The Effect of Place-based Development Policies: Evidence from Indian SEZs*

Yeseul Hyun† Shree Ravi‡

November 25, 2017

JOB MARKET PAPER

LINK TO THE LATEST VERSION

Abstract

We investigate the influence of Indian Special Economic Zones (SEZs) on the local economy. Using nighttime lights data and firm- and worker-level survey data, we find that SEZs boost local economic activity. Moreover, analyzing the impact over a range of distances around the SEZs, we show that the local effects are not due to zero-sum relocations but represent net gains up to the level of a district. Comparing districts with operating SEZs and those with approved SEZs pending operation, our work also reveals an important structural transformation: While firms in the formal sector gain in size and productivity, SEZs instigate a shift in economic activity from the informal sector to the formal sector. This economic formalization is an integral part of the development process. However, we find that only workers at the higher end of the income and education distributions gain from this development, thus raising a concern about its inclusiveness.

*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 the Institute for Economic Development at Boston University. Taeyeon Kim and Wenran Fan provided excellent research assistance, supported by the BU MA-RA mentor program.

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

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

1 Introduction

Place-based development programs such as Special Economic Zones (SEZs) are an increasingly popular policy tool to induce economic development in a region. India introduced the 2005 Special Economic Zones (SEZ) Act as one of the biggest pushes to industrial development in its history. For over ten years from 2005, the country invested 0.5 per cent of its annual GDP into this program, amounting to 62 billion US dollars- substantial by both Indian and international standards.[1] 300 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.[2] However, the observed economic activity of SEZs alone is not evidence of the policy's success. The chief concern is that these observed benefits to regions receiving SEZs could be offset by losses elsewhere in the economy through resource relocations. This would produce little, if any, aggregate gains at considerable costs.

The literature on such place-based development policies measures their cost-effectiveness based on the programs' ability to form linkages with the local economy.[3] Stronger linkages between highly productive and export-oriented firms concentrated within zones, and the non-zonal economy, can increase local competition, produce denser input and labor markets and knowledge spillovers among firms and workers. These can drive long-run development and net positive effects even at an aggregate regional level. Given the magnitude of investment into the Indian SEZ policy, it is important to understand if it had a similar influence on the surrounding economy and to evaluate the extent of its benefits.

We provide one of the first empirical evaluations of the influence of Indian SEZs on the non-SEZ economy. In doing so, we add to the limited understanding of the effects of place-based development policies in the context of developing countries. 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 SEZ neighborhoods but confirms that there are net positive benefits to regions several times the size of the zones- up to the sub-national level of a district. Additionally, we make novel contributions to understanding how the SEZ policy interacts with *both* the informal and formal sectors of the economy, a dual economy framework which is unique to low income countries. Our analysis reveals that SEZs instigate a structural transformation of the economy, increasing firm size (in terms of production, employment and investment) and productivity in the formal sector while crowding out production and causing a shift of resources out of the informal sector. While this 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.

The scale of effects we record for the Indian SEZ policy is striking given its departure from the traditional mould of SEZs across the world. The Indian government incentivized the participation of the

[1] According to Shenoy (2016), 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. Kline and Moretti (2013) and Busso et al. (2013) also estimate the investments in the Tennessee Valley Authority program and the Empowerment Zones to be lower- around 20 billion and 3 billion dollars respectively.

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

[3] See Farole et al. (2011) and Aggarwal (2011) for a discussion.

private sector in zonal development and allowed SEZs to be of substantially smaller physical sizes than found elsewhere in the world. These features cast doubt on the policy's effectiveness and were debated in the Indian media among social, political and academic authorities.[4] Small zones were doubted to be capable of producing significant additional economic activity. 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 favouring large companies over small ones. These valid concerns notwithstanding, our analysis shows that the zones did bring benefits at an aggregate level.[5]

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.[6] 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 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 comparisons 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 thirteen-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 1200 square kilometers, one-fourth the size of a median district in our data-set.[7]

We then use a simple difference-in-differences framework to explore the aggregate effects of these spillovers on the real economy with a richer set of variables on firms and workers drawn from nationally representative firm and worker surveys. We compare outcomes between regions that have at least one operating SEZ (treated regions) and those with at least one SEZ that passes the penultimate round of approval (control regions). 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 chosen as treatment

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

[5] Plausibly a result of the co-locating pattern of SEZs within narrow regions (which strengthened agglomerative forces) and the private sector's ability to target high potential regions and thus greater efficiency in resource allocation.

[6] Due to private sector participation in program location, we face the opposite concern of studies generally involving zones in developed economies which usually target under-industrialized regions.

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

and control. In choosing control regions in this fashion, we are also similar to Busso et al. (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.

We find that average labor productivity of formal sector manufacturing firms in treated disticts 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 resultant agglomeration spillovers from SEZs structurally transform the economy away from informal lines of production towards greater formality. This is especially true in the 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%. We partially attribute this trend to 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.[8] 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 of increasing inequality. The effect on overall worker wages is non-uniform. 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 education) distribution seem not to gain significantly.

Due to the unique nature of Indian SEZs and our focus on how they affect the little-explored dual economic structure of the Indian economy, we provide new insights to the literature on place-based development policies. Our finding of increased formalization is similar to Magruder (2013)who finds that the change in minimum wage rule in Indonesia acted as a big-push mechanism leading to greater formalization.Works studying SEZs and similar programs in India and China, such as Wang (2013), Alder et al. (2016), Chaurey (2016) and Shenoy (2016), do not touch upon the effect of such programs on the formal-informal sector dynamics.

Informality refers to several firm- and worker-level characteristics, the most pertinent ones in the Indian context being 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%

[8]SEZ presence 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. We test the quality channel in an upcoming paper.

of the total labor force.[9] Greater formality is considered more desirable 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.[10] In this respect, the shrinking informal sector due to an increase in SEZ activity is a positive effect. An increase in inequality within treated districts, however, suggests that workers are left out of the wage benefits due to their inability of being absorbed by the formal sector. This insight corroborates the current concern in both developed and developing countries about a "skill gap" where the workforce is unable to fulfill the demand for skilled labor thus holding back further prospects of development. It also provides a cautionary note for countries that start zonal development programs and have a large pool of informal or unskilled workers as it highlights the need for supplementary policies to improve the quality of human capital for inclusive economic growth.

Our analysis of the formal sector is more comparable to the work of others in studying the impact of place-based programs in developed countries that do not have prominent informal sectors. Here we contribute on a more positive note to the mixed evidence on the role of firm agglomeration in boosting productivity and development of a region. Greenstone et al. (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. 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.

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 et al. (2013), Wang (2013) and Chaurey (2016). Given the traditionally low level of labor mobility in India,[11] 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.[12]

While we provide a comprehensive overview of the benefits from the Indian SEZ policy, explicit cost-

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

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

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

[12]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.

benefit calculations are beyond the scope of this current work. 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 concludes 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 his observations of 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). These zones were instead 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 intentions to be: (a) generation of additional economic activity, (b) promotion of exports of goods and services, (c) 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: the minimum size requirement was much lower resulting in physically smaller SEZs compared to municipalities declared as 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 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.[13] These features resulted in two main trends in zone location:[14] 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 B.1 in Appendix B 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

[13] Authors' estimates.

[14] See Aggarwal (2011) for a detailed survey of SEZ developers on issues including zone location and development.

SEZ activity.[15] Additionally, both developers and SEZ firms were exempted from paying the Minimum Alternate Tax (MAT), currently 18.5% of book profits in India.[16] 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)- 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. Labour laws were also made more flexible for SEZ companies which were permitted to lay off redundant workers without seeking the permission of the government which is not the case for firms of similar size outside the zones.[17] All of these incentives helped to create a relatively hassle-free environment for firms which wished to operate in a country not known for its ease of doing business.[18]

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[19] and prior approvals of the state government.[20] 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 B.1 in Appendix B illustrates the approval process of the SEZs.

In our analysis, we only consider those SEZs that pass the penultimate stage of notification. The Ministry of Commerce and Industry provides us with 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

[15]The corporate tax rate is 35% in India.

[16]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.

[17]This is seen as a pro-firm policy by Besley and Burgess (2004) who show that hiring and firing rigidities negatively affect output, investment and employment in the formal sector. Other works that point to the rigid labor regulation in India as a hindrance to growth include Hsieh and Klenow (2009), Hasan and Jandoc (2012) and Kochhar et al. (2006).

[18]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.

