Rural Infrastructure Development and Economic Activity*

Ritam Chaurey†
Duong Trung Le‡

September 2019

Abstract

There is universal consensus that physical infrastructure investments are important for economic growth. However, estimating their causal effects has remained challenging, especially in rural settings. We evaluate the impact of a rural infrastructure development scheme directed towards India's most "backward" districts. Using a Regression Discontinuity Design and multiple data sets covering the entire firm size distribution, as well as household employment surveys and nighttime light data, we show evidence on the effectiveness of the program on local economic activity. We find that villages in treated districts had higher employment and number of firms. This was entirely driven by increases in the employment of microenterprises as well as the number of microenterprises, with no impact on formal firms. We also find increases in workers' wages, the number of days worked, and monthly household consumption expenditure. There is suggestive evidence that both rural electrification and connectivity were important mechanisms driving our results. We find stronger impacts in electricity and road-intensive industries, and in villages that had paved roads and electricity prior to the program. Overall, our paper suggests that improving infrastructure conditions can boost economic activity, especially by stimulating microenterprises.

JEL Codes: O12, O18, O25, R11

*We would like to thank Ama Baafra Abeberese, Sam Asher, Siddharth Hari, Gaurav Khanna, Amit Khandelwal, Solomon Polachek, Martin Rotemberg, David Slichter, Eric Verhoogen, Susan Wolcott, and participants at the Young Economists Symposium (Yale 2017), Northeastern Association of Business, Economics & Technology (Penn State 2017), ACEGD ISI (New Delhi 2018), Eastern Economic Association (Boston 2018), George Mason Schar School of Policy and Government, EPED (Montreal 2018), and Ce2 workshop (Warsaw 2019) for helpful comments. We also thank Karan Singh Bagavathinathan for excellent research assistance. This paper subsumes an earlier version, "Infrastructure Grants and the Performance of Microenterprises." All remaining errors are our own.

†Email: rchaurey@jhu.edu. Johns Hopkins University, School of Advanced International Studies and IZA.
‡Email: duong.le@duke.edu. Duke University, The Nicholas School of Environment.

1 Introduction

There is universal consensus that investments in physical infrastructure (roads, electricity, telecommunications, fast Internet, dams, irrigation, etc.) are important determinants of economic growth (World Bank, 1994). These infrastructural investments are inherently place-based, and are often directed either to economically lagging regions to incentivize growth or to potentially fast-growing regions to further accelerate growth. This non-random placement often complicates the analysis of the causal effects of infrastructure investments. Studying the impacts of infrastructure is especially harder in the context of rural and low-income populations where demand for these services may be low. For example, rural households may not experience the benefits of electricity grid connections if they do not have complementary appliances (Lee et al., forthcoming). Similarly, households in less populated villages may not be able to fully utilize the benefits of rural roads owing to their remoteness (Asher and Novosad, forthcoming). However, the demand for infrastructure investments are more likely to be higher for firms because these services are both production inputs, and are important for the supply chain. Therefore, analyzing responses of firms is of central importance in order to understand the efficacy of infrastructure provision on local economic activity.

In this paper, we study the effects of a rural infrastructure grants program, Rashtriya Sam Vikas Yojana (RSVY), launched by the Government of India, that were extended to districts using a score-based assignment mechanism. We are thus able to address the primary concern of non-random placement of infrastructure investments, by using a regression discontinuity design. We combine multiple data sets covering the entire firm-size distribution, household surveys, as well as data on night-time lights to analyze the effects of RSVY on local economic outcomes, and the mechanisms driving these effects.

RSVY was launched in the fiscal year of 2003-04 with the main goal of facilitating physical infrastructure development in the most economically "backward" districts in India. This program was one of the first direct attempts carried out by the central government to identify and support India's economically lagging districts to reduce regional economic imbalance and speed up development. Under RSVY, each eligible district was entitled to receive 450 million Indian Rupees ("Rs.") ($\approx$ 10 million USD (2010 exchange rates)), over the course of three years, to address "critical gaps" in physical and social infrastructure. This amounted to around 1.15% of the average "backward" district's GDP between 2003-04 and 2005-06. The policy guidelines mandated that the infrastructure gaps should be identified in a decentralized manner at the district-level and involve the community and key stakeholders. This essentially meant that each district, after consultation with various stakeholders could decide where (and on what) they spent the RSVY funds. According to the guidelines, the funds were to be utilized to improve or make complementary investments to existing infrastructure rather than on completely new projects. Therefore, RSVY could be considered as an infrastructure-enhancing grant. For example, for physical infrastructure, i.e. rural connectivity and electrification, RSVY funds could be spent on widening and strengthening of roads, to build small bridges, to build vital road links to connect to the marketplace, and for strengthening the rural electricity transmission and distribution infrastructure. Districts had the flexibility to spend the RSVY funds on multiple small infrastructure-related projects. This flexibility in the use of funds made RSVY different than other programs that are often focused on building a specific infrastructural facility such as new roads and highways, or dams, or electricity grids.

The specific guidelines used by the Government of India to prioritize the treated districts makes RSVY an ideal natural experiment. The central government first allocated to each of the 17 major states in the country a pre-specified quota of districts based on the states' poverty headcount ratios. Next, each state government designated the districts within their state that they deemed fit to receive the grant. However, the central government's guidelines for RSVY specifically requested that the most backward districts - based on an official district-level "Backwardness Index" - must be prioritized as beneficiaries of RSVY grants.

Our empirical strategy relies on the identification of RSVY-eligible "backward" districts using official documentation of the Planning Commission. We reconstruct the "Backwardness Index" scores for each district and use the distance to the score of the cutoff district in each state as the running variable in a Regression Discontinuity Design (RD) framework. We run RD regressions on various economic outcomes at the village, firm, household, and taluk (sub-district) levels, using multiple data sets.

We find a number of results on the effectiveness of RSVY in the short-run. First, we find that overall village employment in RSVY districts increased by 11% relative to control districts. This was driven entirely by village employment in microenterprises, with no effect on employment in formal firms. Next, we corroborate these village-level results with firm-level regressions separately for microenterprises and formal firms. We find, on average, a 5%-13% increase in microenterprise employment, corresponding to an increase of 0.2-0.35 workers per firm. However, formal firms did not experience an increase in employment. This is not surprising given that RSVY was a rural infrastructure program and most formal firms in India locate in or near urban areas,[1] and 36.5% of these formal firms own a generator.[2] On the extensive margin, we find an increase in the number of firms in villages in RSVY districts, again entirely driven by an increase in the number of microenterprises. Thus, overall we find that RSVY led to the growth of microenterprises. Next, we look at the effect of RSVY on individual and household outcomes, and find increases in individual wages (10%-13%), number of days worked (3%), and household monthly consumption expenditure (8.7%), which corroborate our village and firm-level results. All our results are robust across various specifications and bandwidths.

Since RSVY grants could be used across multiple infrastructure projects, it is important to understand the possible channels through which the policy affected the performance of small firms. First, we find an increase in the overall infrastructure development following RSVY in the treated districts, as measured by nighttime light intensity. Next, we find that there was a significant reduction in the probability of experiencing a power cut and problems with access to raw materials. We interpret these results as suggestive evidence for improvements in rural electrification and connectivity in RSVY districts. To provide additional supporting evidence for these two channels, we focus on firms in electricity and road-intensive industries. The reductions in probability of power cut and problems with access to raw materials are concentrated in the most electricity and road-intensive industries respectively. We also find that firms in the most electricity and road-intensive industries had the largest increases in employment. Furthermore, the purpose of RSVY funds was to fill critical infrastructure gaps rather than on providing roads to unconnected areas or electricity grids in unelectrified areas. As such, the effects of this infrastructure-enhancing policy should be larger in villages where paved roads and electricity already existed before RSVY compared to villages without paved roads and electricity. We find evidence consistent with this claim. We show that after the implementation of RSVY, employment in microenterprises as well as number of microenterprises increased in villages that had both paved roads and electricity prior to the policy, and no effects in villages without paved roads and electricity.

[1] Only one in five villages in our sample had a formal firm.

[2] Authors' calculation using Annual Survey of Industries data.

Our results withstand multiple placebo and robustness checks. First, we show graphically that district-level observable characteristics, including geographic (time-invariant) and baseline socio-demographic attributes, are smooth functions around the RD cutoff. Second, using pre-treatment data (3 years before RSVY introduction), we find no effect on any of the main outcome variables before the introduction of the policy. Finally, we find no effect of the policy when the eligibility threshold is hypothetically moved to a different point along the distribution of the backwardness-distance scores (running variable).

Our paper directly contributes to the literature that seeks to establish causal links of different types of infrastructure, for example – rural roads (Aggarwal (2018); Adukia et al. (forthcoming); Asher and Novosad (forthcoming)), highways (Ghani et al. (2016); Faber (2014)), railroads (Donaldson (2018)), bridges (Brooks and Donovan (2019)), electricity (Abeberese (2017); Allcott et al. (2016); Dinkelman (2011); Rud (2012); Lipscomb et al. (2013); Chakravorty et al. (2014); Burlig and Preonas (2016); Lee et al. (forthcoming); Hardy and McCasland (2017); Lenz et al. (2017)), dams (Duflo and Pande (2007)), telecommunications and Internet (Jensen (2007); Aker and Mbiti (2010); Hjort and Poulsen (2019)), and water and sanitation (Alsan and Goldin (2019); Devoto et al. (2012)), on economic outcomes in developing countries. A majority of these papers find positive effects of infrastructure on various measures of economic development. However, recent work on rural electrification in Kenya (Lee et al., forthcoming) and India (Burlig and Preonas, 2016), and rural roads (Asher and Novosad, forthcoming) in India, have questioned the effectiveness of large-scale rural infrastructure programs in benefiting the rural poor.

Our analysis differs in several dimensions from the above papers. First, in contrast to the studies that examine the effects of new electricity connections/grids or new rural roads in places that were unelectrified or unconnected, in our context, RSVY funds were spent on filling up critical gaps in infrastructure in the district. This is an important distinction because it is more likely that firms would operate in areas with pre-existing infrastructure than in remote areas with no electricity. To this extent, our findings that RSVY led to higher employment and number of firms in villages that already had paved roads and electricity prior to the policy, and no effect on unelectrified and unconnected villages, are consistent with these papers. Furthermore, in these papers, the authors mainly look at household or village-level outcomes,[3] whereas our main focus is on industrial outcomes. Since lack of electricity and roads are production constraints for firms, it is more likely that the demand for infrastructure will be higher for firms than for rural households.[4] In this regard, we find that the effects on microenterprises is larger in electricity and road-intensive industries, where the demand for infrastructure is substantial.

Our paper also relates to another strand of literature - the economics of microenterprises. Despite the important role played by microenterprises in developing economies,[5] the existing evidence on them is scant, partially due to the lack of data. Studies on microenterprises thus mainly rely on data collected from randomized controlled trials (RCT) and field experiments, which limits the generalizability of the results. Furthermore, a majority of the studies focus on relaxing financial (Banerjee, 2013; Banerjee et al., 2015;

[3] Asher and Novosad (forthcoming) also look at the impact of rural roads on employment in nonfarm village firms and find a small positive effect. For rural electrification, both Lee et al. (forthcoming) and Burlig and Preonas (2016) conclude that it is important to look at the effects on industrial consumers/firms.

[4] Lee et al. (forthcoming) find that in their rural Kenyan sample, unconnected households have extremely low electrical appliance ownership and conclude that this may be one reason for low demand for electrification.

[5] Using comprehensive data on both formal and informal firms, Hsieh and Olken (2014) show that in India, Indonesia, and Mexico, 98%, 97%, and 92% of firms have fewer than 10 employees, and these small firms employ 65%, 54%, and 22% of the labor force.

de Mel et al., 2008; Karlan and Zinman, 2009; Karlan et al., 2014, 2015; McKenzie, 2017; Rotemberg, forthcoming) and managerial constraints (Bloom and van Reenen, 2007; Bruhn and Zia, 2013; Bruhn et al., 2018; Cole et al., 2011; Drexler et al., 2014; McKenzie and Woodruff, 2017) on firm performance. We contribute to this body of research by analyzing the effect of relaxing a different production constraint related to improving the infrastructure environment in which small firms operate. Our results suggest that rural infrastructure investments are important for the growth of microenterprises.

Finally, given that most infrastructure programs are directed towards particular locations, our paper directly contributes to the literature on place-based policies. While the existing literature has primarily focused on place-based policies that provide tax or other financial incentives (such as wage or capital investment subsidies) in promoting regional economic growth,[6] we examine a place-based policy that solely focused on infrastructural development and did not offer financial incentives to firms. The only other papers focusing on such infrastructure schemes are in the U.S. and Europe: Kline and Moretti (2014) on the Tennessee Valley Authority initiative, Glaeser and Gottlieb (2008) on the 1963 Appalachian Regional Commission, and Becker et al. (2010, 2012) on European Structural Funds. Furthermore, in contrast to previous work that looked at medium and large firms, we focus on the effects on the entire firm-size distribution. This includes microenterprises, which in our context employ close to three-quarters of the workforce and are vital to rural economic activity.[7]

The rest of the paper proceeds as follows: Section 2 provides a detailed description of RSVY, its objectives, and the assignment algorithm. Section 3 explains the data used for the analysis. We describe our empirical strategy in Section 4. Section 5 presents and discusses the empirical results. Finally, Section 6 concludes.

2 Rashtriya Sam Vikas Yojana (RSVY)

The Government of India launched the Rashtriya Sam Vikas Yojana (RSVY) in 2003-04 with the main objectives to "remove barriers to economic growth, accelerate the development process, and improve the quality of life of the people" (Planning Commission, 2003). The program was one of the first direct attempts carried out by the central government to identify and support India's backward districts. RSVY covered a total of 147 backward districts, out of approximately 600 districts in the country. Under the policy guidelines, each district was entitled to receive rural infrastructure grant amounts of 450,000,000 Rupees (approximately 10 million USD) over the course of 3 fiscal years: 2004-05, 2005-06, and 2006-07. The proposed transfer mechanism was equal payments of 150,000,000 Rupees, i.e. one-third of the total fund, per year. Figure 1

[6]For example, in the United States, Neumark and Kolko (2010); Greenbaum and Engberg (2004); Bondonio and Greenbaum (2007); Ham et al. (2011); Busso et al. (2013) provide evidence on two well-known place-based programs: Federal Empowerment Zones (EZ) and State Enterprise Zones (ENTZ). In Europe, there are studies evaluating the effects of "Regional Selective Assistance" in the United Kingdom (Criscuolo et al., 2019), the French ZFUs (Mayer et al., 2015; Givord et al., 2013) and Italy's Law 488/1992 (Bronzini and de Blasio, 2006). See Neumark and Simpson (2015) for a more complete discussion on prior work on place-based policies. Recently, the literature on evaluation of place-based interventions has shifted towards developing economies. Several studies have shown that Chinese Special Economic Zones (SEZs) generated positive effects (Wang, 2013; Lu et al., forthcoming; Cheng, 2014; Alder et al., 2016). In India, Chaurey (2017); Shenoy (2018); Hasan et al. (2017) have found beneficial effects of tax exemption schemes on firms and local economic activity.

[7]Authors' calculation based on the Economic Census of 2005.

Panel A shows the details of the 115 districts that were selected specifically based on a transparent assignment mechanism discussed in the next sub-section.

As per the central government's instructions, all RSVY funds were to be utilized in addressing "critical gaps" in physical and social infrastructure to alleviate the problems of infrastructure deficits, low agricultural productivity, and excessive unemployment (Planning Commission, 2003). To identify these critical gaps, the policy guidelines mandated a decentralized district-level bottom-up planning approach that involved the community and key stakeholders, such as, Panchayati Raj Institutions (village-level institutions), community-based organizations, line departments, etc. This was done in order to ensure that the plan was representative of the needs of the district. District Perspective Plans (DPPs) were then prepared by the District Administration, enlisting the project proposals on which the RSVY funds would be spent. According to Planning Commission guidelines, in the physical infrastructure sector (rural connectivity and electrification), RSVY funds could be spent on road upgradation, to build bridges and culverts, especially vital road links to connect to the marketplace, and for strengthening the rural electricity transmission and distribution infrastructure. Furthermore, the guidelines also mentioned that investment in agriculture or irrigation related programs should be accompanied by important forward and backward linkages such as rural connectivity and electrification wherever possible.[8] The District Perspective Plans with details on the characteristics of programs undertaken at the district level are not publicly available. However, according to an official evaluation study which surveyed a representative sample covering 15 districts from 11 states, approximately 77% of the transferred funds were invested in infrastructural interventions, including rural connectivity, electrification, agricultural and irrigation improvement projects (Program Evaluation Organization, 2010).

In February 2007, the Government of India launched the Backward Regions Grant Fund (BRGF) that subsumed RSVY and extended it to 250 backward districts across 27 states of India.

2.1 Assignment Mechanism

Unlike most place-based programs that are subject to non-random placement, RSVY had a uniquely complete and transparent allocation procedure that was explicitly documented by the Government of India. Following the allocation algorithm, the eligibility of districts under RSVY, i.e. treatment assignment, was based on a two-step process. In the first step, the Central Government determined the number of treatment districts that would be assigned to each of the 17 major Indian states.[9] The quotas were worked out on the basis of state-level prevalence of poverty. In the second step, each state government, in accordance with the assigned quota, chose the specific districts to allocate the RSVY grants. The selection was based on an existing development ranking referred to as the Backwardness Index. This ranking index was public information, and a composite level of districts' economic underdevelopment was constructed from three historical parameters with equal weights: (i) value of output per agricultural worker (1990-1993); (ii) agriculture wage rate (1996-1997); and (iii) districts' percentage of low-caste populations - Scheduled Castes/ Scheduled Tribes (1991). The Backwardness Index ranked a total of 447 districts in the 17 major states with available data for all three parameters above. More details on the construction of the index are discussed in the Appendix.

In addition to the above algorithm, the government had a separate list of 32 districts that were heavily

[8]See http://www.planningcommission.nic.in/plans/stateplan/guid_rsvy.pdf.

[9]These 17 states are the "non-special category" states that comprised more than 97% of India's population in 2005.

affected by Maoist/Naxalite violence. These districts were automatically selected into the RSVY program.

