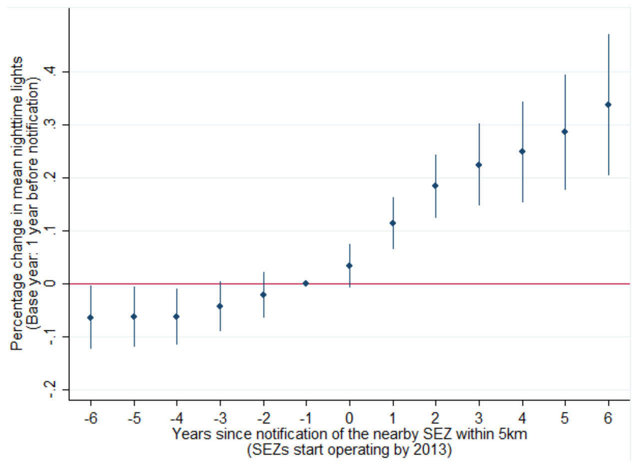
(c) Long Run Effect: NTL per population



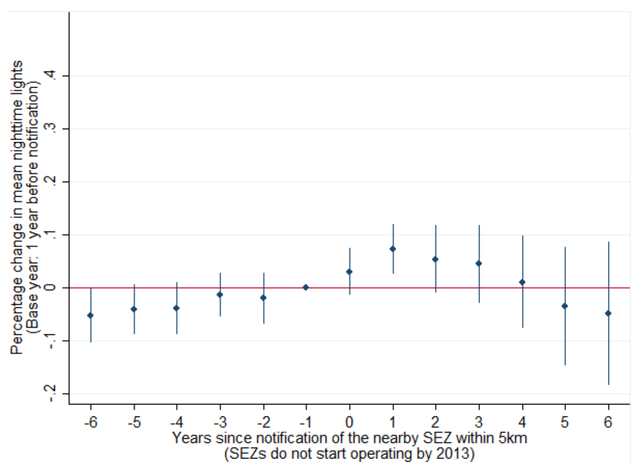
(d) Short Run Effect: NTL per population

Note: Figures (a) and (c) plot 0.01*percentage changes in NTL and NTL per population, respectively, backed out from δ_k's in Specification 1, and figures (b) and (d) present 0.01*percentage change in NTL and NTL per population, respectively, obtained from γ_k's in Specification 1. The base period is the pre-notification period of an SEZ in distance ring x. Cell and year fixed effects included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

Figure 4: Testing the Strength of Identification: Using NTL



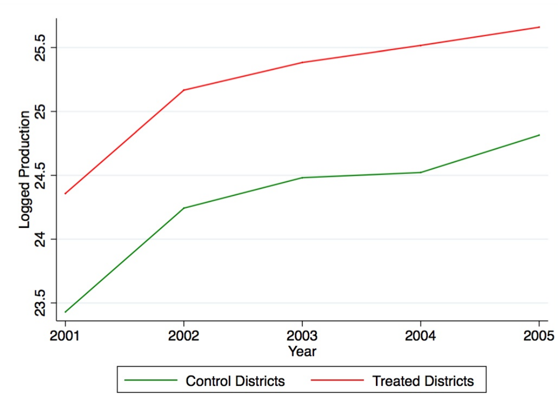
(a) Notified SEZs that became operational eventually



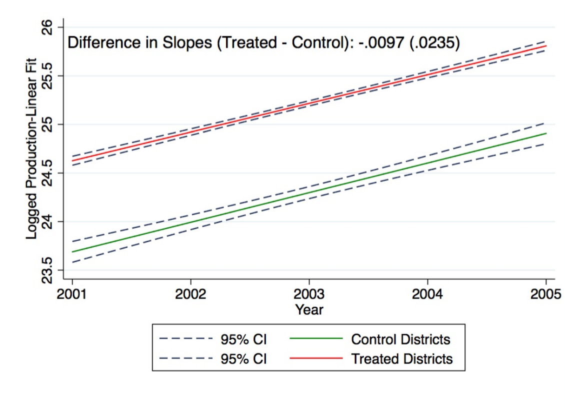
(b) Notified SEZS that did not reach operation

Note: Figures (a) and (b) compare 5 kilometer neighborhoods around SEZs that have begun operation before 2013, and those that have been notified but not yet operational by 2013. The figures plot 0.01*percentage changes in NTL with the the year before notification being the base year. Cell and year fixed effects are included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

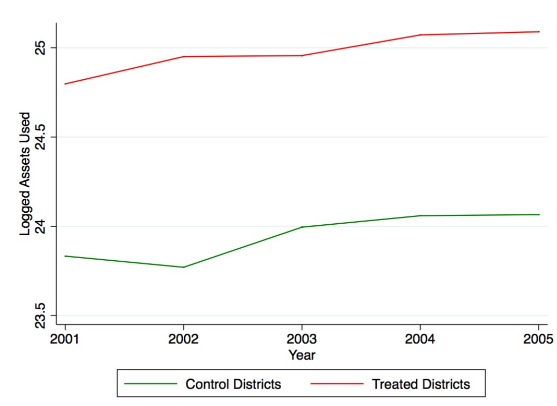
Figure 5: Testing for Pre-trends in District-level Manufacturing Outcomes



