Land Rezoning and Structural Transformation in Rural India

Evidence from the Industrial Areas Program

David Blakeslee
Ritam Chaurey
Ram Fishman
Samreen Malik

Abstract

Zoning laws that restrict rural land to agricultural production pose an important institutional barrier to industrial development. This paper studies the effects of the Industrial Areas (IA) program in Karnataka, India, which rezoned agricultural land for industrial use, but without the economic incentives common with other place-based policies. This paper finds that the program caused a large increase in firm creation and employment in villages overlapping the IAs. Moreover, the surrounding areas experienced spillover effects, with workers shifting from agricultural to non-agricultural employment, and entrepreneurs establishing numerous small-scale service sector and agricultural firms.

Land Rezoning and Structural Transformation in Rural India: Evidence from the Industrial Areas Program[*]

David Blakeslee[†]   Ritam Chaurey[‡]

Ram Fishman[§]   Samreen Malik[¶]

Key words: Industrial Areas, Place-based Policies, Spillovers, Labor Market.

JEL code: 012, 025, R2

[†]Email: david.blakeslee@nyu.edu. New York University (AD).

[‡]Email: rchaurey@jhu.edu. Johns Hopkins University.

[§]Email: ramf@post.tau.ac.il. Tel Aviv University.

[¶]Email: samreen.malik@nyu.edu. New York University (AD).

1   Introduction

Place-based policies are used by governments around the world in an attempt to stimulate localized economic activity. These policies usually employ generous financial incentives in order to attract firms to desired locations,[1] and are justified by the presence of coordination and market failures, which would otherwise lead to inefficiently low levels of industrial agglomeration (Kline and Moretti, 2014b). They have also been promoted as a means of generating broader economic development in economically marginalized areas (Greenstone et al., 2010a; Greenstone and Looney, 2010; Kline and Moretti, 2014a). The evidence on the effectiveness of such policies remains mixed, and is relatively scarce for developing countries.

In this paper, we study a different type of place-based policy which is motivated by regulatory barriers on land-use that impede economic activity, rather than market failures. Such barriers are thought to constrain firm creation, productivity, and growth in many developing countries (Duranton et al., 2015). For example, in India, the difficulty of procuring large parcels of land for industrial use has been frequently cited as a particularly important bottleneck (Rajan, 2013). This is due to land-use laws that reserve most rural land for agricultural production, and impose considerable bureaucratic obstacles to its conversion for industrial production. Because systemic reform of such regulations is rendered extremely difficult by India's political economy, policy makers have attempted to circumvent them through localized interventions that involve land acquisition and rezoning by the government itself (Kazmin, 2015). It is unclear, however, whether such policies are sufficient for attracting firms, or what their broader effects are on the local economy.

We study the Industrial Areas (IA) program of the Indian state of Karnataka.[2] Under this program, the state government acquires and consolidates contiguous parcels of privately-held agricultural land, rezones it for non-agricultural activities, and makes it available to private firms for sale or lease at market rates (Government of India, 2009). Most significantly, no financial incentives are offered to firms to locate their operations in the IAs, in stark contrast to other place-based policies, leading policymakers to describe the IA program in a technical manual as "essentially a piece of real estate promotion" (Government of India, 2009). The intention of the policy, therefore, is to harness market forces to promote industrialization, with the government acting primarily as the facilitator of the necessary rezoning to enable non-agricultural production.

Our analysis addresses two fundamental questions. First, we examine whether such

---

[1]The popular form of place-based policies across the world is to provide financial incentives such as tax exemptions, wage subsidies, hiring credits, land grants, infrastructure grants in a particular region to incentivize firms to locate there.

[2]Karnataka is the seventh largest state in India and the eighth most populous, with a population of roughly 60 million people.

limited incentives succeed in drawing large manufacturing firms to the IAs. In contrast to other place-based policies that offer financial benefits to attract firms, it is far less clear that the availability of land, by itself, will suffice to attract firms to rural locations that may have other disadvantages. Insofar as such a policy is successful, it would suggest that land scarcity represents a binding constraint on local manufacturing activity, as is often claimed by policy analysts in India.

Second, we test whether the IAs generate economic spillovers to surrounding villages. Even if IAs succeed in attracting firms, it is not clear that these firms will recruit a substantial share of their labor force from surrounding areas. Given the low levels of human capital in the areas in which IAs are established, firms may prefer to bring in labor from elsewhere, or may be less labor intensive to begin with. The effect of IAs on local labor markets therefore provides important insights on whether characteristics of the rural labor force are an important factor impeding the structural transformation of rural economies.

For similar reasons, the effect on firm activity in areas near the IAs is also unclear. The literature often hypothesizes that the arrival of firms in a given location will trigger firm creation and increased productivity in surrounding areas through agglomeration economies (Greenstone et al., 2010b). However, local entrepreneurs may lack the necessary skills and resources for establishing firms with substantial input-output linkages to IA-based manufacturing firms. In addition, given that land regulations are not relaxed outside the IAs, the establishment of large, non-agricultural firms will continue to face significant regulatory obstacles.

Evaluating the effects of place-based policies poses two key empirical challenges: (i) the construction of valid counterfactuals to deal with their non-random placement; and (ii) accounting for possible negative and positive spillovers to nearby areas.[3] Our data is particularly well suited to dealing with these issues, as the location of the industrial areas is known with precision, and the key outcomes and explanatory variables are observed at the village level, allowing for an analysis at a high level of spatial resolution.

Our analysis makes use of a difference-in-differences identification strategy to identify the effects of IAs created between the years 1991–2011. Nearly 50 IAs were established in Karnataka during this time. In order to identify the appropriate control group in the presence of possible spillovers, we first conduct a semi-parametric analysis of the impacts of IA creation in the villages overlapping the IA, and extending up to 30 kms away, in a flexible and spatially precise manner. This approach provides evidence that spillovers occur up to a distance of 5 kms from the IA. We therefore designate villages located more than 5 kms from an IA as the control sample, and use separate indicators to estimate treatment effects both within the IA, and in villages located within 5 kms of the IA.

---

[3]See Neumark and Simpson (2015) and Ham et al. (2011) for a discussion of these issues.

2

We document two main results capturing the effects of IAs on economic activity. First, we find that IAs have been highly successful in promoting economic development within the IA zones. The establishment of an IA between 1991–2011 led to a two-fold increase in night-light density, and the creation of roughly 40 new firms and 940 new jobs within each IA.

Second, we find evidence of substantial economic spillovers. We find increases in the number of firms and workers in areas outside the IAs up to a distance of 5 kms. In the local labor markets, there is an increase in the share of male workers engaged in non-agricultural activities. The magnitude of these changes is largest in villages overlapping the IA, but also extends farther out, falling monotonically with distance up to 5 kms from the boundary of the IA.

To better understand the factors driving the local spillovers, we study the types of firms being established within and around the IA. Within the IA, the majority of job and firm growth is in the manufacturing sector, and includes both large and small firms. In stark contrast, newly created firms outside the IAs are mostly in the agricultural[4] and service sectors, and virtually all of these firms employ fewer than 10, and generally only 1 or 2 employees. These patterns lead us to cautiously hypothesize that firm creation outside of IAs is mostly driven by increases in demand for goods and services from workers employed in large, IA-based manufacturing firms, as well as the relaxation of credit constraints for agricultural producers.

We provide additional insights into the mechanisms at play through a heterogeneity analysis based on baseline village characteristics. First, we find that employment growth is larger in IAs located closer to major highways and cities, potentially reflecting the importance of market access. Second, amongst villages located close to IAs, impacts on firm growth and labor force composition are smaller where baseline levels of agricultural productivity are higher, perhaps due to the higher opportunity cost of exiting agriculture in such locations.

We also find important heterogeneities according to the the socio-economic composition of both village populations and firm ownership. First, we find impacts to be larger in villages with higher baseline fractions of Scheduled Castes (SCs), a particularly disadvantaged group in Indian society. In addition, we find that the IAs are associated with somewhat greater firm creation amongst SCs and women, suggesting that such groups have benefited disproportionately from the new economic opportunities created by IAs.

One challenge to our identification strategy is that treatment and control villages differ at baseline, particularly with respect to their distance from cities and main roads. To address this possibility, we control for baseline levels of these variables interacted with time

---

[4]The Economic Census includes both firms producing agricultural goods and firms producing animal products under the rubric of "agriculture."

indicators. We also test the robustness of our results to alternative and demanding choices of the control group that make it comparable to the treatment group, including: villages in the same administrative unit (sub-district);[5] villages that are not more than 10 kms away from the IA; and villages matched with the treatment across primary observables. The results are not significantly affected by these modifications.

A second important challenge to our identification strategy is that we lack village-level data prior to the study period (pre-1990) necessary for comparing pre-trends. As an alternative, we show that villages receiving IAs later in the study period (2001–2011) displayed temporal trends in the preceding decade (1991–2001) that were indistinguishable from their respective comparison groups.

In addition, we use an event study design to show that light density (the only variable for which we have year-by-year data) displayed similar trends in the treatment and control groups prior to the establishment of the IAs, and then experienced a sharp divergence after their creation. Finally, we conduct placebo regressions in which treatment status is assigned to villages prior to the establishment of nearby IAs, and show that these specifications yield null findings.

Our findings contribute to the growing literature on place-based policies in developing countries. Several papers have documented substantial effects for special economic zones (SEZs) in developing countries (Wang, 2013; Cheng, 2014; Alder et al., 2016; Lu et al., 2018). A smaller body of research has explored the effects of other types of place-based policies (Chaurey, 2016; Shenoy, 2018; Abeberese and Chaurey, 2019). While such policies have generally proven effective, their high pecuniary costs and administrative obligations may be prohibitive for many developing countries. As such, our paper makes an important contribution in understanding whether place-based policies consisting primarily of local institutional reforms through land zoning can be successful in promoting industrialization.

Another crucial question in this literature is the nature and extent of the spillovers generated by place-based policies. Such spillovers may take the form of traditional Marshallian agglomeration economies (Ellison and Glaeser, 1999; Rosenthal and Strange, 2004; Ellison et al., 2010; Greenstone et al., 2010b; Kline and Moretti, 2014b); or, alternatively, may operate on the demand side through income channels (Rosenstein-Rodan, 1943; Murphy et al., 1989). The evidence for spillovers from program sites to surrounding areas is mixed. For example, Criscuolo et al. (2019) (UK regional selective assistance), Neumark and Kolko (2010) and Freedman (2013) (California and Texas enterprise zones, respectively), and Martin et al. (2011) (clusters in France) find no local spillovers outside of program areas. On the other hand, Zheng et al. (2017) and Alder et al. (2016) find evidence for positive spillovers of Chinese SEZs and industrial parks, and Greenstone et al. (2010b) find large agglomeration

---

[5]There are 175 sub-districts in Karnataka.

effects on incumbent plants in US counties that attracted a large manufacturing plant. Our findings on spillovers represent an important contribution to this literature, demonstrating both the substantial spillovers generated by IAs, and the way in which these spillovers are constrained by regulations and the structure of the local economy.

Our results also speak to one of the most important themes in development economics: the relationship between the agricultural and manufacturing sectors. Since Lewis (1954), the absorption of (low-productivity) agricultural workers by (high-productivity) manufacturing firms has been viewed central to the development process (see Gollin, 2014 for an overview). Because the IAs generated exogenous variation in the presence of large manufacturing firms in rural areas, our findings shed light on the effects of industrial production on the structure of agrarian economies. This paper provides some of the first empirical evidence for the role of land rezoning in triggering the structural transformation of an agrarian economy and has important implications for the spatial distribution of economic activity in India (Desmet et al., 2015; Amirapu et al., 2019).

The remainder of the paper is organized as follows. In Section 2 we give background details on land-use regulations and IAs in Karnataka. Section 3 presents our data sources and empirical specification. Section 4 presents our results and we conclude in Section 5.

2 Background

In the last twenty years, industrial production in India has increasingly shifted from urban to rural areas, with a disproportionate share of this movement accounted for by firms in the formal sector (Ghani et al., 2012). This trend towards rural production has been impeded, however, by a variety of rules and regulations limiting the use of agricultural land for non-agricultural activities (Morris and Pandey, 2007).[6] The IA policy represents one of several approaches that state governments have employed for overcoming these barriers. We first provide information on land-use policies, and then discuss the Karnataka state industrial areas programs in greater detail.

2.1 Land-use in Karnataka

Karnataka's land-use rules were laid out in the Karnataka Town and Country Planning Act (hereafter, KTCPA) of 1961. Though a variety of amendments have been made to the

---

[6]The common All-India Law for Preservation of the Agricultural Lands, instituted at the time of independence (1947) and revised several times since, places numerous restrictions on the transfer of agricultural land to a non-agriculturist, where the latter is defined as an individual not involved in the cultivation of crops and lacking family ties to agriculture. However, the transfer of land and the changing of land usage is strictly under the jurisdiction of state governments, giving states significant power to acquire land but by compensating owners in a fair manner and using it for various non-agricultural projects.

the Act, the principal rules persist with only minor modifications. Land-use rules can be summarized as follows.

First, the KTCPA invokes the national Land Acquisition Act to establish the power of the state to acquire land as deemed necessary for the purpose of planning and development. To ensure fairness for landowners, an amendment was made to this rule requiring that compensation for any acquired land be based on market value on the date of publication of improvement or development schemes. In addition, the government must provide a "grant of solatium," increasing the compensation by 15% in light of the compulsory nature of acquisition.

Second, the KTCPA also references the national Land Revenue Act in stipulating that permission must be obtained from the Deputy Commissioner in order to use agricultural land for non-agricultural purposes, and defines the fees for land-use conversion. This act reflects the power of the state in determining if the change of land-use is to be granted. However, given the political economy of India, where agricultural interests are fiercely protected, such changes in land-use are difficult to achieve, even for large businesses.[7] In addition, the associated fees and taxes can represent a substantial cost to small- and medium-size businesses, discouraging them from pursuing a change in land-use.