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

[20]Information derived from the published BoA meeting minutes. While these factors were repeatedly highlighted during the decision making process, the relative importance of each, and whether this list is exhaustive, is unclear.

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. Our sample includes all 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. While this is our general strategy, the following sections will elaborate on the modifications made to analyze the effects of the policy at various levels of spatial aggregation.

3 Patterns of SEZ Influence Across Space and Time

In this section, we track the effects of SEZs closely through time and space using highly disaggregated nighttime lights (henceforth referred to as NTL) data as a proxy for economic activity.

3.1 Key Variables

Our main data-set for analyzing the pattern of spillovers from SEZs is the NTL time series (from 2000 to 2013). This is obtained from the National Centers for Environmental Information. The resolution is 30 arc-seconds which covers approximately 854 meters around the center of India, and the data values range from 0 (background noise) to 63. 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. Compared to most countries' releases of socioeconomic data, which are infrequently available at higher levels of spatial aggregation, the NTL will be more helpful in confirming the effects of an SEZ that appear precisely in the year it becomes notified or operational, over narrow as well as wider neighbourhoods.

NTL has been previously used to proxy economic growth or development , degree of urbanization and population density.[21] However, its usage for measuring *intra*-country economic development has been criticized. Mellander et al. (2015) find that, in the case of Sweden, NTL is correlated more with population and establishment density rather than with wages. Even if this were the case in the Indian context, it would provide us with meaningful interpretation since we are also interested in effects related to the relocation of human resources. Moreover, Bhandari and Roychowdhury (2011) find that NTL significantly explain variations in GDP down to the district level in the case of India. We refer the reader to Appendix A for technical caveats on the usage of NTL and a description of how we construct our panel of 1 kilometer-wide cells.

We additionally use the Gridded Population of the World (GPW) data series in order to measure the extent to which the effects recorded by our analysis of the NTL data is driven by population movements.

[21] See Henderson et al. (2012), Ma et al. (2012) and Sutton et al. (1997) respectively. Hodler and Raschky (2014) even use NTL in conjunction with data on the birthplace of political leaders to study regional favoritism.

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.[22] 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, we find it 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 44 soccer fields. The mean is much bigger 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, this size distribution is not surprising. 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 neighborhoods of radius 3 kilometers around SEZ boundaries. These neighborhoods have 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 top-coded nature of the NTL data, we may be capturing the lower bound of the actual growth that took place. As discussed in Appendix A, the data is top-coded at 63 (and bottom-coded at 0) and it is likely that we suffer more from right censoring because SEZs tend to locate in urban areas with already high values of NTL. Figure A.1 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.

3.2 Strategy

Our methodology relies on using the thirteen-year panel data-set of 62386 cells, the construction of which is detailed in Appendix A.2. We compare cells based on their proximity to SEZs and measure the differential effects on them through time due to an SEZ's presence. Additionally, we use cell and year

[22] 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.

fixed effects on de-trended data to add robustness to our findings.

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 a cell is potentially affected by.[23] For such a cell i that is situated outside of SEZs in year t,

$$log(light_{it}) = \alpha_i + \beta_t + \sum_{-6 \leqslant k \leqslant 6}^{k \neq -1} \gamma_k * D_{ikt} + \varepsilon_{it}, \tag{1}$$

where the outcome variable is logged luminosity of cell i at time t, which is basically a logged transformation of 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 correct the standard errors for spatial autocorrelation following the specification of Conley (1999) up to a cutoff of 30 kilometers.

The estimates of interest are γ_k's, for $k \in [-6, 6]$. 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 of 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 which is dependant 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:

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

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

[23] 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 the vicinity, with overlapping notification and operation timelines (especially given the tendency of SEZs to cluster).

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.[24] The average difference in the effects of period 2 and period 0 SEZs (which is 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 shows that we control for all farther away SEZs affecting the cell but not within the d-th ring in whatever period of life they may be.[25] 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. This 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.[26] 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.[27]

3.3 Findings

Activity within SEZs: Direct Effects

An event-study of areas *inside* SEZs confirms the relevance of specification 1. Figure 1a 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, this activity takes on average 2 to 3 years before the SEZ can begin operation. Figure 1b, which studies the event of notification, confirms that the pre-operation growth in Figure 1a may be driven by post-notification activities.

[24] 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.

[25] Appendix A contains a pictorial representation of specification 2 for greater clarity.

[26] Since the population data is available at 5 year intervals, the period of life of an SEZ is updated every 5 years.

[27] The magnitudes should 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.

Spillovers of SEZs Over Time: Event Study Approach

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 1c displays a similar trend as the within-SEZ analysis. When an SEZ starts operating, the immediate neighborhood also experiences a significant increase in NTL which is, as to be expected, lower than the increase within SEZs. It is persistent in magnitude but loses significance over time.[28] 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. A good example 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 Figure1d shows.[29] However, we can still argue that regions around notified SEZs experience a noticeable increase in the *slope* of NTL or growth rates, upon notification. Appendix A contains 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. This does not, however, produce significantly different results from those above which adds to the robustness of our finding.[30]

Spatial Extent of SEZ Spillovers

Figures 2a and 2b follow specification 2 to illustrate the spatial extent of SEZ spillovers. The resulting trends are in line with our expectations. In the long- and short-run scenarios, we observe a sharp increase in NTL in 1 kilometer-neighborhoods of SEZs, with the positive impact continuing to be significant but diminishing with distance. Also in line with our expectations, the magnitude of effects due to operating SEZs is greater than those due to notified SEZs at least within the 2 kilometer-neighborhood. We carry out an alternate, less restrictive specification by considering all cells, and controlling for both nearer and farther away SEZs when studying the effects of SEZs at a particular distance ring x around a cell. Details of this and the corresponding results can be found in Appendix A.

In Figures 2c and 2d, we show that population movements do seem to drive some, but not all, of the increase in NTL.[31] This suggests that there are other channels at play, especially in the case effects produced by operating SEZs.

The chief takeaway from the analysis in this section is that we do not find evidence of a zero-sum relocation

[28]This could be due to the small number of SEZs older than 3 or 4 years old by 2013, considering the mean year of initial operation is 2009 according to Table 1.

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

[30]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.

[31]*NTL per population* is the outcome variable in this case and we use specification 2.

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. 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 holds true when we check patterns over a wider area of 20 kilometers' radius that spans 1200 square kilometers- quarter the size of a median district in our data-set. Hence the impact on the economy has been positive on areas several times the actual size of the zones. This finding motivates a deeper look into the changes occuring in the aggregate economy.

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, and the effect is net positive up to a level of geographical aggregation which is of interest to political and administrative authorities- that of the district. A district, which is the main level of local governance below the state, is divided further into sub-districts that consist of villiages and towns. In our sample, there are 68 districts, each of which, is on average, divided into 19 sub-districts, which is further divided into an average of 9 villages per sub-district. 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. We employ multiple data sets at different levels of administrative units- the village, sub-district and district.[32]

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). 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 service, 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.

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,[33] 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.[34] It comprises of a complete enumeration of firms above a 100 in worker strength with an annual survey of a

[32]The results of village-level analysis of firm acivity and sub-district level analysis of population movements are provided in Appendix B.

[33]Nationally representative surveys on formal service firms have not been conducted so far in India.

[34]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.

repeated random cross-section of smaller firms. This data contains more information than the EC that 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 30000 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 NSS Unorganized Manufacturing and Services quinquennial survey data that covers firms in the unregistered sector of the Indian economy. Each survey round in our study (2000,2005 and 2010) contains information on around 35000 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 set 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 104000 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.

Appendix B.3 gives 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 them employing below 10 workers. This indicates the highly skewed firm size distribution and the vast size of the informal sector.[35] 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 Strategy

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

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

our difference-in-differences estimate of effects upwards.

Our solution is to adapt a similar strategy to the one commonly used in the literature on place-based development policies- we use the the approval process of SEZs as a source of quasi-experimental variation. We compare regions in which there is at least one SEZ that has passed notification (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.[36] In the case of Busso et al. (2013) and Kline and Moretti (2013), control regions were formed out of rejected areas whereas, in our case, the control areas were never disqualified by administrative authorities and were expected to have operating SEZs in the near future.[37]

The immediately obvious point in favour of this 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 potential for growth. Table 2 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 latter districts are also clearly higher and the proportion employed in informal household businesses around 10 percentage points lower. From this table, we get the impression 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.