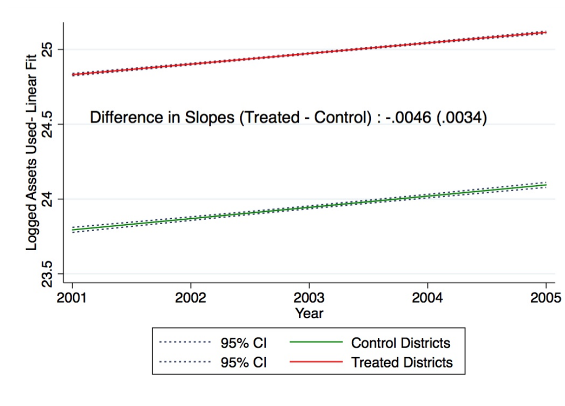
(a) Actual Trend: Production



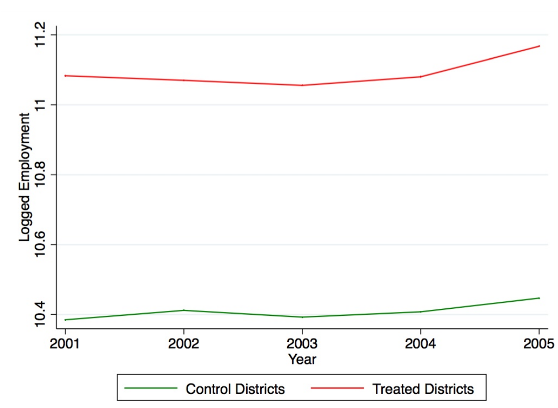
(b) Linear Fit: Production



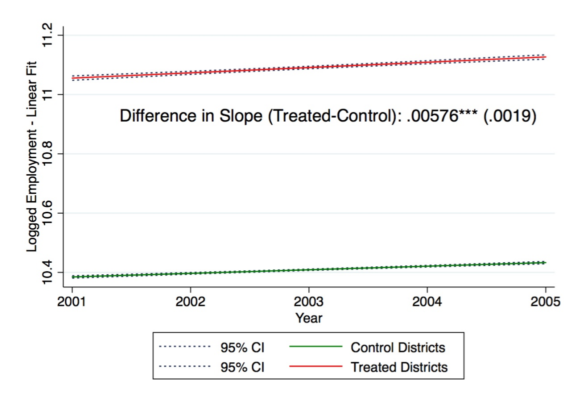
(c) Actual Trend:Total assets



(d) Linear Fit: Total assets



(e) Actual Trend: Employment



(f) Linear Fit: Employment

Note: Treated districts are those with at least one operational SEZ before 2011 while control districts are those with at least one notified SEZ, none of which are operational before 2011. Production, assets and employment are means across district totals, in logged values. Standard errors for differences in trend are reported in parentheses.

Table 1: Summary of SEZ & NTL Data

	Mean	Median	Std. Dev.
Panel A: All SEZs			
Year of notification	2008.1	2008.0	1.89
Area (sq. km)	1.47	0.27	4.83
Public	0.29	0.00	0.21
Manufacturing	0.24	0.00	0.43
IT/electronics/engineering	0.69	1.00	0.47
Neighborhood NTL (2000)	20.0	11.6	18.6
Neighborhood NTL (2005)	19.8	11.1	19.2
Neighborhood NTL per population (2005)	0.030	0.017	0.048
Neighborhood NTL per population (2010)	0.054	0.034	0.064
Number of observations		251	
Panel B: Operational SEZs			
Year of notification	2007.3	2007.0	1.30
Year of operation	2009.7	2010.0	2.23
Number of observations		133	

Note : An SEZ is considered public if any district or state agency was involved in the development process. Neighborhood of an SEZ is defined as the area within 3km away from the boundary of the SEZ.
Source : Ministry of Commerce & Industry, National Centers for Environment Information, NASA

Table 2: Falsification Test: Impact on Districts before SEZ Notification

	Eventual Number of SEZs:	
	Operating (γ)	Notified (δ)
Production	0.009	0.023
	(0.03)	(0.02)
Assets used	0.004	0.032
	(0.03)	(0.02)
Employment	0.015	0.006
	(0.01)	(0.01)
Labour Productivity	-0.002	0.016
	(0.02)	(0.01)
Wage	0.015	0.005
	(0.02)	(0.01)
Industry FE	Y	
Year FE	Y	

Note : This table contains results from Specification 1. All outcome variables are average logged firm-level variables in the formal sector of a district-industry group. Industry and year fixed effects included, and standard errors are clustered at district level and are reported in parentheses.

Source : ASI data on formal manufacturing firms (2000-2005)

Table 3: Comparison of Pre-Treatment Averages Across Districts

Variable	Control	Treated	Other Districts
Demographics			
Density(Sq. Km)	4,059	3,988	431.8
Primary and Below (%)	66.8	61.7	74.8
Higher Secondary and Below (%)	95.8	92.4	97.4
Worker Composition			
Formal Employment[1)]	0.21	0.27	0.16
HH Employment	0.47	0.46	0.57
Manufacturing	0.67	0.59	0.28
Trade	0.11	0.13	0.03
Services	0.15	0.22	0.05
Agriculture	0.07	0.06	0.64
Firm Composition			
Formal Firm[1]	0.07	0.03	0.01
Firms with No Hired Workers	0.38	0.40	0.64
With Power	0.40	0.34	0.24
Average Firm Size	6.20	3.60	2.41
Manufacturing	0.14	0.16	0.2
Services	0.36	0.32	0.24
Trade	0.47	0.49	0.39
Income			
Monthly Earnings (Rs.)	1,263.1	1,665.5	934.5
Districts	28	40	167
Year of Earliest Formal Approval	2007	2006.1	-
Year of Earliest Notification	2008.4	2006.7	-
Year of Earliest Operation	2012.3	2008.3	-

Note : All values are from 2005 (or 2001 in the case of Census variables).
[1)] Formal Employment takes the value of 1 if firm employs over 10 workers, 0 otherwise.
Source : ASI, NSS Employment Unemployment Surveys, NSS Unorganized Manufacturing, Economic Census, Census Digital Library, Ministry of Commerce & Industry