3 Data and Variables Formation

We use several data sources for the analysis. We use the Fourth (1998) and Fifth (2005) rounds of the Economic Census (EC) for the village-level outcomes on employment and number of firms. For the microenterprise-level outcomes, we use information from rounds 56 (2000-01) and 62 (2005-06) of the National Sample Survey - Manufacturing Enterprises Schedule (NSS - Schedule 2.2), and for large formal firms we use the Annual Survey of Industries (ASI) data set between 2001 and 2005. We use NSS – Employment and Unemployment Schedule (NSS - Schedule 10) rounds 55 (1999-2000) and 62 (2005-06) for household and individual outcomes. We control for baseline and time-invariant covariates at the village and district-level by utilizing information from the 2001 Population Census and GIS-processed shapefiles for the country. Since RSVY was a rural infrastructure development program, we restrict our analysis to the rural sample across all data sources. Finally, we proxy for the overall infrastructure environment in the district with a measure of night-time light intensity processed from NASA's satellite transmitted data. We discuss each of these sources below.

3.1 Economic Censuses (EC)

The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production, defense and government administration, conducted by the Ministry of Statistics and Program Implementation (MoSPI), India. Establishments are defined as any location, commercial or residential, where an economic activity is carried out. Both formal and informal establishments are enumerated, irrespective of firm size, including people working out of their houses. We use the Fourth (1998) and Fifth (2005) rounds of the Economic Censuses for our analysis. The Economic Census only provides information on the number of employees by firm, but does not provide any other information on the inputs or output. We restrict the sample to firms engaged in non-agricultural activities. We use employment at the firm-level and also aggregate the microdata to the village level to get measures of employment and number of firms. We also divide the sample into formal firms and microenterprises to look at employment and number of firms. To be consistent with the other data sets (as well as the Factories Act 1948) we define microenterprises as those with less than 10 workers and that use electrical power, or less than 20 workers and do not operate with electrical power. The remaining therefore, are formal firms.

3.2 National Sample Surveys (NSS)

The NSS - Schedule 2.2 is a nationally representative survey in India that provides detailed information on manufacturing microenterprises' business activities and performance. Only small, "unorganized" firms with less than 10 workers and that use electrical power, or less than 20 workers and do not operate with electrical power, are included in this survey.[10] Micro firms meeting these employment criteria account for nearly 80% of India's manufacturing employment (Nataraj, 2011).[11] We use rounds 56 and 62 of the NSS manufacturing enterprise surveys. Since RSVY was introduced in June 2004, information from round 62 (2005-06) captures the short-run, post-treatment effects of this policy. Data from round 56 (2000-01) serves as the baseline period and allows us to perform falsification/placebo tests.[12]

Sampled firms are asked questions regarding their cash flows and operating activities such as employment, wage bill, sales revenues, total value of inputs, sources of capital, as well as various types of investments. Quantitative questions are often asked on the basis of one reference month prior to the survey, e.g. the firm's business performance during the last month prior to the survey date. Besides, there are related questions on firms' subjective perceptions of growth and overall local business environment during the year.

For our analysis, we mainly use information on firm-level employment. To test for mechanisms, we use the subjective infrastructure-related questions that ask enterprises whether the firm experienced power cut during production in the previous year. We use this question to proxy for the availability and/or quality of electricity supply. We also use information from another question that asks firms if they experienced problems with availability of raw materials. We use the responses to this question as a proxy for rural connectivity.

To study the effects of the policy on household and individual outcomes, we use data from the National Sample Survey (Schedule 10) employment-unemployment rounds 55 (1999-2000) and 62 (2005-06). These are nationally representative surveys covering all districts of India, where households in each district are sampled on a rolling basis over the agricultural year (July to June). The survey elicits daily employment and wage information for each household member over the 7 days preceding the interview. We mainly use information on individual-level wages and total number of worker-days, as well as household-level monthly consumption expenditure.

3.3 Annual Survey of Industries (ASI)

We use firm-level data from the Annual Survey of Industries (ASI) between 2001-2005, conducted by the Ministry of Statistics and Program Implementation (MoSPI) in India. The ASI covers all registered industrial units, which includes units with 10 or more workers and use electricity, or have least 20 workers but do not use electricity. Hence, the ASI data set covers large formal firms that are not covered in the NSS manufacturing surveys. The ASI frame is divided into census (surveyed every year) and sample (sampled every few years) sectors. Large firms (greater than 100 employees) are covered in the census sector, whereas the rest of the firms are covered in the sample sector with a third of these firms randomly selected in the survey each year. The reference year for the ASI is the accounting year from 1st April of the previous year to 31st March of the next year. For example, data from 2004 to 05 will include the period from 1st April 2004 to 31st March 2005. For our analysis, we only focus on one outcome variable in the ASI data set – employment.

[10] Essentially, small firms meeting these criteria are not required to register with the state governments under India's 1948 Factories Act, hence often referred to as "unregistered", "unorganized", or "informal" firms.

[11] Nataraj (2011) and Hsieh and Klenow (2014) are previous papers that have used the NSS Sch. 2.2.

[12] The National Sample Survey Organization (NSSO) did not conduct any other similar survey (Schedule 2.2) between Round 56 and 62.

3.4 Population Census (2001) & Geographic data

We use the Primary Census Abstract (PCA) and Village Directory (VD) for the 2001 Population Census to construct pre-RSVY village and district-level covariates. Specifically, we include baseline socio-demographic information for all districts in our sample, such as information on total population, total households, the population share of SC/ST, as well as access to representative public goods such as paved roads, electricity, and irrigation facilities.

In addition, it is important to also control for district's geographic characteristics. We thus utilize the Geographic Information System (GIS) software to process the country's shapefiles provided by the Global Administrative Areas organization (www.gadm.org), and use the relevant district's geographic indicators such as area (in square kilometers), boundary perimeter (in kilometers), elevation (in meters),[13] and distance (in kilometers) to the nearest metropolitan cities.

3.5 Night-time Light Intensity

Besides documenting the reduced-form effect of infrastructure development grants on enterprise performance, our analysis also provides evidence on the underlying mechanism through which the effect takes place. Particularly, we are interested in the direct impacts of the policy on the overall progress of infrastructural environment in the treated districts. A complete and reliable measure of a district's infrastructure development from the government's surveys and censuses is hard to obtain. There is no official documentation on infrastructure and public goods spending that is consistent across all districts in our sample, at least for the period of analysis. The most relevant source, which we also utilize, are the Population Censuses that provide information on certain public goods. However, they are conducted decennially, and do not provide information on changes in districts' infrastructure environment in the interim period. In this paper, we overcome this limitation by adopting night-time light intensity measure as a proxy for district's infrastructure development. Nightlight luminosity is obtained from satellite imagery of the earth at night, recording light output at the 30 arc-second level, equivalent to approximately 1 square kilometer at the equator.[14]

For our empirical analysis, we further process the raw GIS digital light raster to obtain taluk-level[15]

[13]For topographic information, we use the GTOPO 30 Arc-Second Elevation global raster data set developed and maintained by U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS).

[14]Satellite images on luminosity at night is collected by the United States Air Force Defense Meteorological Satellite Program (DMSP)'s Operational Linescan System, and then maintained and processed by the National Oceanic and Atmospheric Associations (NOAA). According to the technical description of data collection from NASA, satellites orbit the earth fourteen times a day with a nighttime overpass between 20:30 and 22:00, sending images of every location spanning -180 to 180-degree longitude and -65 to 75-degree latitude at a resolution of 30 arc-seconds. In terms of data processing, the night light images observed for places experiencing the bright half of the lunar cycle, the summer months when the sun sets late, aurora activity (the northern and southern lights), forest fires, or obscured by cloud cover were all excluded from final aggregation. These restrictions effectively remove intense sources of natural light, leaving mostly man-made light. The final product for analysis is a full global set of light intensity pixels, each storing a coded digital number as an integer between 0 (no light) and 63 (top-coded, brightest level). In addition, for the years with more than one satellite orbiting earth and reporting information, we simply average light outcomes across all satellites.

[15]A taluk is an administrative unit below the district level. The analysis at this level thus allows us to still capture within-district variation. At the same time, it also ensures that nightlight raster is still averaged into a sufficiently spanned geographic unit (as opposed to smaller geographic units such as village-level), which reduces the potential existence of spatial gross outliers.

population-weighted light intensity.[16] By design, we assign more weight to light intensity in populated areas where the majority of infrastructure development may take place. Unlit segments that also had low levels of inhabitation receive lesser weight.

The use of night-light as a proxy for economic and infrastructure activities has become popular among economists.[17] There is an overwhelming consensus that light intensity and economic activity are closely related. In addition, Min (2008) shows that there is a strong association between nightlight luminosity and public-goods provision, especially across low-income countries. Particularly in India, Baskaran et al. (2015) further show that nighttime light emission is suitable as a proxy measure for public-service provisions such as electricity. Burlig and Preonas (2016) also use changes in nighttime brightness as an indicator for electrification under RGGVY, a national rural electrification scheme in India.

3.6 Summary Statistics

Table 1 presents the summary statistics of the important outcome and control variables used in this analysis. Across different data sets, we only focus on the rural sample for our analysis. For illustrative purposes, we employ a common bandwidth that consists of all districts located 0.03 backward-score distance around the RD assignment threshold ($|z| \leq 0.03$). As shown in Section 5, this bandwidth is among those for which we report all regression outcomes. It also encompasses the Calonico et al. (2014) data-driven optimal bandwidth associated with each individual outcomes of interest. Our RD restrictive sample covers 115 districts. According to the 2005 Economic Census, there were a total of 6,283,987 enterprises operating in 97,571 villages in our sample. Over 99 percent of the enterprises are categorized as microenterprises. For example, a typical village in our sample had around 50 microenterprises, and only 1 in 5 villages had a formal firm. Furthermore, around 87.5% of total village employment in our sample is concentrated in microenterprises, with 83.39 people working in microenterprises and 11.83 people working in formal firms. We focus on microenterprises in Panel B from the Unorganized Manufacturing frame (Schedule 2.2) of the NSS 2000-01. On average, a microenterprise in our sample employs 2.49 workers, with a standard deviation of 3.04. The small scale of microenterprises should be kept in mind while interpreting the results. With respect to infrastructure-related problems, 17 percent of firms in the sample reported to have experienced a power cut in the year of the survey, and 15 percent reported having no access to raw materials. An average worker in our sample (Panel C) earns 417 Rs/week and works 65 hours/week, and the monthly household consumption expenditure is around 2,500 Rs.

According to the 2001 Demographic Census, 44 percent of villages in an average district in the sample do not have any paved roads, and approximately two-thirds have electricity coverage. The average luminosity levels of night-light emission is also low, recorded at about 2.8 in our sample of districts (out of the top-coded indicator of 63). It is worth mentioning that even though the RD estimates are supposed to be informative about the sub-population at the discontinuity (Lee and Lemieux, 2010), the large number of firms and employment around the threshold provides some generalizable conclusions.

[16] We collect the population raster dataset named Gridded Population of the World, Version 3 (GPWv3) from the Socioeconomic Data and Applications Center (SEDAC) - a Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) - hosted by CIESIN at Columbia University.

[17] See Henderson et al. (2012); Alesina et al. (2016); Chen and Nordhaus (2011); Hodler and Raschky (2014); Klomp (2016); Shenoy (2018).

4 Empirical Strategy

4.1 Reconstruction of Backwardness Score Index

Since RSVY selection process followed a transparent score-based rule, we evaluate the effects of the program using a Regression Discontinuity Design (RD). First, we take the actual number of districts allocated to each of the 17 major states as given. Our main analysis ultimately relies on within-state comparisons of the marginal districts around the state-specific cutoff scores. Therefore, our approach is internally valid when we take the number of districts assigned to each state as-is. Furthermore, we also control for state fixed effects in our empirical specifications. This helps account for any unobserved variation at the state level that might be jointly correlated with both the outcome variables and the district's treatment status.

Next, we reconstruct the entire selection criteria based on Backwardness Index rankings of districts in each state from the second step of the assignment algorithm.[18] Provided with the allotted number of districts by the central government (from the first step), the state governments were supposed to choose the most backward districts for selection, based on the publicly available "Backwardness" Ranking Index. This composite index was constructed from three historical parameters with equal weights i) value of output per agricultural worker (1990-1993); (ii) agricultural wage rate (1996-1997); and (iii) districts' percentage of low-caste populations - Scheduled Castes/ Scheduled Tribes (1991) (Planning Commission, 2003). We perfectly reconstruct the composite score for each district in our sample. We then rank all district scores within each state and generate two important elements: (i) the cutoff score for each state - that is, the score associated with the least backward district that would receive the RSVY grant assigned to the state; and (ii) each district's score distance to the state-specific cutoff, which we refer to as the "re-centered distance score".

From (i) we obtain the full list of districts that should have been granted RSVY funding had there been perfect compliance with the central government's guidelines. This list includes all districts with state-specific backwardness scores below their state's cutoff. More explicitly, the "Backwardness" Index ranking data is available for 447 districts for the 17 major states in India.[19] In our sample of the 147 districts that actually received the RSVY grants, 32 districts were affected by left-wing extremist violence, and their selection was not based on the backwardness index. Having removed these 32 districts from our sample, we are left with 115 districts that received the grants. These 115 districts are shown in Figure 1's Panel A. Out of 115 RSVY districts, 19 (12.9%) belong to states with missing ranking data. To the extent that the actual RSVY assignment to these 19 districts was endogenous, i.e. they were funded without having Backwardness Index information, we remove them from our estimation sample. This leaves us with 96 districts that received RSVY grants and had ranking data available. For these 96 districts, the assignment algorithm had a prediction accuracy of 80.2%, i.e. we correctly predicted 77 of them. These predicted districts are shown in Figure 1's Panel B. Our prediction accuracy is distinctly different from a random draw of districts from the overall pool (21.48%),[20] and provides credence to our approach. Quantitatively, our estimates should

[18] See Appendix for more details.

[19] Data on economic under-development parameters was unavailable for the remaining Indian states classified as "special category" or union territories. Therefore, it is unclear how these state governments selected eligible RSVY districts.

[20] Randomly drawing 96 districts from the pool of 447 districts for which ranking data is available results in a prediction accuracy of 21.48%.

therefore provide a lower-bound of the actual impact of RSVY.

We use the re-centered distance score as the running variable for our RD design. Formally, the re-centered distance score for each district in the sample is defined as follows:

1. For each of the 17 states with available backwardness index data, we use each district's score and denote it as x_{ds}. Subscript d denotes "district" and s denotes "state." x_{ds} is thus a composite index score that is constructed from available under-development parameters. The lower the composite score, the more backward the district.

2. Denoting the state's delegated number of RSVY-eligible districts as k_s, we obtain the cutoff score in state s, which is the index score associated with the k_s^{th} district (i.e. the "cutoff" district) in that state in ascending order of x_{ds}. We denote the cutoff score for state s as x_{ds}^k.

3. We re-center the sequence x_{ds}, so that the cutoff district in the sequence would receive a re-centered distance score of 0. That is:
$$z_{ds} = x_{ds} - x_{ds}^k \tag{1}$$

The district's state-specific re-centered distance score, z_{ds}, serves as the running variable in our subsequent RD regressions. By design, districts to the left of the cutoff – those with non-positive distance scores – are more backward than the state's cutoff district, and should be RSVY-eligible according to the selection rule.

It is worth noting that this process of replicating the central government's assignment formula has been adopted in several papers that study the impacts of NREGA – an employment guarantee program implemented in 2006-07, three years after RSVY (Zimmermann and Khanna, 2017; Zimmermann, 2017; Bhargava, 2014; Hari and Raghunathan, 2017). Compared to these studies, our approach differs in one important dimension. Instead of utilizing the state-specific districts' ordinal ranks as the running variable, we adopt the districts' backwardness scores, using the score distance to cutoff as our running variable. The main advantages of our approach are twofold. From a technical perspective, continuous score distances allow us to deviate from using a discrete running variable in the RD framework. Adopting discrete rankings as a running variable essentially limits the available choices of bandwidth size in estimation, and/or the ability to obtain reliable estimates of the Average-Treatment-Effect (ATE) or the associated standard errors (Lee and Card, 2008; Kolesár and Rothe, 2018). The second advantage of employing the distance score as a running variable pertains to sample selection of districts close to the cutoff for estimation purposes. An important identifying assumption in our context requires that districts with similar composite backwardness scores are comparable in both observed and unobserved characteristics, in the absence of RSVY grants. It is possible that the ordinal rank variable may not adequately satisfy this identification assumption. For instance, a district A might possess a composite score significantly higher than the score of the "cutoff" district B in the same state. It might be the case that there are no other districts with the backwardness score in between A and B, so that the re-centered under-development rank of district A becomes +1 (i.e. one ordinal rank above the cutoff district in the state). This consequently means that the unsuitable district A would always be included in the estimation sample's control group, even when using the most conservative bandwidth using re-centered rank as the RD running variable (e.g. ±1 rank from the cutoff). However, adopting distance scores as the running variable with restrictive bandwidth around the cutoff would allow for the exclusion of this unsuitable district A from the estimation sample, thereby providing cleaner causal estimates.

4.2 Empirical Design

Our empirical analysis follows the parametric Regression Discontinuity Design functional form as suggested by Imbens and Lemieux (2008):

$$y_{idst} = \alpha_0 + \alpha_1 RSVY_{ds} + \delta(z_{ds}) + \alpha_2 X^1_{dt-1} + \alpha_3 X^2_d + \gamma(X_{isdt}) + \pi_s + \varepsilon_{idst} \tag{2}$$

where the subscripts refer to a firm (or village/household/individual)–level observation i, in district d, in state s, in year t. Thus, y_{idst} is the firm (or village/household/individual)–level outcome variables of interest. $RSVY_{ds}$ is an indicator representing actual treatment status, that equals one, if the district was selected to receive the RSVY grant. z_{ds} is the constructed re-centered score distance discussed in the previous section, which serves as the running variable in our RD design. Following Gelman and Imbens (2019), $\delta(z_{ds})$ is a polynomial function of the score variable that allows for both linear and quadratic specifications. Also, since cut-offs are state-specific, we control for π_s, the state fixed effects, in all specifications.