One may be concerned that treated regions seem to be disproportionately among those with earlier notified SEZs, as shown in Table 2 . This could imply that the order of notification is correlated with unobservables relevant for the outcomes studied. This concern about timing is mitigated by two supporting factors. Firstly, we find that 60 percent of the control districts were targeted earlier by SEZ developers that did *not* manage to reach the 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 2 . We see 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. Secondly, the locations were not chosen solely based on profitability

[36]Refer to Appendix B.2 for an overview.

[37]Busso et al. (2013) further strengthened their strategy using the set of selection criteria for Empowerment Zones and matching treated and control zones through propensity score weighting. In this, they had the advantage of a longer time seies on all the socioeconomic indicators that were known to influence zone selection. In our case, the exact set of conditions used by the BoA to deem an SEZ notification-worthy, is unknown. What we know from the meeting minutes is that these included other criteria than economic indicators for the region- such as the ability of the developer to posesss the land and propose a viable development plan.

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 choose locations within the state or district in which they are headquartered.[38] 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 potential. There is a possibility, however, that treated regions differ in the number of 'capable' developers than control regions given that they have earlier operating SEZs; we address this problem with the addition of region fixed effects to our framework which would take into account non-time varying differences in 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 3a and 3b both kinds of regions did not experience significantly different pre-trends 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 3b 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.

Additionally we take advantage of the ASI data to analyze pre-trends in district-industry totals of production, assets used and employment in formal manufacturing industries prior to the SEZ Act. Figure 4 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:

$$\begin{aligned} log(y_{fidt}) = \alpha_0 + \alpha_i + \alpha_d + \beta_t + \gamma_d * No.EventuallyOperatingSEZs_d \\ + \delta_d * No.EventuallyNotifiedSEZs_d + \varepsilon_{fidt} \end{aligned} \quad (3)$$

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_d and δ_d ,should not show any significance and we find this to be true. The results can be found in Appendix B.4 .

[38] From interviews with SEZ developers in Tamil Nadu.

Given the supporting evidence for our identification strategy drawn from disparate sources of information, we proceed to adopt it and use the following difference-in-differences framework:

$$y_{firt} = \alpha_0 + \alpha_1 T_r + \alpha_2 T_r 1_{AFTER} + \alpha_{ir} + \beta_t + \varepsilon_{firt} \tag{4}$$

yf_{irt} is the 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 is 1 for regions that were treated with at least one operational SEZ before the post-treatment period as stated in Appendix B.2. The value is 0 for regions that have at least one SEZ notified before this time but none operational yet. 1_{AFTER} is the time indicator which takes the value 1 for the post-treatment period, and 0 otherwise. Region, industry (and in an alternate specifications, region-industry fixed effects) 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.

The annual data on the formal manufacturing sector also allows us to use 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. 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:

$$y_{fidt} = \alpha_0 + \alpha_1 * No.OperatingSEZs_{dt} + \alpha_2 * No.NotifiedSEZs_{dt} + \alpha_{id} + \beta_t + \varepsilon_{fidt} \tag{5}$$

y_{fidt} is the outcome variable of a formal manufacturing firm f in district d, industry i 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-industry and year fixed effects are included with the standard errors being clustered at the district level. In the above specifications, we are assuming 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. 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 which could reveal the mechanism behind the increase in light activity following SEZ notification in the previous section.

4.3 Findings

Effects on Formal Firms in the non-SEZ economy

To study the changes in an average firm in formal manufacturing, we use a 10-year district-industry panel and follow specification 5. Column 2 of Table 3 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 .7 % and 1.2% depending on the value of output elasticity of capital.[39]

Note that the presence of an additional *notified* SEZ (the values of which are presented in Table B.5) also seems to positively impact firm investment and the wage level. This is consistent with us observing 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 3. 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 3, we see that this is not the case. The 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. 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 SEZs 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

[39] α in $Y=A(K)^{\alpha}(L)^{1-\alpha}$ where A, K and L are total factor productivity, capital and labour respectively.

operating SEZ firms.

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).

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 4 exhibits the values of coefficient α_2 from specification 4 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 4 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. Since services are not often industrializable and tradeable, increases in local demand has to be satiated by local service firms, both informal and formal. Hence it is reasonable to expect 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 framework by considering two common pre-treatment years and one post-treatment year for firms from both sectors. 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 4 are shown in Table 5. 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.

The bottom panel of Table 5 further highlights the movements of resources between the formal and informal sectors of treated district economies. We observe an 8.2% increase in the proportion of people employed in the formal sector and a decrease of 27% in the total number of informal sector firms with no impact on the overall number of firms. These trends indicate that SEZs encourage the expansion of the formal sector and tend to crowd out activity in the informal sector.

Given a productivity distribution within the informal sector, its decline as observed in Tables 4 and 5 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 generate by SEZs, which would explain the decrease in average productivity and total output in the informal sector. Informal sector crowding out could also happen due to less productive firms shutting down because they are unable to cope with a more competitive formal sector or because the workers no longer have to resort to subsistence activites ("forced informality") with more job opportunities in services or formal manufacturing. From the bottom panel of Table 5, we provide evidence of an 8% decrease in the proportion of small household manufacturing businesses that are generally less productive, providing some support for this. In an upcoming paper, we weigh the relative importance of firm deaths and firm switching in explaining informal sector decline.

Effects on the Overall Wage Distribution

Figure 5 reveals the results of quantile regressions on worker wages in the 10th, 25th, 50th, 75th and 90th percentile of income distribution. We clearly see that the wage increase is not uniform across the distribution of workers: workers in the upper end of the distribution gain the most, with the 90th percentile wage earners experiencing the maximum wage increase of around 42%. There is no significant increase in wages among workers in the lower end of the wage distribution, with the 10th percentile workers in treated districts even appearing to experience a decrease in wages, albeit statistically insignificant. Table 6 which uses the 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 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 consistent with the generally accepted positive correlation between education and formality of occupation, and with the fact that SEZs increase wages in the formal sector and not in the informal sector.

We also observe a slight but significant decrease in proportion of people in treated districts that are educated only up to the primary level, the results of which are in Appendix Table B.6. Since there is some evidence of net in-migration to treated districts, driven by our analysis of NTL per population, this change may occur due to the in-migration of more educated workers to take advantage of the well-paying labour market in treated districts. 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 Conclusion

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 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. Thus SEZs seem to have driven a structural transformation 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 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 et al. (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. 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. 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 which case weighted survey data on income cannot be used in conjunction with the tax rate to calculate a tax bill. 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 interesting to compare the effects of the different zonal development styles and draw conclusions on the optimal design for the Indian context.

Aradhna Aggarwal. Special economic zones: Revisiting the policy debate. *Economic and Political Weekly*, 41(43/44), 2006.

Aradhna Aggarwal. Economic impacts of sezs: Theoretical approaches and analysis of newly notified sezs in india. 2010.

Aradhna Aggarwal. Promoting agglomeration economies and industrial clustering through sezs: Evidence from india. *Journal of International Commerce, Economics and Policy*, 2(2), 2011.

Aradhna Aggarwal et al. *Impact of special economic zones on employment, poverty, and human development*. Indian Council for Research on International Economic Relations New Delhi, 2007.

Simon Alder, Lin Shao, and Fabrizio Zilibotti. Economic reforms and industrial policy in a panel of chinese cities. *Journal of Economic Growth*, 21:305–349, 2016.

Amrit Amirapu and Michael Gechter. Indian labor regulations and the cost of corruption: Evidence from the firm size distribution. *Working Paper*, 2014.

Timothy Besley and Robin Burgess. Can labor regulation hinder economic performance? evidence from india. *Quarerly Journal of Economics*, 119(1):91–134, 2004.

Laveesh Bhandari and Koel Roychowdhury. 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, 2011.

Matias Busso, Jesse Gregory, and Patrick Kline. Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947, 2013.

Ritam Chaurey. Location-based tax incentives: Evidence from india. *Journal of Public Economics, forthcoming*, 2016.

Xi Chen and William D Nordhaus. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594, 2011.

Timothy G Conley. Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1): 1–45, 1999.

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

Thiemo Fetzer. Can workfare programs moderate violence? evidence from india. 2014.

Edward L Glaeser and Joshua D Gottlieb. The economics of place-making policies. Technical report, National Bureau of Economic Research, 2008.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying agglomeration spillovers: Evidence from winners and losers of large plant opernings. *Journal of Political Economy*, 118(3):536–598, 2010.

Rana Hasan and Karl Robert H. Jandoc. Labor regulation and the firm size distribution in indian manufacturing. *Working Paper*, 2012.