Finally, the KTCPA states that there is no need for change of land-use if the new economic activity is undertaken by the current land owner, and the original economic activity also continues to occur. For example, if a farmer wants to establish a small mechanic shop on a share of his agricultural land, then this would be permitted. These rules, therefore, establish a land-use regime in which the greatest regulatory friction arises from the conversion of agricultural land to non-agricultural activities, with allowances made for small-scale, non-agricultural economic activities undertaken by farmers/dwellers. This feature of the land-use regulations will be important for interpreting the results presented later.

2.2   Industrial Programs and IAs in Karnataka

Since independence, the Indian government has played a large role in shaping the economy via various industrial policies. The main objective of these policies is to provide regulations and procedures for the development and management of industrial undertakings throughout the country, with close control over the respective roles of the public and private sectors. One approach to promoting industrialization has been through the creation of a variety of Industrial Estates (IE), a general label subsuming a number of place-based policies. Included in this are: IAs, export processing zones (EPZs), special economic zones (SEZs), and industrial parks and complexes. The various types of of IEs differ according to their economic

---

[7]A recent, well-publicized example of these hurdles was the failure by Tata to secure land for a major production plant in the state of West Bengal.

objectives, the incentives offered, and the economic activities they promote.

These programs began in 1955 with the founding of the first IE in Rajkot, Gujarat,[8] and soon spread to the other states of India. Competition between states has led to a broad convergence over time in industrial policy, with most states providing similar promotions and incentives.[9] Despite the relative uniformity of industrial policy, however, the execution and implementation of policy has been far more uneven, and may have contributed to the extreme regional imbalances that characterize industrial production in India.

In this paper, we study the effects of IAs in Karnataka between the years 1991–2011. IAs represent one of the industrial policies pursued by the state, relying primarily on the operation of market forces, with mainly regulatory support from the state government via rezoning the land use from agriculture to non-agriculture activities. During 1991–2011, 47 IAs were established, and 18 additional IAs were established after our sample period ends in 2016 (Figure 1).

A central challenge in this program is to determine a suitable site for the IA, the responsibility for which lies with the Karnataka Industrial Areas Development Board (KIADB). Selection of the site is based on a few criteria that includes the presence of suitable infrastructure, proximity to towns, and the promotion of backward areas. Once a site has been selected, the government uses the Land-Acquisition Act to acquire land from the current owners and re-zone the area to allow industrial activities. The plot is then equipped with basic utilities and infrastructure, including power and recycling facilities; and then leased or sold to firm owners.

While there is some lag between the announcement of the IA and the year of its establishment, the lag between establishment and the operation of arriving firms is minimal. This is because during the interim period between announcement and establishment, the government begins the process of finding firms that wish to establish operations in the IA, so that when the IA is finally opened a number of firms immediately begin operations. In particular, according to the technical manual, "[a]n attraction for a prospective occupier is the time saved in finding a site and in preparing the land" (Government of India, 2009). We therefore use the date of establishment of the IA as the date of treatment in our analysis.

Overall, the crucial benefit offered by IAs for firms is that the re-zoning of land by the state obviates the need for individual firms to engage in the costly and time-consuming

---

[8]Industrial Estates were not an Indian innovation, but were instead borrowed from the British, and had indeed long existed in various forms in the advanced, industrial economies. These would include such areas as IAs, parks, zones, districts, and so on, all of which refer to geographical units set aside for primarily industrial activity, though with significant variation in terms of incentives offered across various types of industrial estates as well as across countries.

[9]As noted by Saez (2002), the inter-jurisdictional competition between states of India is not only in terms of implementing industrial policies but is pervasive on various dimensions and primarily stemming from the economic liberalization policies of 1990s in India.

Figure 1: Timing of IA Establishment
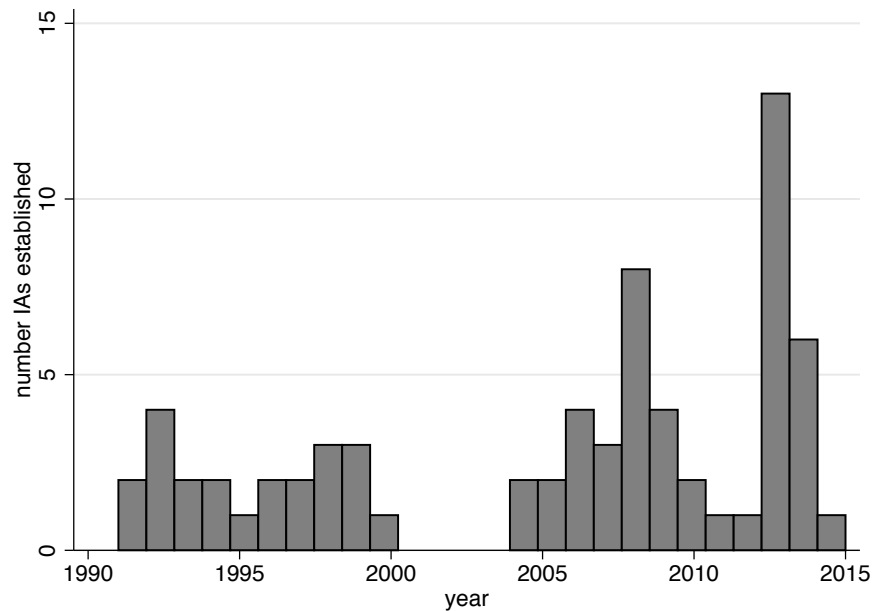
Notes: Figure 1 shows the number of IAs established in each year.
Source: Authors' own collected data containing the location and date of implementation of Industrial Area in the state of Karnataka.

efforts necessary for identifying a suitable plot of land, and securing the necessary approvals for converting it for non-agricultural activities.

3   Empirical Approach

We begin by describing the data used in this study (Section 3.1). We then present a semi-parametric analysis that guides our main empirical specification (Section 3.2). We go on to examine balance in baseline levels and pre-trends between the treatment and control groups (Section 3.3).

3.1   Data

Our analysis employs several sources of administrative data. The Karnataka Industrial Areas Development Board (KIADB) provides us with the year and location in which each IA was set up. We match the information on these IAs to the Economic and the Demographic Censuses at the village-level. We restrict the sample of IAs to those not proximate, at baseline, to pre-existing IAs or other hubs of manufacturing activity.[10] The Economic Census of India

---

[10]Our main results are insensitive to this restriction.

is a complete enumeration of all economic establishments except those engaged in non-commercial crop production, and includes both formal and informal firms irrespective of firm size. It provides us with village-level information on the number of firms by industrial classification and size, the overall number of workers in these firms, and the social caste and gender of firm owners. We use the Economic Censuses from the years 1990, 1998, 2005 and 2013. Two of our main outcome variables are derived from this source: (1) the number of firms based in each village; and (2) the number of workers employed in these firms.

The Demographic Census provides us with village-level information on the shares of the population working in various sectors, including cultivation of own farms, agricultural wage labor and non-agricultural employment of several types, their literacy rates, and the presence of various public goods (paved roads, banking facilities, etc.). We use the Demographic Censuses of 1991, 2001, and 2011. Our third main outcome variable from this source is the share of the labor force in each village which is employed in non-agricultural activities. Note that the number of workers reported in the Economic Census refers to workers employed in firms belonging to a particular village, whereas the number of workers reported in the Demographic Census refers to residents of the village, regardless of where they are employed.

We also make use of night-time lights data at the village level. The satellite data on night-time lights are collected by the National Aeronautics and Space Administration's (NASA) Defense Meteorological Satellite Programs Operational Linescan System (DMSP-OLS) via a set of military weather satellites that have been orbiting the earth since 1970. In the night-time lights data, each pixel is encoded with a measure of its annual average brightness on a 6-bit scale from 0 to 63. In our analysis, we use as night light density as our outcome of interest, which is the average luminosity per pixel. This means that the village luminosity measure is divided by the surface area of the village. This night-time light data covers the years 1992–2013.

3.2 Empirical Specification

Our primary empirical strategy for identifying the direct effects and spillovers of the IAs is based on a difference-in-differences design. We first estimate the direct effects of IAs in the villages in which they are located, and then estimate the associated spillovers to villages in the vicinity. The unit of analysis is the village, denoted by v at time t, where $t \in \{1991, 2011\}$ for variables from the Demographic Census, and $t \in \{1990, 2013\}$ for those from the Economic Census. The regression is specified as:

$$y_{v,i,t} = \alpha + \beta(IA_v \times post_t) + (post_t \times X_v)\Gamma + \delta_{i,t} + \eta_v + \varepsilon_{v,i,t}. \qquad (1)$$

The subscript i indicates the IA to which village v is closest. IA_v is a dummy variable indicating that village v's boundaries overlap with those of an IA, and $post_t$ is a dummy taking a value of 1 for $t = 2011$ or $t = 2013$, depending on the outcome. We primarily rely on a long-differences estimation strategy, in order to capture the long-run effects of IAs. However, in some specifications we include the intermediate rounds of the Demographic and Economic Censuses, and assign the $post_t$ variable a value of 1 for all years following the establishment of an IA.

X_v is a vector of control variables that may have influenced the choice of IA locations or affect economic development, observed at baseline and interacted with time. These include the (log) distance to the nearest highway and nearest city, baseline nightlight density, infrastructure (the presence of railway stations, post offices, and telephone connectivity), the presence of a primary school, (log) population, the share of the village land which is covered by forest, the share of the population that belong to the scheduled castes, the share of male workers employed in agriculture, and the (log) distance to IAs established prior to 1991. The coefficients η_v denote village fixed effects. We also include time-interacted IA fixed effects ($\delta_{i,t}$) for the IA to which each village is closest, so that identification is based on comparisons of growth in villages which are proximate to the same IA. Standard errors are clustered at the level of the IA established between 1991–2011 to which each village is nearest, in order to account for potential spatial correlation in unobservables.

The location identifier in the Economic Census gives the village in which a firm is located, but does not indicate whether the firm is located within an IA. To identify firms located within the IAs and those located in nearby villages, we use maps which show the boundaries of each village and IA. Villages whose boundaries overlap those of an IA are assigned a value of 1 for the IA_v indicator (treatment village), and all other villages are assigned a value of 0. Since spillovers to nearby areas would contaminate our control group, we make use of the high spatial resolution of our data to identify the spatial extent of economic spillovers to neighboring areas, and then exclude these villages from the control group.

To that end, we estimate a similar difference-in-differences specification to that presented above, but which accounts for distance to the IA semi-parametrically through the inclusion of indicator variables for 1 km distance bins, each of which is interacted with the *post* indicator:

$$y_{v,i,t} = \alpha + \sum_{j=1}^{n} \beta_j(1[dist_v \in bin_j] \times post_t) + (post_t \times X_v)\Gamma +$$
$$\delta_{i,t} + \eta_v + \varepsilon_{v,i,t}. \tag{2}$$

In Figure 2, we plot the point estimates and 95% confidence intervals of the β_j coefficients, with villages 15–20 kms from the IA as the omitted group. The figure provides important

Figure 2: Effects of IAs on Light density, Firms and Employees
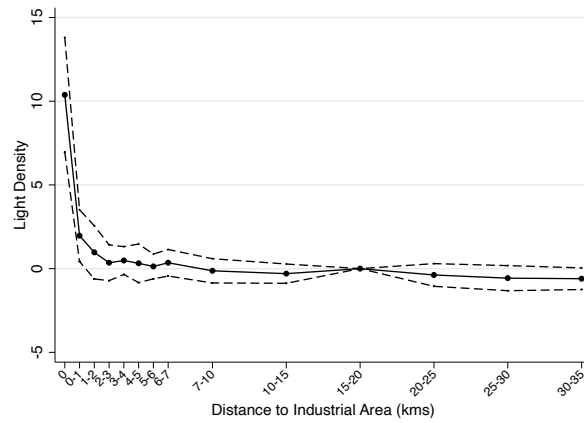
2.1: Light Density
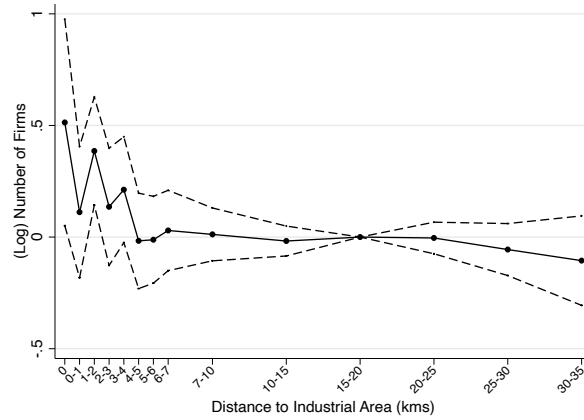
2.2: Firms
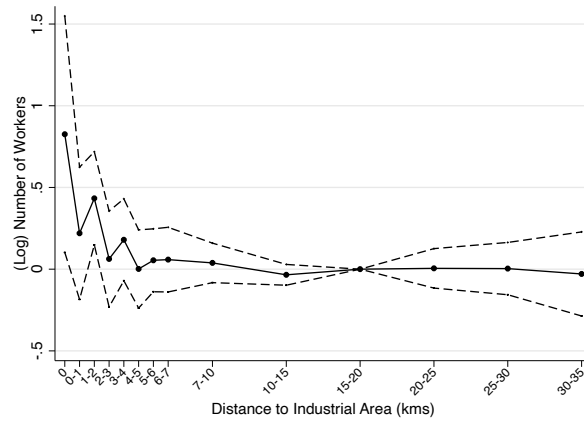
2.3: Employees

Notes: Figure 2 plots the estimated coefficients (β_j) of the distance-post interaction terms ($1[dist_v \in bin_j] \times post_t$ where j is each distance bin) from the difference-in-differences regression given in Specification (2). In Figure 2.1 the outcome is the level of light density, in Figure 2.2 the outcome variable is the (log) number of firms, and in Figure 2.3 the (log) number of employees. The x-axis measures the distance (in kms) of the village from the IA, where "0" refers to villages whose boundaries overlap those of the IA, and the omitted category is villages 15–20 kms from the IA. Dashed lines indicate 95% confidence intervals.