Table 4: Effect of Additional Operating SEZ on Formal Manufacturing Firms

	All Firms		Size 1: <=10		Size 2: (10,20]		Size 3: (20,100]		Size 4: >100	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	Mean	No. Operational	Mean	No. Operational	Mean	No. Operational	Mean	No. Operational	Mean	No. Operational
Production	28,630,983	0.022	2,649,814	0.015	9,626,209	0.019	33,598,769	0.028	303,229,348	0.006
		(0.011)		(0.007)		(0.003)		(0.004)		(0.003)
Assets Used	22,976,901	0.015	169,397	0.035	627,814	-0.002	2,446,087	0.015	27,508,346	0.004
		(0.008)		(0.015)		(0.007)		(0.007)		(0.007)
Employment	43.8	0.009	6.4	0.001	14.6	0.001	41.7	0.008	284.3	0.005
		(0.004)		(0.002)		(0.002)		(0.001)		(0.002)
Labour Productivity	579,546	0.018	545,796	0.014	873,270	0.015	1,045,494	0.021	1,289,803	0.002
		(0.011)		(0.005)		(0.004)		(0.003)		(0.002)
Wages	59,928	0.012	39,340	0.011	47,099	0.007	54,721	0.010	74,608	0.016
		(0.003)		(0.003)		(0.003)		(0.003)		(0.007)
New Firm[1)]	.024	0.005	.019	0.009	.025	0.009	0.030	0.011	0.020	0.007
		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)
District FE	Y		Y		Y		Y		Y	
Industry FE	Y		Y		Y		Y		Y	
Year FE	Y		Y		Y		Y		Y	
Observations	122,624		18,554		21,919		39,533		42,618	

Note : This table reports the results from running Specification 2 separately for each dependent variable and size category. All means are in Rupees (except for Employment and New) and are reported in numbers and proportions respectively. Dependant variables listed in Column 1 are at the firm-level in logged values. District, industry and year fixed effects are included, and standard errors are clustered at district level and are reported in parentheses. See Table A.4 in the Appendix for coefficient on number of notified and not yet operational SEZs.

[1)] New Firm takes the value of 1 if firm formed after first SEZ gets notified in the district, 0 otherwise.

Source : ASI data (2000-2009)

Table 5: Effect of Additional Operating SEZ on New vs. Old Formal Firms

	Full Sample	Old	New
Production	0.022	0.013	0.185
	(0.011)	(0.004)	(0.044)
Assets Used	0.015	0.013	0.015
	(0.008)	(0.006)	(0.008)
Employment	0.009	0.010	-0.005
	(0.004)	(0.003)	(0.028)
Labour Productivity	0.018	0.004	0.171
	(0.011)	(0.002)	(0.041)
Wage	0.012	0.004	0.051
	(0.003)	(0.001)	(0.028)
District FE	Y	Y	Y
Industry FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	122,624	105,454	17,170

Note : This table reports the results from running Specification 2 for the logged firm-level values of the dependent variable separately for two groups of firms- old and new. Firms are categorized as old if they have been operating for 4 years or longer. Industry, district, and year fixed effects are included. Standard errors are clustered at district level and are reported in parentheses.
Source : ASI Data (2000-2009)

Table 6: Impact on the Informal Sector

	Manufacturing			Services		
Dependent Variable	Mean	(1)	(2)	Mean	(3)	(4)
Production	73,967	-0.765	-0.826	71,264	0.243	0.177
		(0.26)	(0.38)		(0.151)	(0.25)
Gross Value Added	18,657	-0.578	-0.646	38,715	0.228	0.213
		(0.18)	(0.24)		(0.108)	(0.18)
Assets used	200,938	-0.401	-0.419	106,424	0.423	0.444
		(0.14)	(0.19)		(0.194)	(0.30)
Employment	3.9	-0.192	-0.157	1.8	0.182	0.192
		(0.08)	(0.11)		(0.060)	(0.10)
Gross Value Added per worker	4,836	-0.372	-0.452	21,216	0.044	0.022
		(0.12)	(0.17)		(0.084)	(0.14)
Labor Productivity	18,506	-0.545	-0.618	39,054	0.058	-0.017
		(0.20)	(0.30)		(0.122)	(0.20)
Wage	3,064	-0.574	-0.626	1,586	0.412	0.435
		(0.17)	(0.25)		(0.168)	(0.24)
District FE		Y			Y	
Industy FE		Y			Y	
District-Industry FE			Y			Y
Year FE		Y	Y		Y	Y
Observations		59,233			42,056	

Note : This table reports the estimated coefficients to the indicator for district treated with at least one operational SEZ before 2011 (α_2) from Specification 3. Dependant variables listed in Column 1 are at the firm-level and in logged values. All means in Rupees, except for employment reported in numbers. Columns (1) and (3) include district, industry, and year fixed effects. Columns (2) and (4) include district-industry and year fixed effects. Standard errors are clustered at district level and are reported in parentheses.
Source : NSS Unorganized Manufacturing & Services

Table 7: Impact on Total Informal and Formal Economic Activity

Panel A. Formal Sector[1)]

	Dependant Variable				
	Production	Employment	Investment	Wage	Labor Productivity
Manufacturing	0.385	0.166	0.316	0.130	0.214
	(0.12)	(0.08)	(0.14)	(0.04)	(0.08)

Panel B. Informal Sector[1)]

	Dependant Variable				
	Production	Employment	Investment	Wage	Labor Productivity
Manufacturing	-0.694	-0.278	-0.416	-0.574	-0.545
	(0.237)	(0.149)	(0.163)	(0.17)	(0.20)
Services	0.605	0.555	0.732	0.412	0.058
	(0.356)	(0.288)	(0.382)	(0.168)	(0.122)

Panel C. Overall

	Dependant Variable			
	Household Employment	Employed in Firms>10	Number of Firms[2)]	Number of Informal Firms[2), 3)]
Manufacturing	-0.082	0.079	0.078	-0.238
	(0.036)	(0.036)	(0.463)	(0.137)
Services	-0.039	0.056	0.117	-0.200
	(0.034)	(0.022)	(0.414)	(0.103)