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

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

Chang-Tai Hsieh and Peter J. Klenow. Misallocatiion and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4), 2009.

Chang-Tai Hsieh and Benjamin A. Olken. The missing "missing middle". *Journal of Economic Perspectives*, 28(3):89–108, 2014.

Patrick Kline and Enrico Moretti. 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, 2013.

Kalpana Kochhar, Utsav Kumar, Raghuram Rajan, Arvind Subramanian, and Ioannis Tokatlidis. India's pattern of development: What happened, what follows? *Journal of Monetary Economics*, 53(5), 2006.

Zhe Liang, Simon Appleton, and Lina Song. Informal employment in china: Trends, patterns and determinants of entry. *IZA Working Paper*, (No. 10139), 2016.

Ting Ma, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan. 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, 2012.

Jeremy R. Magruder. Can minimum wages cause a big push? evidence from indonesia. *Journal of Development Economics*, 100(1):48–162, 2013.

Alfred Marshall. Principles of economics, i, london 1898. *subsequent editions Marshall altered his verdict and he remarks (in a footnote to the passage, in which he describes the Physiocrats' achievement as the first attempt at a London 1916, Appendix B 2, p. 756), that Cantillon has some claim to being considered systematic*, page 53, 1890.

Charlotta Mellander, José Lobo, Kevin Stolarick, and Zara Matheson. Night-time light data: A good proxy measure for economic activity? *PloS one*, 10(10):e0139779, 2015.

Stelios Michalopoulos and Elias Papaioannou. Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152, 2013.

Brian Min, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. Detection of rural electrification in africa using dmsp-ols night lights imagery. *International journal of remote sensing*, 34(22):8118–8141, 2013.

Enrico Moretti. Local labor markets. *Handbook of Labor Economics*, 4b, 2010.

Arpita Mukherjee and Bhavook Bhardwaj. Imposition of mat on sezs: Concerns and the way forward. 2016.

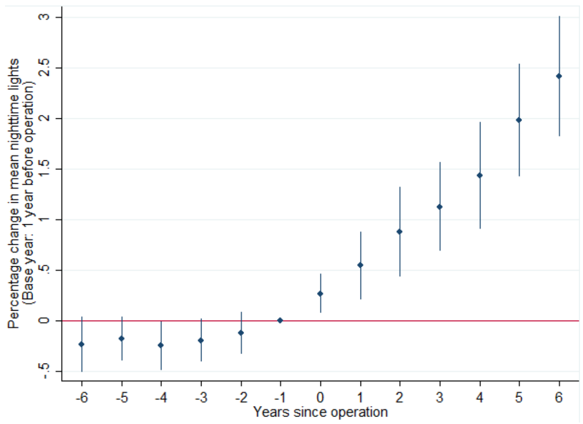
Ajay Shenoy. Regional development through place-based policies: Evidence from a spatial discontinuity. *Working Paper*, 2016.

Paul Sutton, Dar Roberts, Chris Elvidge, and Henk Meij. A comparison of nighttime satellite imagery and population density for the continental united states. *Photogrammetric Engineering and Remote Sensing*, 63(11):1303–1313, 1997.

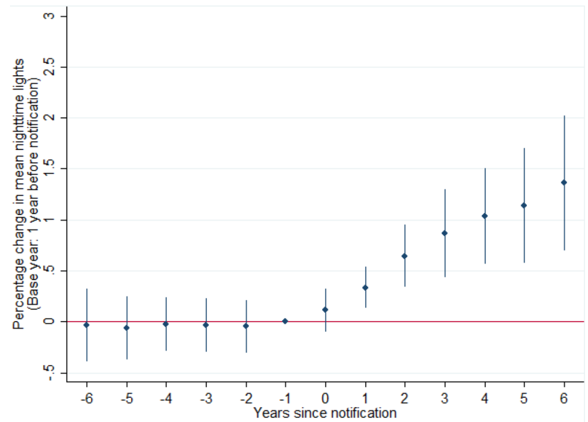
Petia Topalova. Factor immobility and regional impacts of trade liberalization: Evidence on poverty from india. *IMF Working Paper*, 2010.

Jin Wang. The economic impact of special economic zones: Evidence from chinese municipalities. *Journal of Development Economics*, 101:133–147, 2013.

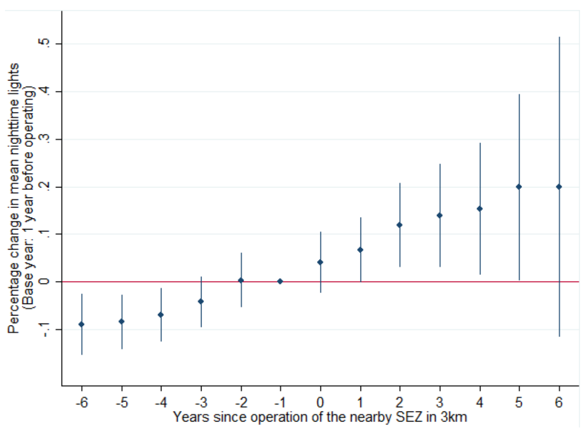
Figure 1: Event Study of SEZs and SEZ Neighborhoods



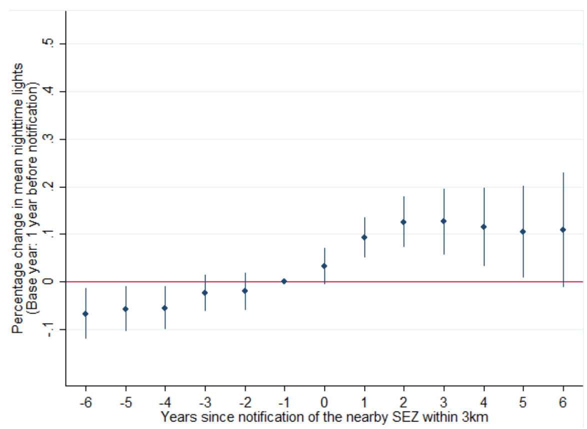
(a) Event: Operation, Inside SEZs



(b) Event: Notification, Inside SEZs



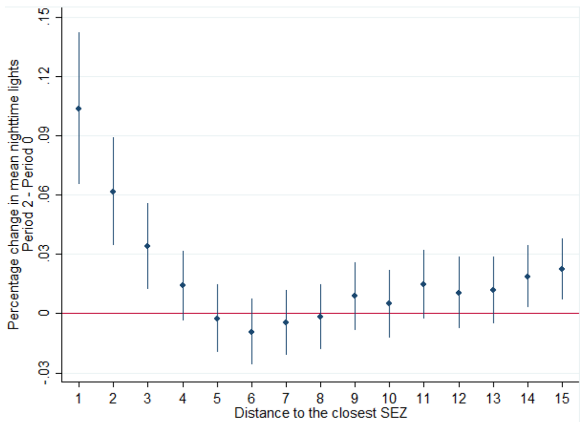
(c) Event: Operation, SEZ Neighborhood



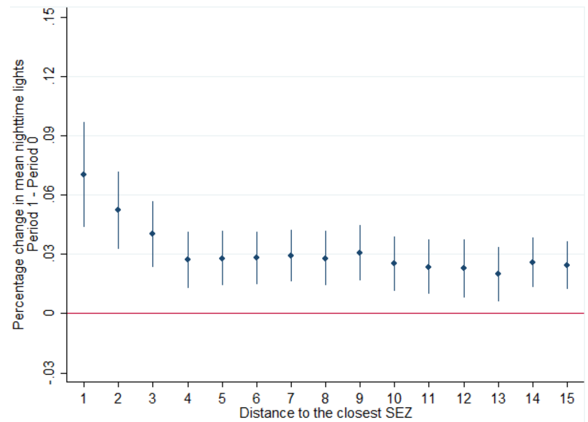
(d) Event: Notification, SEZ Neighborhood

Note: Figures plot .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.