Source: The data obtained from demographic and economic censuses and remotely sensed night-lights, merged at the village level.

insights that guide our empirical estimation. First, there is a large and statistically significant increase in light density, as well as the (log) number of firms and workers within the IAs. We also observe that there are (monotonically declining) spillovers at distances up to 5 kms away for most of our outcome variables. In the remainder of the paper, we therefore construct the control group from villages located more than 5 kms from an IA. In the baseline specifications which are focused on impacts occurring within the IAs, we omit the "spillover group" (i.e. villages which do not intersect the IAs but are located up to 5 kms away from it) from the analysis. In other specifications, which estimate spillover impacts, we represent spillover effects in the regression model using either a single indicator, or separate indicators for villages in the distance bins of (0-1], (1-2], (2-3], (3-4], and (4-5] kms from the IA.[11] We also perform robustness tests in which we restrict the control group to villages which lie within the same sub-district, or which are located no farther than 10 kms from the IA.

As noted above, we lack information on the precise location of firm activity, which leads us to attribute some of the spillovers induced by the IA in adjacent areas to the IA itself. This means that the estimated coefficient on the indicator variable for distances of (0-1] kms will be an underestimate of the true magnitude of the immediate spillover from the IA.

3.3 Baseline Balance

Table 1 presents summary statistics for village-level baseline (early 1990s) characteristics of our sample, disaggregated by the treatment status of the village. Column (1) gives the mean level of the indicated variable in control villages; and column (2) gives the difference between treatment and control villages, estimated using a regression of the indicated variable on a dummy for treatment (intersection with the IA). Column (3) includes IA fixed effects in this regression. While the samples are mostly similar, there are statistically significant differences between treatment and control villages in terms of baseline nightlights, forest cover and the distances to cities and highways. As mentioned above, we control for time interactions of these variables in all specifications estimated in the paper. We also test the robustness of our results to alternative definitions of the control group that make them more similar to treatment villages in these observables.

Columns (4)–(6) of Table 1 tests for parallel pre-trends. Because we lack the village-level data prior to the study period, we are unable to test for trends before the establishment of our initial IAs. As an alternative, we test the parallel trends assumption for IAs established between 2001–2011, for which we are able to use data from the 1990s (i.e., 1990 for Economic

---

[11]We also perform robustness tests in which we allow for spillovers to extend beyond 5 kms, and find null effects for spillovers at distances beyond 5 kms for virtually all outcomes, and similar impacts at smaller distances to those estimated in the main spillover specifications.

Table 1: Summary Statistics

	Baseline Levels, 1990/91			Change, 1991 − 2001		
	Control Mean	Treatment − Control		Control Mean	Treatment − Control	
	(1)	(2)	(3)	(4)	(5)	(6)
Demographics						
Log Population	6.410	-0.166 (0.133)	0.058 (0.125)	0.121	0.055 (0.073)	0.069 (0.074)
Pct Population Scheduled Caste	0.193	0.052** (0.025)	0.037 (0.023)	0.001	-0.003 (0.009)	-0.007 (0.010)
Pct Male Literacy	0.487	0.011 (0.023)	-0.009 (0.016)	0.118	0.039** (0.016)	0.031* (0.018)
Pct Male Workers, Non-Agr	0.020	0.013* (0.007)	0.012 (0.007)	0.056	0.006 (0.035)	0.007 (0.034)
Pct Male Workers, Agr	0.807	0.012 (0.021)	0.012 (0.025)	-0.077	-0.029 (0.025)	-0.029 (0.023)
Infrastructure (unrelated to IA sites)						
Primary School Present	0.860	-0.027 (0.033)	0.022 (0.033)	-0.027	-0.137 (0.092)	-0.125* (0.068)
High School Present	0.395	-0.061 (0.037)	-0.007 (0.034)	0.155	-0.064 (0.045)	-0.045 (0.048)
Bus Stand Present	0.670	-0.087*** (0.030)	0.007 (0.030)	0.056	0.071 (0.067)	0.016 (0.071)
Post Office	0.317	-0.099*** (0.027)	-0.030 (0.021)	0.013	-0.008 (0.043)	-0.016 (0.045)
Telephone	0.168	-0.073*** (0.019)	-0.015 (0.016)	0.410	0.052 (0.068)	0.055 (0.073)
Economic Indicators						
Log Employment	3.573	-0.216 (0.199)	0.164 (0.136)	0.034	0.075 (0.170)	-0.036 (0.145)
Log Firms	2.965	-0.206 (0.166)	0.136 (0.101)	-0.139	0.118 (0.179)	0.107 (0.174)
Any Enterprise >99 Workers	0.011	0.000 (0.006)	0.004 (0.007)	-0.003	-0.011 (0.051)	-0.013 (0.051)
Any Enterprise 10–99 Workers	0.251	-0.004 (0.038)	0.052 (0.034)	-0.039	0.030 (0.095)	0.008 (0.092)
Land Use						
Pct Land Cultivated	0.658	0.055** (0.021)	0.046*** (0.015)	0.003	0.012 (0.023)	0.008 (0.023)
Pct Land Uncultivated	0.130	-0.003 (0.010)	-0.004 (0.007)	-0.000	-0.004 (0.023)	0.001 (0.024)
Pct Land Waste	0.115	0.018 (0.016)	-0.013 (0.012)	-0.003	-0.017 (0.027)	-0.013 (0.029)
Pct Cultivated Land Irrigated	0.191	0.017 (0.034)	0.050 (0.032)	0.066	0.084 (0.053)	0.087 (0.051)
Pct Land Forest	0.098	-0.071*** (0.019)	-0.029** (0.012)	0.001	0.009 (0.007)	0.004 (0.008)
Infrastructure (related to IA sites)						
Log Distance from City	4.008	-0.292** (0.129)	-0.297** (0.114)			
Log Distance from Highway	2.470	-1.020*** (0.211)	-0.800*** (0.183)			
Paved Road	0.645	-0.076* (0.039)	-0.026 (0.034)	0.054	-0.020 (0.067)	-0.067 (0.072)
Railroad	0.008	0.007 (0.007)	0.007 (0.006)	0.003	0.024 (0.019)	0.024 (0.020)
Tap Water	0.179	-0.016 (0.026)	0.036 (0.027)	0.346	-0.102 (0.100)	-0.109 (0.110)
Electrified	0.947	0.009 (0.022)	0.002 (0.020)	0.024	0.022 (0.051)	0.030 (0.048)
Light Density (1992)	1.787	2.233*** (0.506)	1.578*** (0.466)	2.183	0.379 (0.657)	0.262 (0.688)
IA F.E.s			Yes			Yes
Number of Treatment Villages		74			50	

Note: Column (1) gives the mean value of the indicated variable for control villages (all villages located more than 5 kms from the nearest Industrial Area (IA)). Columns (2) and (3) give estimated coefficients of the indicated variable on a treatment indicator (treatment villages are those villages whose boundaries overlap those of the IA). Estimates in Column (3) include nearest-IA fixed effects. All variables are measured in 1990 or 1991, except light density, which is measured in 1992. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Columns (4)–(6) give estimates for trends (changes) in the indicated variables between 1991–2001 (1990–1998 for Economic Census Indicators). The treatment sample is limited to villages that received an IA between 2001–2011, and the control sample to those control villages for which the nearest IA was established between 2001–2011. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Indicators and 1991 for Demographic Indicators) in order to generate pre-trends. In column (4) we report the mean of the trend in the control sample, which includes all the villages which are more than 5 kms from the nearest IA. In column (5) we present the difference in trends between control and treatment villages, estimated using a regression of the trend of the indicated variable on a dummy for IA villages, controlling for distance to 1991–2001 IAs. In column (6) we include IA fixed effects. We are unable to reject equality of trends for most variables of interest, including the principal outcome variables (light density, firms, enterprises, and labor force composition), although we note this test may have low power (Roth, 2020).

4   Results

We first report, in Section 4.1, estimates of the direct impacts of IAs on the villages in which they are located ("within IA" impacts), using specification 1. We conduct robustness tests using alternative control groups (Section 4.1.2); and use an event study framework and placebo regressions to establish that the treatment preceded the observed treatment effect (Section 4.1.3). In Section 4.2, we estimate spillover effects using specification 2.

4.1   Impact within IAs

4.1.1   Baseline Results

Our main results for the direct effects of IAs are presented in Table 2. The outcomes of interest are night-light density, the number of firms in the village, and the number of employees in these firms. In addition, because the IA policy sought to attract large firms, we also estimate the effect of IAs on the number of firms in different size categories, as measured by their number of workers (columns (4)–(6)). Due to the high incidence of zeros for several of the outcomes of interest, particularly night-time light density and the number of medium and large firms, we present the results in both levels (panel A) and logs (panel B).

The regressions indicate that IAs have been associated with large increases in the level of light density (10.35), employment in firms (375 workers per village), and the number of firms (16 firms per village, though it is imprecisely estimated). The results are similar in direction when using levels and logs, though the use of logs generally yields more precisely estimated coefficients. IAs also lead to the creation of 0.65 large firms (greater than 99 employees), 2.8 medium-sized firms (10–99 employees), and 12.793 small firms (less than 10 employees) per village. Because there are an average of 2.5 villages overlapping each IA, the magnitude of the reported coefficients must be scaled up by a factor of 2.5 to estimate the total increase in employments and firms within each IA.

14

Table 2: Effect of IAs on Outcomes

	Light Density	Employees	Firms	Firms Number of Employees: >99	10–99	<10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Levels						
within IA	10.350***	374.986***	16.348	0.650***	2.816*	12.793
	(1.727)	(133.939)	(15.293)	(0.227)	(1.534)	(13.903)
Control Mean	7.206	130.278	64.267	0.019	0.638	63.559
	(7.162)	(359.202)	(105.668)	(0.276)	(2.359)	(104.523)
R-squared	0.230	0.001	0.090	0.015	0.010	0.093
N	47914	38630	38630	38630	38630	38630
Panel B: Logs						
within IA	0.352*	0.811***	0.557**	0.321***	0.371**	0.612***
	(0.185)	(0.242)	(0.208)	(0.098)	(0.170)	(0.197)
R-squared	0.246	0.013	0.011	0.017	0.008	0.010
N	47914	37656	37298	38630	38630	38630

Note: Regression results are coefficients (β) of $treatment \times post$ interaction terms ($IA_v \times post$) from the difference-in-differences regression given in Specification (1). The outcome variables are the light density in Column (1), the number of employees in Column (2), the number of firms in Column (3), and the number of firms disaggregated by size in Columns (4)–(6). In Panel A the outcome variables are measured in levels, and in Panel B in logs. For levels, we also provide the endline control mean. For logarithmic transformations, we use $log(1 + x)$ for variables in Columns (4)–(6), and the asinh ($log(x + \sqrt{x^2 + 1})$) for light density. Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

4.1.2 Alternative Control Groups

In our main specification, the control group includes all villages that are located more than 5 kms away from the nearest IA. Such villages may be quite distant from the treated villages, raising potential concerns about differences between control and treatment villages. We therefore test the robustness of our results to the use of a more geographically proximate control group, based on villages located in the same sub-district, or located within 10 kms of the same IA. We also control for the (time-interacted) log of the distance to the nearest of the ten largest cities in the state and its square.[12] The results are given in Table 3 and are found to be similar to those obtained through our main specification, albeit somewhat smaller in magnitude in comparison to the baseline results for (log) employees and firms.

As an additional robustness check, we re-estimate our main regressions using a coarsened exact matching (CEM) algorithm (Blackwell et al., 2009). The gist of this methodology is to construct a control group which is similar in key observables to the treatment group.[13] We implement the CEM using distance to the nearest of the 10 largest cities in the state; distance to the nearest highway; fraction of land with forest cover; share of men in non-agricultural wage labor; population; and light density.[14] We present the results in Table 4, and find them to be similar to, and sometimes stronger than those reported in Table 2.

4.1.3 Timing

An important limitation of our empirical approach is that the temporal resolution of the data is generally insufficiently fine-grained to demonstrate that the creation of the IA precedes the growth in economic activity observed at the end of the study period. To examine the timing in the case of nightlight density, the only variable for which we have year-by-year data, we conduct an event study analysis that compares trends in night-time light density across control and treatment villages before and after the establishment of IAs.[15]

For this purpose, we run the following regression:

$$y_{v,i,t,z} = IA_v \left(\sum_{j=-8}^{-2} \beta_j 1[z=j] + \sum_{j=0}^{4} \beta_j 1[z=j] \right) + \tag{3}$$
$$(\mu_t \times X_v)\Gamma + \eta_v + \delta_{i,t} + \varepsilon_{v,i,t}.$$

---

[12]The ten largest cities in Karanataka are: Bengaluru, Hubballi-Dharwad, Mysuru, Kalaburagi, Mangaluru, Belagavi, Davanagere, Ballari, Vijayapura and Shivamogga.

[13]This methodology has been used in other papers, such as Ganguli (2015) and Adukia (2017).

[14]The control sample is restricted to villages more than 10 kms from the nearest IA, to ensure spillover effects not bias the control group.

[15]Henderson et al. (2012), Hodler and Raschky (2014), Michalopoulos and Papaioannou (2013), and Storeygard (2016) use night-time lights data as a proxy for economic development in contexts where income data is unavailable or of low quality; and Pinkovskiy and Sala-i Martin (2016) show that night-time light density is a robust proxy of economic activity.