Notes : This table reports the estimated coefficients to the indicator if district is treated with an operational SEZ before post-treatment year, from Specification 3. Regressions are carried out separately for manufacturing and service sectors, and dependent variables are district-industry totals in logged values. District and year fixed effects are included. Standard errors are clustered at district level and are reported in the parentheses.
[1)] Post-treatment period for formal sector analysis at district level includes until 2009; For informal sector and worker level analysis, the analysis covers until 2011. [2)] Analysis uses the Economic Census data with post treatment year 2012. [3)] Informality is defined according to the 10-worker rule.
Source : Source: ASI, NSS (Unorganized Manufacturing & Services, Employment survevy), Economic Census

Table 8: Wage Effects across the Income Distribution: Quantile Regression Results

	Quantile Regression:					OLS
	90%	75%	Median	25%	10%	
Treatment × After	0.351	0.195	0.139	0.052	-0.042	0.165
	(0.101)	(0.100)	(0.108)	(0.118)	(0.106)	(0.082)
District FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	63,782	63,782	63,782	63,782	63,782	63,782

Notes : This table reports the average and quantile treatment effects from regressions run on logged monthly worker wages in the 90th, 75th, 50th, 25th and 10th percentiles of the distribution. District, 2-digit industry and year fixed effects are included. Standard errors are clustered at the district level and reported in parentheses.
Source : NSS Employment and Unemployment Surveys

Table 9: Drivers of Wage Effect

	Worker Characteristics (Education)			Firm Characteristics (Size/Formality)[1]	
	(., Primary]	(Primary, Higher Secondary]	(Higher Secondary,.]	Informal	Formal
Panel A. Manufacturing					
Treatment × After	-0.203	0.153	0.512	0.111	0.407
	(0.264)	(0.194)	(0.21)	(0.28)	(0.13)
Mean Wage (Rs.)	906.9	1,772.2	5,486.2	1,510.2	2,697.3
District FE	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Observations	15,388	6,686	9,727	5,746	5,314
Panel B. Service					
Treatment × After	0.340	-.041	0.394	0.022	0.329
	(0.296)	(0.137)	(0.19)	(0.13)	(0.17)
Mean Wage (Rs.)	1,465.6	2,565.7	6,310.7	1,737.1	4,722.1
District FE	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Observations	5,167	6,962	8,531	8,909	8,189

Notes : This table presents the estimated coefficient to the Treatment × After term from Specification 3. Dependent variable is logged worker wages. Separate regressions are run for sectors formed from every combination of row and column headings. District, industry and year fixed effects are included. Standard errors are clustered at district level and are reported in the parentheses.

[1] Firm considered formal if it employs greater than 10 workers, otherwise informal

Source : NSS Employment Unemployment Surveys

Table 10: Effect of SEZs on Composition of Educated Workforce

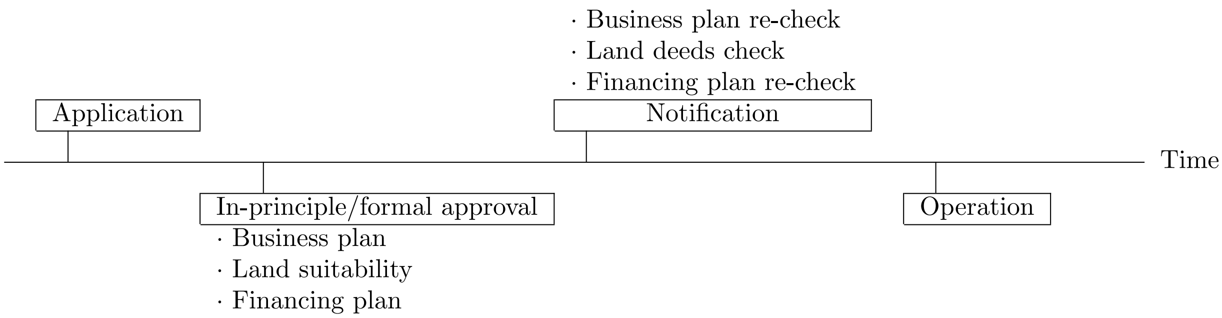
Logged Worker Wages	Proportion of workers with	
	Below Primary	Below Secondary
Treatment × After	-0.054	-0.034
	(0.03)	(0.04)
Mean Proportion	0.59	0.92
District-industry FE	Y	Y
Year FE	Y	Y
Observations	303,055	303,055

Notes : This table reports the estimated coefficients to Treatment × After from Specification 3. District-industry and year fixed effects are included. Standard errors are clustered at district level and are reported in the parentheses.
Source : NSS Employment Unemployment Survey