We further include two district-level vectors of predetermined variables X^1_{dt-1} and X^2_d. Vector X^1_{dt-1} includes a series of district's socio-demographic characteristics at the baseline. It always includes the three components from which the backwardness rankings were calculated – value of output per agricultural worker (1990-1993), agricultural wage rate (1996-1997), and districts' percentage of low-caste populations - Scheduled Castes/ Scheduled Tribes. The other covariates in X^1_{dt-1} depend on whether the regression is at the village, firm, household, or individual level. For our village-level regressions using the Economic Census, we include village-level controls from the 2001 Population Census such as total village population, whether the village had paved roads, whether the village had electricity, and the share of area of the village that is irrigated. For our firm-level regressions using the NSS manufacturing surveys, and the Annual Survey of Industries, or the individual and household regressions using the NSS employment-unemployment schedule,[21] we include district-level variables from the 2001 Population Census such as district population, share of villages in the district that had paved roads, share of villages in the district that had electricity, and the share of irrigated area in the district. Vector X^2_d further includes the district's time-invariant covariates: area, boundary perimeter, elevation, and distance to the nearest metropolitan cities.[22] Additionally, X_{isdt} includes different sets of covariates depending on the regression. For the NSS manufacturing firm-level regressions, covariates include the microenterprises' physical operating structure (inside or outside the household, whether with fixed premises or not), and owner's gender as well as highest education level. Furthermore, in the NSS firm-level regressions we also control for industry fixed effects. For the individual and household level regressions we control for age, religion, education, occupation, social group, and religion. Finally, ε_{idst} is a stochastic error term clustered at the district-level.

Estimating equation (2) would likely produce biased estimates of the effects of the policy, since actual treatment $RSVY_{ds}$ may be endogenous. We therefore use the predicted treatment indicator $\mathbf{1}\{z_{ds} \leq 0\}$.

[21] In the NSS surveys and the ASI data set, the lowest level at which the unit is identified is the district.

[22] We define a metropolitan area to be any city in India with a total population of at least 500,000 based on the 2001 Census. We use two measures for a district's nearness to metro areas: (i) distance to the nearest metro city, and (ii) average distance to the nearest 5 cities.

This binary treatment variable is assigned a value of one to a district with a non-positive state-specific score distance to the cutoff, hence economically backward enough to be eligible for RSVY under the assignment guideline. We run regressions of the form:

$$y_{idst} = \beta_0 + \beta_1 \mathbf{1}\{z_{ds} \leq 0\} + \delta(z_{ds}) + \beta_2 X^1_{dt-1} + \beta_3 X^2_d + \gamma(X_{isdt}) + \pi_s + \varepsilon_{idst} \tag{3}$$

The main coefficient of interest is β_1, which is associated with the predicted treatment status $\mathbf{1}\{z_{ds} \leq 0\}$. This coefficient represents the discontinuous changes in outcomes between treated and comparison districts located close to the cutoff. Under the standard RD identification assumption that marginal districts around the discontinuity are as good as random, β_1 represents the Local Average Intent-to-Treat ("ITT") effect of the policy.

In the Appendix section, we also estimate the Treatment-on-the-Treated ("TOT") effects. Specifically, we run instrumental variable regressions in the form of Fuzzy Regression Discontinuity, instrumenting $RSVY_{ds}$ in equation (2) with $\mathbf{1}\{z_{ds} \leq 0\}$. The first stage of the Fuzzy RD requires that there is a discontinuity in the probability of receiving RSVY at the cutoff. Figure 2 shows this discontinuity graphically. It plots the probability of receiving RSVY as a function of the running variable (re-centered score distance). The graph also provides quadratic fitted curves and the corresponding 95 percent confidence intervals on both sides of the cutoff. It is visually clear that the average probability of receiving RSVY decreases discretely to the right of the cutoff.

Another key aspect related to regression discontinuity designs is the choice of bandwidth. We test for the sensitivity of our RD estimates on bandwidth selection by reporting the coefficients across three bandwidths, (0.02, 0.025, 0.03), around our RD threshold. We further show in our sensitivity analysis that the Calonico et al. (2014) data-driven optimal bandwidth associated with all of our main outcomes of interest lie within this [0.02, 0.03] range.

4.3 Validating the identification assumptions

Treatment assignment at the threshold is only "as good as random" when the polynomial function of the running variable is smooth, or continuous, across the RD threshold. In essence, districts must not be able to manipulate their relative backwardness scores so as to determine their treatment status. This assumption is reasonable because the backwardness score index was constructed using historical development parameters collected in the early 1990s, roughly a decade before the introduction of the RSVY program, thus limiting the possibility of districts strategically misreporting information. Regardless, we visually check for treatment status manipulation by looking at Figure 3. This figure plots the distribution of districts over the re-centered distance score measure. We also conduct a McCrary (2008) density test for potential manipulation of the running variable. If there was strategic manipulation, we should have seen visual evidence of "bunching" in the density of the assignment variable to the left of the treatment cutoff. Figure 3 shows no such bunching and the kernel density function of the re-centered distance scores is smooth around the threshold. Consistent with the visual evidence, the McCrary (2008) test does not reject the null hypothesis of no discontinuity in the density of districts.

Another potential threat to identification would be if there were contemporaneous public programs with a similar development focus that were also implemented on the basis of the district backwardness ranking index. To the best of our knowledge, no such district-level program existed during this time. The RSVY program was the first national public infrastructure development initiative that the Government of India introduced, that adopted a transparent assignment formula on the basis of the backwardness index. The other large-scale public/development projects that used the backwardness index to determine eligibility of districts were the Backward Regional Grant Fund (BRGF), and the National Rural Employment Guarantee Act (NREGA). BRGF in fact, subsumed the RSVY program, and was introduced in 2007. It extended the total number of eligible districts for infrastructure cash grants to 250 districts. The first phase of NREGA was implemented in April 2006, covering the 200 most backward districts. Both programs started at least two years after the introduction of RSVY. Hence, these programs do not contaminate our results at least in the sample considered in our paper.

However, two other village-level infrastructure programs were also introduced by the Government of India in the 2000s – (i) Pradhan Mantri Gram Sadak Yojana (Prime Minister's Village Road Program, or PMGSY) introduced in 2000 and (ii) Rajiv Gandhi Grameen Vidyutikaran Yojana (Prime Minister's Rural Electrification Program or RGGVY) introduced in 2005.[23] Both these programs were implemented based on village-level population cutoffs. For example, PMGSY targeted roads to villages with population exceeding two discrete thresholds (500 and 1,000), and RGGVY targeted electrification to villages with a population larger than 300 people. Multiple reasons suggest that these two programs do not affect our empirical setting and results. First, the source of identifying variation for RSVY is the district-level distance to cutoff score, which is different than village-level population cutoffs for PMGSY and RGGVY. Secondly, in our regressions, we control for village population. Finally, for PMGSY and RGGVY to affect our results, it must be the case that the number or share of villages with these population cutoffs (300, 500, 1000) must be differentially higher in RSVY treated districts than in the control districts. We check for this in Appendix Table A1 and do not find any such differences with respect to PMGSY and RGGVY eligible villages.

5 Results

We begin by presenting results on the effects of RSVY on employment at the village and firm level from the Economic Census (EC). Then, we discuss results on the effects of the policy on microenterprises (NSS) and on large firms (ASI). Next, we present results on the extensive margin, i.e. the number of firms in villages. We then corroborate the effect of RSVY on firm and village level outcomes with evidence on individuals and households. After discussing the main findings, we provide evidence on plausible mechanisms driving these effects. Finally, we conduct robustness tests for our results. For all the outcome variables, we present intent-to-treat (ITT) estimates using linear and quadratic functions of the running variable, across three alternate bandwidths – 0.02, 0.025, and 0.03. The tables for the main results also show the mean and standard deviations for the control group, and all results are interpreted with respect to these. In the Appendix, we show the corresponding treatment-on-treated (TOT) estimates. For brevity, we present our TOT results for one representative bandwidth (0.025) across linear and quadratic specifications.

[23] See Asher and Novosad (forthcoming) and Aggarwal (2018) for more details on PMGSY and Burlig and Preonas (2016) for more details on RGGVY.

5.1 Village and Firm-level Results (EC 2005)

In Table 2 Panel A, we look at the ITT effects of RSVY on employment at the village-level from the Economic Census 2005. Note that these are short-term impacts of the policy, measured approximately a year and a half after the introduction of RSVY. We find that total employment in all firms in villages increased by 11.1%-17.5% across columns 1 through 5 in RSVY-treated districts relative to control districts. The results are quantitatively similar across the different bandwidth choices and specifications (linear and quadratic), and statistically significant for bandwidths of 0.025 and 0.03. This is equivalent to an increase of 11-18 jobs in firms per village (or an increase of 0.07-0.11 standard deviations in village employment in firms). To understand whether the effects of RSVY are different across the firm size distribution, we further divide village employment in all firms into employment in microenterprises and in formal firms. We find that the village level employment in all firms is almost completely due to employment in microenterprises. Village employment in microenterprises increases in RSVY districts relative to control districts by 11.5%-18.3% across the columns and is statistically significant for bandwidths of 0.025 and 0.03. This is equivalent to an increase of 10-16 jobs in microenterprises per village (or an increase of 0.08-0.12 standard deviations in village employment in microenterprises). However, there is no statistically significant change in village employment in formal firms between treated and control districts. This is not surprising given that RSVY was a rural infrastructure program, and in our rural sample, a village in the RSVY districts only had about 0.22 formal firms. The corresponding TOT results for village-level employment across the firm-size distribution is shown in Appendix Table A2. We find that both village-level employment in all firms and microenterprises increase, but there is no change in employment in formal firms. The right-hand panel of Figure 4(a) presents the graphical representations of the results from Table 2 Panel A. Each scattered point in the graph represents bin-averaged values of log of village-level employment after partialling out the state fixed effects and the district-level covariates used in equation (3).[24] The left-hand panel shows the visual representation of the results using pre-RSVY data from 1998, where we find no discontinuous jumps at the cutoff.

In Panel B, we look at the results from firm-level regressions using EC 2005. We find that on average, firm-level employment in RSVY districts increases by 5.5%-6.6%. The results are statistically significant across bandwidths and specifications. Next, we look at the effect of RSVY on microenterprises. We find that, on average, employment in a microenterprise increases by 5.2%-6.2%. These results are also statistically significant across bandwidths and specifications. Finally, we find no change in employment for formal firms. Taken together, we find that employment in RSVY districts increased relative to control districts, but were only driven by increases for microenterprises. There was no impact on employment for formal firms. We look at firm-level results in more detail next.

5.2 Firm-level Results (NSS Schedule 2.2)

In Table 3, we look at the effects of RSVY on microenterprises from the NSS manufacturing surveys (round 62). In Panel A, we look at firm-level employment (in logs) for microenterprises, about two years after RSVY was introduced. We find that microenterprises in RSVY districts saw an increase in employment by 8.5%-13.2%. These results are qualitatively similar to the firm-level results from the Economic Census

[24]Specifically, we residualize the firm's outcomes on the terms X^1_{dt-1}, X^2_d and π_s which were included in equation (3).

previously discussed. These coefficients translate to an increase of 0.22-0.34 employees per microenterprise. In Panel B, we look at the effects on employment (in levels) of microenterprises. We find that firm-level employment increased by 0.465-0.675 workers in RSVY districts compared to control districts. For both employment measures, the coefficients are statistically significant across bandwidths and specifications. The corresponding TOT results are also statistically significant in Appendix Table A2. Finally, the right-hand panel of Figure 4(b) visually represents the results along with the discontinuity at the cutoffs. The left-hand panel shows the visual representation of the results using pre-RSVY data from 2001, where we find no discontinuous jumps at the cutoff.

5.3 Firm-level Results (ASI)

The results using the EC discussed previously suggest that there was an increase in employment in firms in RSVY districts, driven primarily by microenterprises. The NSS results also confirmed that microenterprises in RSVY districts increased in size, in terms of employment. Additionally, we also saw from the EC results that formal firms saw no increase in employment. To provide further credence to this, we investigate the effects of RSVY on formal firms in the ASI. In Table 4 we look at the effects of the policy on employment for formal firms for all years between 2001-2005.[25] Across all the years, bandwidths, and specifications (linear or quadratic), we do not find any effects on formal firm employment in RSVY districts compared to control districts. These results support our main claim that RSVY caused an increase in overall employment that was completely due to increases in size of microenterprises. Formal firms across treated and control districts remained unaffected.

5.4 Extensive margin – number of firms

Thus far, we have have seen that there was a microenterprise-driven growth in districts that received the RSVY grants. However, these results are only related to the intensive margin of the policy. Districts that received RSVY grants may also have experienced an increase in the number of new firms – the extensive margin. We look at the results on the extensive margin due to the policy next. In Table 5 Panel A we use the EC and look at the total number of firms in villages, the number of microenterprises in villages, as well as the number of formal firms in villages. We find that the total number of firms differentially increased in villages in RSVY districts relative to control districts by 8.9%-18%. In numbers, these coefficients translate to an increase of 4.7-9.4 new firms in villages in RSVY districts. Not surprisingly, similar to our earlier results, this growth was completely driven by an increase in the number of microenterprises. The number of microenterprises in RSVY districts increased by 8.8%-18.1%. These correspond to an increase in 4.7-9.4 microenterprises in villages in RSVY districts. In contrast, we find no effects of RSVY on the number of formal firms. The graphs corresponding to these results are shown in the right-hand panel of Figure 5. The left-hand panel shows the visual representation of the results using pre-RSVY data from 1998.

Next, in Panel B, we turn to the NSS manufacturing survey on microenterprises. In this survey, there is one particular question that sheds light on the age of the microenterprise. This question asks whether the firm

[25] ASI is a yearly panel data set for formal firms.

had been established less than 3 years ago. Since the data covers 2005-06, i.e. two years after RSVY began, these results are a slight overestimate of the effects of the policy on new firms. However, we find evidence corroborating the results from Panel A. We find that the proportion of microenterprises in RSVY districts that were established less than 3 years ago was 7.8%-10.7% higher than in control districts. These results are also significant across bandwidths and specifications. The corresponding TOT estimates for results from Panel A and B are reported in Appendix Table A2.

5.5 Individual and Household Results (NSS Schedule 10)

The firm and village level results show that RSVY led to an increase in employment for microenterprises. Next, we look at the effects of RSVY on individual wages, and the number of days worked in the last 7 days in Panels A and B of Table 6. Consistent with our previous results, we find that wages increased by 10.1%-13.5%, and number of days worked increased by 3%-3.3% for individuals in RSVY districts relative to control districts in the short run (about 2 years after the policy). These results are statistically significant across bandwidths and specifications. Furthermore, since microenterprises are very often based out of the household, an increase in firm size should directly translate in to an increase in the household consumption expenditure. We test for this in Panel C. We find that across bandwidths and specifications, monthly household consumption expenditure increases by 8.7%-12.2% and is statistically significant. These coefficients translate in to an increase of 41–55 Rupees/week in wages and about 210–301 Rupees/month increase in monthly household consumption expenditure. The individual and household results thus provide additional evidence on the effects of RSVY, especially on microenterprises.

5.6 Mechanisms

RSVY cash grants were provided to foster infrastructural development in the backward districts. Therefore, one would expect that a main channel driving the effects on microenterprises is through direct improvements in the overall infrastructural environment in the treated districts. First, we proxy for districts' level of infrastructure development by night-time light luminosity. We then estimate the RD coefficients for each year between 1998 and 2013 separately. We find positive and significant growth in nightlight density in treated districts, almost immediately after the introduction of RSVY. The effects of RSVY on nightlights is graphically shown in Figure 6, with each point representing the coefficient on the RD estimate for the given year. The corresponding estimates are reported in Appendix Table A5, where due to space constraints we only show the coefficients between 2001 and 2010. We find that night-lights started to grow differentially for the treated districts, almost immediately after policy introduction in 2004. The statistically significant impact lasted for four to five subsequent years and dissipated around 2009. This coincides with the period when RSVY was in effect. The reversal in trends after 2008 is most likely due the introduction of the Backward Region Grants Fund (BRGF), another program with grants for infrastructure, that followed RSVY after 2007, and increased coverage to more backward districts. BRGF followed an identical selection process as RSVY, and essentially converted a majority of the control districts in our analysis into treated districts under the new policy.

Since the main beneficiaries of RSVY were microenterprises, we focus on them and check whether they benefited from these overall infrastructure improvements. To answer this, we focus on two measures from the NSS manufacturing survey in Table 7. First, we look at firms' responses to the question on the nature of problems faced, if any, during the reference year with respect to power cuts. We interpret the responses to this question as a measure of electrification. Second, we look at whether firms had problems with access to raw materials. Access to raw materials could increase either because there is an increase in firms supplying these raw materials in the same location or because there is better connectivity to suppliers in other locations. We interpret the responses to this question as a measure of road connectivity. In Panel A, we find that in RSVY districts, the probability of a firm experiencing a power cut decreased by 7.2%-13.9%. These coefficients are statistically significant for bandwidths of 0.02 and 0.03. In Panel B, we find that there is a 10.3%-18.3% decline in the probability that firms had no access to raw materials. These coefficients are statistically significant for bandwidths of 0.02 and 0.025. Furthermore, the TOT in Appendix Table A2 estimates are also statistically significant. Overall, the results from Table 7 provide suggestive evidence that RSVY led to improvements in electrification and rural connectivity, that directly impacted microenterprises.

If RSVY-induced improvements in rural electrification and connectivity were indeed important channels for microenterprise growth, the effects should have been differentially higher for firms in electricity and road intensive industries. We test for these channels in Table 8. We use a measure of electricity intensity from Abeberese (2017)[26] in a 3 digit-industry defined as the average kilowatt-hours of electricity consumed per rupee of output by firms in that industry. For road-intensity, we use transport (traveling and freight) expenses at the 3 digit-industry level. Next, we divide industries by their degree of electricity and road intensity into terciles. We expect firms in the highest tercile of electricity intensity to experience a decline in problems related to power cuts and firms in the highest tercile of road intensity to experience a decline in problems of having no access to raw materials the most. In Table 8, we find evidence consistent with our expectations. In Panel A, we find that for firms in the highest tercile of electricity-intensive industries in RSVY districts, problems of experiencing a power cut declined by 19.9%. Firms in the middle and bottom terciles of electricity intensity industries, also saw a decline in power cuts, though the effects are statistically insignificant. In Panel B, we find that firms in industries in the highest tercile of road-intensity saw a decline of 26.2% with respect to not having access to raw materials. The coefficients for firms in the middle and bottom terciles are negative but not statistically significant. These results bolster our claim that RSVY-funded improvements in rural electrification and connectivity were important channels in driving microenterprise growth.