Table 3: Effect of IAs on Firm Outcomes, Alternative Specifications

	Levels				Logs			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A Light Density								
within IA	11.064***	9.491***	11.484***	10.076***	0.439***	0.359**	0.503***	0.383**
	(1.698)	(1.772)	(1.709)	(1.696)	(0.087)	(0.158)	(0.083)	(0.166)
R-squared	0.154	0.267	0.152	0.264	0.241	0.269	0.239	0.271
N	47914	10378	47914	10378	47914	10378	47914	10378
Panel B: Employees								
within IA	338.453**	300.379**	353.583***	320.874**	0.550*	0.457	0.592*	0.556*
	(130.292)	(132.610)	(129.798)	(129.250)	(0.307)	(0.326)	(0.303)	(0.319)
R-squared	0.000	0.005	0.000	0.004	0.012	0.025	0.012	0.022
N	38626	8142	38626	8142	38626	8142	38626	8142
Panel C: Firms								
within IA	13.862	4.900	13.658	8.160	0.352*	0.281	0.375*	0.352*
	(14.114)	(14.960)	(13.994)	(14.732)	(0.208)	(0.219)	(0.206)	(0.206)
R-squared	0.078	0.104	0.078	0.103	0.008	0.022	0.008	0.020
N	38626	8142	38626	8142	38626	8142	38626	8142
Sub-District X Year F.E.s	Yes		Yes		Yes		Yes	
IA X Year F.E.s		Yes	Yes	Yes		Yes	Yes	Yes
<10 kms from IA		Yes		Yes		Yes		Yes
Quadratic Distance Control			Yes	Yes			Yes	Yes

Note: Regression results are coefficients (β) of $treatment \times post$ interaction terms ($IA_v \times post$) from difference-in-differences regression given in Specification (1). Columns (1)-(4) take the outcome variable in levels, and Columns (5)-(8) in logs. For logarithmic transformations, we use $log(x)$ for employees and firms, and the asihnh ($log(x + \sqrt{x^2 + 1})$) for light density. Control villages are those located more than 5 kms from the nearest IA. The regressions include village and time fixed effects and a vector of time-interactions with baseline controls. Estimations in Columns (1), (3), (5), and (7) also include time-interacted sub-district fixed effects. Estimations in Columns (2), (4), (6), and (8) restrict the sample of control villages to within 5-10 kms of an IA, and include time interaction with the nearest-IA fixed effects. Estimations in Columns (3), (4), (7) and (8) include the log of distance to the nearest of the 10 largest cities in the state of Karnataka, and its square. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 4: Coarsened Exact Matching

	Levels			Logs		
	Light Density	Employees	Firms	Light Density	Employees	Firms
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline CEM						
within IA	13.173***	509.319***	37.632**	0.684***	1.132***	0.894***
	(1.755)	(185.369)	(16.241)	(0.116)	(0.323)	(0.239)
R-squared	0.750	0.171	0.298	0.896	0.240	0.296
N	822	734	734	822	734	734
Panel B: Expanded CEM						
within IA	14.776***	504.524***	38.023**	0.857***	1.118***	0.892***
	(2.047)	(186.555)	(16.827)	(0.113)	(0.325)	(0.240)
R-squared	0.756	0.177	0.285	0.904	0.261	0.324
N	520	662	648	520	662	648

Note: Regression results are coefficients (β) of $treatment \times post$ interaction terms ($IA_v \times post$) from difference-in-differences regression given in Specification (1). The outcome variables are the light density, the number of employees, and the number of firms. In Columns (1)-(3) the outcome variables are measured in levels, and in Columns (4)-(6) in logs. All logarithmic transformations are based on $log(x)$, except for light density which is transformed using the asinh ($log(x + \sqrt{x^2 + 1})$). All specifications use the coarsened exact matching (CEM) method, and use all villages within the state to determine the best match for control villages. In Panel A, the comparison villages are selected based on the following variables measured at baseline: distance to the nearest of the 10 largest cities in the state; distance to the nearest highway; fraction of land with forest cover; share of men in non-agricultural wage labor; population; and light density. In Panel B, we additionally include the baseline measure of the outcome variable (in levels) to the vector of matching variables. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

where $y_{v,i,t,z}$ is the night-time light density in village v at year t, which occurs z years after the establishment of the nearest IA i. IA_v is a dummy indicating that a village overlaps the boundaries of the IA and μ_t are year fixed effects, which are interacted with the vector of baseline controls (X_v). The rest of the terms in the regression are as in our baseline specification.

Figure 3: Event study using Light Density: Within IA
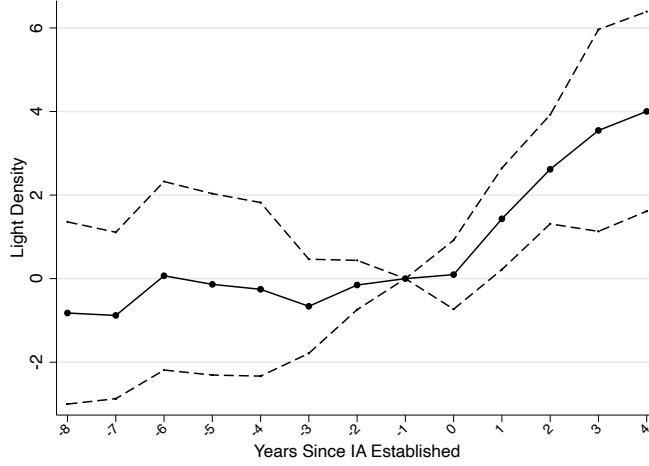
Notes: Figure 3 plots estimated coefficients (β_j) for the event study regression specified in Specification 3. The coefficients capture the treatment effect of within IA on nightlight density at various years, relative to the date of IA establishment, specified in the horizontal axis, with control villages located more than 5 kms from the IA. The regression includes village and year fixed effects, as well as interactions of baseline control variables with time. Dashed lines indicate 95% confidence intervals, accounting for clustering of errors at the nearest-IA level; dashed lines indicate 95% confidence intervals.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

The results, plotted in Figure 3, indicate that night lights display similar trends across control and treatment villages prior to the establishment of the IA, but then begin to diverge rapidly within 1–2 years of the IA's establishment. The patterns of light-density growth clearly demonstrate that the creation of IAs precedes the take-off of local growth.[16]

As an alternative strategy for testing whether IA creation precedes the increase in economic activity, we use additional rounds of data to conduct falsification tests, in which we assign treatment status to villages prior to the actual establishment of their IA and test whether these placebo treatments yield significant impacts. Specifically, we (separately) test whether treatment effects were evident in 2005 for IAs that were established between 2005–2011, and whether treatment effects were evident in 2013 for IAs that were established after

---

[16]In the results not shown here, we find that the result is robust to a specification which excludes villages which have 0-nightlights eight years prior to the establishment of the nearest IA.

2012. For these regressions, we additionally include intermediate rounds (1998 and 2005) of the Economic Census from these year). For example, in the placebo regressions for IAs created between 2005–2011, we use the 1990, 1998, and 2005 rounds of the Economic Census. The treatment variable takes a value of 0 in 1990 and 1998, and a value of 1 in 2005 for villages receiving an IA between 2005–2011.

The results are given in Table 5, and provide no indication of impacts that appear prior to IA establishment. While it is important to note that this test will have low power due to the small number of IAs created during these short time intervals, the statistical insignificance and relatively small magnitudes of the estimates argue against the possibility that our results are driven by differential pre-trends.

4.2   Spillovers

In this section, we present the results for the spillover effects on villages which do not overlap the IAs but are located in its vicinity. To do so, we estimate specification 2 for our three principal outcome variables and present our results in Table 6 and 7.

4.2.1   Nightlights

Table 6 presents the results for night-light density. In column (1) the outcome variable is light density, measured in levels, and in column (2) using the asinh logarithmic transformation (there is a large number of observations taking a value of 0). There is a statistically significant increase in light density within the IAs, with a level increase of 10.4, and an asinh increase of 35%. This increase in light density extends out for several kms from the IAs when measured in levels; but when measured in logs shows more ambiguous effects. In columns (3)–(5) we limit the sample to observations that lacked light at the baseline, and take as the outcomes the same two measures as before, as well as an indicator taking a value of 1 for any light. In columns (6) and (7) we limit the sample to villages that had light at baseline, and measure the outcome in levels and asinh. The treatment effects were similar according to baseline light density, but are measured more precisely for villages lacking light at baseline.

4.2.2   Employees and Firms

In Table 7, we estimate spillover effects on the numbers of employees and firms in columns (1) and (2), respectively. The estimated impacts within the IA are similar to those estimated in Section 4.1. In addition, we find that there are significant spillover effects on both variables up to a distance of 4 kms from the IA. The spillovers are smaller in magnitude than the direct effect within the IA, and are declining with distance from the IA.

Table 5: Falsification Tests

	Light Density				Log Employees				Log Firms			
	Year of IA				Year of IA				Year of IA			
	2005-11		2012-15		2005-11		2012-15		2005-11		2012-15	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Within IA	-0.594	-0.239	0.898	0.914	0.087	0.044	0.104	0.142	0.009	-0.079	-0.010	0.068
	(0.610)	(0.621)	(0.690)	(0.937)	(0.232)	(0.270)	(0.139)	(0.154)	(0.196)	(0.221)	(0.111)	(0.075)
R-squared	0.108	0.131	0.263	0.251	0.034	0.048	0.035	0.045	0.022	0.036	0.019	0.024
N	71871	12609	95828	15516	64557	11099	87020	13957	64557	11099	87020	13957
IA X Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
≤10 kms from IA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Regression results are coefficients (β) of the *treatment* × *post* interaction terms ($IA_v \times post$) from the difference-in-differences regression given in Specification (1). The sample is limited to the years 1990–2005 for Columns (1)–(2), (5)–(6), and (9)–(10); and the *treatment* variable takes a value of 1 for villages in which an IA was established from 2006–2011. The full sample is for Columns (3)–(4), (7)–(8), and (11)–(12); and the *treatment* variable takes a value of 1 for villages in which an IA was established from 2012–2015. The outcome variables are light density in Columns (1)–(4), log number of employees in Columns (5)–(8) and log number of firms in Columns (9)–(12). Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. Controls are also included for the distance-post interaction terms ($\beta_{1,j}(1[dist \in bin_j] \times post_t)$) for IAS established during the respective study periods. Village fixed effects are included, as well as nearest-IA fixed effects interacted with time dummies. The nearest-IA is defined as the nearest IA established in the year indicated at the column head. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 6: Effect of IAs on Night Lights, Spillovers

	Full Sample		0-Light			> 0 Light	
	Level	Asinh	Any (0/1)	Level	Asinh	Level	Asinh
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
within IA	10.382***	0.352*	0.057	10.292***	1.039***	10.503***	0.482***
	(1.736)	(0.186)	(0.078)	(3.476)	(0.202)	(1.810)	(0.098)
0–1 kms	1.959***	0.044	0.103	1.692***	0.472**	1.900**	0.085
	(0.676)	(0.116)	(0.082)	(0.348)	(0.201)	(0.839)	(0.057)
1–2 kms	0.980	-0.024	0.058	1.015***	0.299***	0.979	0.017
	(0.636)	(0.083)	(0.041)	(0.170)	(0.094)	(0.820)	(0.053)
2–3 kms	0.415	-0.009	0.002	0.709**	0.120	0.208	-0.012
	(0.459)	(0.049)	(0.020)	(0.299)	(0.075)	(0.638)	(0.045)
3–4 kms	0.580*	0.062	0.053**	0.309*	0.137***	0.586	0.037
	(0.324)	(0.042)	(0.024)	(0.163)	(0.048)	(0.502)	(0.037)
4–5 kms	0.421	0.019	0.055**	0.243**	0.137***	0.488	-0.012
	(0.450)	(0.047)	(0.021)	(0.119)	(0.036)	(0.610)	(0.031)
R-squared	0.230	0.246	0.131	0.183	0.198	0.249	0.173
N	47914	47914	21856	21856	21856	26058	26058

Note: Regression results are coefficients (β_j) of the *distance* × *post* interaction terms ($1[dist_v \in bin_j] \times post_t$) in the difference-in-differences regression given in Specification (2). The outcome variable is light density, measured in levels in Columns (1), (4), and (6); in logs using logarithmic asinh transformation ($log(x + \sqrt{x^2 + 1})$) in Columns (2), (5), and (7); and as a binary indicator for access in Column (3). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1. Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

In columns (3)–(5) we estimate the spillover effects on the number of firms in different size categories. Within villages overlapping the IA ("within IA" effects), the main impact is on the creation of large firms (more than 99 employees). In contrast, outside the IAs growth is concentrated in small firms (less than 10 employees). In fact, in results not shown, we find that most of the new firms in the latter size category have only 1 or 2 employees. Table 8 shows the same patterns hold for employment: most employment growth occurs in large firms within the IA, and in small firms outside the IA.

The clear difference in the pattern of firm and employment creation within the IA (direct effect) and in its vicinity (spillovers) reflect both the success of the IA program in achieving its stated intention to attract large firms to rural areas, and the limits on its ability to trigger similar changes outside of the IA itself. This difference could be driven by a number of potential factors: agglomeration effects may be too weak; firms may be credit constrained;

Table 7: Effect of IAs on Firm Outcomes, Spillovers

			Firms		
			Number of Employees:		
	Employees	Firms	>99	10–99	<10
	(1)	(2)	(3)	(4)	(5)
Panel A: Levels					
within IA	376.702***	16.664	0.651***	2.826*	13.098
	(134.099)	(15.311)	(0.227)	(1.537)	(13.916)
0–1 kms	54.280	12.492	-0.025	0.376	12.100
	(45.183)	(8.329)	(0.041)	(0.349)	(8.298)
1–2 kms	62.884*	14.883**	-0.022	0.173	14.748**
	(35.829)	(7.184)	(0.019)	(0.268)	(7.197)
2–3 kms	0.716	5.422	-0.047**	-0.055	5.539
	(17.485)	(4.786)	(0.018)	(0.241)	(4.817)
3–4 kms	11.982	2.549	-0.022	0.051	2.505
	(14.829)	(5.088)	(0.021)	(0.131)	(5.106)
4–5 kms	6.378	4.981	-0.033	0.177	4.846
	(11.538)	(4.549)	(0.024)	(0.149)	(4.513)
R-squared	0.001	0.090	0.015	0.010	0.093
N	38630	38630	38630	38630	38630
Panel B: Logs					
within IA	0.978***	0.652***	0.250***	0.312**	0.564***
	(0.352)	(0.206)	(0.076)	(0.130)	(0.184)
0–1 kms	0.340*	0.234*	-0.008	0.155**	0.213*
	(0.187)	(0.122)	(0.020)	(0.071)	(0.110)
1–2 kms	0.545***	0.498***	-0.006	0.066	0.450***
	(0.142)	(0.118)	(0.011)	(0.055)	(0.109)
2–3 kms	0.162	0.245**	-0.024**	0.010	0.228**
	(0.130)	(0.113)	(0.009)	(0.068)	(0.102)
3–4 kms	0.257**	0.296***	-0.010	0.015	0.268**
	(0.120)	(0.109)	(0.012)	(0.046)	(0.102)
4–5 kms	0.072	0.066	-0.013	0.037	0.059
	(0.108)	(0.092)	(0.010)	(0.044)	(0.082)
R-squared	0.018	0.010	0.017	0.006	0.009
N	38630	38630	38630	36350	38630

Note: Regression results are coefficients (β_j) of the *distance* × *post* interaction terms ($1[dist_v \in bin_j] \times post_t$) in the difference-in-differences regression given in Specification (2). The outcome variables are number of employees in Column (1), the number of firms in Column (2) and the number of firms disaggregated by size in Columns (3)–(5) (measured in levels in Panel A and logs in Panel B). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 8: Effect of IAs on Number of Workers by Firm Size

	Levels			Logs		
	Firm Size			Firm Size		
	>99	10–99	<10	>99	10–99	<10
	(1)	(2)	(3)	(4)	(5)	(6)
within IA	241.736***	85.975*	52.324	1.215***	0.679*	0.641***
	(77.596)	(44.962)	(32.517)	(0.332)	(0.348)	(0.216)
Spillovers	8.666	-0.790	11.835	-0.073	0.092	0.239**
	(10.044)	(3.733)	(8.518)	(0.048)	(0.106)	(0.098)
Control Mean	8.688	11.862	111.300			
	(277.188)	(52.667)	(185.112)			
R-squared	0.000	0.013	0.038	0.013	0.006	0.012
N	38630	38630	38630	38630	38630	38630

Note: Regression results are coefficients (β_j) of the *distance* × *post* interaction terms ($1[dist_v \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2$ corresponds to the distance bin (0-4] and denoted as "spillover"), in the difference-in-differences regression given in Specification (2). The outcome variable is the number of workers by firm size in Column (1)–(3) and the log number of workers by firm size in Columns (4)–(5). For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** $p<0.01$, ** $p<0.05$, and * $p<0.1$.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

the returns to scale in the provision of goods and services to firms within the IAs could be low; and land use restrictions which still hold outside the IA might continue to impede large firm creation.[17] However, the available data does not allow us to determine the relative importance of these potential factors.