A Appendix

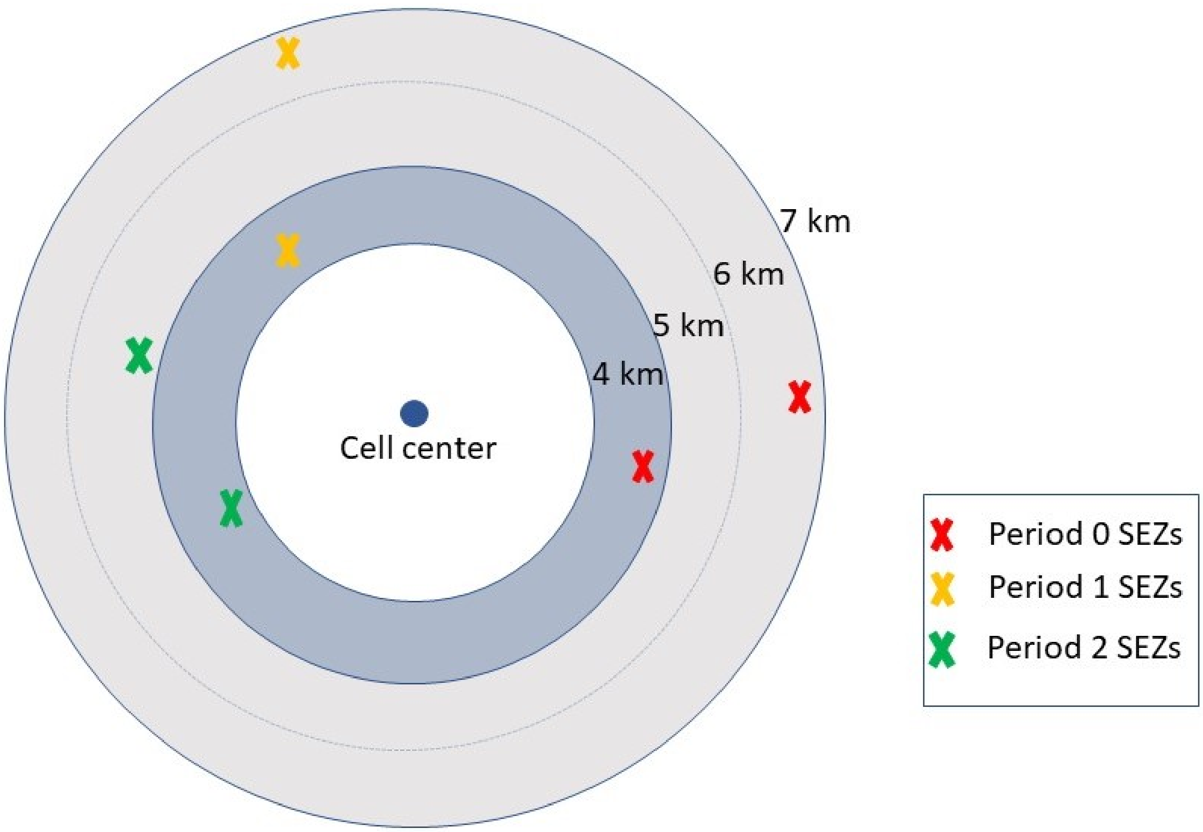
A.1 Supplementary Figures and Tables

Figure A.1: Timeline of SEZ Approval Procedure



Note : Figure shows the different stages that SEZs pass through before beginning operation. Refer to. Section 2 for detailed stage-wise explanation

Figure A.2: Pictorial representation of the Distance Dimension Specification



Example: Analyzing the Effects of an SEZ at a distance of 4 to 5 kilometers
For conducting the 4 kilometer distance ring analysis, we select all cells that have their closest SEZ between 4 and 5 kilometers, such as the one depicted above. We then control for all the other SEZs, in rings farther away, and in all stages of their life- before they become notified (period 0), after notification (period 1) and after operation (period 2). Specification 2 then isolates the effect of a notified but not yet operational SEZ in the 4 kilometer distance ring through the coefficient γ_4 (the baseline is the effect of a period 0 SEZ in the same ring). Similarly the effect of an operational SEZ in the 4th kilometer ring is captured by the coefficient δ_4, our main coefficient of interest.

Table A.1: Benefits to SEZ Developers and SEZ Units

	Developers	Units
Administrative	Single window clearance for Central and State level approval	
Tax	Exemption from Minimum Alternate Tax Exemption from Central and State Sales Tax Service and Dividend Distribution Tax Duty-free domestic procurement of goods, services	
	100% Tax exemption for 10 consecutive years since SEZ notification	Year 1-5: 100% tax exemption Year 6-10: 50% tax exemption Year 11-15: 50% of reinvested profits
Others		Infrastructural support Upper limit extended for managerial remuneration, external commercial borrowings allowed, etc. Flexible hiring and firing practices

Source: Department of Commerce, Government of India

Table A.2: Summary of Key Variables of Workers and Firms before 2005

	mean	**sd**	**25th pctl**	**50th pctl**	**75th pctl**	**min**	**max**
Demographics							
Density(/sq. Km)	3,989	2,606	2,279	3,597	4,845	250	19,865
Literacy	63.48	10.60	56.67	63.37	69.68	37.49	83.40
Primary and below	0.59	0.49	0			0	1
Secondary and below	0.92	0.28	1			0	1
Amenities[1)]							
Bank Density	1.50	0.59	1.10	1.37	1.73	0.11	3.22
Primary School Density	3.74	1.60	2.61	3.52	4.71	1.27	10.07
Secondary School Density	1.30	0.59	0.93	1.15	1.58	0.48	3.48
Wage[2)]							
Informal Manufacturing	8.5	1.7	7.4	8.1	9.9	3.8	12.5
Informal Services	8.1	1.4	7.3	8.2	9.0	2.0	11.0
Formal Manufacturing	11.0	0.8	10.5	11.0	11.5	4.9	14.2
Per capita Consumption[2)]	7.0	0.6	6.6	6.9	7.4	3.3	11.0
Firm Composition							
Greater than 10 Workers	0.04	0.20				0	1
Any Registration	0.53	0.50				0	1
Own Account Enterprise	0.41	0.49				0	1
Unincorporated	0.93	0.25				0	1
Operating with Power	0.34	0.47				0	1
No External Finance	0.90	0.30				0	1
Informal Finance—Financed	0.41	0.49				0	1
Manufacturing	0.15	0.36				0	1
Trade & Services	0.82	0.38				0	1
Services	0.33	0.47				0	1
Infrastructure	0.02	0.14				0	1
Firm Size							
Formal Manufacturing							
Employment[3)]	122.7	596.4	9.0	24.0	114.0	1.0	45,481
Average Labour Productivity[2)]	13.7	1.5	12.9	13.8	14.7	0.4	19.9
Informal Manufacturing							
Employment	5.4	4.8	2.1	3.7	7.0	1.0	35.0
Average Labour Productivity	10.3	2.4	8.8	9.7	12.2	3.6	15.6
Informal Services							
Employment	2.0	1.1	1.3	1.8	2.2	1.0	10.0
Average Labour Productivity	11.3	0.7	10.8	11.3	11.8	7.6	13.7