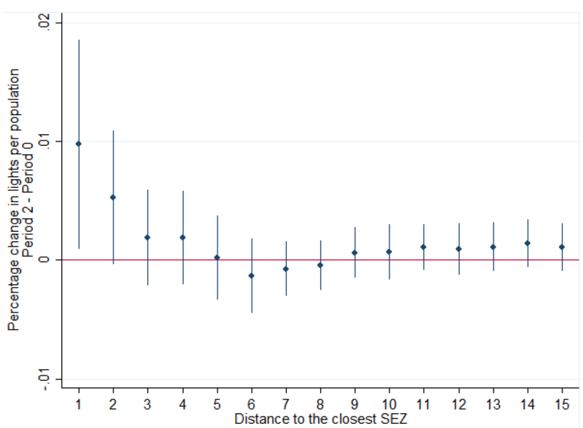
Figure 2: Effect of SEZs across Distances



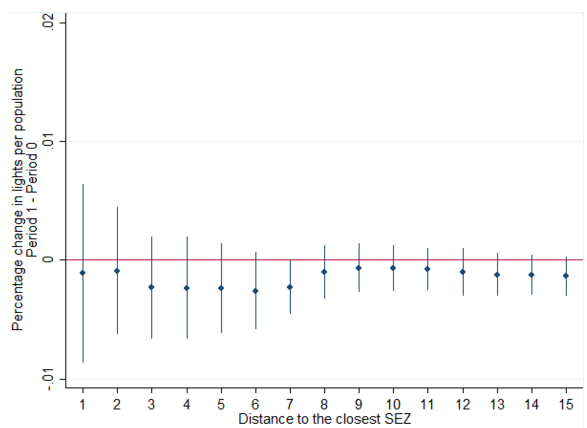
(a) Long Run Effect: NTL



(b) Short Run Effect: NTL



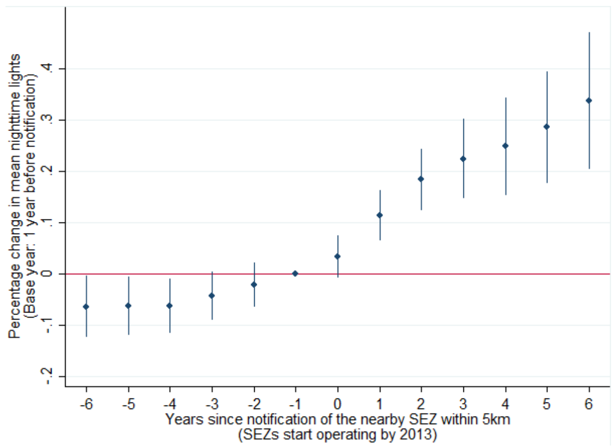
(c) Long Run Effect: NTL per population



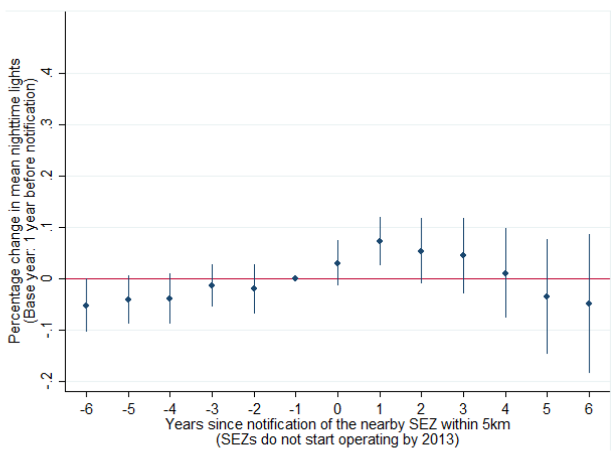
(d) Short Run Effect: NTL per population

Note: Figures plot .01*percentage change in NTL and NTL/population backed out from $\delta_x' s$ in specification 2 due to the presence of an additional notified/operating SEZ in the 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.

Figure 3: Testing Strength of Identification: Using NTL



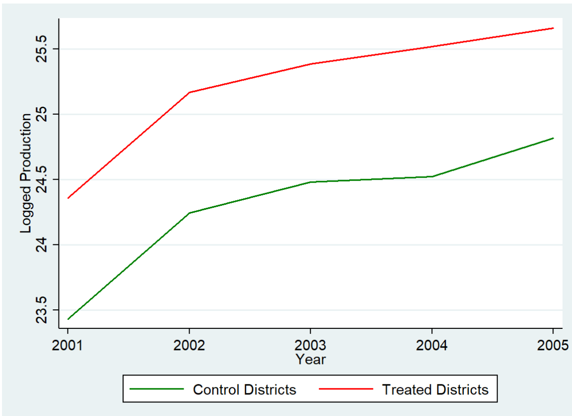
(a) Notified SEZs that became operational eventually



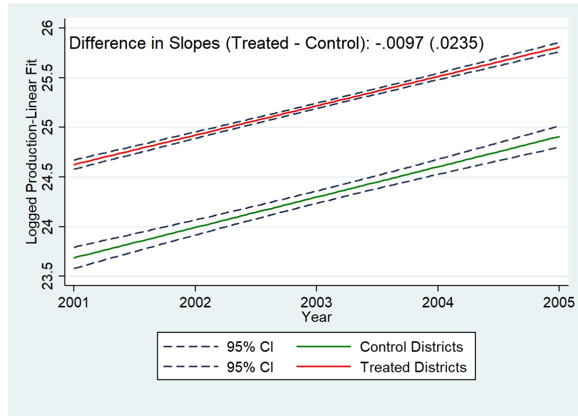
(b) Notified SEZS that did not reach operation

Note: Figures (a) and (b) show no significant signs of pre-trends before the year of notification indicating that regions treated with SEZs that ultimately became operational were not already growing, and at a different trend than those with SEZs that remained unoperational. Figure (b) further shows an initial increase in activity after notification of unoperational SEZs, indicating developer seriousness. The year before operation 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.

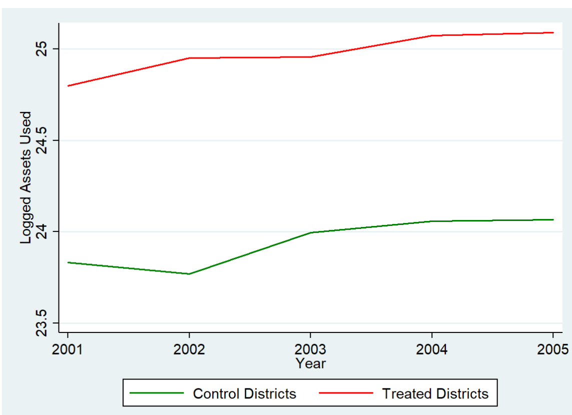
Figure 4: Testing for Pre-trends in District-level Manufacturing Outcomes