Having seen that firms in the most electricity and road-intensive industries experienced the greatest reductions in infrastructure-related problems, we now check whether these firms also experienced higher growth due to RSVY. In Table 9, we look at the employment responses of firms across terciles of electricity and road-intensive industries to RSVY. In both Panels A (electricity-intensive) and B (road-intensive), we find a consistent pattern. Looking across columns (terciles of electricity and road-intensive industries) we find that employment for firms is the largest in the highest tercile, followed by the middle tercile, and finally the smallest for the lowest tercile, in response to RSVY. This provides further evidence that the effects of RSVY were the largest in electricity and road-intensive industries.

As mentioned earlier, we consider RSVY an infrastructure-enhancing program, with the grants supposed to be based on District Perspective Plans that would identify critical gaps in infrastructure within a district. For this reason, RSVY funds were meant to strengthen rural electricity transmission and distribution rather

[26]We thank Ama Baafra Abeberese for sharing this data with us.

than building new grids in unelectrified villages. Similarly, the funds were to be used to build critical road links rather than building new roads in villages that had no roads. Hence, if RSVY funds were spent effectively, we would expect that village employment in microenterprises as well as number of microenterprises to be higher in villages that already had paved roads and electricity in the baseline (in 2001) compared to villages that were not endowed with such infrastructure. We check for this in Table 10. The results match our expectations. In Panel A, we look at village employment in microenterprises and in Panel B at the number of microenterprises in the village. The sample in columns 1 and 2 is restricted to villages with no paved roads or electricity at the baseline.[27] The sample in columns 3 and 4 includes villages that had paved roads and electricity at the baseline. We find that RSVY did not statistically affect either village employment or the number of microenterprises in villages with no roads or electricity (columns 1 and 2 across Panels A and B). However, RSVY resulted in an increase in village employment in microenterprises by 19%-19.9% (Panel A) and increased the number of microenterprises by 17.9%-18.6% (Panel B) in villages endowed with both paved roads and electricity. This provides further credence that RSVY funds were utilized in filling up critical infrastructure gaps within the district and this directly led to microenterprise growth.

Although RSVY was successful in improving economic outcomes in treated districts, we have shown some evidence of heterogeneous effects across industries (electricity and road-intensive), as well as villages with different levels of infrastructure at baseline. Given that RSVY was a district level infrastructure grants program, another important dimension of heterogeneity is related to "administrative remoteness" – distance from the administrative headquarter. Asher et al. (2018) show that distance to administration matters for the provision of public goods and services, and in turn, affects economic development. In Table 11, we perform a similar test to look at whether there were differential effects of RSVY across villages that were at various distances from the district headquarters. In these regressions, we include district fixed effects, a measure of the distance to the headquarters, and an interaction of RSVY indicator and a measure of distance to the headquarters. The inclusion of district fixed effects implies that our estimates come from variation within districts across villages located at different distances in the treated and control districts. In the table, our main focus is on the interaction of RSVY indicator and a measure of distance to the headquarters. We use two measures of distance – (a) standardized distance from the district headquarters[28] and (b) log distance from the district headquarters. In Table 11, columns 1 through 3 report the results for employment in microenterprises in the village and columns 4 through 6 report the number of microenterprises in the village. Panels A and B include the two different measures of distance mentioned earlier. In columns 1 through 3, across Panels A and B, we find that the coefficient on the interaction of RSVY indicator and a measure of distance to the headquarters is negative and statistically significant across bandwidths of 0.02 and 0.025. We find similar results for number of microenterprises in villages. In columns 4 through 6, across Panels A and B, we find that the coefficient on the interaction of RSVY indicator and a measure of distance to the headquarters is negative. Employment in microenterprises and number of microenterprises was lower in villages further away from the district headquarters than in villages that were closer to it. This provides suggestive evidence on the "administrative remoteness" channel (Asher et al. (2018)) where villages further away from the district headquarters saw fewer RSVY investments and consequently lower microenterprise

[27]We use data from the Village Directory for the 2001 Population Census to define baseline infrastructure.

[28]For each village in a district, we calculate standardized distance from headquarters = (Actual distance of village from headquarters - mean distance for all villages from district headquarters)/standard deviation of distances from headquarters for villages in the district.

growth.

5.7 Robustness Tests

To check for the robustness of our findings, we conduct a sensitivity analysis across various bandwidths, including the Calonico et al. (2014) data-driven optimal bandwidth. The coefficients across the bandwidths for our outcomes of interest are shown in Figures 7 and 8. Both figures illustrate that our results are not sensitive to the choice of bandwidth. Having discussed the results in detail, we also perform two falsification tests, and show that effect of RSVY becomes statistically indistinguishable from zero under counterfactual events. Specifically, we show that the policy had no effect in districts before RSVY was implemented or in districts that did not receive RSVY grants. First, we run the RD regressions with pre-RSVY data for our various outcomes of interest – the fourth round of the Economic Census (1998), NSS manufacturing survey round 56, schedule 2.2 (2000-2001), and NSS employment-unemployment round 55, schedule 10 (1999-2000). In Appendix Table A3, we show the results for this analysis. We find no significant effects of the policy on all main outcomes before RSVY was implemented. The graphical representation of these results are provided in the left-hand panels of Figures 4(a), 4(b), and 5. In the second test, we replicate our regressions by adopting a hypothetical cutoff that is constructed identically to our baseline specifications, but after removing all the treated districts from the sample. Essentially, in this exercise we test whether RSVY grants had any effect on districts that did not actually receive the grants. Overall, in Appendix Table A4, we find no statistically significant effects at these hypothetical cutoffs.[29] Apart from these tests, in Appendix Figures A1 and A2, we also graphically show that there are no discontinuous jumps at the cutoff for district and village-level observable characteristics. This includes the three parameters on which the backwardness index was calculated. These tests provide credibility to our claim that the main effects are indeed caused by the RSVY grants.

6 Conclusion

This paper studies the effects of a rural infrastructure development program, Rashtriya Sam Vikas Yojana (RSVY) on the performance of firms in India's backward districts. We exploit the transparent treatment-assignment mechanism of RSVY, that uses a district backwardness score index allowing us to reconstruct state-specific score cutoffs. We estimate the effects of the policy using a regression discontinuity design across the entire firm-size distribution. We find that RSVY led to increases in village employment, entirely driven by employment in microenterprises. We find corresponding increases in the number of microenterprises in villages, again driven by increases in the number of microenterprises. Our firm-level results similarly suggest that microenterprises in RSVY districts saw an increase in employment. There were no changes for formal firms.

Our empirical results also shed light on potential mechanisms underlying the effect on firm outcomes. We

[29] Note that in these regressions, after removing our original the treated districts, we are essentially comparing outcomes in districts that are not similar in terms of observables. In fact, districts to the right of the hypothetical cutoff are better off than districts on the left of the cutoff. Hence, as expected, for most outcomes we find a negative coefficient.

show that RSVY cash grants directly improved the infrastructural development of treated districts, measured using night-light luminosity. This improvement was in turn realized by firms, who reported significantly lower likelihood of power cuts and problems with no access to raw materials. Furthermore, our microenterprise-level results are stronger for firms in electricity and road-intensive industries. We also find that employment in microenterprises and number of enterprises are higher in villages that had roads and electricity prior to RSVY. Overall, the results suggest that both RSVY-funded rural electrification and connectivity were important channels that led to microenterprise growth.

Although far from a rigorous welfare analysis, we calculate a back-of-the-envelope cost per job for the program. Using our most conservative estimates, we find that RSVY had a "cost per job" of USD 3,751 (at 2010 prices).[30] This is comparable to the cost per job for the Regional Selective Assistance program in the United Kingdom of USD 3,541 (Criscuolo et al., 2019), but smaller than other government spending programs in the US (Suárez Serrato and Wingender, 2016). This also suggests that rural infrastructure programs in developing countries may be a cost-effective way to generate employment, especially in small firms.

A limitation of our analysis is that we are only able to look at the short-term effects of infrastructure grants. Especially, for developing countries with large infrastructure gaps, studying the long-run effects of infrastructure investments is critical. Firms could gain from public investments in roads, electrification, dams, better telecommunication (Internet, mobile telephone networks), and other investments. Among the plethora of options available to policymakers, which ones should be prioritized is a very important question. These are promising avenues for future research.

References

Abeberese, A. B. (2017): "Electricity Cost and Firm Performance: Evidence from India," *Review of Economics and Statistics*, 99, 839–852.

Adukia, A., S. Asher, and P. Novosad (forthcoming): "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction," *American Economic Journal: Applied Economics*.

Aggarwal, S. (2018): "Do Rural Roads Create Pathways out of Poverty? Evidence from India," *Journal of Development Economics*, 133, 375–395.

Aker, J. C. and I. M. Mbiti (2010): "Mobile Phones and Economic Development in Africa," *Journal of Economic Perspectives*, 24, 207–32.

Alder, S., L. Shao, and F. Zilibotti (2016): "Economic Reforms and Industrial Policy in a Panel of Chinese Cities," *Journal of Economic Growth*, 21, 305–349.

[30] Our coefficient estimate from Table 2's column 5 (bandwidth of 0.03) suggests a 14.4% increase in village employment. Using the lower bound of the 90% confidence interval, we find suggestively an increase of 1.73% (0.144-1.645*(0.077)). This translates to 1.77 jobs per village. In our sample, with a bandwidth of 0.03, the number of villages per RSVY district is 1,781. Multiplying the two gives us approximately 3,149 additional jobs per district. To calculate the total costs, we use the amount received by each RSVY district (450 million Rupees) and multiply it by a marginal cost of public funds of 1.2 (Chaurey, 2017). Finally, we convert it to 2010 USD at an average exchange rate of 45.71 INR/USD, to make our estimates comparable to (Criscuolo et al., 2019). Cost per job, then, is simply the total cost/additional jobs.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016): "Ethnic Inequality," *Journal of Political Economy*, 124, 428–488.

Allcott, H., A. Collard-Wexler, and S. O'Connell (2016): "How Do Electricity Shortages Affect Industry? Evidence from India," *American Economic Review*, 106, 587–624.

Alsan, M. and C. Goldin (2019): "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 18801920," *Journal of Political Economy*, 127, 586–638.

Asher, S., K. Nagpal, and P. Novosad (2018): "The Cost of Distance: Geography and Governance in Rural India," *Working Paper*.

Asher, S. and P. Novosad (forthcoming): "Rural Roads and Structural Transformation," *American Economic Review*.

Banerjee, A. (2013): "Microcredit Under the Microscope: What Have We Learnt in the Last Two Decades, What Do We Need to Know," *Annual Review of Economics*, 5, 487–519.

Banerjee, A., D. Karlan, and J. Zinman (2015): "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, 7, 2015.

Baskaran, T., B. Min, and Y. Uppal (2015): "Election Cycles and Electricity Provision: Evidence from a Quasi-Experiment with Indian Special Elections," *Journal of Public Economics*, 126, 64–73.

Becker, S. O., P. H. Egger, and M. von Ehrlich (2010): "Going NUTS: The effect of EU Structural Funds on regional performance," *Journal of Public Economics*, 94, 578 – 590.

——— (2012): "Too much of a good thing? On the growth effects of the EU's regional policy," *European Economic Review*, 56, 648 – 668.

Bhargava, A. K. (2014): "The Impact of India's Rural Employment Guarantee on Demand for Agricultural Technology," *IFPRI Discussion Paper 01381*.

Bloom, N. and J. van Reenen (2007): "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 122, 1351–1408.

Bondonio, D. and R. Greenbaum (2007): "Do local tax incentives affect economic growth? What mean impacts miss in the analysis of enterprise one policies," *Regional Science of Urban Economics*, 37, 121–136.

Bronzini, R. and G. de Blasio (2006): "Evaluating the Impact of Investment Incentives: The Case of Italy's law 488/1992," *Journal of Urban Economics*, 60, 327–349.

Brooks, W. and K. Donovan (2019): "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua." *Working Paper*.

Bruhn, M., D. Karlan, and A. Schoar (2018): "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico," *Journal of Political Economy*, 126, 635–687.

Bruhn, M. and B. Zia (2013): “Stimulating managerial capital in emerging markets: the impact of business training for young entrepreneurs,” *Journal of Development Effectiveness*, 5, 232–266.

Burlig, F. and L. Preonas (2016): “Out of the Darkness and Into the Light? Development Effects of Rural Electrification,” *Working Paper*.

Busso, M., J. Gregory, and P. Kline (2013): “Assessing the Incidence and Efficiency of a Prominent Place Based Policy,” *American Economic Review*, 103, 897–947.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 82, 2295–2326.

Chakravorty, U., M. Pellib, and B. U. Marchand (2014): “Does the Quality of Electricity Matter? Evidence from Rural India,” *Journal of Economic Behavior & Organization*, 107, 228–247.

Chaurey, R. (2017): “Location-Based Tax Incentives: Evidence from India,” *Journal of Public Economics*, 156, 101–120.

Chen, X. and W. D. Nordhaus (2011): “Using Luminosity Data as a Proxy for Economic Statistics,” *Proceedings of the National Academy of Sciences*, 108, 8589–8594.

Cheng, Y. (2014): “Place-based Policies in a Development Context - Evidence from China,” *Working Paper, UC Berkeley*.

Cole, S., T. Sampson, and B. Zia (2011): “Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets,” *Journal of Finance*, 66, 1933–1967.

Criscuolo, C., R. Martin, H. G. Overman, and J. Van Reenen (2019): “Some Causal Effects of an Industrial Policy,” *American Economic Review*, 109, 48–85.

de Mel, S., D. McKenzie, and C. Woodruff (2008): “Returns to Capital in Microenterprises: Evidence from a Field Experiment,” *Quarterly Journal of Economics*, 123, 1329–1372.

Devoto, F., E. Duflo, P. Dupas, W. Parient, and V. Pons (2012): “Happiness on Tap: Piped Water Adoption in Urban Morocco,” *American Economic Journal: Economic Policy*, 4, 68–99.

Dinkelman, T. (2011): “The Effects of Rural Electrification on Employment: New Evidence from South Africa,” *American Economic Review*, 101, 3078–3108.

Donaldson, D. (2018): “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, 108, 899–934.

Drexler, A., G. Fischer, and A. Schoar (2014): “Keeping It Simple: Financial Literacy and Rules of Thumb,” *American Economic Journal: Applied Economics*, 6, 1–31.

Duflo, E. and R. Pande (2007): “Dams,” *The Quarterly Journal of Economics*, 122, 601–646.

Faber, B. (2014): “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” *The Review of Economic Studies*, 81, 1046–1070.

Gelman, A. and G. Imbens (2019): "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37, 447–456.

Ghani, E., A. G. Goswami, and W. R. Kerr (2016): "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 126, 317–357.

Givord, P., R. Rathelot, and P. Sillard (2013): "Place-based Tax Exemptions and Displacement Effects: An Evaluation of the Zones Franches Urbaines Program," *Regional Science Urban Economics*, 43, 151–163.

Glaeser, E. and J. Gottlieb (2008): "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 155–239.

Greenbaum, R. and J. Engberg (2004): "The Impact of State Enterprise Zones on Urban Manufacturing Establishments," *Journal of Policy Analysis and Management*, 23, 315–339.

Ham, J., C. Swenson, A. Imrohoroglu, and H. Song (2011): "Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Communities," *Journal of Public Economics*, 95, 779–797.

Hardy, M. and J. McCasland (2017): "Light off, Light on: The Effects of Electricity Shortages on Small Firms," *Working Paper*.

Hari, S. and K. Raghunathan (2017): "Providing more than just Employment? Evidence from NREGA in India," *Working Paper*.

Hasan, R., Y. Jiang, and R. M. Rafols (2017): "Place-Based Preferential Tax Policy and Its Spatial Effects: Evidence from India's Program on Industrially Backward Districts," *Working Paper*.

Henderson, V., A. Storeygard, and D. N. Weil (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102, 2012.

Hjort, J. and J. Poulsen (2019): "The Arrival of Fast Internet and Employment in Africa," *American Economic Review*, 109, 1032–79.

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

Hsieh, C.-T. and P. J. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 129, 1035–1084.

Hsieh, C.-T. and B. A. Olken (2014): "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 28, 89–108.

Imbens, G. W. and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615 – 635.

Jensen, R. (2007): "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector*," *The Quarterly Journal of Economics*, 122, 879–924.

Karlan, D., R. Knight, and C. Udry (2015): "Consulting and Capital Experiments with Microenterprise Tailors in Ghana," *Journal of Economic Behavior & Organization*, 118.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014): "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics*, 129, 597–652.

Karlan, D. and J. Zinman (2009): "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts," *Review of Financial Studies*, 23, 433–464.

Kline, P. and E. Moretti (2014): "Local Economic Development, Agglomeration Economies, and the Big Push: 100 years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 275–331.

Klomp, J. (2016): "Economic Development and Natural Disasters: A Satellite Data Analysis," *Global Environmental Change*, 36, 67–88.

Kolesár, M. and C. Rothe (2018): "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108, 2277–2304.

Lee, D. S. and D. Card (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 655–674.

Lee, D. S. and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Lee, K., E. Miguel, and C. Wolfram (forthcoming): "Experimental Evidence on the Economics of Rural Electrification," *Journal of Political Economy*.

Lenz, L., A. Munyehirwe, J. Peters, and M. Sievert (2017): "Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwandas Electricity Access Roll-Out Program," *World Development*, 89, 88 – 110.

Lipscomb, M., A. M. Mobarak, and T. Barham (2013): "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil," *American Economic Journal: Applied Economics*, 5, 200–231.

Lu, Y., J. Wang, and L. Zhu (forthcoming): "Place-based Policies, Creation, and Displacement: Evidence from China's Economic Zone Program," *American Economic Journal: Economic Policy*.

Mayer, T., F. Mayneris, and L. Py (2015): "The Impact of Urban Enterprise Zones on Establishment Location Decisions and Labor Market Outcomes: Evidence from France," *Journal of Economic Geography*, 17, 709–752.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

McKenzie, D. (2017): "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition," *American Economic Review*, 107, 2278–2307.

McKenzie, D. and C. Woodruff (2017): "Business Practices in Small Firms in Developing Countries," *Management Science*, 63, 2967–2981.

Min, B. (2008): "Democracy and Light: Electoral Accountability and the Provision of Public Goods," *Working Paper, mimeo UCLA*.