Notes : All averages are at the district-level, and when possible, at the village-level. Items under Amenities are values of 2001.
[1)] per 10,000 population, [2)] logged real values, [3)] in absolute numbers
Source : Census Digital Library of India, NSS and ASI surveys, EC data

Table A.3: Treatment & Control Group Formation

Data-set	Level of analysis	Pre-treatment Year(s)	Post-treatment Year(s)	Considered Set	Treated Set
Economic Census	Village/Town	2005	2013	1 or more notified SEZs by 2012	1 or more operational SEZs by 2012
NSS Unorganized Firms	District	2000-01, 2004-05	2010-11	1 or more notified SEZs by 2010	1 or more operational SEZs by 2010
NSS Worker Survey	District	2000-01, 2004-05	2010-11	1 or more notified SEZs by 2010	1 or more operational SEZs by 2010
ASI Formal Manufacturing	District	2001 to 2005	2006 - 2009	1 or more notified SEZs by 2012	1 or more SEZs notified or operating each year

Table A.4: Effect of Every Additional Notified SEZ on a Firm in Formal Manufacturing

	Firm size				
Dependent Variable	All Firms	Size 1: <=10	Size 2: (10,20]	Size 3: (20,100]	Size 4: >100
Production	0.003	0.017	0.002	0.006	0.004
	(0.005)	(0.013)	(0.005)	(0.004)	(0.002)
Assets Used	0.018	-0.010	0.017	0.020	0.001
	(0.008)	(0.011)	(0.007)	(0.005)	(0.003)
Employment	0.007	-0.002	-0.000	0.000	0.005
	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)
Labour Productivity	0.002	0.010	0.002	0.003	0.004
	(0.005)	(0.007)	(0.004)	(0.004)	(0.002)
Wages	0.005	0.008	0.004	0.005	0.020
	(0.001)	(0.002)	(0.002)	(0.002)	(0.007)
New[1)]	0.002	0.002	0.002	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
District-industry FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Observations	122,624	18,554	21,919	39,533	42,618

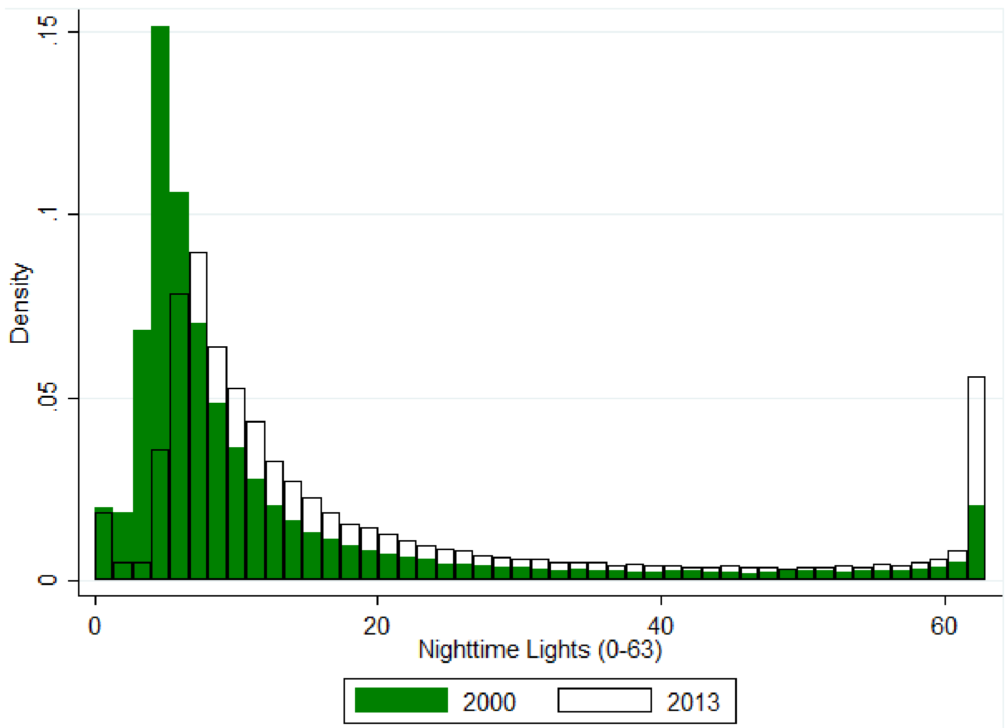
Notes : This table reports the estimated coefficients to the number of notified, non-operational, SEZs in the district of the firm in a year from Specification 2. Dependant variables listed in Column 1 are at the firm-level, and enters in logged values, and the regressions are run separately for each dependent variable and size category. District-industry and year fixed effects are included. Standard errors are clustered at district level and are reported in the parentheses.

[1)] New takes the value of 1 if firm formed after first SEZ gets notified in the district, 0 otherwise.

A.2 Caveat to NTL Usage

NTL cannot be a perfect substitute of traditional socioeconomic data. Each cell has a value between 0 and 63, which means that some lights are bottom- or top- coded. Studies that adopt nighttime lights usually suffer from the fact that non-negligible portion of their data is bottom-coded as most of them focus on underdeveloped countries. In our case, on the other hand, there is a high probability that we suffer from top-coded observations. This is because we focus on area that are more likely to be more developed within India, which is relatively developed among developing countries. Although the right-censoring might affect the empirical results, this would only underestimate the positive effects of SEZs on the neighborhood, if there are any. We, therefore, argue that the estimated spillover effects of SEZs are conservative.