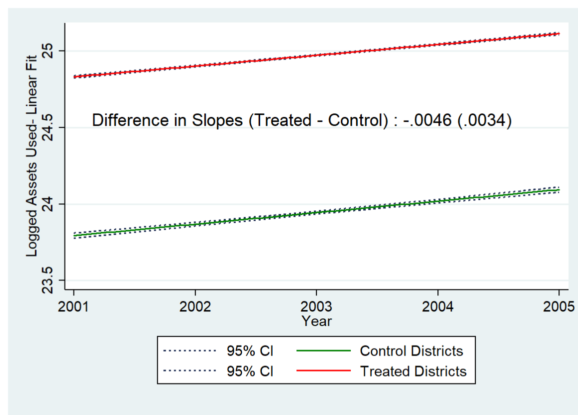
(a) Actual Trend: Production



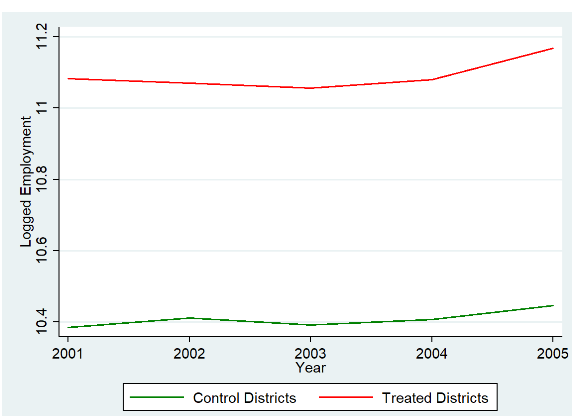
(b) Linear Fit: Production



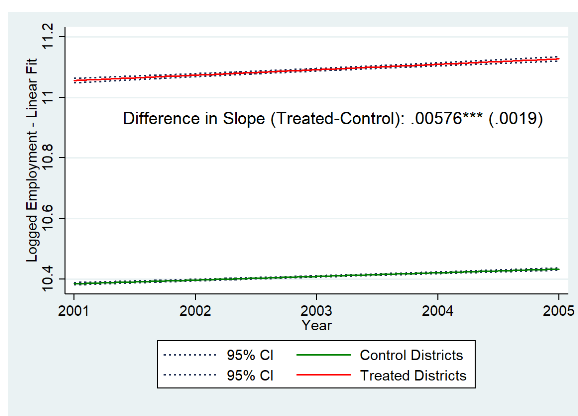
(c) Actual Trend:Total assets



(d) Linear Fit: Total assets



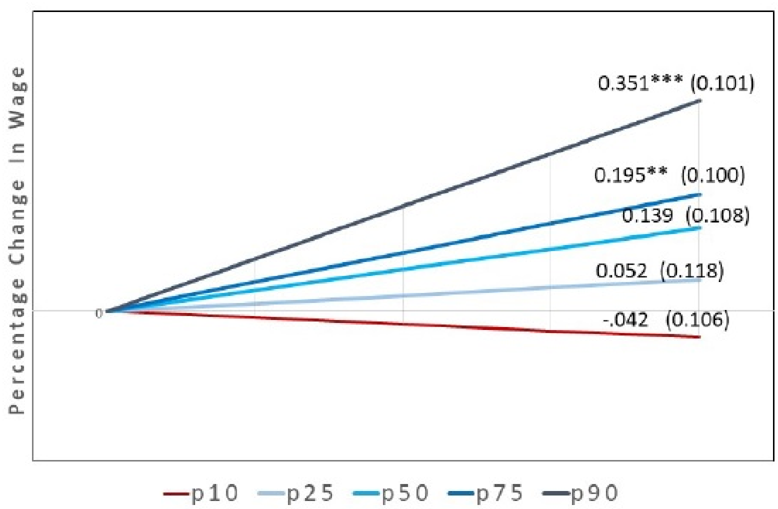
(e) Actual Trend: Employment



(f) Linear Fit: Employment

Note: Treated districts: Those with at least one operational SEZ before 2011. Control districts: Those with at least one notified SEZ, none of which are operational before 2011. Production, assets and employment are district-2 digit industry totals, in logged values. Standard errors for differences in trend in parantheses.

Figure 5: Effect of SEZs on the Wage Distribution



Note: Figure shows the percentage change in monthly wages, from quantile regressions of logged worker wages (10th, 25th, 50th, 75th, and 90th percentile), in treated districts relative to control districts. District, 2-digit industry and year fixed effects included, standard errors are clustered at the district level and reported in parantheses. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: Figure shows the percentage change in monthly wages, from quantile regressions of logged worker wages (10th, 25th, 50th, 75th, and 90th percentile), in treated districts relative to control districts. District, 2-digit industry and year fixed effects included, standard errors are clustered at the district level and reported in parantheses. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 1: Summary of SEZ & NTL Data

	Mean	Median	Std. Dev.
Panel A: All SEZs			
Year of notification	2008.1	2008.0	1.892
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	1.303
Year of operation	2009.7	2010.0	2.226
Number of observations		133	

· An SEZ is labeled as 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.

Table 2: Comparison of Pre-Treatment Averages Across Treatment, Control and Other Districts

Variable	Control	Treated	Other Areas Within States
Demographics			
Density(/sq. Km)	4059	3988	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.)	1263.1	1665.5	934.5
Districts	24	40	167
Year of Earliest Formal Approval	2006.96	2006.125	-
Year of Earliest Notification	2008.44	2006.707	-
Year of Earliest Operation	2012.3	2008.317	-

1- Value 1 if firm employs over 10 workers, 0 otherwise

Table 3: Effect of Every Additional Operating SEZ on a Firm in Formal Manufacturing

Dependent Variable	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)
	Mean	No. Operational (α_1)	Mean	No. Operational (α_1)	Mean	No. Operational (α_1)	Mean	No. Operational (α_1)	Mean	No. Operational (α_1)
Production	28630983	0.022**	2649814	0.015**	9626209	0.019***	33598769	0.028***	303229348	0.006*
		(0.011)		(0.007)		(0.003)		(0.004)		(0.003)
Assets Used	22976901	0.015*	169397	0.035**	627814	-0.002	2446087	0.015**	27508346	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	579546	0.018*	545796	0.014***	873270	0.015***	1045494	0.021***	1289803	0.002
		(0.011)		(0.005)		(0.004)		(0.003)		(0.002)
Wages	40538	0.012***	39340	0.011***	47099	0.007**	54721	0.010***	74608	0.016**
		(0.003)		(0.003)		(0.003)		(0.003)		(0.007)
New	.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)
N	122624		18554		21919		39533		42618	

· All means in Rupees, except for Employment and New, reported in numbers and proportions respectively
· Dependant variables listed in column 1 are at the firm-level, and enters in logged values
· New- Takes value 1 if firm formed after first SEZ gets notified in the district, 0 otherwise
· Main RHS variable is the number of operational SEZs in the district of the firm in a year
· The respective coefficient α_1 is obtained from running specification 5 separately for each dependent variable and size category
· See Appendix for coefficient on number of notified and not yet operational SEZs, α_2
· District- industry and year fixed effects included with errors clustered at district level
· standard errors in parantheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Effect of SEZs on the Informal Sector

	Manufacturing			Services		
Dependent Variable	Mean	(1)	(2)	Mean	(3)	(4)
Production	73967	-0.765***	-0.826**	71264	0.243	0.177
		(0.26)	(0.38)		(0.151)	(0.25)
Gross Value Added	18657	-0.578***	-0.646***	38715	0.228**	0.213
		(0.18)	(0.24)		(0.108)	(0.18)
Assets used	200938	-0.401***	-0.419**	106424	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	4836	-0.372***	-0.452***	21216	0.044	0.022
		(0.12)	(0.17)		(0.084)	(0.14)
Labor Productivity	18506	-0.545***	-0.618**	39054	0.058	-0.017
		(0.20)	(0.30)		(0.122)	(0.20)
Wage	3064	-0.574***	-0.626**	1586	0.412**	0.435*
		(0.17)	(0.25)		(0.168)	(0.24)
Observations	59233	59233			42056	42056

· All means in Rupees, except for employment reported in numbers
· Dependant variables listed in column 1 are at the firm-level and in logged values
· Main RHS variable is indicator for district treated with at least one operational SEZ before 2011 and corresponding coefficient (α_2) reported from specification 4
· Panels (I): District, industry, year fixed effects included
· Panels (II): District- industry, year fixed included
· standard errors clustered at district level, in parantheses
· * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Impact on Total Economic Activity: Informal and Formal Sectors

			Formal[1]		
	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)
			Informal		
	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)
			Overall		
	Household Emp	Emp in Firms>10	No. Firms[2]	No. Informal Firms[2,3]	
Manufacturing	-0.082**	0.079**	0.078	-.238*	
	(0.036)	(0.036)	(0.463)	(0.137)	
Services	-0.039	0.056**	0.117	-.200**	
	(0.034)	(0.022)	(0.414)	(0.103)	

· Row headings are the dependent variables, district-industry totals, in logged values
· Regressions are carried out separately for manufacturing and service sectors
· Cells report α_2, coeffecient on main RHS variable,
· Indicator if district is treated with an operational SEZ before post-treatment year, from specification 4
· 1-post treatment year for formal sector analysis at district level up to 2009, Informal sector and worker level analysis up to 2011
· 2-Analysis using the Economic Census data with post treatment year 2012
· 3-Informality defined according to the 10-worker rule
· standard errors clustered at district level, in parantheses, district industry and year fixed effects included
· * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Drivers of Wage Effect

		Worker Characteristics (Education)		
		(., Primary]	(Primary, Higher Secondary]	(Higher Secondary,.]
Manufacturing	α_2	-0.203	0.153	0.512**
		(0.264)	(0.194)	(0.21)
	Mean Wage (Rs)	906.9	1772.2	5486.2
	N	15388	6686	9727
Services	α_2	0.340	-.041	0.394**
		(0.296)	(0.137)	(0.19)
	Mean Wage (Rs)	1465.6	2565.7	6310.7
	N	5167	6962	8531
		Firm Characteristics (Size/Formality)[1]		
		Informal	Formal	
Manufacturing	α_2	0.111	0.407***	
		(0.28)	(0.13)	
	Mean Wage (Rs)	1510.2	2697.3	
	N	5746	5314	
Services	α_2	0.022	0.329*	
		(0.13)	(0.17)	
	Mean Wage (Rs)	1737.1	4722.1	
	N	8909	8189	