Nataraj, S. (2011): "The Impact of Trade Liberalization on Productivity: Evidence from India's Formal and Informal Manufacturing Sectors," *Journal of International Economics*, 85, 292–301.

Neumark, D. and J. Kolko (2010): "Do Enterprise Zones Create Jobs? Evidence from California's Enterprise Zone Program," *Journal of Urban Economics*, 68, 1–19.

Neumark, D. and H. Simpson (2015): "Chapter 18 - Place-Based Policies," in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. C. Strange, Elsevier, vol. 5 of *Handbook of Regional and Urban Economics*, 1197 – 1287.

Planning Commission (2003): "Backward Districts Initiative - Rashtriya Sam Vikas Yojana - The Scheme and Guidelines for Preparation of District Plans," Tech. rep., Government of India.

Program Evaluation Organization (2010): "Evaluation Study on Rashtriya Sam Vikas Yojana (RSVY)," Tech. rep., Government of India.

Rotemberg, M. (forthcoming): "Equilibrium Effects of Firm Subsidies," *American Economic Review*.

Rud, J. P. (2012): "Electricity Provision and industrial Development: Evidence from India," *Journal of Development Economics*, 97, 352–367.

Shenoy, A. (2018): "Regional Development through Place-Based Policies: Evidence from a Spatial Discontinuity," *Journal of Development Economics*, 130, 173–189.

Suárez Serrato, J. C. and P. Wingender (2016): "Estimating Local Fiscal Multipliers," Working Paper 22425, National Bureau of Economic Research.

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

World Bank (1994): "World Development Report 1994: Infrastructure for Development," Tech. rep., World Bank.

Zimmermann, L. (2017): "Why Guarantee Employment? Evidence from a Large Indian Public-Works Program," *Working Paper*.

Zimmermann, L. and G. Khanna (2017): "Guns and Butter? Fighting Violence with the Promise of Development," *Journal of Development Economics*, 124, 120–141.

Table 1: Summary Statistics for the main variables (Representative Bandwidth of $|z| \leq 0.03$)

	Observations	Mean	SD	Source
Panel A: Village Employment and Firms				
1. Employment Outcomes (per village):				
Total Employment	97,571	95.21	232.22	EC 2005
Formal Employment	97,571	11.83	80.44	EC 2005
Informal Employment	97,571	83.39	192.03	EC 2005
2. Firm Count (per village):				
Total Firms	97,571	49.82	109.70	EC 2005
Formal Firms	97,571	0.22	1.34	EC 2005
Informal Firms	97,571	49.60	109.20	EC 2005
3. Employment Outcomes (per firm):				
Total Employment	6,283,987	1.94	7.40	EC 2005
Formal Employment	6,283,987	53.95	91.46	EC 2005
Informal Employment	6,283,987	1.69	1.42	EC 2005
Panel B: Firms				
1. Microenterprises:				
Employment (level)	8,912	2.49	3.04	NSS 62 - Sch. 2.2
Problem of Experiencing Power Cut (%)	8,912	0.17	0.38	NSS 62 - Sch. 2.2
Problem with Access to Materials (%)	8,912	0.15	0.36	NSS 62 - Sch. 2.2
2. Formal Enterprises:				
Employment (level)	2,505	127.01	224.84	ASI 2005
Panel C: Household/Individual Outcomes				
Workload (hours/week)	40,655	65.11	11.84	NSS 62 - Sch. 10
Wage (Rs./week)	40,655	416.87	455.80	NSS 62 - Sch. 10
Consumption Expenditure (Rs./week)	40,655	2,480.81	1,615.35	NSS 62 - Sch. 10
Panel D: Village-Level variables				
Population	97,571	1667.4	1962.9	DC 2001
Paved Roads (%)	97,571	0.56	0.24	DC 2001
Electricity Coverage (%)	97,571	0.67	0.29	DC 2001
Panel E: District-Level variables				
1. Geographic Characteristics:				
Area (km sq.)	115	5,328.36	4,112.39	GADM
Elevation (m)	115	214.80	192.28	GTOPO30
Distance to nearest city (km)	115	116.51	57.13	GADM
Night-light density	115	2.80	3.65	DMSP-OLS
2. Socio-Demographic Characteristics:				
Number of villages per district	115	1,699.41	1,202.35	DC 2001
Share of SC/ST population (% 1991)	115	26.55	12.48	PC 2003
Output per Agricultural Worker (Rs. 1990-93)	115	5,750.84	4,350.64	PC 2003
Agricultural Wage Rate (Rs. 1996-97)	115	32.76	8.42	PC 2003
3. RD Running Variable				
Backwardness Composite Score	115	0.329	0.069	PC 2003
Distance to Cutoffs (z)	115	0.006	0.013	PC 2003

Note: This table shows summary statistics for the main outcomes and control variables used in the analysis. The sample includes all firms operating in the RD restricted bandwidth of districts with the re-centered Backwardness Index Scores (z) within 0.03 point from the cutoff, i.e. $|z| \leq 0.03$. Sources: EC 2005: Economic Census 2005; NSS 62 - Sch. 2.2: National Sample Survey, Round 62 (2005-06) – Unorganized Manufacturing Enterprises; ASI 2005: Annual Survey of Industries (2005); NSS 62 - Sch. 10: National Sample Survey, Round 62 (2005-06) – Employment-Unemployment; DC 2001: Demographic (Population) Census 2001; GADM: Database of Global Administrative Areas; DMSP-OLS: Defense Meteorological Program-Operational Linescan System; PC 2003: Planning Commission (2003).

Table 2: Employment (village-level and firm-level) – Economic Census 2005

	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Village Employment						
1. Total Employment						
RD Estimate	0.120	0.111	0.122	0.139*	0.144*	0.175**
S.E.	(0.0851)	(0.0817)	(0.0805)	(0.0740)	(0.0770)	(0.0733)
Observations	73,335	73,335	83,356	83,356	92,677	92,677
Control Group Mean [SD]	100.78 [249.31]		93.68 [237.27]		102 [257.82]	
2. Microenterprise Employment						
RD Estimate	0.123	0.115	0.129	0.145**	0.149*	0.183**
S.E.	(0.0827)	(0.0792)	(0.0786)	(0.0727)	(0.0756)	(0.0716)
Observations	73,302	73,302	83,313	83,313	92,633	92,633
Control Group Mean [SD]	88.58 [213.77]		81.80 [199.02]		88.37 [213.83]	
3. Formal Employment						
RD Estimate	0.0550	0.0649	0.0170	0.0160	0.0332	0.0184
S.E.	(0.0698)	(0.0728)	(0.0683)	(0.0722)	(0.0573)	(0.0577)
Observations	7,100	7,100	7,627	7,627	8,942	8,942
Control Group Mean [SD]	12.2 [76.25]		11.88 [80.89]		13.63 [89.72]	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Firm-level Employment						
1. All Firms						
RD Estimate	0.0587**	0.0563**	0.0553**	0.0549**	0.0584**	0.0661**
S.E.	(0.0272)	(0.0263)	(0.0264)	(0.0259)	(0.0233)	(0.0252)
Observations	5,642,855	5,642,855	4,921,316	4,921,316	4,602,895	4,602,895
Control Group Mean [SD]	1.91 [7.00]		1.92 [7.29]		1.98 [7.53]	
2. Microenterprises						
RD Estimate	0.0569**	0.0546**	0.0521*	0.0517**	0.0546**	0.0621**
S.E.	(0.0271)	(0.0263)	(0.0264)	(0.0258)	(0.0232)	(0.0249)
Observations	5,617,755	5,617,755	4,900,207	4,900,207	4,582,755	4,582,755
Control Group Mean [SD]	1.67 [1.42]		1.67 [1.43]		1.70, [1.44]	
3. Formal Firms						
RD Estimate	-0.0324	-0.0310	-0.0587	-0.0583	-0.0541	-0.0427
S.E.	(0.0446)	(0.0446)	(0.0491)	(0.0491)	(0.0492)	(0.0497)
Observations	25,100	25,100	21,109	21,109	20,140	20,140
Control Group Mean [SD]	51.02 [84.45]		52.01 [87.98]		54.33 [87.99]	
RD Bandwidth ($\|z\|$)	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: Panels A and B respectively report village-level and firm-level regression results on employment impacts. Odd columns show estimates from first-order RD polynomial specifications and even columns second-order polynomial. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) and baseline backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population). Village covariates include baseline total population (log), share of irrigated area, paved road coverage, and electricity coverage. Standard errors are clustered at the district level.

Table 3: Effects on Microenterprises – NSS (Schedule 2.2) 2005-06

	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Employment (log)						
RD Estimate	0.124**	0.132***	0.0975*	0.0939*	0.0921**	0.0856**
S.E.	(0.0500)	(0.0490)	(0.0511)	(0.0530)	(0.0410)	(0.0419)
R-square	0.345	0.346	0.342	0.342	0.349	0.350
Observations	6,758	6,758	7,579	7,579	8,580	8,580
Panel B: Employment (level)						
RD Estimate	0.675***	0.672***	0.558***	0.566***	0.479***	0.465***
S.E.	(0.212)	(0.201)	(0.200)	(0.209)	(0.165)	(0.169)
R-square	0.230	0.230	0.237	0.237	0.240	0.241
Observations	6,758	6,758	7,579	7,579	8,580	8,580
Control Group Mean [SD]	2.55 [3.11]		2.58 [3.11]		2.61 [3.09]	
Bandwidth	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables are microenterprises' employment in log-transformed (Panel A) and level (Panel B). Odd numbered columns show estimates from a linear RD specification and even numbered columns include a second-order polynomial specification. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Firm-specific covariates include firm owner's education level and ownership status. Standard errors are clustered at the district level.

Table 4: Effects on Formal Enterprises – Annual Survey of Industries

	Dependent Variable: Employment (log)					
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Year: 2001						
RD Estimate	0.0927	0.0884	0.0136	0.0145	-0.0475	-0.0684
S.E.	(0.208)	(0.205)	(0.195)	(0.198)	(0.160)	(0.188)
Observations	1,500	1,500	1,554	1,554	1,975	1,975
Year: 2002						
RD Estimate	0.176	0.162	0.0497	0.0639	-0.0904	0.00768
S.E.	(0.269)	(0.260)	(0.247)	(0.251)	(0.176)	(0.201)
Observations	1,579	1,579	1,646	1,646	2,028	2,028
Year: 2003						
RD Estimate	0.154	0.150	-0.0373	-0.00615	-0.284*	-0.175
S.E.	(0.178)	(0.177)	(0.151)	(0.154)	(0.143)	(0.142)
Observations	1,607	1,607	1,709	1,709	2,043	2,043
Year: 2004						
RD Estimate	-0.0976	-0.162	-0.251	-0.326*	-0.245	-0.264
S.E.	(0.197)	(0.174)	(0.178)	(0.165)	(0.150)	(0.172)
Observations	1,996	1,996	2,124	2,124	2,566	2,566
Year: 2005						
RD Estimate	-0.103	-0.117	-0.188	-0.188	-0.174	-0.193
S.E.	(0.159)	(0.156)	(0.150)	(0.150)	(0.124)	(0.136)
Observations	1,795	1,795	1,918	1,918	2,377	2,377
Bandwidth /z/	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is formal enterprises' employment (log), using ASI data sets. RD estimates reported for every year between 2001 and 2005. Odd numbered columns show estimates from a linear RD specification and even numbered columns include a second-order polynomial specification. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Firm-specific covariates include owner's age and education level. Standard errors are clustered at the district level.

Table 5: Extensive Margin – Number of firms

	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Village level – Number of firms (log) [Economic Census 2005]						
1. All Firms						
RD Estimate	0.0985	0.0890	0.130*	0.141*	0.151**	0.180**
S.E.	(0.0855)	(0.0807)	(0.0737)	(0.0715)	(0.0744)	(0.0710)
R-square	0.599	0.600	0.593	0.594	0.600	0.600
Observations	73,335	73,335	83,356	83,356	92,677	92,677
Control Group Mean [SD]	53.64 [28.91]		49.44 [119.57]		52.46 [123.78]	
2. Microenterprises						
RD Estimate	0.0980	0.0885	0.130*	0.142*	0.151**	0.181**
S.E.	(0.0854)	(0.0807)	(0.0737)	(0.0716)	(0.0744)	(0.0710)
R-square	0.599	0.599	0.593	0.593	0.599	0.600
Observations	73,302	73,302	83,313	83,313	92,633	92,633
Control Group Mean [SD]	53.4 [128.44]		49.22 [119.11]		52.22 [123.25]	
3. Formal Firms						
RD Estimate	0.0465	0.0445	0.0279	0.0275	0.0188	0.00864
S.E.	(0.0384)	(0.0402)	(0.0391)	(0.0407)	(0.0314)	(0.0335)
R-square	0.134	0.134	0.131	0.131	0.141	0.141
Observations	7,100	7,100	7,627	7,627	8,942	8,942
Control Group Mean [SD]	0.24 [1.28]		0.22 [1.27]		0.25 [1.35]	
Panel B: Microenterprises – Established less than 3 years ago(%) [NSS (Schedule 2.2) 2005-06]						
RD Estimate	0.0784**	0.0519	0.107***	0.108***	0.0998***	0.104***
S.E.	(0.0327)	(0.0318)	(0.0313)	(0.0309)	(0.0320)	(0.0323)
R-square	0.183	0.187	0.173	0.173	0.158	0.159
Observations	6,528	6,528	7,349	7,349	8,350	8,350
Control Group Mean [SD]	0.10 [0.29]		0.09 [0.29]		0.10 [0.30]	
Bandwidth ($\|z\|$)	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables include village's number of firms (log), number of microenterprises, and number of formal enterprises (log) (Panel A), and microenterprises' dummy indicating establishment less than three years (Panel B). Odd numbered columns show estimates from a linear RD polynomial specification and even numbered columns include a second-order polynomial specification. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) and baseline backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population). Village covariates include baseline total population (log), share of irrigated area, paved road coverage, and electricity coverage. Standard errors are clustered at the district level.

Table 6: Individual and Household outcomes – NSS (Schedule 10), 2005-06

	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Wage (log)						
RD Estimate	0.123*	0.123*	0.114*	0.101*	0.135**	0.128**
S.E.	(0.0728)	(0.0728)	(0.0619)	(0.0585)	(0.0538)	(0.0544)
R-square	0.360	0.360	0.360	0.361	0.358	0.358
Observations	4,914	4,914	5,422	5,422	6,232	6,232
Control Group Mean [SD]	401.1 [447.3]		408.83 [460.09]		411.02 [456.82]	
Panel B: Days worked (in the last 7 days) (log)						
RD Estimate	0.0331**	0.0333**	0.0326*	0.0299*	0.0318**	0.0320**
S.E.	(0.0159)	(0.0157)	(0.0177)	(0.0175)	(0.0145)	(0.0144)
R-square	0.051	0.051	0.050	0.050	0.049	0.049
Observations	31,290	31,290	34,818	34,818	39,143	39,143
Control Group Mean [SD]	6.58 [1.13]		6.61 [1.09]		6.61 [1.09]	
Panel C: MHCE (log)						
RD Estimate	0.0883*	0.0867*	0.122**	0.114**	0.122***	0.122***
S.E.	(0.0489)	(0.0512)	(0.0477)	(0.0466)	(0.0423)	(0.0419)
R-square	0.191	0.191	0.182	0.184	0.162	0.162
Observations	6,602	6,602	7,357	7,357	8,249	8,249
Control Group Mean [SD] ('000Rs)	2.42 [1.66]		2.47 [1.62]		2.45 [1.40]	
Bandwidth	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables are an individual's monthly wage (log-transformed; Panel A), weekly workdays (log-transformed, Panel B), and monthly household consumption expenditure (MHCE) (log-transformed; Panel C). Odd numbered columns show RD estimates from linear specifications and even numbered columns show quadratic specifications. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Individual-specific covariates include age, sex, religion, education level, and the social group that the individual belongs to. Standard errors are clustered at the district level.

Table 7: Microenterprises – Mechanisms

	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Firm experiencing power cut (%)						
RD Estimate	-0.139**	-0.126**	-0.0723	-0.0834	-0.0881*	-0.0860*
S.E.	(0.0562)	(0.0555)	(0.0556)	(0.0565)	(0.0459)	(0.0468)
R-square	0.211	0.214	0.195	0.199	0.185	0.185
Observations	6,758	6,758	7,579	7,579	8,580	8,580
Control Group Mean [SD]	0.19 [0.39]		0.20 [0.40]		0.17 [0.38]	
Panel B: Firm has no access to raw materials (%)						
RD Estimate	-0.178**	-0.183**	-0.171**	-0.171**	-0.110	-0.103
S.E.	(0.0796)	(0.0804)	(0.0791)	(0.0780)	(0.0823)	(0.0790)
R-square	0.253	0.253	0.233	0.233	0.210	0.211
Observations	6,758	6,758	7,579	7,579	8,580	8,580
Control Group Mean [SD]	0.14 [0.35]		0.15 [0.36]		0.15 [0.36]	
Bandwidth	0.02		0.025		0.03	
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variables are binary indicators for whether the microenterprise experienced power shortage during production (Panel A) and had no acccess to raw materials (Panel B). Odd numbered columns show estimates from a linear RD specification and even numbered columns include a second-order polynomial specification. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Firm-specific covariates include firm owner's education level and ownership status. Standard errors are clustered at the district level.

Table 8: Microenterprises – Heterogeneity by Industry

	(1)	(2)	(3)
Panel A: Electricity-intensive industries			
Firm experiencing power cut (%)			
RD Estimate	-0.199**	-0.0723	-0.0276
S.E.	(0.0759)	(0.0556)	(0.111)
R-squared	0.245	0.192	0.327
Observations	2,389	7,575	1,858
Degree of electricity intensity (tercile)	>66th	33rd to 66th	<33rd
Panel B: Road-intensive industries			
Firm has no access to raw materials (%)			
RD Estimate	-0.262***	-0.0321	-0.113
S.E.	(0.0986)	(0.0981)	(0.0814)
R-square	0.406	0.181	0.411
Observations	2,599	2,828	2,152
Degree of road intensity (tercile)	>66th	33rd to 66th	<33rd
State & Industry Fixed Effects	Yes	Yes	Yes
District & Firm Controls	Yes	Yes	Yes

Note: The dependent variables are binary indicators for whether the microenterprise experienced power shortage during production (Panel A) and had no access to raw materials (Panel B). District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Firm-specific covariates include firm owner's education level and ownership status. All estimates are reported using a linear RD polynomial specification and a restricted sample using a representative bandwidth of $|z| \leq 0.025$. Standard errors are clustered at the district level.