4.2.3 Labor force composition

We next explore the effects of the IA on the composition of the labor force in villages overlapping and surrounding the IA. This analysis makes use of data from the Demographic Census, which gives the fraction of workers involved in different categories of activity. We focus on non-agricultural employment, which may include both salaried and informal (usually

---

[17]With respect to the latter possibility, it may be significant that land use regulations do not apply to smaller firms, which are permitted to operate within homes and other small buildings, and are not required under the KTCPA to secure alterations of land zoning as do larger enterprises.

daily wage labor) employment of different types, as the key outcome variable.[18]

Table 15 gives the effects of the IAs on the share of workers employed in non-agricultural sectors (as defined above) disaggregated by gender. The size of the non-agricultural labor force is reported in three ways: as a share of working adults (columns 1 and 4), in levels (columns 2 and 5) and in logs (columns 3 and 6). The establishment of IAs results in a 12 percentage point increase in the share of male non-agricultural employment. There are also substantial spillover effects, which start at 4.8 percentage points in villages less than 1 km from the IA, and decline monotonically up to a distance of 5 kms away from the IA. Figure 4 presents these findings graphically. The results for the log of the non-agricultural labor force paint a similar picture; but, when they are measured in levels, are imprecisely estimated.

Figure 4: Effects of IAs on Workers
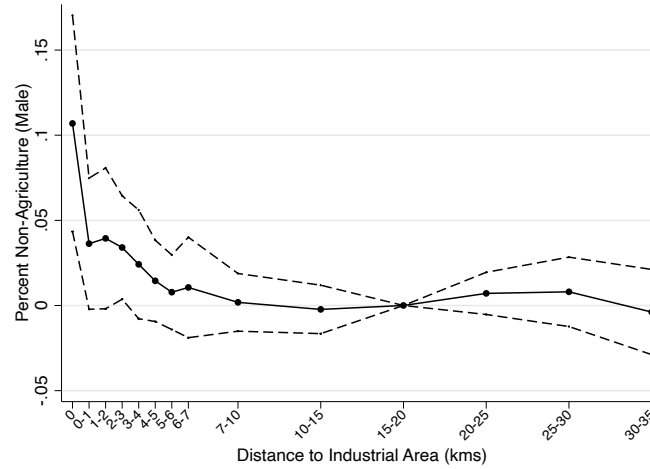
Notes: Figure 4 plots the coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$ where j is each distance bin) from the difference-in-differences regression given in Specification (2). The outcome variable is the percent of male workers in non-agricultural wage labor. The x-axis measures the distance (in kms) of the village from the IA, where "0" refers to villages whose boundaries overlap those of the IA, and the omitted category is villages 15–20 kms from the IA. Dashed lines indicate the 95% confidence interval.
Source: The data obtained from demographic and economic censuses and remotely sensed night-lights, merged at the village level.

Female non-agricultural employment increases by a similar magnitude in villages over-lapping the IA (12 percentage points). However, due the lower labor force participation of women, the level increase is roughly 23 female workers, which is approximately a third of the

---

[18]A finer disaggregation of occupational activity is made difficult by the fact that the 1991 and 2011 censuses report different sets of categories of occupation. While both include agricultural labor and own-farm cultivation, as well as household-based business, the 2011 census aggregates all other non-agricultural activities outside the household into a single category, while the 1991 census disaggregates it into finer categories, including: livestock, forestry, and fishing; mining and quarrying; manufacturing and processing; construction; trade and commerce; transportation, store, and communication; and other.

increase estimated for male workers. Up to 1 km away, the effect for women remains similar to that for males, albeit imprecisely estimated. Farther away, however, spillover effects are substantially weaker.

Interestingly, in results not shown, we find that the increase in non-agriculture employment of male workers is driven primarily by a reduction in the share of workers cultivating land they own or rent, with only a small decline in the share of individuals working as agricultural wage laborers on others' land, who tend to be much weaker socio-economically. The own-farm cultivators who are shifting to non-agricultural employment are likely to be small holder farmers, for whom agricultural incomes tend to be low and precarious (Blakeslee et al., 2020).

Changes in the labor force composition could be brought about either by shifts in the occupations of the baseline village labor force, or by a change in the composition of the labor force, perhaps through an increase in the labor force participation rate within the village or by selective migration into villages surrounding the IAs. To partially distinguish between these possibilities, we examine whether IAs have any impact on the size of the population, the fraction belonging to scheduled castes, the literacy rate, and the labor force participation rate. We focus on the male population, as, in India, men are far more likely to migrate in response to employment opportunities. As seen in Table 9, we find no evidence for any such impacts, which suggests the results are driven primarily by shifts in occupations within the village.

4.2.4 Additional Specifications

Our primary estimation strategy consists of a long-differences regression using the years 1990 and 2013 (1991 and 2011 for demographic census variables) as the starting and ending points. This approach is motivated by an attempt to capture the long-term effects of the IA program. As a robustness test, we conduct an additional analysis which also uses additional, intermediate rounds of the Economic and Demographic Censuses.

We use data from census rounds in the years 1990, 1998, 2005, and 2013 for outcomes observed in the Economic Census; and data from 1991, 2001 and 2011 for outcomes observed in the Demographic Census. Treatment is assigned the value of 1 for every round of data occurring after the establishment of the IA. The results are given in Table 10, and are consistent with out main findings.

We also estimate specifications that allow for spillovers that extend beyond 5 kms from the IA. For this analysis, we include indicator variables for 1 km distance intervals extending up to 10 kms from the IA and construct the control group from villages located more than 10 kms from the IA. The results are given in Table 11. There is little evidence that treatment effects extend beyond 5 kms, and the coefficients on nearer distances are largely unaffected

26

Table 9: Effect of IAs on Demography and Labor Force Participation

		Male			
				Labor Force Participation Rate	
	Log Population	Literacy Rate	Percent SC	Percent: Full Time	Part-time
	(1)	(2)	(3)	(4)	(5)
within IA	0.034	0.009	-0.004	-0.008	0.001
	(0.083)	(0.013)	(0.009)	(0.016)	(0.001)
0–1 kms	0.021	0.008	0.009	-0.020	-0.000
	(0.046)	(0.012)	(0.007)	(0.013)	(0.001)
1–2 kms	0.027	0.019	0.013*	-0.005	0.001*
	(0.036)	(0.012)	(0.007)	(0.009)	(0.001)
2–3 kms	-0.031	0.007	0.006	0.003	0.001***
	(0.026)	(0.008)	(0.004)	(0.006)	(0.001)
3–4 kms	0.053**	0.013*	-0.001	-0.003	-0.000
	(0.025)	(0.007)	(0.005)	(0.009)	(0.000)
4–5 kms	0.018	0.007	0.006	0.005	0.000
	(0.027)	(0.008)	(0.006)	(0.007)	(0.001)
Control Mean		0.691	0.206	0.549	0.070
		(0.114)	(0.204)	(0.123)	(0.099)
R-squared	0.053	0.037	0.041	0.006	0.005
N	43130	43130	43130	43130	43130

Note: Regression results are coefficients (β_j) of the distance-post inter-action terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (2). The outcome variables are (male) log adult population in Column (1), literacy rate in Column (2), share of the adult population that is SC in Column (3), and labor force participation rate in Columns (4)–(5). For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 10: Panel Specification

	Light Density	Log Employees	Log Firms	Pct Non-Agr Labor
	(1)	(2)	(3)	(4)
Within IA	8.319***	0.607***	0.276**	0.100***
	(1.427)	(0.205)	(0.120)	(0.021)
0–1 kms	1.694***	0.150*	0.032	0.049**
	(0.554)	(0.087)	(0.074)	(0.018)
1–2 kms	0.359	0.258***	0.240***	0.048***
	(0.345)	(0.082)	(0.063)	(0.014)
2–3 kms	0.177	-0.030	0.010	0.027**
	(0.257)	(0.065)	(0.053)	(0.011)
3–4 kms	0.402***	0.127**	0.142***	0.027***
	(0.145)	(0.057)	(0.049)	(0.009)
4–5 kms	0.413*	-0.011	0.017	0.019**
	(0.223)	(0.064)	(0.058)	(0.008)
R-squared	0.231	0.033	0.018	0.026
N	95828	87129	87129	68938

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (2). The sample includes the 4 rounds of the Economic Census (1990, 1998, 2005, and 2013). The outcome variables are the light density in Column (1), the log number of employees in Column (2), the log number of firms in Column (3), and the share of male workers in non-agricultural wage labor in Column (4). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

by this change in the control group.

4.3   Heterogeneities

In order to shed light on the mechanisms driving the results, we next turn to a heterogeneity analysis based on firm and village characteristics. A substantial literature posits that spillovers of the type observed above may be driven by either agglomeration economies or demand-side factors. Examining the *types* of firms arising in areas outside the IA may help distinguish between these. Heterogeneities according to village characteristics—such as the education of the work force, proximity to input and output markets, and the productivity of agricultural activities—may also help in understanding the mechanisms driving spillovers.

In addition, there may be important heterogeneities in the socio-economic status of groups benefiting from the establishment of IAs. For example, persistent inequities in India based on caste and gender may affect the ability or desire of these groups to take advantage of the new economic opportunities created by IAs. We explore these questions in detail in the following sections.

4.3.1   Firm characteristics

Table 12 reports the estimated impacts of IAs on the number of firms disaggregated by both size (panels A, B and C for firms of more than 99, 10-99 and less than 10 workers, respectively) and sector. The outcome variables are the number of firms in manufacturing, agriculture, retail, restaurants, transport, banking, construction, and storage, respectively. For brevity, we aggregate impacts in all distance bins up to 4 kms away from the IA into a single spillover category, and focus on impacts on the logarithm of the number of firms.

The results indicate that the IAs led to the creation of large manufacturing firms within the IA, and small service sector (retail and restaurants) and agricultural firms both within and around the IAs.[19] It is worth reiterating that the within-IA effects will conflate economic activity occurring within the IA zone and that occurring in villages whose boundaries overlap those of the IA. This could imply that the within-IA increase in service sector firms may reflect activity taking place in close proximity to, but outside, the IAs. There is no evidence of an increase in large firms other than manufacturing, either within IAs or in the vicinity.

The patterns of the results lead us to cautiously hypothesize that the observed spillovers are mostly driven by demand-side factors, rather than agglomeration effects. The creation of small retail and restaurant firms could be explained by an increase in demand for such services from the workforce employed within the IAs. The creation of small agricultural firms—which provide inputs and services for crop and animal production—is more ambiguous, but could

---

[19]Impacts on medium size firms are somewhat similar to those on small firms, but are imprecisely estimated

Table 11: Expanded Spillovers

	Light Density	Log Employees	Log Firms	Pct Non-Agr Labor
	(1)	(2)	(3)	(4)
within IA	11.054***	1.084***	0.729***	0.120***
	(2.048)	(0.387)	(0.226)	(0.034)
0–1 kms	2.369***	0.409**	0.270*	0.049**
	(0.822)	(0.202)	(0.137)	(0.021)
1–2 kms	0.874	0.622***	0.554***	0.039
	(0.652)	(0.171)	(0.141)	(0.023)
2–3 kms	0.428	0.249*	0.306**	0.037**
	(0.459)	(0.144)	(0.132)	(0.015)
3–4 kms	0.697*	0.282*	0.320**	0.021
	(0.364)	(0.150)	(0.132)	(0.014)
4–5 kms	0.565	0.110	0.089	0.014
	(0.534)	(0.126)	(0.110)	(0.012)
5–6 kms	0.411	0.168	0.099	0.006
	(0.424)	(0.109)	(0.099)	(0.013)
6–7 kms	0.687	0.155	0.130	0.006
	(0.436)	(0.113)	(0.092)	(0.013)
7–8 kms	0.441	0.129	0.127	0.006
	(0.405)	(0.121)	(0.101)	(0.012)
8–9 kms	0.097	0.144	0.118	0.004
	(0.418)	(0.102)	(0.089)	(0.010)
9–10 kms	0.151	0.073	0.062	-0.011
	(0.293)	(0.090)	(0.075)	(0.013)
R-squared	0.224	0.016	0.009	0.091
N	42180	36350	36350	43122