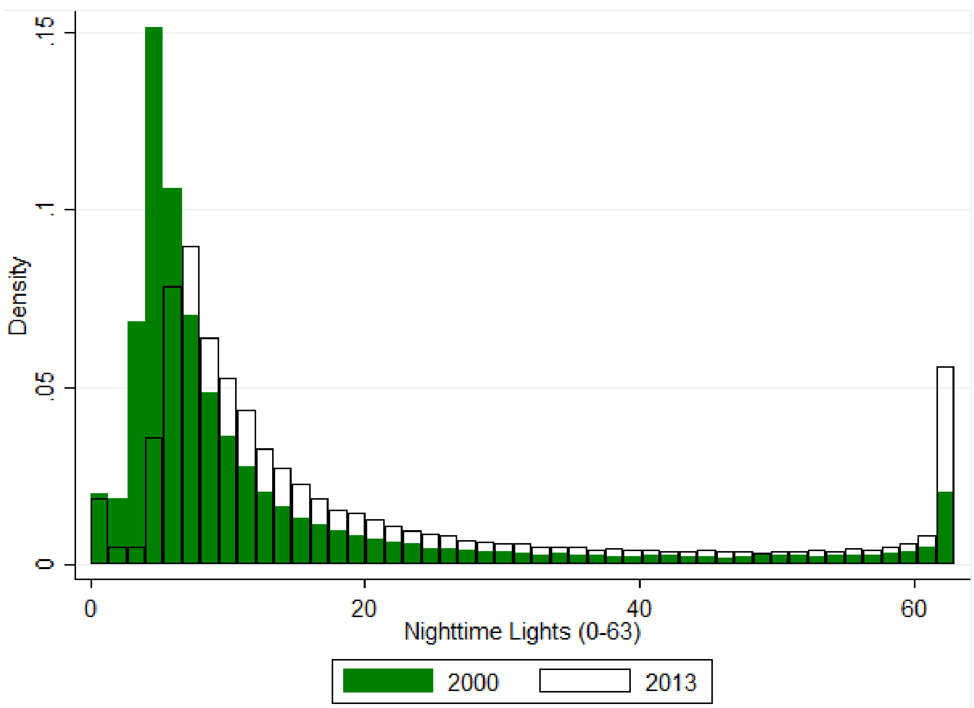
· 1- Firm considered formal if it employs greater than 10 workers, otherwise informal
· Dependent variable is logged wage
· Separate regressions run for sectors formed from every combination of row and column headings
· $\alpha_2's$, coefficients on indicator from equation 4,
· if district is treated with an oeprational SEZ before 2011, are reported
· District, industry and Year FE, errors clustered at district level
· standard errors in parantheses
· * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A Appendix to Section 4

A.1 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 since 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.

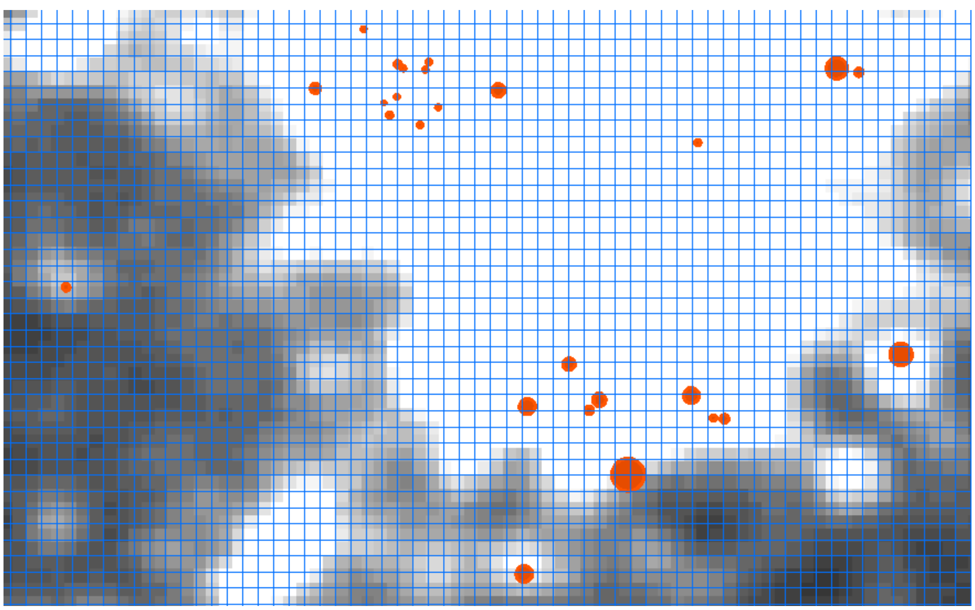
Figure A.1: Distribution of Nighttime Lights



A.2 Construction of the cell-level panel

We overlay a fine grid 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/spillover effects of SEZs on the cells, we restrict our attention to cells that are believed to be strictly outside of SEZs using the geocoded location data of the SEZs. We assume that SEZs are circularly shaped since the exact shape of SEZs are unknown. Then, 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 some 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\%. We then record for each cell, the NTL reading (after removal of ephemeral events and gas flares)[40] 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 valid observations (cells) is 62,386 per year. See \hyperref[image_clustering]{Figure \ref{image_clustering}} containing the fine grid cells and circles of SEZs.

Figure A.2: SEZs and Grid Layer

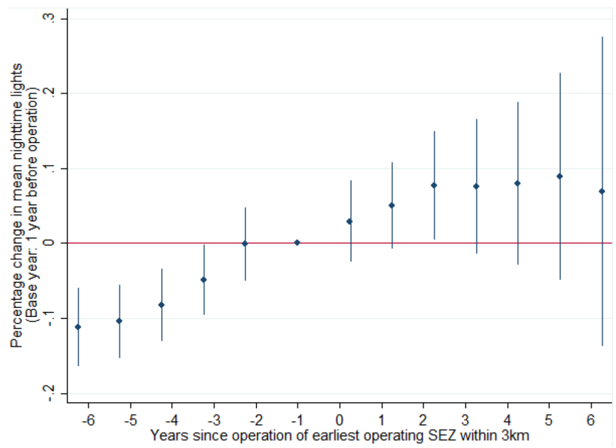


[40] In cases there were two satellites collecting data for a cell, we take the average of the two data.

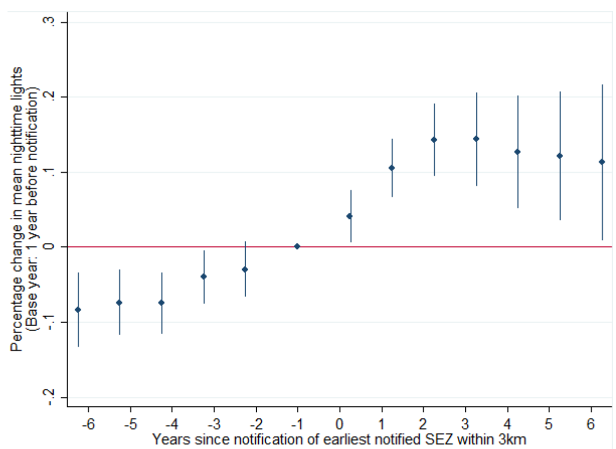
A.3 Alternate specification for Events Study Analysis