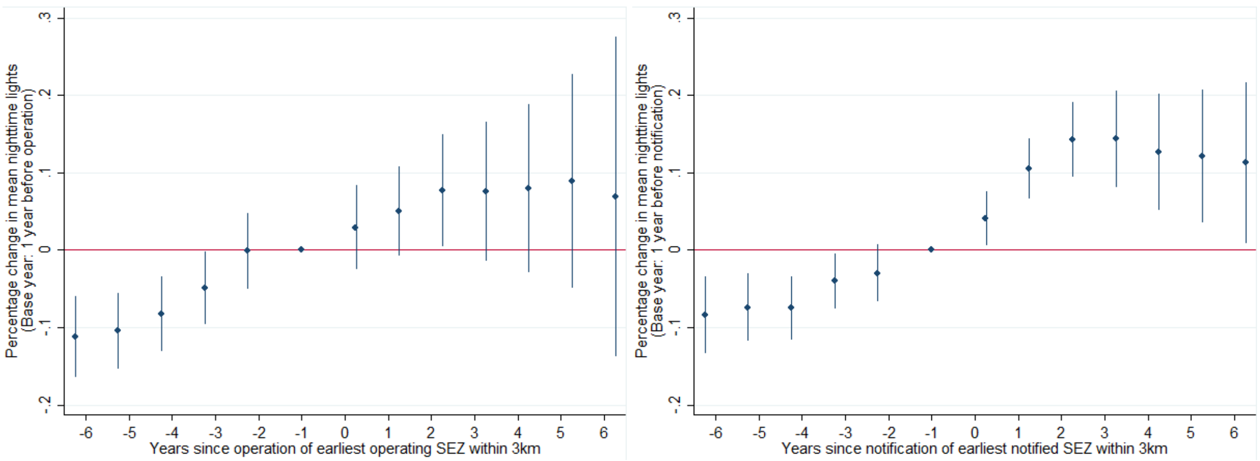
Figure A.3: Distribution of Nighttime Lights



A.3 Alternate specification for Events Study Analysis

We carry out an alternate specification where we consider all cells that have at least one SEZ within 3 kilometers, where the event is the *earliest* notification/operation of an SEZ in the 3 kilometer-neighborhood. We run Specification 1 using this alternate definition of the event, and find no significant differences from the results of the specification in the main paper:

Figure A.4



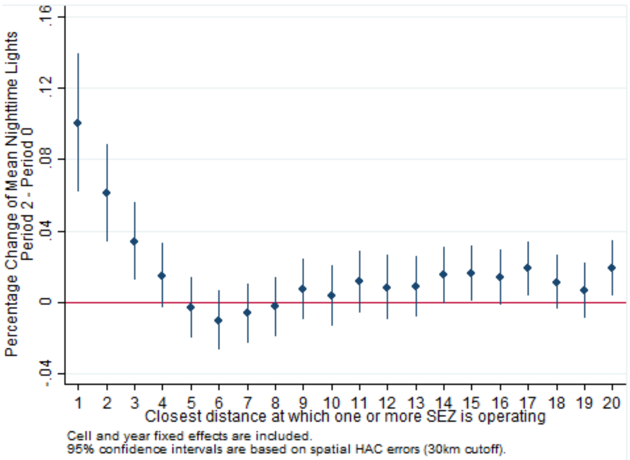
(a) Event: Operation, SEZ neighborhood

(b) Event: Notification, SEZ neighborhood

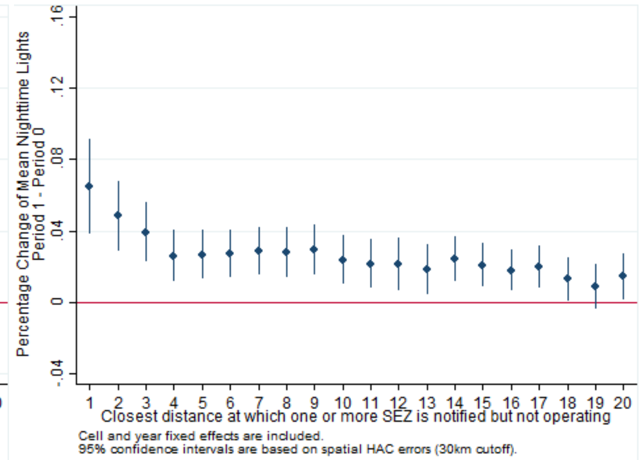
Note: Figures plot .01*percentage change in NTL backed out from $\gamma_k' s$ in Specification 1. The year before the event (operation/notification), year -1, is the base year. Cell and year fixed effects are included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

A.4 Extending Distance Analysis to 20 kilometers

Figure A.5



(a) Long-run effect: NTL



(b) Short-run effect: NTL

Note: Figures plot .01*percentage change in NTL backed out from $\delta'_x s$ in Specification 2 where the analysis extends beyond 15 kilometers and up to 20 kilometers away from SEZs. The number of observations (cells) is 95,239 per year. The base period is the pre-notification period of an SEZ in distance ring x. Cell and year fixed effects included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