Table 9: Microenterprises – Heterogeneous Effects in Electricity and Road-intensive Industries

	Dependent Variable: Microenterprise's Employment (log)		
	(1)	(2)	(3)
Panel A: Electricity-intensive industries			
RD Estimate	0.220***	0.0975*	0.0661
S.E.	(0.0836)	(0.0511)	(0.0828)
R-squared	0.423	0.342	0.432
Observations	2,389	7,575	1,858
Degree of electricity intensity (tercile)	>66th	33rd to 66th	<33rd
Panel B: Road-intensive industries			
RD Estimate	0.125	0.0839	0.0410
S.E.	(0.0851)	(0.0619)	(0.0443)
R-squared	0.376	0.323	0.332
Observations	2,599	2,828	2,152
Degree of road dependency (tercile)	>66th	33rd to 66th	<33th
State & Industry Fixed Effects	Yes	Yes	Yes
District & Firm Controls	Yes	Yes	Yes

Note: The dependent variables are microenterprises' employment (log). District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Firm-specific covariates include firm owner's education level and ownership status. All estimates are reported using a RD polynomial specification and a restricted sample using a representative bandwidth of $|z| \leq 0.025$. Standard errors are clustered at the district level.

Table 10: Heterogeneity by Baseline Village Infrastructure

	Linear	Quadratic	Linear	Quadratic
	(1)	(2)	(3)	(4)
Panel A: Village employment in microenterprises (log)				
RD estimate	-0.0227	-0.0183	0.190**	0.199**
S.E.	(0.155)	(0.154)	(0.0895)	(0.0821)
R-squared	0.368	0.368	0.576	0.577
Observations	13,477	13,477	45,956	45,956
Sample	No roads or electricity	No roads or electricity	Roads and electricity	Roads and electricity
Panel B: Number of microenterprises in the village (log)				
RD estimate	-0.0170	-0.0163	0.179**	0.186**
S.E.	(0.150)	(0.150)	(0.0844)	(0.0800)
R-squared	0.383	0.383	0.586	0.587
Observations	13,477	13,477	45,956	45,956
Sample	No roads or electricity	No roads or electricity	Roads and electricity	Roads and electricity
State Fixed Effects	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes

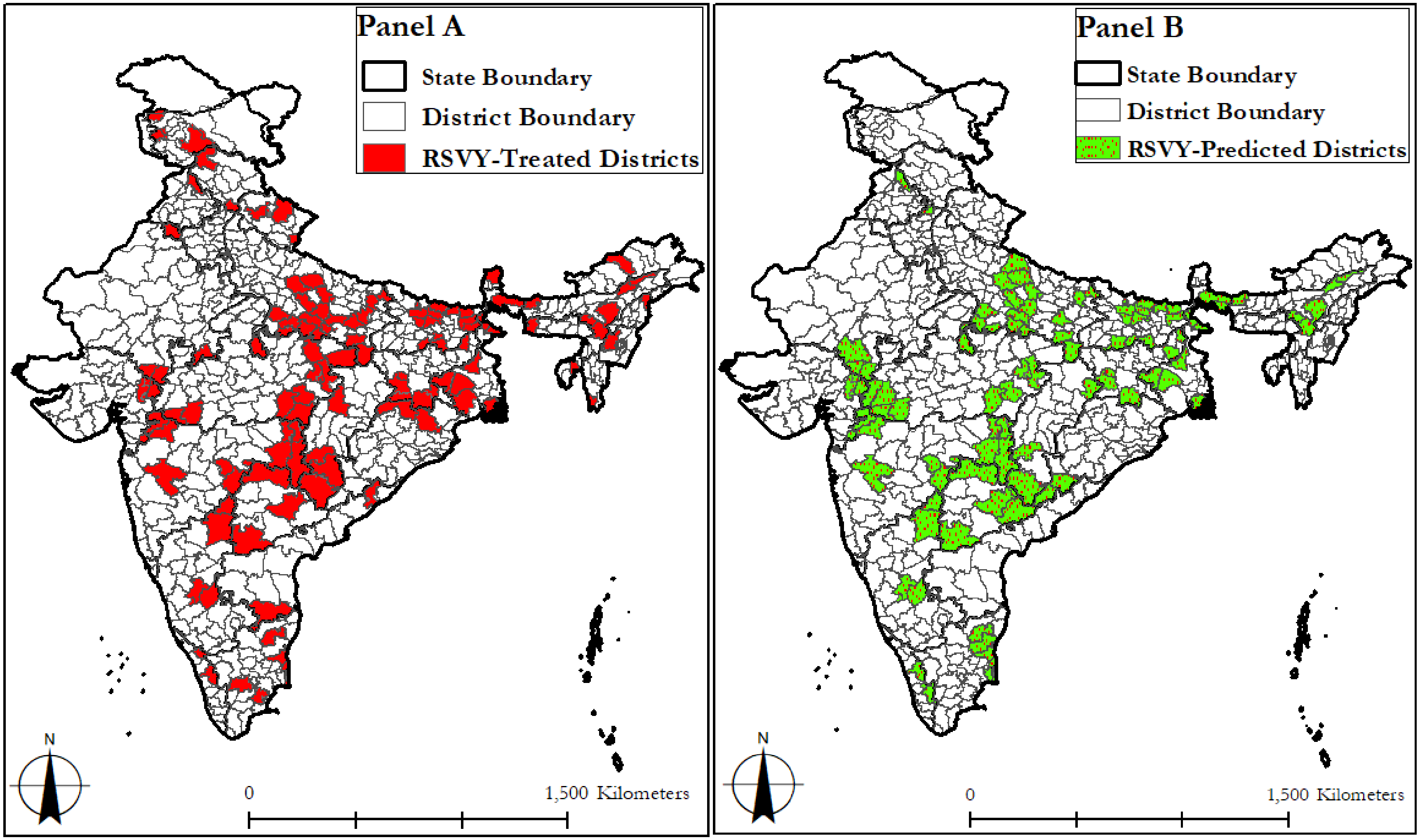
Note: The dependent variables are village-level microenterprises' employment (intensive margin; Panel A) and number of microenterprises (extensive margin; Panel B). Odd columns show estimates from first-order RD polynomial specifications and even columns second-order polynomial. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) and baseline backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population). Village covariates include baseline total population (log), share of irrigated area, paved road coverage, and electricity coverage. All estimates are reported under a restricted sample using a representative bandwidth of $|z| \leq 0.025$. Standard errors are clustered at the district level.

Table 11: Heterogeneity by Proximity to District Headquarters – Economic Census 2005

	Microenterprise Employment			Number of Microenterprises		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standardized distance interaction						
standardized distance to headquarter	-0.0274*	-0.0258	-0.0210	-0.0200	-0.0168	-0.0133
	(0.0141)	(0.0157)	(0.0177)	(0.0139)	(0.0149)	(0.0169)
RSVY X standardized distance	-0.0348	-0.0413*	-0.0487*	-0.0343	-0.0442*	-0.0494**
	(0.0221)	(0.0235)	(0.0252)	(0.0219)	(0.0225)	(0.0242)
R-squared	0.603	0.594	0.601	0.617	0.609	0.614
Observations	88,598	79,278	69,267	88,598	79,278	69,267
Panel B: Log distance interaction						
log(distance to headquarter)	-0.0587***	-0.0526**	-0.0440*	-0.0436**	-0.0344*	-0.0271
	(0.0193)	(0.0218)	(0.0243)	(0.0184)	(0.0197)	(0.0218)
RSVY X log(distance)	-0.0477	-0.0606*	-0.0753*	-0.0457	-0.0651*	-0.0761**
	(0.0339)	(0.0363)	(0.0385)	(0.0326)	(0.0334)	(0.0355)
R-squared	0.603	0.594	0.601	0.617	0.609	0.614
Observations	88,600	79,280	69,269	88,600	79,280	69,269
Bandwidth	0.03	0.025	0.02	0.03	0.025	0.02
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes

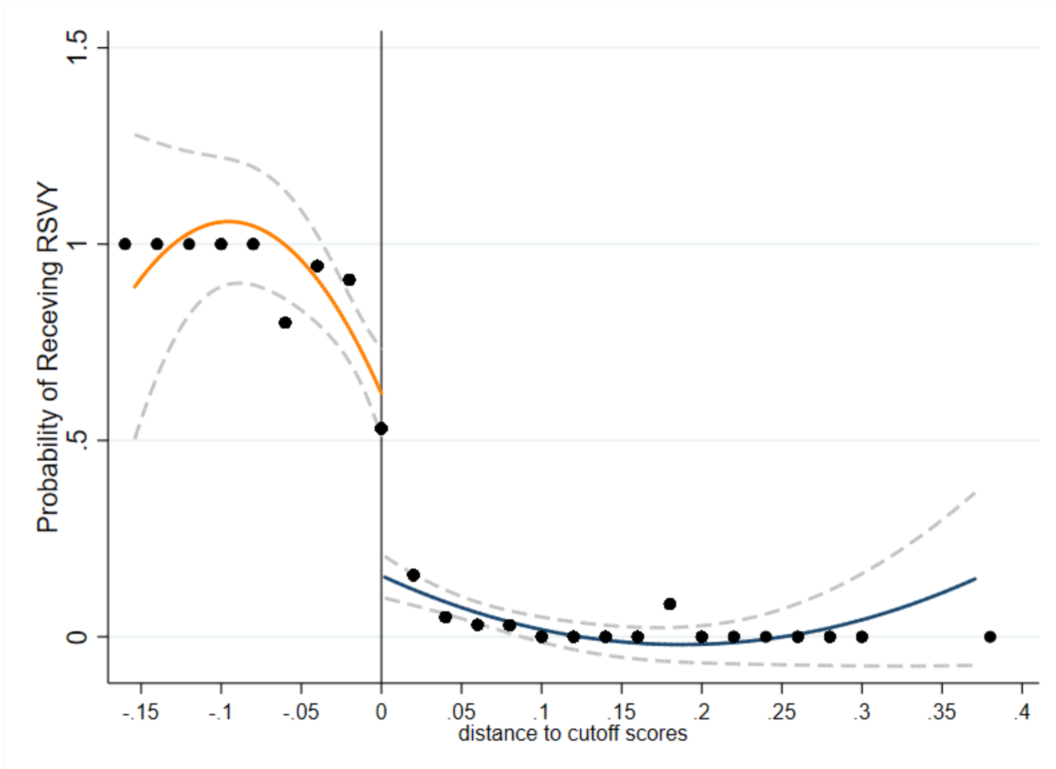
Note: The dependent variables are village-level microenterprises' employment (i.e. intensive margin; columns (1)-(3)) and number of microenterprises (i.e. extensive margin; columns (4)-(6)). District covariates include geographic (area, elevation, boundary length, and proximity to big cities) and baseline backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SC/ST population). Village covariates include baseline total population (log), share of irrigated area, paved road coverage, and electricity coverage. All estimates are reported using a linear RD specification. Standard errors are clustered at the district level.

Figure 1: Maps of RSVY Treated and Predicted Districts



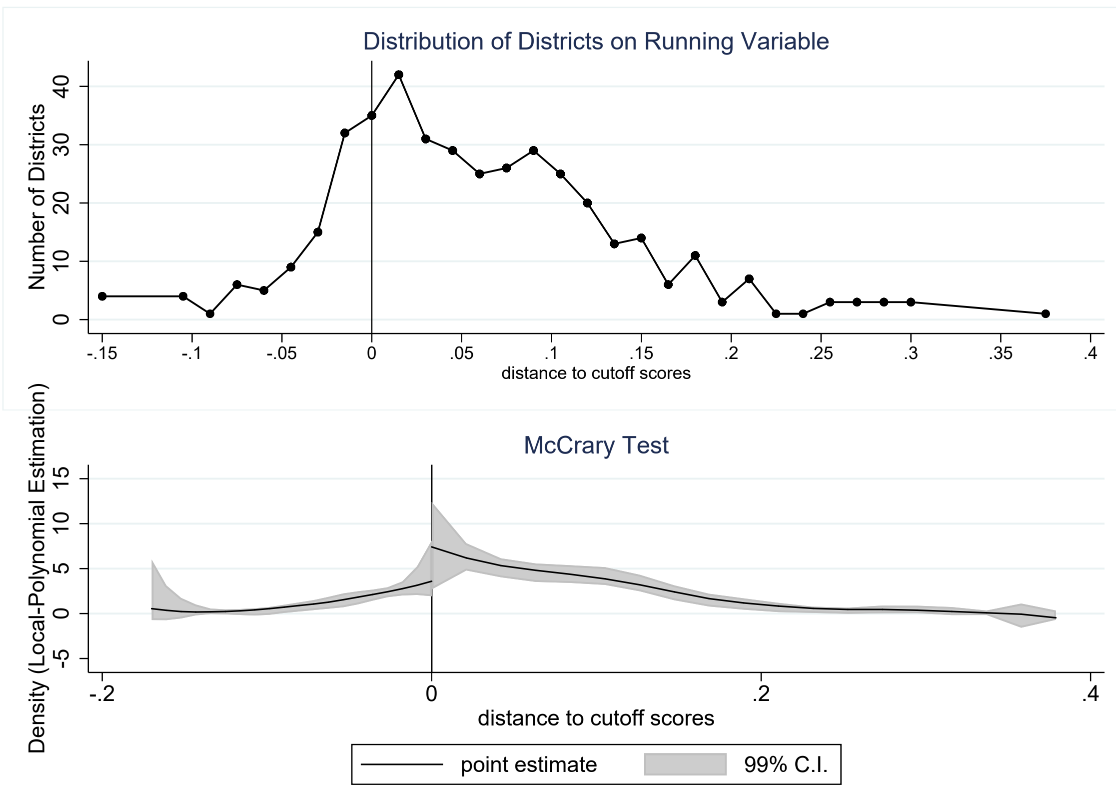
Note: Panel A highlights the districts that received the RSVY grants (115 total; in red). Panel B highlights all districts predicted to receive the RSVY grants based on their Backwardness Ranking scores (96 total; in green). Selection criteria is discussed in section 4. Thick black lines represent state boundaries. Thin black lines represent district boundaries.

Figure 2: Discontinuity in Treatment Probability (First Stage)

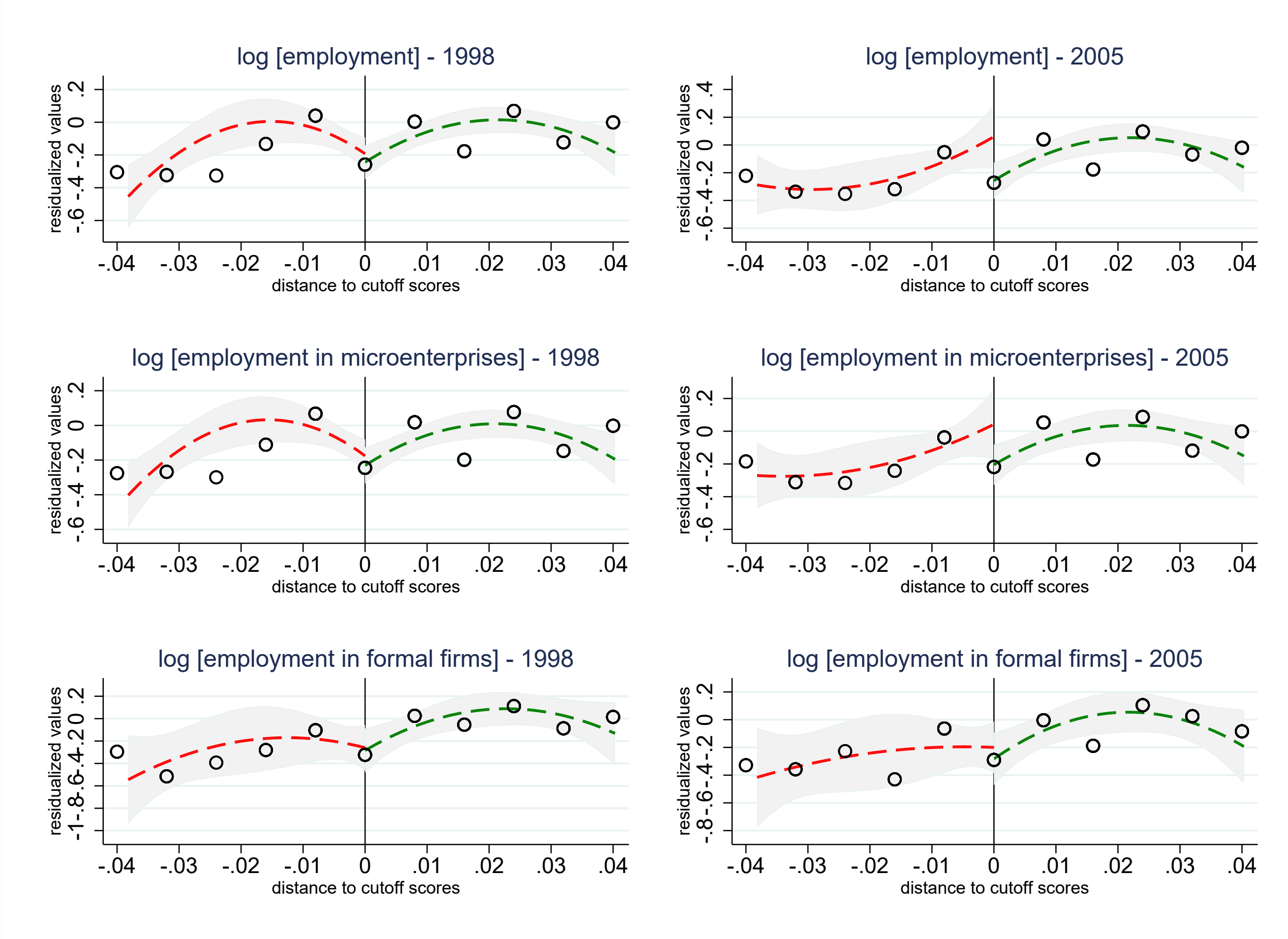


Note: The graph plots the probability of receiving RSVY treatment by districts over the RD running variable (district's standardized distance scores from the cutoff). Quadratic fitted curves on each side of the cutoff as well as 95% confidence interval bands are also included.