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2, \ldots, j = 9]$ correspond to the distance bins (1-2], ..., (9-10] and denoted as "spillover"), in the difference-in-differences regression given in Specification (2). The outcome variables are the light density in Column (1), the log number of employees in Column (2), the log number of firms in Column (3), and the share of male workers in non-agricultural wage labor in Column (4). Control villages are those located more than 10 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 12: Effect of IAs by Firm Sector and Size

	Sector							
	Manu	Agr	Retail	Restaur	Transp	Bank	Constr	Storage
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)
Panel A: >99 Employees								
Within IA	0.163**	0.013	0.006	-0.000	0.001	-0.000	0.013	-0.000
	(0.069)	(0.013)	(0.005)	(0.000)	(0.000)	(0.000)	(0.014)	(0.000)
Spillovers	-0.010**	-0.000	-0.002	-0.000	0.001	-0.000	0.001	-0.000
	(0.005)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
R-squared	0.028	0.001	0.001	0.001	0.007	0.002	0.003	0.000
N	38970	38970	38970	38970	38970	38970	38970	38970
Panel B: 10–99 Employees								
Within IA	0.164	0.103	0.026	-0.019	0.026*	0.013	0.024	-0.002
	(0.110)	(0.067)	(0.035)	(0.017)	(0.015)	(0.025)	(0.032)	(0.013)
Spillovers	0.017	0.010	0.004	0.003*	0.001	-0.003	-0.007	0.000
	(0.031)	(0.018)	(0.007)	(0.002)	(0.002)	(0.004)	(0.012)	(0.004)
R-squared	0.047	0.004	0.022	0.016	0.033	0.002	0.002	0.017
N	38970	38970	38970	38970	38970	38970	38970	38970
Panel C: <10 Employees								
Within IA	0.223	0.461	0.440**	0.227***	-0.039	-0.052	0.052	-0.047
	(0.193)	(0.346)	(0.168)	(0.048)	(0.064)	(0.068)	(0.134)	(0.068)
Spillovers	0.076	0.372**	0.146***	0.081***	-0.026	0.003	0.024	-0.013
	(0.052)	(0.159)	(0.043)	(0.029)	(0.018)	(0.016)	(0.027)	(0.016)
R-squared	0.057	0.028	0.558	0.444	0.010	0.097	0.000	0.053
N	38970	38970	38970	38970	38970	38970	38970	38970

Note: Regression results are coefficients (β_j) of the *distance* $\times$ *post* interaction terms ($1[dist_v \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2$ corresponds to the distance bin (0-4] and denoted as "spillover"), in the difference-in-differences regression given in Specification (2). The outcome variables are the log of the number of firms in different size categories (Panel A, B and C) and sectors (Columns (1)–(9)). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

potentially reflect the relaxation of local credit constraints. Specifically, the greater incomes being earned in IAs may allow farmers to procure rather than produce key inputs and services, or to themselves establish firms providing these outputs. The data does not enable us to test these hypotheses conclusively.

To make further progress in understanding the mechanisms driving these spillovers, we next look at the mediating effect of characteristics of the local economy.

4.3.2   Village characteristics

We next explore associations between baseline village characteristics and the magnitude of the effects of the IAs on firms and workers in the local economy. We focus on five village-level factors that are all measured in 1991: literacy (share of literate population); the share of scheduled castes (SC) in the population; irrigation (share of cultivated area which is irrigated); proximity to a large city (log distance to the nearest of the 10 largest cities in the state), and proximity to a highway (log distance). We add interactions of these variables with the *post* variable in each distance bin ($1[dist \in bin_j] \times post_t \times var_i$) to the regression, as well as $post_t \times var_i$ to account for account for growth patterns in control villages. All interaction terms are estimated jointly in a single regression.[20]

The results are given in Table 13. In columns (1)–(2) the outcome is the log number of workers, in columns (3)–(4) the log number of firms, and in columns (5)–(6) the share of male workers in non-agricultural employment.

First, we find that impacts on firm creation and employment are higher within IAs that are located closer to highways or to cities, perhaps due to proximity to input and output markets (Donaldson and Hornbeck, 2016). Second, irrigation plays an important role in the size of the spillover effects of IAs: where the share of irrigated land is higher, firm and employment growth are lower, and fewer workers shift to non-agricultural employment. This may reflect the reduced wage premium for non-agricultural employment in more agriculturally productive locations, which can attenuate the shift to non-agricultural employment, whether in IA-based firms or as local entrepreneurs (Blakeslee et al., 2020). Third, higher literacy rates are somewhat associated with increased firm creation in spillover villages, consistent with models stressing the correlation of human capital and entrepreneurship (Lucas Jr, 1978; Moretti, 2004). Finally, we find that there is a greater shift to non-agricultural employment in villages with higher SC population. While it is difficult to determine the precise reason for this striking result, we cautiously note that SCs typically have lower earnings, and therefore would face a higher wage premium from IA-based manufacturing employment. In addition, it may be the case that SCs have limited opportunities for income growth through more

---

[20]Because each of these variables might be correlated with the population size, we include interactions of the latter (logged) with the treatment variables and time dummies (i.e. post $\times$ distance bin $\times$ log(population)).

Table 13: Heterogeneous Effects by Village Characteristics

	Log Employees		Log Firms		Pct Non-Agr Labor	
	(1)	(2)	(3)	(4)	(5)	(6)
Within IA X						
Literacy Rate	1.836	0.855	-0.553	-0.899	-0.027	-0.028
	(2.438)	(2.047)	(1.322)	(1.333)	(0.206)	(0.214)
Pct SC	0.771	1.499	-0.300	-0.046	-0.083	-0.066
	(1.173)	(1.252)	(1.013)	(1.099)	(0.129)	(0.138)
Pct Land Irrigated	-2.829	-2.131	-1.009	-0.763	-0.059	-0.022
	(2.016)	(2.307)	(1.396)	(1.509)	(0.182)	(0.181)
Ln(Distance City)	-0.478	-0.831**	-0.403	-0.525*	-0.066	-0.075
	(0.385)	(0.338)	(0.276)	(0.287)	(0.060)	(0.053)
Ln(Distance Highway)		-0.705***		-0.249		-0.021
		(0.246)		(0.175)		(0.030)
Spillovers X						
Literacy Rate	0.498	0.452	0.748*	0.720	-0.002	-0.004
	(0.405)	(0.475)	(0.396)	(0.433)	(0.053)	(0.054)
Pct SC	0.385	0.511**	0.102	0.182	0.072**	0.076**
	(0.241)	(0.231)	(0.211)	(0.196)	(0.033)	(0.033)
Pct Land Irrigated	-0.682*	-0.603*	-0.484*	-0.433	-0.070**	-0.066*
	(0.369)	(0.325)	(0.284)	(0.261)	(0.033)	(0.034)
Ln(Distance City)	0.056	0.094	0.040	0.064	-0.017**	-0.016**
	(0.069)	(0.090)	(0.073)	(0.086)	(0.006)	(0.007)
Ln(Distance Highway)		-0.139		-0.088		-0.006
		(0.090)		(0.076)		(0.007)
R-squared	0.019	0.020	0.011	0.011	0.037	0.037
N	38446	38446	38446	38446	42946	42946

Note: Regression results are coefficients (β_j) of the interaction of the indicated variables in rows (measured at baseline) with the the $distance \times post$ interaction terms ($1[dist_v \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2$ corresponds to the distance bin (0-4] and denoted as "spillover"), in the difference-in-differences regression given in Specification (2). Each column corresponds to a single regression where all indicated variables are simultaneously estimated. The outcome variables are the log of the number of employees in Columns (1)–(2), the log number of firms in Columns (3)–(4), and the percentage of non-agricultural labor in the labor force in Columns (5)–(6). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1. Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

traditional pathways, and face fewer such obstacles in the modern manufacturing sector.

It is important to note, of course, that none of these associations can be interpreted as causal, given the lack of exogenous variation in the variables of interest.

4.3.3 Social Ownership

We next examine whether the economic effects documented in this paper have been socially inclusive. In India, many state programs include explicit policies to encourage the participation of minority groups and vulnerable populations, lest existing social exclusions be perpetuated in the program's implementation. Because the IA program lacked any such targeting for marginalized groups, it is interesting to know whether, and to what extent, members of these communities benefited from it. We therefore examine the effect of the IA program on the creation of firms owned by two particularly salient marginalized communities: women and Scheduled Castes (SCs). The results are given in Table 14.

Panel A shows the results for female-owned firms. In columns (1) and (2), the outcome variables are the number of firms owned by women and the number of employees working in such firms, respectively; in columns (3) and (4) the outcomes are similar, but measured in logs; and in columns (5) and (6), the outcomes are reported as shares of total firms and employment (i.e. the variables are divided by total firms and employment, respectively). We find that there is a substantial increase in the (log) number of female-owned firms and employment in such firms, both within IAs (38% and 53%, respectively) and in spillover villages (26% for both variables).[21] In addition, the shares of female-owned firms and employment in spillover villages increases by about 2.5 percentage points. However, the results are imprecise when measured in levels.

Panel B shows similar effects on SC-owned firms and employment. SC firm ownership increases by 55% within the IA, and by 26% in spillover villages; and the number of employees in such firms increases by 69% within the IA, and by 30% in spillover villages. There is also a statistically significant increase of about 2 percentage points in the share of firms owned by SCs and employment therein in spillover villages.

5   Discussion and Conclusion

Our findings indicate that the IA program has been effective in attracting firms to rural locations, and that this resulted in changes in the local and surrounding economy. It is important to note, however, that this does not necessarily indicate that the policy resulted in

---

[21]The 1990 economic census excluded information on female firm ownership, preventing the use of the difference-in-difference estimator with 1990 as the baseline. We therefore estimate a similar difference-in-differences model which uses 1998 and 2013 as the baseline and endline periods.

Table 14: Effect of IAs on Female- and SC-owned firms

	Levels		Logs		Percent	
	Firms	Employees	Firms	Employees	Firms	Employees
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Female-Owned Firms						
within IA	5.456	59.291	0.377	0.532*	0.019	0.019
	(4.685)	(51.669)	(0.268)	(0.281)	(0.036)	(0.035)
Spillovers	1.907	0.005	0.260***	0.260**	0.025***	0.022***
	(1.248)	(1.873)	(0.086)	(0.107)	(0.008)	(0.008)
Control Mean	14.858	22.483			0.156	0.135
	(44.257)	(105.632)			(0.175)	(0.166)
R-squared	0.054	0.019	0.052	0.037	0.005	0.005
N	33842	33842	33842	33842	33842	33842
Panel B: SC-Owned Firms						
within IA	3.682	14.128	0.551**	0.685**	0.035	0.006
	(2.571)	(9.081)	(0.218)	(0.270)	(0.030)	(0.029)
Spillovers	1.597***	5.202	0.256***	0.303***	0.019***	0.021***
	(0.361)	(3.331)	(0.047)	(0.057)	(0.007)	(0.007)
Control Mean	6.308	10.372			0.093	0.081
	(14.424)	(27.445)			(0.146)	(0.138)
R-squared	0.024	-0.000	0.016	0.009	0.003	0.001
N	34316	34316	34316	34316	34316	34316

Note: Regression results are coefficients (β_j) of the *distance* $\times$ *post* interaction terms ($1[dist_v \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2$ corresponds to the distance bin (0-4] and denoted as "spillover"), in the difference-in-differences regression given in Specification (2). The outcome variables are the number of firms and employees for female-owned firms in Panel A and SC-owned firms in Panel B. The outcomes are in levels in Columns (1)–(2), logs in Columns (3)–(4), the share of all the firms and employees in Columns (5)–(6). For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

net aggregate increases in employment or firm activity at larger spatial scales. The available data does not make it possible to empirically determine whether, in the counterfactual scenario in which an IA was not established, the firms setting up within the IA would have been created or not; and, if they were, whether they would have located elsewhere in the district, state, or country. Nonetheless, these findings offer important insights into the general role of land zoning regulations in determining the patterns of rural economic development in India.

With this caveat in mind, a back-of-the-envelope calculation suggests that each IA was responsible for the creation of approximately 940 jobs in the villages overlapping the IA, and approximately 400 jobs in villages in its vicinity. This is also reflected in almost 500 local workers shifting their income generation from agriculture to the manufacturing and service sectors.[22]

This suggests that jobs created in new firms in and around IAs are filled by both commuters and the local village labor force. While we are unable to observe village-level output or incomes, the large increase in nightlights within IAs and surrounding areas suggests this change was accompanied by substantial boosts to local income growth. This is consistent with suggestive correlations that we find between IAs and village-level asset ownership, housing amenities, and the utilization of banking services (Table 16).[23]

We interpret the effects of IAs as being driven primarily by the provision of contiguous plots for non-agricultural economic activity, which relaxed the land constraint faced by firms in rural areas. However, it is possible that firms were also attracted by complementary infrastructure investments made by the government. To test this possibility, in Table 17 we estimate the effect of IA creation on infrastructure indicators. There is little evidence that the government made targeted investments in areas adjacent to the IA. However, it is important to note that we can only observe extensive measures of infrastructure availability in the census, and are therefore unable to rule out public investments that may have improved the quality of infrastructure.

Two revealing disparities between the pattern of the impacts within IAs and in their vicinity are worth mentioning. First, while within IAs new firms are mostly large manufacturing firms, outside the IAs they are mostly very small service sector and agricultural firms. This leads us to cautiously hypothesize that demand-side and/or credit-related channels are driving the spillover results, rather than agglomeration effects. Second, the impact

---

[22]These calculations are based on our estimated coefficients for the within-IA and the disaggregated distance bins (in Tables 7 and 15) multiplied by the average number of villages in each of the specified distance bins.

[23]These estimates are based on a single cross-sectional regression using data from the 2011 Economic Census, as these variables were not collected in earlier years of the census, and should therefore be interpreted with caution.