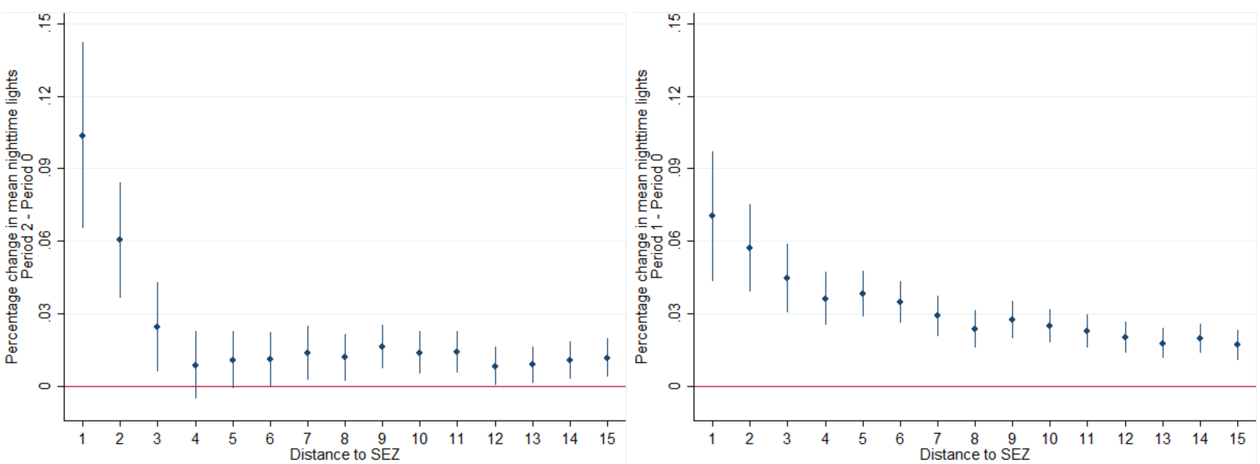
A.5 Alternate specification for Distance Dimension Analysis

We carry out an alternate, less restrictive specification by considering all cells with at least one SEZ in the distance ring x and controlling for both nearer and farther away SEZs when studying the effects of SEZs at that particular distance ring around the cell. For a cell i that is situated outside of SEZs in year t and has one or more SEZ in $x-1$ to x km ($1 \leq x \leq 15$),

$$log(light_{it}) = \alpha_i + \beta_t + \gamma_x * period1_{ixt} + \delta_x * period2_{ixt} + \sum_{1\leq d\leq 15}^{d\neq x} \sum_{\theta=0}^{2} \lambda_d^p period\theta_{idt} + \epsilon_{it},$$

The long.run effects due to operating SEZs are displayed on the left, while the short-run effect due to notified SEZs is in the bottom. Cell and year fixed effects are included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

Figure A.6



(a) Long-run effect: NTL

(b) Short-run effect: NTL

Note: Figures plot .01*percentage change in NTL backed out from $\delta_x^{\prime}s$ in the altered version of Specification 2 where *all* cells with at least one SEZ in distance ring x are considered in the distance ring x analysis (instead of only those cells with their *closest* SEZ in distance ring x). The base period is the pre-notification period of an SEZ in distance ring x. Cell and year fixed effects included. 95% confidence intervals are generated based on spatial HAC errors with 30 kilometer cutoff.

A.6 SEZs' Influence on Village-level Economic Activity

We follow Specification 3 at the village level with the Economic Census data. Table A.5 reveals a large and significant increase in the total working population and an economically, if not statistically, significant increase in the number of firms in a treated village as compared to the control village. We also observe increased hiring among firms, with average size expanding by 13.5% (which translates to roughly one additional worker to an average firm) and with the proportion of firms with no hired workers decreasing by 4.3%. This analysis gives us a preliminary view of an expansion in industrial activity which is consistent with an expansion in NTL recorded over areas of similar dimensions.[45]

Table A.5: Village Level Analysis of Firm and Worker Numbers

Dependent Variable	Mean	Treatment × After
Total Firms	330,675	0.252
		(0.412)
Total Workers	2,190,081	0.630
		(0.341)
Average Workers per Firm	6.2	0.135
		(0.055)
Firms with 0 Hired Worker[1)]	0.38	-0.043
		(0.020)
Observations		2,497,090

Notes : This table reports the estimated coefficients to Treatment × After from Specification 3. All outcome variables are in logged values. District-industry and year fixed effects are included. Standard errors are clustered at district level and are reported in the parentheses.
[1)] Firms with 0 hired workers are also known as Own Account Enterprises (OAEs).

[45] The 5-kilometer neighborhood around an SEZ spans an area of roughly 75 square kilometers, almost twice the area of a typical village in our study.

A.7 SEZs' Influence on Sub-district level Population Movements

We use the Gridded Population of the World (GPW) data in conjunction with sub-district level administrative boundaries.[46] The 251 SEZs in our sample are situated in 126 sub-districts with the number of SEZs per sub-district varying between 1 and 12 (with a mean of 2.02 and a median of 1). We restrict our attention to those with at least one SEZ notified before 2010 and evaluate whether the ones with at least one operating SEZ show faster population growth. For sub-district i in district d at time t:

$$\begin{aligned} log(population_{idt}) &= \alpha_0 + \alpha_1 Year_t + \alpha_2 District_d + \alpha_3 Operating_{idt} \\ &+ \alpha_4 Operating_{idt} \times After_t + \epsilon_{idt}, \end{aligned} \tag{1}$$

where $Operating_{idt} = 1$ if there is at least one operating SEZ in subdistrict i and $After_t = 1$ in year 2010. Standard errors are clustered at district level.

The estimation result reported in Table A.6 suggests that there is no differential trend of population density growth between the sub-districts whose SEZs started operating before 2010 and those whose SEZs are only notified by 2010. In other words, it is not likely that there are population movement across sub-districts. This is consistent with the fact that Indian labor market tends to be spatially restricted, meaning that the labor mobility is low.

Table A.6: Effect of Operating SEZs on Sub-district Population density

Dependent variable	Log of population density
Year 2005	0.875
	(0.012)
Year 2010	0.164
	(0.028)
Operating	0.332
	(0.374)
Operating × After	0.029
	(0.026)
District FE	Y
Number of observations	321
Overall R^2	0.629

Notes : District fixed effects are included, and year 2000 is omitted. Standard errors are clustered at district level and are reported in the parentheses.

[46] Acquired from the Survey of India (http://www.surveyofindia.gov.in/).