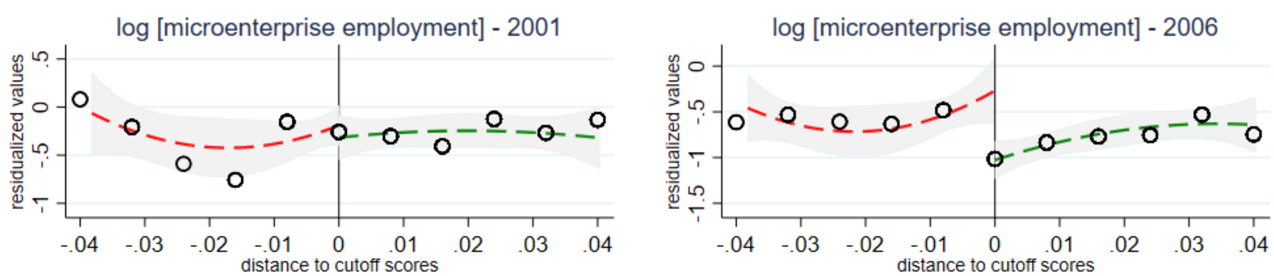
Figure 3: Distribution of Districts over Running Variable



Note: The top panel of the figure plots the distribution of districts over the RD running variable (re-centered distance scores from the cutoff). The bottom panel plots a non-parametric regression to each half of the distribution following McCrary (2008), testing for manipulation of the running variable at the cutoff.



(a) Panel A: Village level employment (Economic Census)

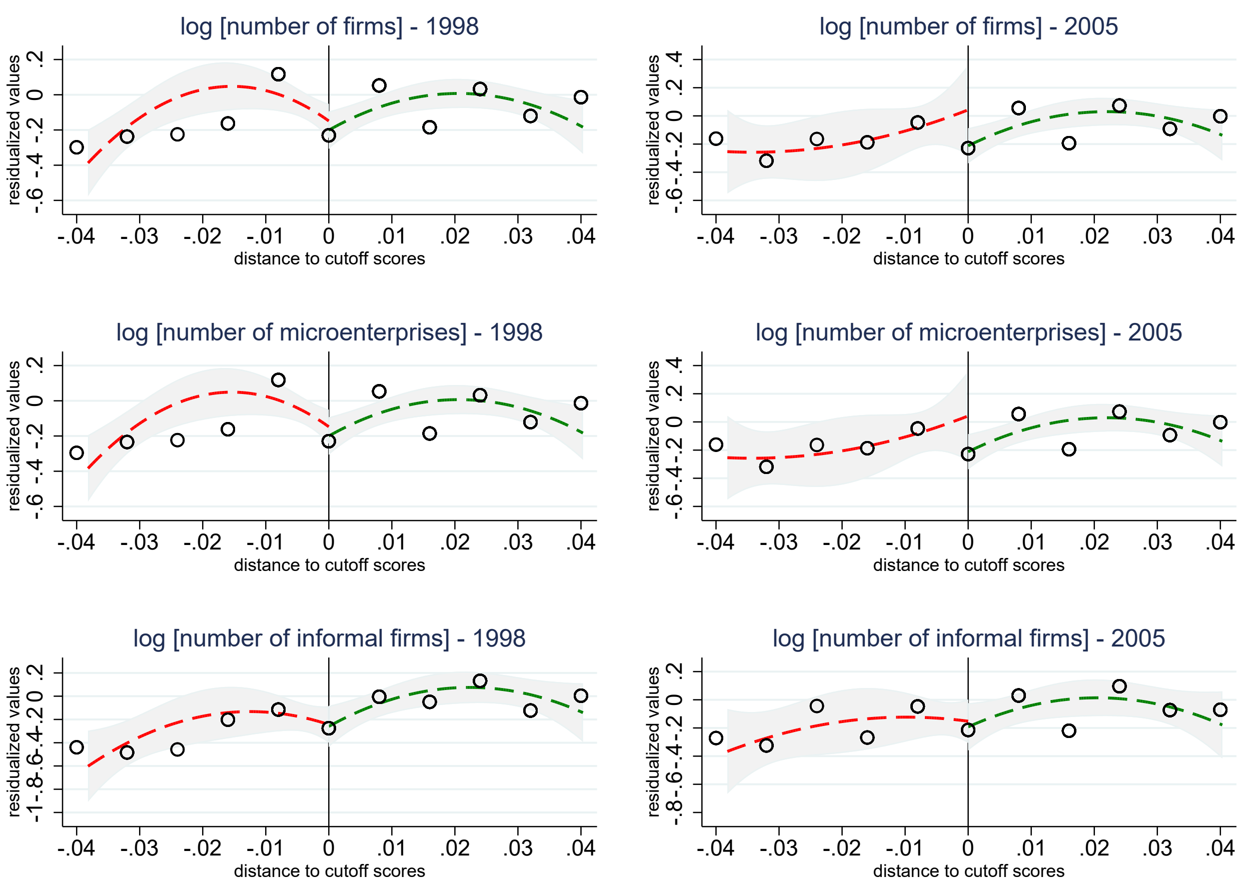


(b) Panel B: Microenterprise Employment (NSS)

Figure 4: Employment (Economic Census and NSS)

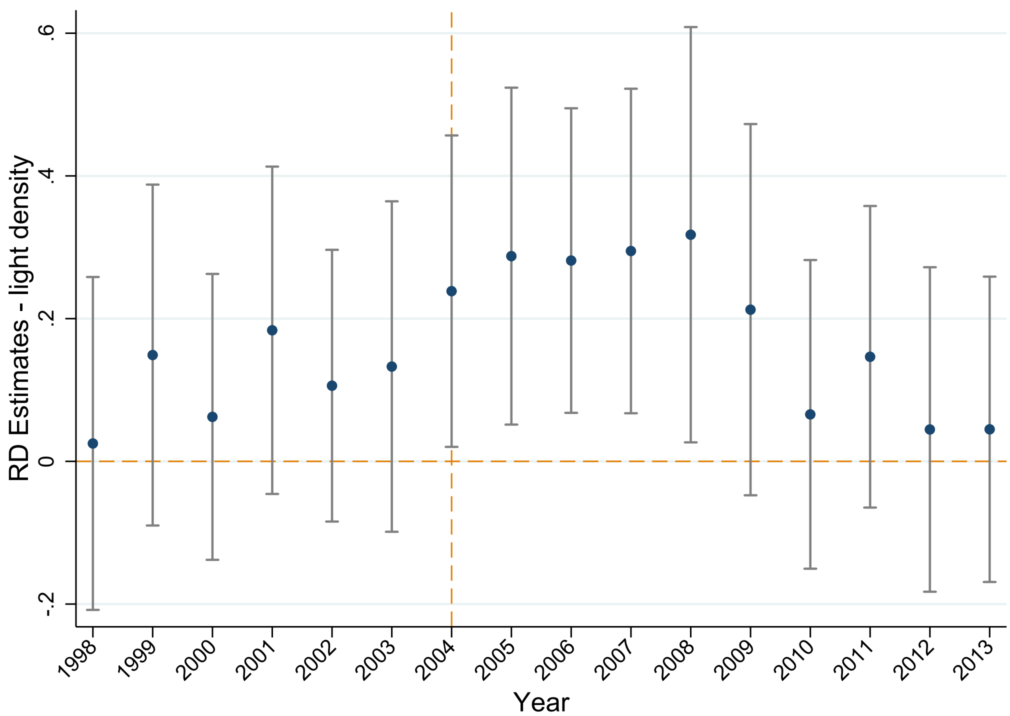
Note: Panel A includes graphs using the 4th and 5th rounds of the Economic Census and Panel B includes graphs using the NSS Schedule 2.2. All graphs on the left correspond to pre-RSVY employment outcomes; all graphs on the right correspond to post-RSVY employment outcomes. In Panel A, the top row shows total village employment in all firms (in logs), the middle row shows village employment in microenterprises (in logs), and the third row shows village employment in formal firms (in logs). Each scatter point represents the bin-average of residualized values of village- (Panel A) and firm-level (Panel B) employment after controlling for all covariates discussed in the main specification. The sample includes all observations in districts with the re-centered Backwardness Index Scores (z) within the restricted 0.04 point from the cutoff, i.e. $|z| \leq 0.04$. In all graphs, the quadratic fitted curves along with the associated confidence intervals (with standard errors clustered at the district level) are presented.

Figure 5: Number of Firms (Economic Census)



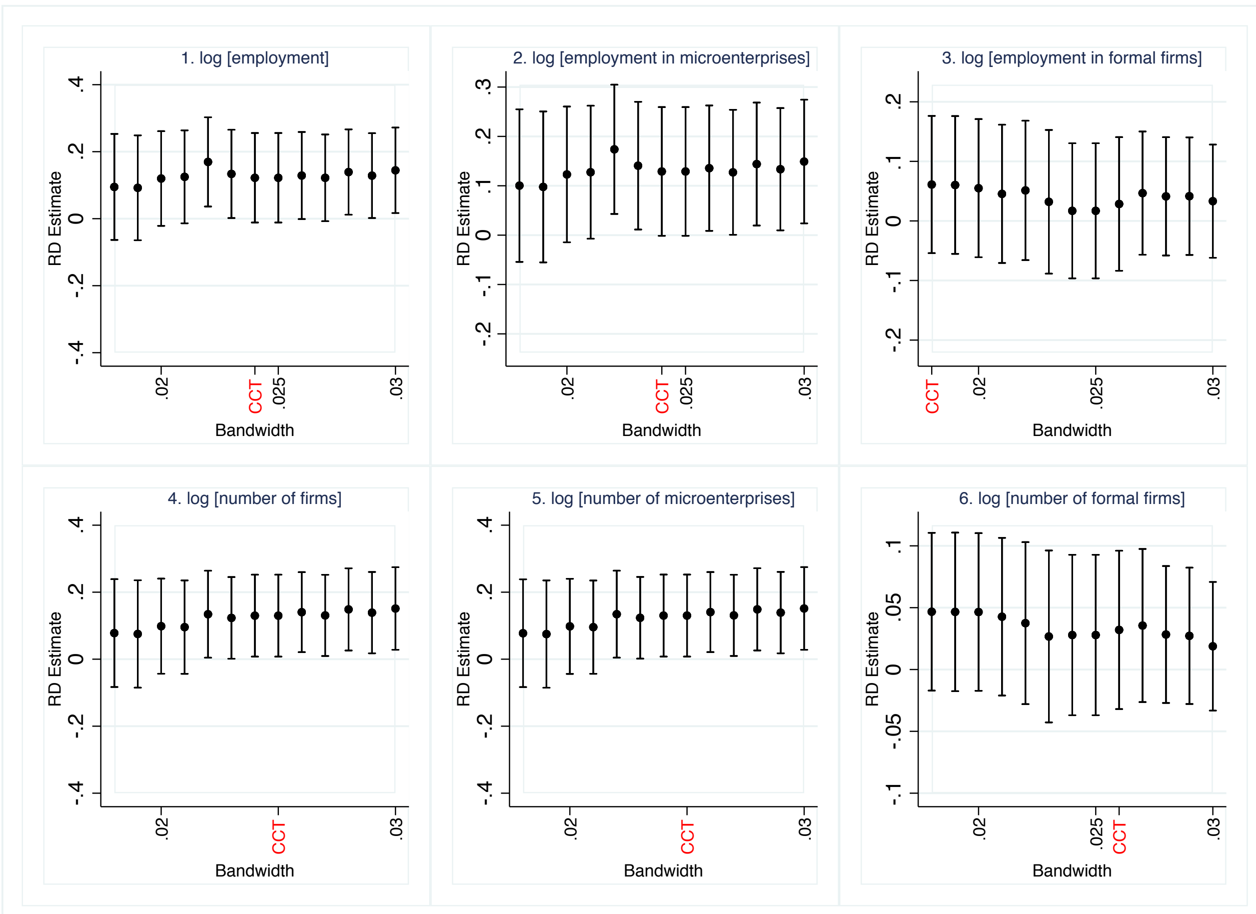
Note: This figure includes three panels and six graphs and uses the 4th and 5th rounds of the Economic Census. All graphs on the left correspond to the outcome variable in 1998 (pre-RSVY) and the graphs on the right correspond to the outcome variables in 2005 (post-RSVY). The top panel represents total number of firms in the village (in logs), the middle panel represents total number of microenterprises (in logs), and the bottom panel represents total number of formal firms (in logs). Each scatter point represents the bin-average of residualized values of firms in the village (all firms, microenterprises, and formal firms) after controlling for all variables in the main specification. The sample includes all villages in districts with the re-centered Backwardness Index Scores (z) within the restricted 0.04 point from the cutoff, i.e. $|z| \leq 0.04$. In all graphs, the linear fitted curves along with the associated confidence intervals (with standard errors clustered at the district level) are presented.

Figure 6: RD Estimate of RSVY Impact - Mechanism Tests



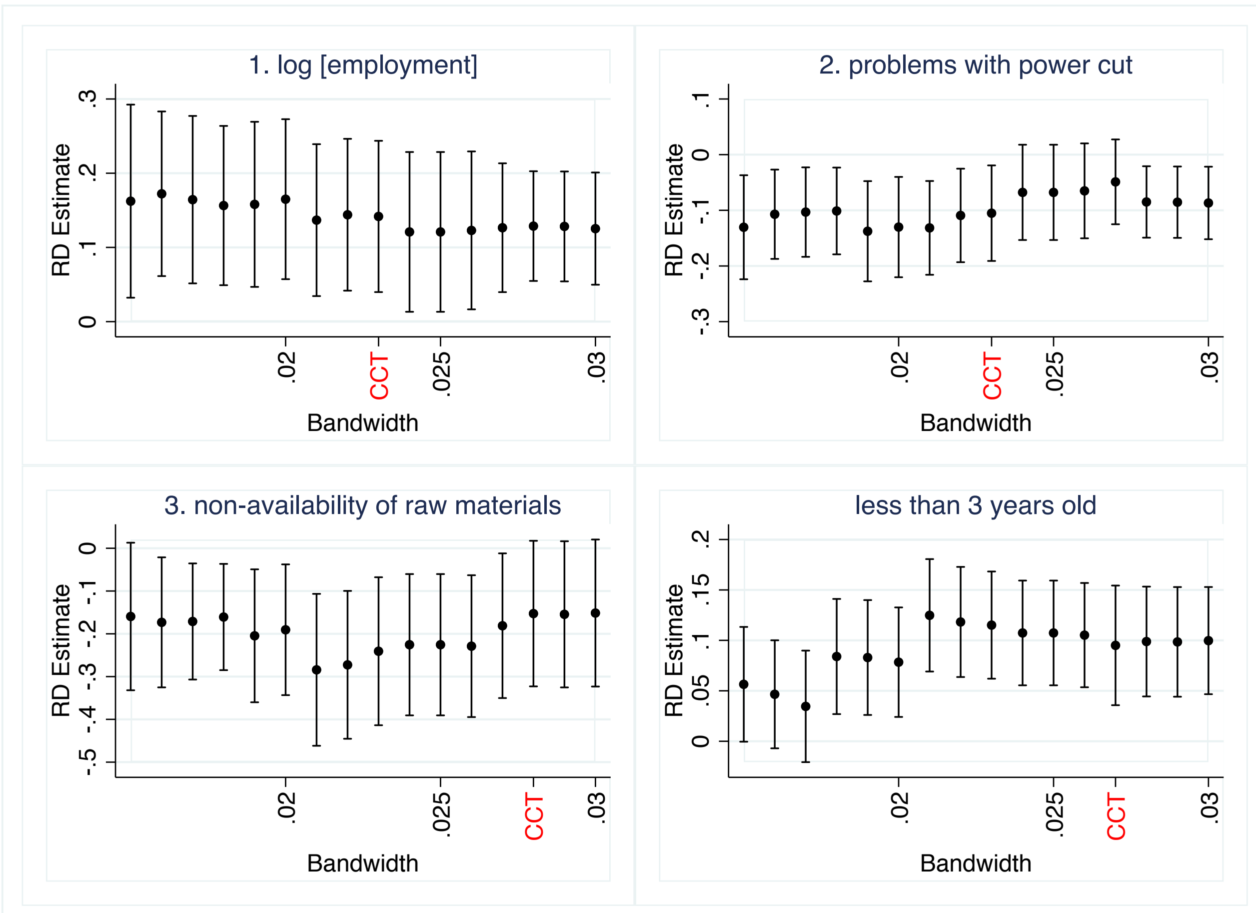
Note: The graph plots RD estimates (β_1 from equation (3)) and corresponding 90% confidence intervals with night-light density (proxy for the level of infrastructural development) as outcome variables, across a 16-year period of both pre- and post-intervention. Robust standard errors are clustered at the district level.

Figure 7: Sensitivity Analysis for Economic Census Outcomes



Notes: The panels show the estimated RD (ITT) point estimates and confidence intervals of all main outcome variables from the Economic Census (2005), under varying bandwidths. The bandwidth selected according to Calonico et al. (2014) criteria are also reported on the horizontal axis (in red).

Figure 8: Sensitivity Analysis for NSS Manufacturing Survey Outcomes



Notes: The panels show the estimated RD (ITT) point estimates and confidence intervals of all main outcome variables from the NSS manufacturing survey (Round 62, Schedule 2.2, 2005-06), under varying bandwidths. The bandwidth selected according to Calonico et al. (2014) criteria (denoted "CCT") are also reported on the horizontal axis (in red).

Appendix

Table A1: Balance on Different Village Population Cutoffs (Sample $|z| \leq 0.03$)

	Treatment districts	Control districts	Treatment-Control	p-value on difference
Number of Villages				
All	877.25	829.30	47.95	0.67
Population $>$ 300	800.35	740.73	59.62	0.52
300 $<$ Population $<$ 450	68.25	77.08	-8.83	0.63
Population $>$ 500	708.12	636.49	71.62	0.35
500 $<$ Population $<$ 750	114.92	120.57	-5.65	0.78
Population $>$ 1000	493.88	420.35	73.53	0.15
Share of Villages				
Population $>$ 300	0.92	0.92	-0.00	0.95
300 $<$ Population $<$ 450	0.070	0.075	-0.005	0.55
Population $>$ 500	0.83	0.82	0.01	0.81
500 $<$ Population $<$ 750	0.12	0.13	-0.01	0.56
Population $>$ 1000	0.59	0.59	0.00	0.88
Number of districts	52	63		

Note: This table presents the means across treatment and control districts for the number and share of villages by population cutoffs important for the eligibility of PMGSY, and RGGVY.