Table 15: Effect of IAs on Labor Force

	Male			Female		
	Non-Agr Workers			Non-Agr Workers		
	Pct	Level	Log	Pct	Level	Log
	(1)	(2)	(3)	(4)	(5)	(6)
within IA	0.119***	66.503**	0.430***	0.122***	22.521*	0.496***
	(0.034)	(28.696)	(0.145)	(0.043)	(12.006)	(0.160)
0–1 kms	0.048**	15.626	0.219**	0.050	-0.064	0.088
	(0.020)	(13.141)	(0.109)	(0.030)	(5.494)	(0.126)
1–2 kms	0.038*	19.458	0.173	-0.006	1.448	0.093
	(0.021)	(13.200)	(0.105)	(0.023)	(4.999)	(0.129)
2–3 kms	0.036***	5.175	0.150**	-0.001	-2.767	0.025
	(0.013)	(5.719)	(0.069)	(0.019)	(2.501)	(0.097)
3–4 kms	0.020*	12.615	0.185***	0.002	1.054	0.149***
	(0.012)	(8.654)	(0.052)	(0.018)	(4.933)	(0.055)
4–5 kms	0.013	5.838	0.109*	-0.011	-0.247	0.023
	(0.010)	(9.342)	(0.056)	(0.016)	(5.938)	(0.079)
Control Mean	0.271	116.02		0.243	55.29	
	(0.249)	(199.60)		(0.277)	(100.09)	
R-squared	0.035	0.191	0.030	0.016	0.161	0.034
N	43122	43122	43122	41700	41700	41700

Note: Regression results are coefficients (β_j) of the *distance × post* interaction terms ($1[dist_v \in bin_j] \times post_t$) in the difference-in-differences regression given in Specification (2). The outcome variables are the share of male non-agricultural workers in the labor force as percentage in Column (1), levels in Column (2), and logs in Column (3). Columns (4)–(6) present analogous results for the share of female non-agricultural workers. For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** $p<0.01$, ** $p<0.05$, and * $p<0.1$.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

Table 16: Effect of IAs on Assets and Credit Access

	Assets					House Quality			
	Pct HHs w/					Pct HHs w/			Pct HHs Use Bank
	TV	Radio	Scooter	Bicycle	Mobile	Brick House	Indoor Toilet	Num Rooms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
within IA	7.930**	-1.912	3.025	1.441	6.962	-1.475	17.453*	-0.016	10.702**
	(3.249)	(2.451)	(2.348)	(2.892)	(5.279)	(6.000)	(8.775)	(0.136)	(4.253)
Spillovers	3.971***	1.863	1.010	1.418	1.270	-0.991	4.732	0.030	5.264***
	(1.333)	(1.989)	(0.885)	(1.564)	(1.815)	(1.637)	(3.648)	(0.067)	(1.654)
Control Mean	43.909	19.138	17.755	35.915	48.784	27.790	24.405	2.621	59.310
	(19.274)	(17.150)	(12.727)	(19.004)	(20.053)	(29.172)	(27.506)	(0.842)	(26.883)
R-squared	0.234	0.045	0.127	0.060	0.076	0.014	0.176	0.048	0.051
N	19348	19348	19348	19348	19348	19655	19348	19348	19348

Note: Regression results are coefficients (β_j) of the distance terms ($1[dist \in bin_j]$), where j is each distance bin, from a cross-sectional regression and omitting time interactions. The direct effect ("within IA") is associated with distance bin $j = 1$ and the spillover effect corresponds to distances of (0-4] kms and is denoted by the distance bin $j = 2$. The outcome variables are the percentage of households owning the assets indicated in Columns (1)–(5); the characteristics of the structure in which households live in Columns (6)–(8); and the share of households who make use of banking services in Column (9). The endline control means are given in the table. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

38

Table 17: Effect of IAs on Infrastructure

	Paved Road	Health Center	Primary School	Tap Water	Phone	Electr
	(1)	(2)	(3)	(4)	(5)	(6)
within IA	0.013	-0.046*	0.012	-0.001	-0.007	-0.002
	(0.038)	(0.024)	(0.019)	(0.032)	(0.020)	(0.002)
0–1 kms	0.036	-0.012	-0.012	0.018	0.029	-0.007
	(0.026)	(0.012)	(0.023)	(0.020)	(0.024)	(0.008)
1–2 kms	-0.010	0.020*	-0.036	0.007	-0.013	-0.001
	(0.027)	(0.012)	(0.022)	(0.019)	(0.024)	(0.005)
2–3 kms	0.001	-0.008	-0.025	0.006	0.007	-0.007
	(0.026)	(0.009)	(0.023)	(0.018)	(0.010)	(0.004)
3–4 kms	-0.029*	-0.000	-0.017	-0.022	-0.001	-0.006
	(0.017)	(0.008)	(0.016)	(0.017)	(0.014)	(0.006)
4–5 kms	-0.025	0.002	-0.013	0.021	-0.011	-0.005
	(0.024)	(0.007)	(0.013)	(0.014)	(0.013)	(0.004)
Control Mean (1991)	0.594	0.057	0.805	0.167	0.157	0.892
	(0.491)	(0.232)	(0.396)	(0.373)	(0.364)	(0.311)
Control Mean (2011)	0.894	0.079	0.894	0.879	0.936	0.994
	(0.307)	(0.270)	(0.308)	(0.326)	(0.244)	(0.075)
R-squared	0.670	0.079	0.371	0.610	0.694	0.884
N	46022	46022	46114	46022	46022	46022

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (2). The outcome variables are binary variables for the presence of paved roads in Column (1), health center in Column (2), primary school in Column (3), tap water in Column (4), access to phone in Column (5) and access to electricity in Column (6). The baseline and endline control means are given in the table. Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The data obtained from demographic and economic censuses and remotely sensed nightlights, merged at the village level.

of IAs appear to be larger in areas closer to major cities and to highways, likely reflecting the importance of other factors, such as market access, for firm creation, even when land constraints are relaxed. Outside the IAs impacts are more sensitive to the local value of agricultural production, potentially reflecting a higher opportunity cost of exiting farming.

The success of the IA program suggests that the extensive agricultural zoning found throughout India, though ostensibly protecting the interests of agriculturalists, represents a significant impediment to local economic development. This program should be seen as complementing more traditional policies facilitating rural development, such as road construction (Asher and Novosad, 2019), electrification (Burlig and Preonas, 2016), and investments in human capital, which have generally yielded modest results for local economies. Given the relatively slow pace of urbanization, and the continuing prevalence of extreme poverty in rural areas, the IA program represents an attractive alternative to traditional policies promoting development through the movement of workers to urban areas (Kline and Moretti, 2014b).

References

Abeberese, A. B. and Chaurey, R. (2019). An unintended consequence of place-based policies: A fall in informality. *Economics Letters*, 176:23 – 27.

Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics*, 9(2):23–59.

Alder, S., Shao, L., and Zilibotti, F. (2016). Economic reforms and industrial policy in a panel of chinese cities. *Journal of Economic Growth*, 21(4):305–349.

Amirapu, A., Hasan, R., Jiang, Y., and Klein, A. (2019). Geographic concentration in indian manufacturing and service industries: Evidence from 1998 to 2013. *Asian Economic Policy Review*, 14(1):148–168.

Asher, S. and Novosad, P. (2019). Rural roads and local economic development. *forthcoming: American Economic Review*.

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). cem: Coarsened exact matching in stata. *The Stata Journal*, 9(4):524–546.

Blakeslee, D., Fishman, R., and Srinivasan, V. (2020). Way down in the hole: Adaptation to long-term water loss in rural india. *American Economic Review*, 110(1):200–224.

Burlig, F. and Preonas, L. (2016). Out of the darkness and into the light? development effects of rural electrification. *Energy Institute at Haas WP*, 268:26.

Chaurey, R. (2016). Location-based tax incentives: Evidence from india. *Journal of Public Economics*.

Cheng, Y. (2014). Place-based policies in a development context - evidence from china. *Working Paper, UC Berkeley*.

Criscuolo, C., Martin, R., Overman, H. G., and Van Reenen, J. (2019). Some causal effects of an industrial policy. *American Economic Review*, 109(1):48–85.

Desmet, K., Ghani, E., O'Connell, S., and Rossi-Hansberg, E. (2015). The spatial development of india. *Journal of Regional Science*, 55(1):10–30.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A market access approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Duranton, G., Ghani, E., Goswami, A. G., and Kerr, W. (2015). *The misallocation of land and other factors of production in India*. The World Bank.

Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? *The American Economic Review*, 89(2):311–316.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *The American Economic Review*, 100(3):1195–1213.

Freedman, M. (2013). Targeted Business Incentives and Local Labor Markets. *Journal of Human Resources*, 48(2):311–344.

Ganguli, I. (2015). Immigration and ideas: What did russian scientists bring to the united states? *Journal of Labor Economics*, 33(S1):S257–S288.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2012). Is India's Manufacturing Sector Moving Away From Cities? Harvard Business School Working Papers 12-090, Harvard Business School.

Gollin, D. (2014). The lewis model: A 60-year retrospective. *Journal of Economic Perspectives*, 28(3):71–88.

Government of India (2009). *Technical EIA Guidance Manual for Industrial Estates*. Ministry of Environment & Forests: Government of India.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010a). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010b). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *The Journal of Political Economy*, 118(3):536–598.

Greenstone, M. and Looney, A. (2010). *An economic strategy to renew American communities*. Hamilton Project, Brookings Institution.

Ham, J., Swenson, C., Imrohoroglu, A., and Song, H. (2011). Government programs can improve local labor markets: Evidence from state enterprise zones, federal empowerment zones and federal enterprise communities. *Journal of Public Economics*, 95:779–797.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *The American Economic Review*, 102(2):994–1028.

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

Kazmin, A. (2015). India: Land in demand. https://www.ft.com/content/2bba915c-18fa-11e5-a130-2e7db721f996.

Kline, P. and Moretti, E. (2014a). 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.

Kline, P. and Moretti, E. (2014b). People, places and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics*.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, 22(2):139–191.

Lu, Y., Wang, J., and Zhu, L. (2018). Place-based policies, creation and agglomeration economies: Evidence from china's economic zone program. *American Economic Journal: Economic Policy*.

Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.

Martin, P., Mayer, T., and Mayneris, F. (2011). Public support to clusters. *Regional Science and Urban Economics*, 41(2):108 – 123.

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

Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *The American Economic Review*, 94(3):656–690.

Morris, S. and Pandey, A. (2007). Towards reform of land acquisition framework in india. *Economic and Political Weekly*, pages 2083–2090.

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1989). Industrialization and the big push. *The Journal of Political Economy*, 97(5):1003–1026.

Neumark, D. and Kolko, J. (2010). Do enterprise zones create jobs? evidence from california's enterprise zone program. *Journal of Urban Economics*, 68(1):1–19.

Neumark, D. and Simpson, H. (2015). Place-based policies. *Handbook of Regional and Urban Economics*, 5:1197–1287.

Pinkovskiy, M. and Sala-i Martin, X. (2016). Lights, camera? income! illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, 131(2):579–631.

Rajan, R. (2013). Why india slowed. http://blogs.reuters.com/india-expertzone/2013/05/01/why-india-slowed/.

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of eastern and south-eastern europe. *Economic Journal*, 53(210/211):202–211.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2119–2171.

Roth, J. (2020). Pre-test with caution: Event-study estimates after testing for parallel trends.

Saez, L. (2002). *Federalism without a centre: The impact of political and economic reform on India's federal system.* Sage Publications.

Shenoy, A. (2018). Regional development through place-based policies: Evidence from a spatial discontinuity. *Journal of Development Economics*, 130:173–189.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-saharan africa. *The Review of Economic Studies*, 83(3):1263–1295.

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

Zheng, S., Sun, W., Wu, J., and Kahn, M. E. (2017). The birth of edge cities in china: Measuring the effects of industrial parks policy. *Journal of Urban Economics*.

For Online Publication Only:
Land Rezoning and Structural Transformation in Rural India: Evidence from the Industrial Areas Program

David Blakeslee[*]   Ritam Chaurey[†]

Ram Fishman[‡]   Samreen Malik[§]

February 4, 2021

---

[*]Email: david.blakeslee@nyu.edu. New York University (AD).

[†]Email: rchaurey@jhu.edu. Johns Hopkins University.

[‡]Email: ramf@post.tau.ac.il. Tel Aviv University.

[§]Email: samreen.malik@nyu.edu. New York University (AD).

Appendix A

This Appendix provides the spatial distribution of Industrial areas (IAs) throughout the state of Karnataka. Figure A1 provides the spatial distribution of IAs in Karnataka, as well as their relation to census towns and geographic features. All of the IAs used in this study are established between 1991–2015 (with no IAs being established during the period 2000-2004) as shown in Figure 1 and have been active since inception.

Figure A1: Spatial Distribution of Industrial Areas and Census Towns in Karnataka
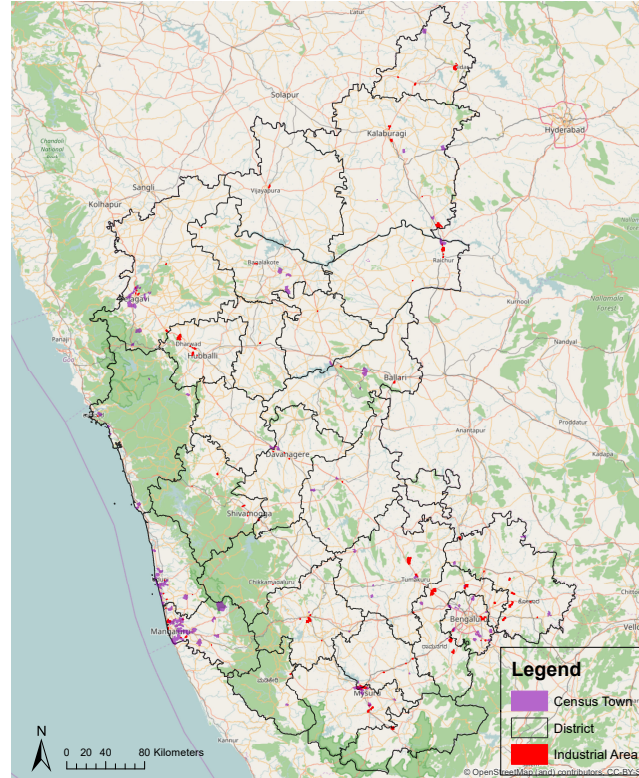
Note: This figure shows the spatial distribution of Industrial Areas as in our sample along with the census town .
Source: http://kiadb.in/industrial-areas/

Appendix B

This Appendix contains all the additional analyses and robustness exercises corresponding to the main results presented in the paper.