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.3



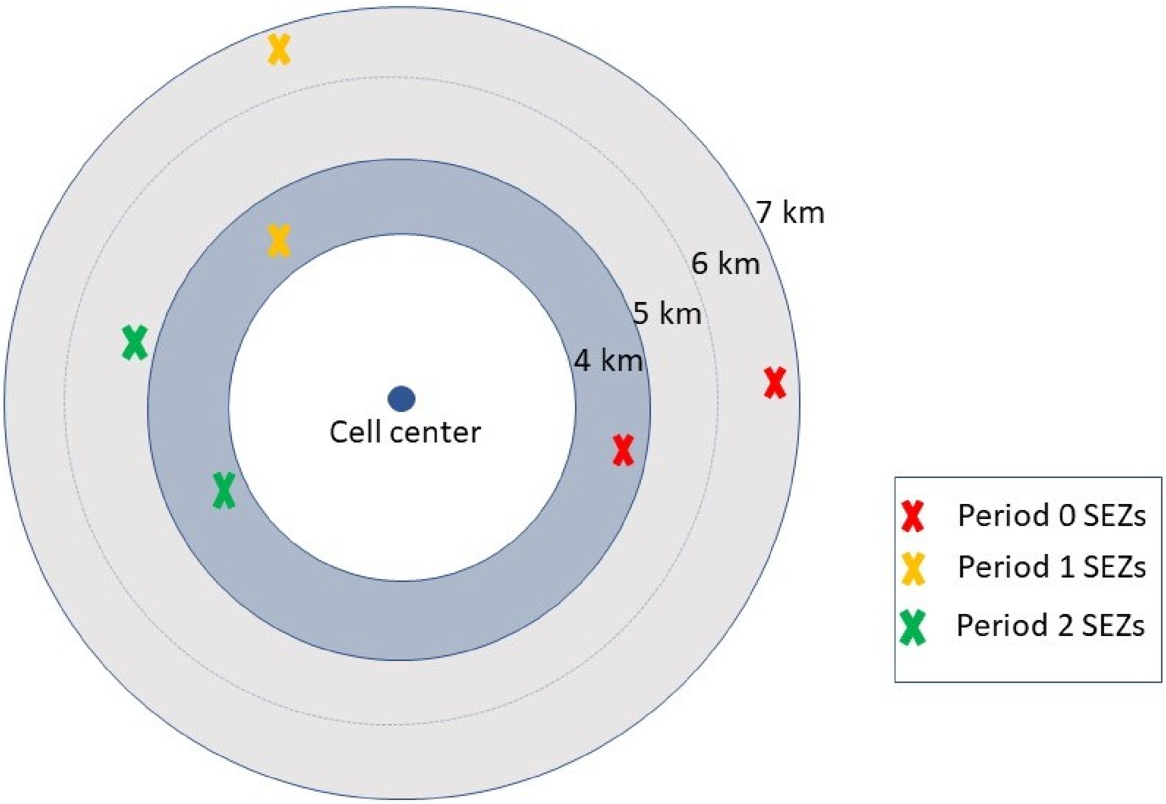
(a) Event: Operation, SEZ neighborhood



(b) Event: Operation, SEZ neighborhood

A.4 Pictorial representation of the Distance Dimension Specification 2

Figure A.4: Analyzing the Effects of an SEZ at a distance of 4 to 5 kilometers



Note: 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.

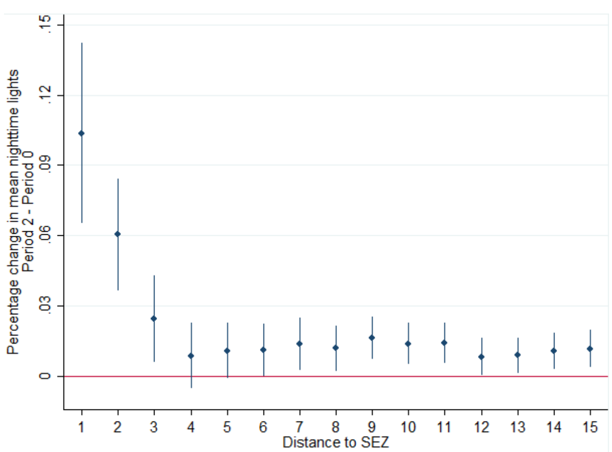
A.5 Alternate specification for Distance Dimension Analysis

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 ($1 \leq x \leq 15$),

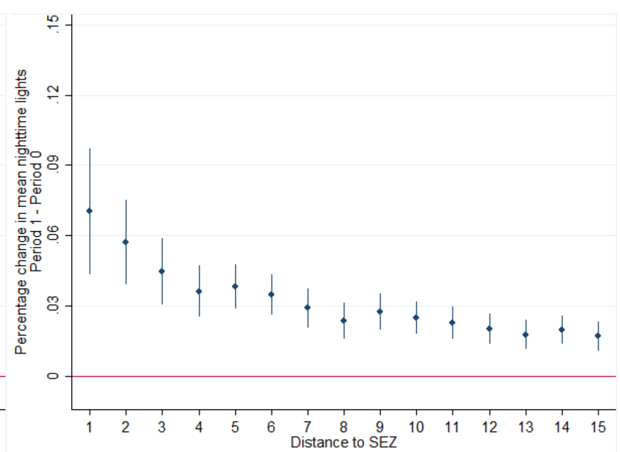
$$\begin{aligned} log(light_{it}) = \alpha_i + \beta_t + \gamma_x * period1_{ixt} + \delta_x * period2_{ixt} \\ + \sum_{1\leq d\leq 15}^{d\neq x} \sum_{\theta=0}^{2} \lambda_d^p period\theta_{idt} + \epsilon_{it}, \end{aligned}$$

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.5



(a) Long-run effects of SEZs across distances



(b) Short-run effects of SEZs across distances

B Appendix to Section 5

Figure B.1: SEZ Timeline

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

Table B.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, Govt. of India

Table B.2: Treatment & Control Group Formation

Data-set	Level of analysis	Pre-treatment	Post-treatment	Considered set	Of which treated:
Economic Census	Village/Town	2005	2013	>=1 notified SEZ on or before 2012	with >=1 operational by 2012
NSS Unorganized Firms	District	2000-01, 2004-05	2010-11	>=1 notified SEZ on or before 2010	with >=1 operational by 2010
NSS Worker Survey	District	2000-01, 2004-05	2010-11	>=1 notified SEZ on or before 2010	with >=1 operational by 2010
ASI Formal Manufacturing	District	2001 to 2005	2006 to 2009	>=1 notified SEZ	total operational & notified every year

Table B.3: Summary of Key Variables of Workers and Firms before 2005

	mean	sd	25th pctl	50th pctl	75th pctl	min	max
Demographics							
Density(/sq. Km)	3989	2606	2279	3597	4845	250	19865
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	45481
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

All averages are at the district-level, and when possible, at the village-level
* values as of 2001
1 -per 10000 population
2 - logged real values
3 - in absolute numbers
Source: Census Digital Library of India, NSS and ASI surveys, EC data

Table B.4: Falsification Test: The Effect of SEZs on Districts before Notification

	Eventual Number of:	
	Operating SEZs	Notified SEZs
	α_1	α_2
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 and year fixed effects included
· All outcome variables are average firm-level variables in the formal sector of a district-industry group, in logged real values
· Errors clustered at district level, in parantheses
· Source: ASI data on formal manufacturing firms (2000-2005)

Table B.5: Effect of Every Additional Notified SEZ on a Firm in Formal Manufacturing

Dependent Variable	All Firms	Size 1: <=10	Size 2: (10,20]	Size 3: (20,100]	Size 4: >100
	(α_2)	(α_2)	(α_2)	(α_2)	(α_2)
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	0.002***	0.002*	0.002	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	122624	18554	21919	39533	42618

· Dependant variables listed in column 1 are at the firm-level, and enters in logged values
· New- Takes value 1 if firm formed after first SEZ gets notified in the district, 0 otherwise
· Main RHS variable is the number of notified, non-operational, SEZs in the district of the firm in a year
· The respective coefficient α_2 is obtained from running specification 5
separately for each dependent variable and size category
· District- industry and year fixed effects included with errors clustered at district level
· standard errors in parantheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.6: Effect of SEZs on Composition of Educated Workforce

Proportion of workers with	α_2	N	Mean Proportion
Below Primary	-0.054*	303055	0.59
	(0.03)		
Below Secondary	-0.034	303055	0.92
	(0.04)		

· Main dependent variable: indicator if worker has below primary, and below secondary education
· $\alpha_2's$, coefficients on indicator from equation 4,
· if district is treated with an operational SEZ before 2011, are reported
· District-industry and Year FE, clustered at district level
· standard errors in parantheses
· * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.1 SEZs' Influence on Village-level Economic Activity

We follow specification 4 at the village level with the Economic Census data. Table B.7 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.[41]

Table B.7: Village Level Analysis of Firm and Worker Numbers

Dependent Variable	Mean	α_2
Total Firms	330675	0.252
		(0.412)
Total Workers	2190081	0.630*
		(0.341)
Avg Workers per Firm	6.2	0.135**
		(0.055)
Pr(Firms with 0 Hired Workers*)	0.38	-0.043**
		(0.020)
Observations	2497090	

* Also known as Own Account Enterprises (OAEs)
Errors clustered at village level
All outcome variables in logged values
α_2 reported as in specification 4
standard errors in parantheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[41]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.

B.2 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.[42] 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:

$$\begin{aligned} log(population_{idt}) = \alpha_0 + \alpha_1 Year_t + \alpha_2 District_d + \alpha_3 Operating_{idt} \\ + \alpha_4 Operating_{idt} \times After_t + \epsilon_{idt}, \end{aligned} \tag{6}$$

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 B.8 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 B.8: 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)
Number of observations	321
Overall R^2	0.629

· District fixed effects are included, and year 2000 is omitted. Standard errors are clustered at district level.
· * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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