Table A2: Estimates for Treatment-on-the-Treated (TOT)

	Linear			Quadratic			
Dependent Variable	RD Estimate	S.E.	R-Square	RD Estimate	S.E.	R-Square	Observations
Panel A: RSVY Impact on Employment – Economic Census 2005							
A1: Village Employment (log)							
Total Employment	0.591*	(0.345)	0.560	0.748*	(0.423)	0.553	92,677
Formal Employment	0.149	(0.144)	0.112	0.116	(0.137)	0.113	8,942
Informal Employment	0.602*	(0.348)	0.568	0.772*	(0.429)	0.560	92,633
A2: Firm-level Employment (log)							
All Firms	0.120**	(0.0606)	0.015	0.120**	(0.0607)	0.015	4,921,316
Formal Firms	-0.147	(0.138)	0.070	-0.147	(0.140)	0.070	21,109
Informal Firms	0.113*	(0.0581)	0.016	0.113*	(0.0581)	0.016	4,900,207
Panel B: RSVY Impact on Microenterprises – NSS (Schedule 2.2) 2005-06							
Employment (log)	0.327**	(0.154)	0.242	0.332**	(0.158)	0.242	7,579
Employment (levels)	1.694**	(0.697)	0.132	1.765**	(0.718)	0.129	7,579
Panel C: RSVY Impact on Household Welfare – NSS (Schedule 10) 2005-06							
Wages (log)	0.306	(0.197)	0.346	0.261	(0.171)	0.352	5,422
Days worked (last 7 days) (log)	0.103	(0.0692)	0.029	0.0996	(0.0684)	0.030	34,818
MHCE (log)	0.390*	(0.220)	0.148	0.360*	(0.205)	0.156	7,357
Panel D: Extensive Margin – RSVY Impacts on Firm Establishment							
D1: Village level – Firm Quantity (log)							
All Firms	0.599*	(0.342)	0.579	0.753*	(0.419)	0.571	92,677
Formal Firms	0.0554	(0.0724)	0.140	0.0338	(0.0731)	0.141	8,942
Informal Firms	0.599*	(0.343)	0.578	0.754*	(0.421)	0.570	92,633
D2: Microenterprises – Established less than 3 years (%)							
	0.279**	(0.126)	0.152	0.292**	(0.137)	0.150	7,579
Panel E: Microenterprises – Evidence on Impact Channels							
Power Cut (%)	-0.213**	(0.107)	0.123	-0.254	(0.181)	0.124	7,579
No Access to Materials (%)	-0.645**	(0.313)	0.052	-0.659**	(0.328)	0.047	7,579

Note: this table presents Treatment-on-the-Treated RSVY impacts for all outcome variables presented previously in the analysis. Estimates are reported under the first-order RD polynomial specification and a restricted sample using a representative bandwidth of $|z| \leq 0.025$. All else remains unchanged from previous exercises.

Table A3: Pre-treatment impacts

	Linear			Quadratic			
	RD Estimate	S.E.	R-Square	RD Estimate	S.E.	R-Square	Observations
Panel A: RSVY Impact on Employment – Economic Census 1998							
A1: Village Employment (log)							
Total Employment	0.101	(0.0978)	0.488	0.112	(0.0965)	0.488	83,695
Informal Employment	0.111	(0.0947)	0.502	0.124	(0.0926)	0.502	83,591
Formal Employment	-0.0897	(0.0960)	0.059	-0.0910	(0.0937)	0.059	7,546
A2: Firm-level Employment (log)							
All Firms	-0.0367	(0.0418)	0.020	-0.0351	(0.0393)	0.024	3,449,092
Formal Firms	-0.0396	(0.0812)	0.065	-0.0342	(0.0739)	0.068	22,333
Informal Firms	-0.0341	(0.0393)	0.022	-0.0326	(0.0374)	0.026	3,426,759
Panel B: RSVY Impact on Microenterprises – NSS (Schedule 2.2) 2000-01							
Employment (log)	0.0809	(0.0542)	0.216	0.0730	(0.0524)	0.216	17,842
Employment (count)	0.192	(0.153)	0.121	0.205	(0.152)	0.121	17,842
Panel C: RSVY Impact on Household Welfare – NSS (Schedule 10) 1999-00							
Wages (log)	0.0810	(0.0673)	0.332	0.0817	(0.0683)	0.332	16,253
Days worked (last 7 days) (log)	0.0466	(0.0407)	0.062	0.0488	(0.0400)	0.062	35,265
MHCE (log)	-0.0404	(0.0581)	0.218	-0.0419	(0.0580)	0.220	6,450
Panel D: Extensive Margin – RSVY Impacts on Firm Establishment							
Village level – Firm Quantity (log)							
All Firms	0.0990	(0.0870)	0.499	0.116	(0.0824)	0.500	83,695
Formal Firms	-0.0224	(0.0604)	0.099	-0.0240	(0.0565)	0.100	7,546
Informal Firms	0.101	(0.0872)	0.498	0.118	(0.0825)	0.499	83,591
Panel E: Microenterprises – Evidence on Impact Channels							
Power Cut (%)	-0.0426	(0.0561)	0.107	-0.0456	(0.0543)	0.107	17,842
No Access to Raw Materials (%)	-0.0393	(0.0596)	0.139	-0.0380	(0.0596)	0.139	17,842

Note: this table replicates the main regressions for all outcome variables, but using pre-treatment data sets. Estimates are reported using a linear RD polynomial specification and a restricted sample with a representative bandwidth of $|z| \leq 0.025$. All else remains unchanged from previous exercises.

Table A4: Hypothetical Eligibility Threshold

	Linear			Quadratic			
	RD Estimate	S.E.	R-Square	RD Estimate	S.E.	R-Square	Observations
Panel A: RSVY Impact on Employment – Economic Census 2005							
A1: Village Employment (log)							
Total Employment	-0.184	(0.128)	0.559	-0.188	(0.126)	0.559	84,665
Informal Employment	-0.199	(0.128)	0.564	-0.202	(0.127)	0.564	84,647
A2: Firm-level Employment (log)							
All Firms	-0.0979	(0.0679)	0.023	-0.111*	(0.0661)	0.023	2,716,904
Informal Firms	-0.0980	(0.0627)	0.025	-0.110*	(0.0615)	0.025	2,704,032
Panel B: RSVY Impact on Microenterprises – NSS (Schedule 2.2) 2005-06							
Employment (log)	-0.111	(0.108)	0.393	-0.0399	(0.107)	0.395	2,854
Employment (count)	-0.320	(0.357)	0.345	-0.178	(0.370)	0.346	2,854
Panel C: RSVY Impact on Household Welfare – NSS (Schedule 10) 2005-06							
Wage	-0.108	(0.104)	0.382	-0.174	(0.136)	0.383	1,880
Days worked (last 7 days) (log)	-0.0801	(0.0563)	0.057	-0.0821**	(0.0330)	0.062	8,053
MHCE (log)	0.203	(0.153)	0.307	0.170	(0.154)	0.307	2,172
Panel D: Extensive Margin – RSVY Impacts on Firm Establishment							
Village level – Number of firms (log)							
All Firms	-0.152	(0.114)	0.587	-0.154	(0.113)	0.587	84,665
Informal Firms	-0.152	(0.114)	0.587	-0.155	(0.113)	0.587	84,647
Panel E: Microenterprises – Evidence on Impact Channels							
Power Cut (%)	-0.00224	(0.133)	0.323	0.0880	(0.123)	0.326	2,854
No Access to Materials (%)	0.462	(0.330)	0.361	0.350	(0.362)	0.362	2,213

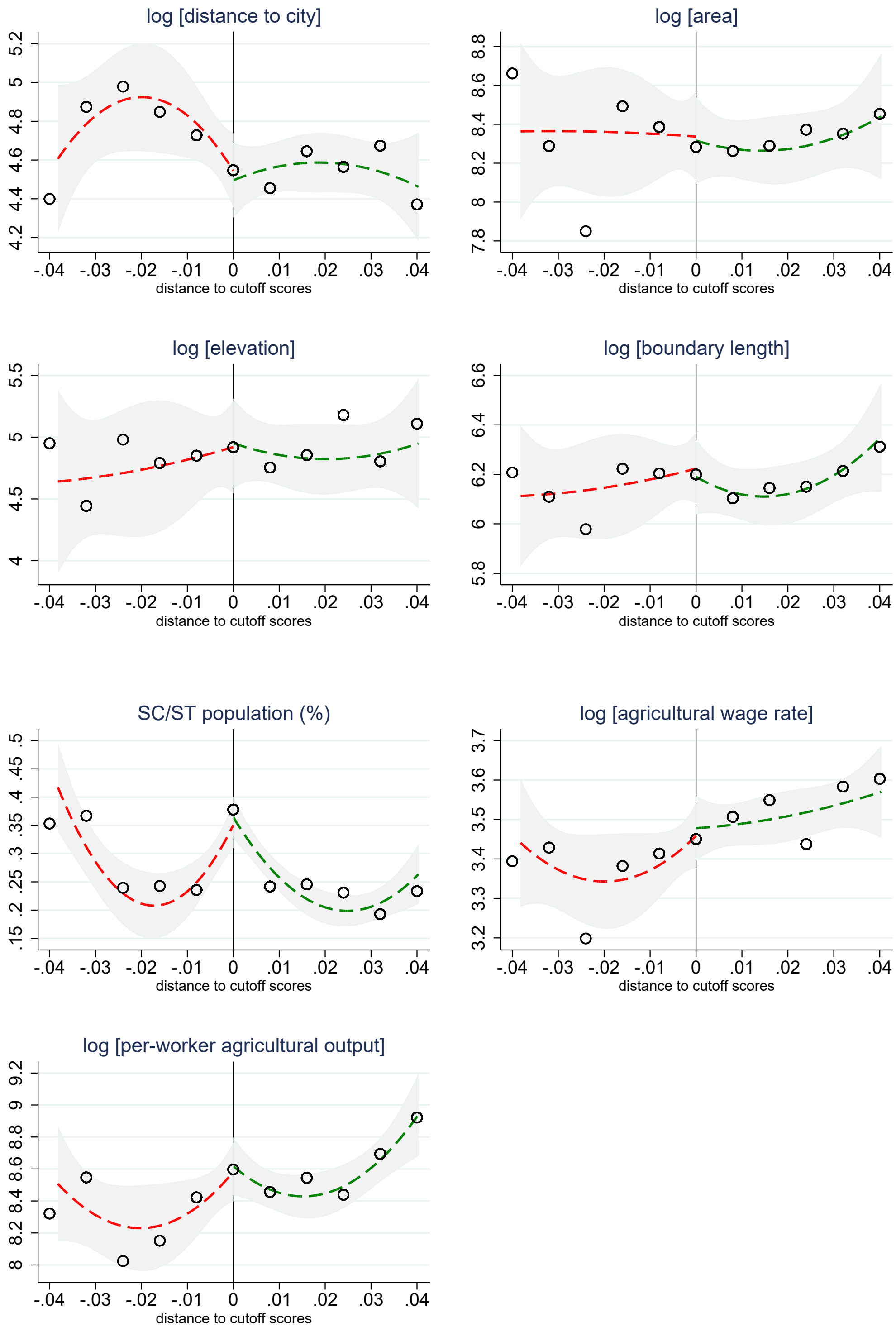
Note: this table replicates the main regressions for all outcome variables, but employing a hypothetical cutoff (i.e. unrelated to RSVY). Estimates are reported using a linear RD polynomial specification and a restricted sample with a representative bandwidth of $|z| \leq 0.025$. All else remains unchanged from previous exercises.

Table A5: Supplementary RD results for night-light impacts

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
RD Estimates	0.0622	0.184	0.106	0.133	0.239*	0.288**	0.282**	0.296**	0.319*	0.213	0.0662
S.E.	(0.120)	(0.137)	(0.115)	(0.141)	(0.133)	(0.142)	(0.132)	(0.139)	(0.186)	(0.159)	(0.132)
R-squared	0.515	0.531	0.525	0.554	0.556	0.541	0.531	0.521	0.435	0.522	0.505
Observations	444	444	443	444	445	444	445	444	444	443	444
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

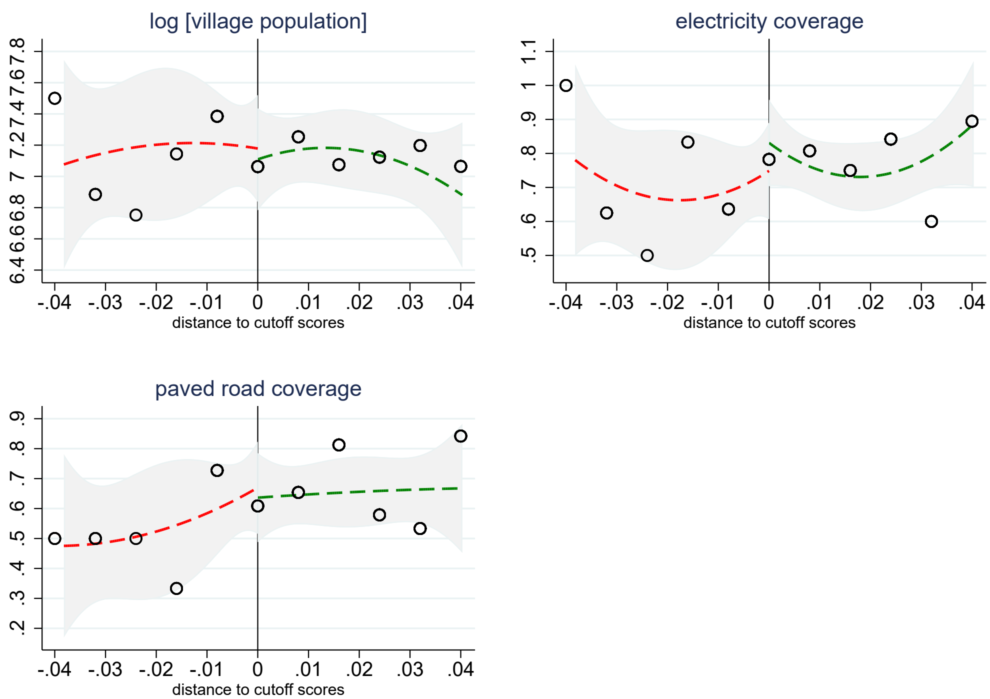
Note: This table presents the supplementary result for annual impacts of RSVY on night-time light density, corresponding to Figure 8. Estimates are reported using a linear RD polynomial specification and a restricted sample with a representative bandwidth of $|z| \leq 0.025$. District covariates include geographic (area, elevation, boundary length, and proximity to big cities) backwardness parameters (per-worker agricultural output, agricultural wage rate, share of SCST population), total population (log), and baseline infrastructural conditions (share of irrigated area, paved road coverage, and electricity coverage). Standard errors are clustered at the district level.

Figure A1: Balance on District-level Baseline Observable Characteristics



Notes: Each panel in the figure corresponds with a district's baseline characteristics, including geographic measures (distance to big city, area, elevation, boundary length; in log), and the district's baseline backwardness index's historical parameters, including 1) the share of SC/ST population, 2) log of agricultural wage rate, and 3) log of per-worker agricultural output. The sample includes all districts with the re-centered Backwardness Index Scores (z) within the restricted 0.04 point from the cutoff, i.e. $|z| \leq 0.04$. In all graphs, the quadratic fitted curves along with the associated confidence intervals (calculated on the basis of standard errors clustered at the district level) are presented.

Figure A2: Balance on Village-level Baseline Observable Characteristics



Notes: Each panel in the figure corresponds with a village's baseline social and infrastructural conditions, including population (log), paved road and electricity coverage. The sample includes all districts with the re-centered Backwardness Index Scores (z) within the restricted 0.04 point from the cutoff, i.e. $|z| \leq 0.04$. In all graphs, the quadratic fitted curves along with the associated confidence intervals (calculated on the basis of standard errors clustered at the district level) are presented.

Detailed Construction of the Planning Commission's Backwardness Index

Data Collection

The backwardness index is constructed by adopting historical parameters with equal weights: (i) value of output per agricultural worker (1990-1993); (ii) agriculture wage rate (1996-1997); and (iii) districts percentage of low-caste populations Scheduled Castes/ Scheduled Tribes (1991 Census). This backwardness index ranks a total of 447 districts in 17 major states with available data for the parameters above. Data on agricultural productivity per worker was available for only 17 states. As a result, the state of Goa, all special category states except Assam were excluded from the analysis. There is, thus, available information for 482 districts of the 17 States. In addition, the Task Force Department further decided to exclude districts with urban agglomerations of over one million population as per the 2001 census. The state capitals were also excluded. The reason for these exclusions is that urban centers would almost always generate economic activates that would obviate the need for public works programs. Consequently, 35 additional districts were further excluded from the analysis. This leaves the backwardness ranking index being confined to 447 districts.

It should also be noted that in most states, the number of districts has increased since 1991 due to division of old districts. In those cases, the Scheduled Caste and Scheduled Tribe (SC/ST) pollution proportion for the original district in 1991 would be applied to the new districts created by the division of the district. This imputation process is done similarly for agricultural wages and agricultural productivity per worker.

Ranking Computation

The index was computed for each variable. For agricultural productivity per worker and agricultural wages, the index was computed as follows:

$$\frac{(ActualValue - MinimumValue)}{(MaximumValue - MinimumValue)}$$

The lower the index value, the more backward the district. In the case of the parameter for SC/ST population, it is presumed *a-priori* that districts with higher proportion of SC/ST population are more backward. To ensure that the index values in the three variables moved in the same direction, the index for SC/ST population was calculated as follows:

$$\frac{(MaximumValue - ActualValue)}{(MaximumValue - MinimumValue)}$$

The districts with higher percentage of SC/ST population would have a lower value for the index.

Next, the three sub-indices were aggregated with equal weights of one-third to each, resulting in a composite index. The Planning Commission used the composite index as the final product to rank districts on their level of backwardness. The districts with low wages, low productivity and high SC/ST population were ranked as more backward on the index, i.e. getting a lower rank value. The discrete ranking, thus, ranges from 1 for the most backward district, to 447 for the least backward, subject to data availability.