Table B1: Coarsened Exact Matching

	Levels			Logs		
	Light Density	Employees	Firms	Light Density	Employees	Firms
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline CEM						
within IA	13.173***	509.319***	37.632**	0.684***	1.132***	0.894***
	(1.755)	(185.369)	(16.241)	(0.116)	(0.323)	(0.239)
R-squared	0.750	0.171	0.298	0.896	0.240	0.296
N	822	734	734	822	734	734
Panel B: Expanded CEM						
within IA	14.776***	504.524***	38.023**	0.857***	1.118***	0.892***
	(2.047)	(186.555)	(16.827)	(0.113)	(0.325)	(0.240)
R-squared	0.756	0.177	0.285	0.904	0.261	0.324
N	520	662	648	520	662	648

Note: Regression results are coefficients (β) of $treatment \times post$ interaction terms ($IA_v \times post$) from difference-in-differences regression given in Specification (**??**). The outcome variables are the light density, the number of employees, and the number of firms. In Columns (1)-(3) the outcome variables are measured in levels, and in Columns (4)-(6) in logs. All logarithmic transformations are based on $log(x)$, except for light density which is transformed using the asinh ($log(x + \sqrt{x^2 + 1})$). All specifications use the coarsened exact matching (CEM) method, and use all villages within the state to determine the best match for control villages. In Panel A, the comparison villages are selected based on the following variables measured at baseline: distance to the nearest of the 10 largest cities in the state; distance to the nearest highway; fraction of land with forest cover; share of men in non-agricultural wage labor; population; and light density. In Panel B, we additionally include the baseline measure of the outcome variable (in levels) to the vector of matching variables. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The merged data including census, demographic, industrial area location and nighlights.

Table B2: Falsification Tests

	Light Density				Log Employees				Log Firms			
	Year of IA				Year of IA				Year of IA			
	2005-11		2012-15		2005-11		2012-15		2005-11		2012-15	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Within IA	-0.594	-0.239	0.898	0.914	0.087	0.044	0.104	0.142	0.009	-0.079	-0.010	0.068
	(0.610)	(0.621)	(0.690)	(0.937)	(0.232)	(0.270)	(0.139)	(0.154)	(0.196)	(0.221)	(0.111)	(0.075)
R-squared	0.108	0.131	0.263	0.251	0.034	0.048	0.035	0.045	0.022	0.036	0.019	0.024
N	71871	12609	95828	15516	64557	11099	87020	13957	64557	11099	87020	13957
IA X Year F.E.s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
≤10 kms from IA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Regression results are coefficients (β) of the *treatment* $\times$ *post* interaction terms ($IA_v \times post$) from the difference-in-differences regression given in Specification (**??**). The sample is limited to the years 1990–2005 for Columns (1)–(2), (5)–(6), and (9)–(10); and the *treatment* variable takes a value of 1 for villages in which an IA was established from 2006–2011. The full sample is for Columns (3)–(4), (7)–(8), and (11)–(12); and the *treatment* variable takes a value of 1 for villages in which an IA was established from 2012–2015. The outcome variables are light density in Columns (1)–(4), log number of employees in Columns (5)–(8) and log number of firms in Columns (9)–(12). Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. Controls are also included for the distance-post interaction terms ($\beta_{1,j}(1[dist \in bin_j] \times post_t)$) for IAS established during the respective study periods. Village fixed effects are included, as well as nearest-IA fixed effects interacted with time dummies. The nearest-IA is defined as the nearest IA established in the year indicated at the column head. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.
Source: The merged data including census, demographic, industrial area location and nighlights.

Table B3: Effect of IAs on Number of Workers by Firm Size

	Levels			Logs		
	Firm Size			Firm Size		
	>99	10–99	<10	>99	10–99	<10
	(1)	(2)	(3)	(4)	(5)	(6)
within IA	241.736***	85.975*	52.324	1.215***	0.679*	0.641***
	(77.596)	(44.962)	(32.517)	(0.332)	(0.348)	(0.216)
Spillovers	8.666	-0.790	11.835	-0.073	0.092	0.239**
	(10.044)	(3.733)	(8.518)	(0.048)	(0.106)	(0.098)
Control Mean	8.688	11.862	111.300			
	(277.188)	(52.667)	(185.112)			
R-squared	0.000	0.013	0.038	0.013	0.006	0.012
N	38630	38630	38630	38630	38630	38630

Note: Regression results are coefficients (β_j) of the $distance \times post$ interaction terms ($1[dist_v \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2$ corresponds to the distance bin (0-4] and denoted as "spillover"), in the difference-in-differences regression given in Specification (**??**). The outcome variable is the number of workers by firm size in Column (1)–(3) and the log number of workers by firm size in Columns (4)–(5). For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.
Source: The merged data including census, demographic, industrial area location and nighlights.

Table B4: Effect of IAs on Demography and Labor Force Participation

		Male			
				Labor Force Participation Rate	
	Log Population	Literacy Rate	Percent SC	Percent: Full Time	Part-time
	(1)	(2)	(3)	(4)	(5)
within IA	0.034	0.009	-0.004	-0.008	0.001
	(0.083)	(0.013)	(0.009)	(0.016)	(0.001)
0–1 kms	0.021	0.008	0.009	-0.020	-0.000
	(0.046)	(0.012)	(0.007)	(0.013)	(0.001)
1–2 kms	0.027	0.019	0.013*	-0.005	0.001*
	(0.036)	(0.012)	(0.007)	(0.009)	(0.001)
2–3 kms	-0.031	0.007	0.006	0.003	0.001***
	(0.026)	(0.008)	(0.004)	(0.006)	(0.001)
3–4 kms	0.053**	0.013*	-0.001	-0.003	-0.000
	(0.025)	(0.007)	(0.005)	(0.009)	(0.000)
4–5 kms	0.018	0.007	0.006	0.005	0.000
	(0.027)	(0.008)	(0.006)	(0.007)	(0.001)
Control Mean		0.691	0.206	0.549	0.070
		(0.114)	(0.204)	(0.123)	(0.099)
R-squared	0.053	0.037	0.041	0.006	0.005
N	43130	43130	43130	43130	43130

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (**??**). The outcome variables are (male) log adult population in Column (1), literacy rate in Column (2), share of the adult population that is SC in Column (3), and labor force participation rate in Columns (4)–(5). For levels, we also provide the endline control mean. Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The merged data including census, demographic, industrial area location and nighlights.

Table B5: Panel Specification

	Light Density	Log Employees	Log Firms	Pct Non-Agr Labor
	(1)	(2)	(3)	(4)
Within IA	8.319***	0.607***	0.276**	0.100***
	(1.427)	(0.205)	(0.120)	(0.021)
0–1 kms	1.694***	0.150*	0.032	0.049**
	(0.554)	(0.087)	(0.074)	(0.018)
1–2 kms	0.359	0.258***	0.240***	0.048***
	(0.345)	(0.082)	(0.063)	(0.014)
2–3 kms	0.177	-0.030	0.010	0.027**
	(0.257)	(0.065)	(0.053)	(0.011)
3–4 kms	0.402***	0.127**	0.142***	0.027***
	(0.145)	(0.057)	(0.049)	(0.009)
4–5 kms	0.413*	-0.011	0.017	0.019**
	(0.223)	(0.064)	(0.058)	(0.008)
R-squared	0.231	0.033	0.018	0.026
N	95828	87129	87129	68938

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (??). The sample includes the 4 rounds of the Economic Census (1990, 1998, 2005, and 2013). The outcome variables are the light density in Column (1), the log number of employees in Column (2), the log number of firms in Column (3), and the share of male workers in non-agricultural wage labor in Column (4). Control villages are those located more than 5 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.

Source: The merged data including census, demographic, industrial area location and nighlights.

Table B6: Expanded Spillovers

	Light Density	Log Employees	Log Firms	Pct Non-Agr Labor
	(1)	(2)	(3)	(4)
within IA	11.054***	1.084***	0.729***	0.120***
	(2.048)	(0.387)	(0.226)	(0.034)
0–1 kms	2.369***	0.409**	0.270*	0.049**
	(0.822)	(0.202)	(0.137)	(0.021)
1–2 kms	0.874	0.622***	0.554***	0.039
	(0.652)	(0.171)	(0.141)	(0.023)
2–3 kms	0.428	0.249*	0.306**	0.037**
	(0.459)	(0.144)	(0.132)	(0.015)
3–4 kms	0.697*	0.282*	0.320**	0.021
	(0.364)	(0.150)	(0.132)	(0.014)
4–5 kms	0.565	0.110	0.089	0.014
	(0.534)	(0.126)	(0.110)	(0.012)
5–6 kms	0.411	0.168	0.099	0.006
	(0.424)	(0.109)	(0.099)	(0.013)
6–7 kms	0.687	0.155	0.130	0.006
	(0.436)	(0.113)	(0.092)	(0.013)
7–8 kms	0.441	0.129	0.127	0.006
	(0.405)	(0.121)	(0.101)	(0.012)
8–9 kms	0.097	0.144	0.118	0.004
	(0.418)	(0.102)	(0.089)	(0.010)
9–10 kms	0.151	0.073	0.062	-0.011
	(0.293)	(0.090)	(0.075)	(0.013)
R-squared	0.224	0.016	0.009	0.091
N	42180	36350	36350	43122

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where $j = 1$ corresponds to a distance of 0 and denoted as "within IA" and $j = 2, \ldots, j = 9]$ correspond to the distance bins (1-2], ..., (9-10] and denoted as "spillover"), in the difference-in-differences regression given in Specification (**??**). The outcome variables are the light density in Column (1), the log number of employees in Column (2), the log number of firms in Column (3), and the share of male workers in non-agricultural wage labor in Column (4). Control villages are those located more than 10 kms from the nearest IA. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1. Source: The merged data including census, demographic, industrial area location and nighlights.

Table B7: Effect of IAs on Assets and Credit Access

	Assets					House Quality			
	Pct HHs w/					Pct HHs w/			Pct HHs
	TV	Radio	Scooter	Bicycle	Mobile	Brick House	Indoor Toilet	Num Rooms	Use Bank
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
within IA	7.930**	-1.912	3.025	1.441	6.962	-1.475	17.453*	-0.016	10.702**
	(3.249)	(2.451)	(2.348)	(2.892)	(5.279)	(6.000)	(8.775)	(0.136)	(4.253)
Spillovers	3.971***	1.863	1.010	1.418	1.270	-0.991	4.732	0.030	5.264***
	(1.333)	(1.989)	(0.885)	(1.564)	(1.815)	(1.637)	(3.648)	(0.067)	(1.654)
Control Mean	43.909	19.138	17.755	35.915	48.784	27.790	24.405	2.621	59.310
	(19.274)	(17.150)	(12.727)	(19.004)	(20.053)	(29.172)	(27.506)	(0.842)	(26.883)
R-squared	0.234	0.045	0.127	0.060	0.076	0.014	0.176	0.048	0.051
N	19348	19348	19348	19348	19348	19655	19348	19348	19348

Note: Regression results are coefficients (β_j) of the distance terms ($1[dist \in bin_j]$), where j is each distance bin, from a cross-sectional regression and omitting time interactions. The direct effect ("within IA") is associated with distance bin $j = 1$ and the spillover effect corresponds to distances of (0-4] kms and is denoted by the distance bin $j = 2$. The outcome variables are the percentage of households owning the assets indicated in Columns (1)–(5); the characteristics of the structure in which households live in Columns (6)–(8); and the share of households who make use of banking services in Column (9). The endline control means are given in the table. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance *** p<0.01, ** p<0.05, and * p<0.1.

Source: The merged data including census, demographic, industrial area location and nighlights.

Table B8: Effect of IAs on Infrastructure

	Paved Road	Health Center	Primary School	Tap Water	Phone	Electr
	(1)	(2)	(3)	(4)	(5)	(6)
within IA	0.013	-0.046*	0.012	-0.001	-0.007	-0.002
	(0.038)	(0.024)	(0.019)	(0.032)	(0.020)	(0.002)
0–1 kms	0.036	-0.012	-0.012	0.018	0.029	-0.007
	(0.026)	(0.012)	(0.023)	(0.020)	(0.024)	(0.008)
1–2 kms	-0.010	0.020*	-0.036	0.007	-0.013	-0.001
	(0.027)	(0.012)	(0.022)	(0.019)	(0.024)	(0.005)
2–3 kms	0.001	-0.008	-0.025	0.006	0.007	-0.007
	(0.026)	(0.009)	(0.023)	(0.018)	(0.010)	(0.004)
3–4 kms	-0.029*	-0.000	-0.017	-0.022	-0.001	-0.006
	(0.017)	(0.008)	(0.016)	(0.017)	(0.014)	(0.006)
4–5 kms	-0.025	0.002	-0.013	0.021	-0.011	-0.005
	(0.024)	(0.007)	(0.013)	(0.014)	(0.013)	(0.004)
Control Mean (1991)	0.594	0.057	0.805	0.167	0.157	0.892
	(0.491)	(0.232)	(0.396)	(0.373)	(0.364)	(0.311)
Control Mean (2011)	0.894	0.079	0.894	0.879	0.936	0.994
	(0.307)	(0.270)	(0.308)	(0.326)	(0.244)	(0.075)
R-squared	0.670	0.079	0.371	0.610	0.694	0.884
N	46022	46022	46114	46022	46022	46022

Note: Regression results are coefficients (β_j) of the distance-post interaction terms ($1[dist \in bin_j] \times post_t$), where j is each distance bin, from the difference-in-differences regression given in Specification (**??**). The outcome variables are binary variables for the presence of paved roads in Column (1), health center in Column (2), primary school in Column (3), tap water in Column (4), access to phone in Column (5) and access to electricity in Column (6). The baseline and endline control means are given in the table. Control villages are those located more than 5 kms from the nearest IA. A vector of time-interacted controls is included for characteristics determining site selection or correlated with potential growth. The regression includes village and time fixed effects, a vector of time-interactions with baseline controls, and nearest-IA fixed effects interacted with time dummies. Robust standard errors (clustered at the nearest-IA level) are shown in parentheses. Stars denote statistical significance: *** p<0.01, ** p<0.05, and * p<0.1.
Source: The merged data including census, demographic, industrial area location and nighlights.