Infrastructure Grants and the Performance of Microenterprises*

Ritam Chaurey†
Duong Trung Le‡

August 2018

Abstract

We evaluate the impact of a place-based infrastructure development scheme directed towards India's most "backward" districts, on the performance of microenterprises. "Backward" districts were selected based on a transparent score-based assignment mechanism. Using a Fuzzy Regression Discontinuity Design, we find that firms in treated districts had higher profits, revenues, and employment. Improvements in electrification was an important channel driving these results, as firms used more electricity and had a lower likelihood of facing power cuts. We also find increases in migrants, and proportion of new firms in treated districts, along with negative spillovers in areas closer to the treated districts.

JEL Codes: O12, O18, O25, R11

---

*We would like to thank Ama Baafra Abeberese, Siddharth Hari, Gaurav Khanna, Amit Khandelwal, Solomon Polachek, David Slichter, 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, and EPED (Montreal 2018) for helpful comments. All remaining errors are our own.

†Email: rchaurey@jhu.edu. Johns Hopkins University, School of Advanced International Studies.

‡Email: dle1@binghamton.edu. Binghamton University - SUNY.

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 often directed to economically lagging regions to incentivize firm growth in order to alleviate regional economic imbalances. Such "place-based" infrastructure programs are popular both in developed and developing[1] countries. Previous work on the effects of place-based policies, and more generally on the effects of infrastructure investments has focused on medium and large firms. As a result, there is almost no evidence on the effects on microenterprises. Since these firms account for a massive share of the non-agricultural employment in developing countries,[2] understanding whether they grow in size, and become more profitable in response to improvements in local infrastructure, are important questions.

In this paper, we study the effects of a place-based infrastructure grants program, Rashtriya Sam Vikas Yojana (RSVY), launched by the Government of India, on the performance of microenterprises. In our setting, RSVY grants were extended to districts using a score-based assignment mechanism. We are thus able to address the primary concern of non-random placement, common to most place-based policies, by using a regression discontinuity design. We combine detailed nationally representative surveys on the operations of small and microenterprises, as well as data on night-time lights, and migration to analyze the direct effects on microenterprises, the mechanisms driving these effects, and spillovers due to these infrastructure grants.

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 for reducing regional economic imbalances and speed up development. Under RSVY, the Central Government provided sizable cash grants for infrastructure investments to the poorest districts in the country. Each eligible district was entitled to receive 450 million Indian Rupees (abbrev. "Rs") (approx. 7 mil-

[1]For example, the EU Structural Funds, comprising the European Regional Development Fund (ERDF) and European Social Fund (ESF), is aimed at reducing regional imbalances across member states through investments in energy, telecommunications, and transport infrastructure. The Tennessee Valley Authority (TVA) development and modernization policy and the Appalachian Regional Commission (ARC) are examples in the United States. TVA involved substantial investments in dams, road networks, canals, and new schools and covered the states of Tennessee, Kentucky, Alabama, and Mississippi. Other examples include the Western Regional Development Strategy (WRDS) in Western China that provided railway, airport, water and power infrastructure construction.

[2]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.

lion USD) over the course of three years to be spent on infrastructure investments. This amounted to around 1.15% of the average "backward" district's GDP between 2003-04 to 2005-06. The specific guideline that the Government of India developed to prioritize the treated districts makes this policy an ideal natural experiment. According to the guideline, the 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 guideline for RSVY implementation 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 government's identification of RSVY-eligible "backward" districts. Specifically, we follow the Planning Commission's official documentation to reconstruct the "Backwardness Index" scores for each district in the country where historical under-development statistics are available. The Backwardness Index was constructed based on 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 population - Scheduled Castes/ Scheduled Tribes (1991 Census). Since for each state we know the number of districts that received RSVY grants, we know the cutoff district that received the grant based on the Backwardness Index scores. Finally, we calculate the state-specific distance to cutoff score for each district in our sample, assigning the standardized score of 0 to the cutoff district in each state.[3] We use this distance to cutoff score as the running variable in a Regression Discontinuity Design (RD)[4] framework. We run our RD regressions on various economic outcomes at both the firm and taluk (administrative units below the district) levels, using data on small and microenterprises collected by the National Sample Surveys (Manufacturing Schedule) as well as night-time lights data.

We find a number of results on the effectiveness of RSVY on small and microenterprises, approximately two years (2005-06) after the policy came in to effect. Firms in districts that received RSVY grants show a large increase in gross profits, of around 24%, relative to firms in districts that just missed receiving the RSVY grants (close to the cutoff). Firms also employed more workers on average (0.325 workers above the mean of 2.44 workers), and their revenues increased by 39%, following RSVY. Our results are robust across various specifications and different data-driven bandwidth selection techniques (Imbens and Kalyanaraman

[3] We provide a detailed description of the index reconstruction in Section 4.

[4] Our RD running variable is thus different from the running variable employed in a few other studies exploiting a similar transparent selection algorithm used to evaluate the National Rural Employment Guarantee Act (Zimmermann, 2017; Zimmermann and Khanna, 2017). In these papers, the running variable is the district's discrete state-specific standardized ranks, and not the score distance as in our RD design.

(2012); Calonico et al. (2014); Cattaneo and Vazquez-Bare (2016); Calonico et al. (2018)).

Since RSVY grants were used for improvements in rural connectivity, electrification, and agricultural and irrigation projects, it is important to understand possible channels through which the policy affected the performance of small firms. Although RSVY grants were invested in a bundle of infrastructure goods, we find strong evidence for one channel. Our findings suggest that improvements in electrification may be one reason why the performance of microenterprises improved drastically. First, we find an increase in the overall infrastructure development following RSVY in the treated districts, as measured by nighttime light intensity. Furthermore, as an immediate consequence, we find that there was a significant increase in a firm's electricity-related expenditure, and a reduction in treated firms' probability of experiencing a power cut.

Any evaluation of place-based policies is incomplete without serious consideration given to spillovers (Glaeser and Gottlieb (2008)). Place-based programs can lead to positive spillovers (agglomeration economies, knowledge spillovers) to nearby areas. They could also lead to negative spillovers in terms of displacement of economic activity from non-treated areas to treated areas. We address the issue of spillovers in two ways. First, we find that RSVY-treated districts saw higher in-migration of people, and an increase in the number of new microenterprises. These increases in in-migration and number of new firms may have been at the expense of nearby areas that did not receive RSVY grants. We specifically address this issue using difference-in-differences estimation by comparing outcomes in districts near the RSVY treated districts to those further away. We find evidence for negative spillovers, as firms in districts closer to RSVY-treated districts saw a decline in employment compared to firms in districts further away. Taken together, this is suggestive evidence that there was a decline in employment in firms in districts closer to the treated districts, and some of these people may have migrated in to the treated districts. However, we find no differential effects on the number of firms. This suggests that there was an actual increase in new firms in treated districts and not a relocation of firms from control to treated districts.

Finally, we conduct a number of 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 policy threshold is hypothetically moved to a different point along the distribution of the distance scores (running variable).

Our paper directly contributes to the literature on place-based policies and extends it in several important dimensions. 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,[5] our paper examines 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. To our knowledge, this is the first paper studying a place-based policy focused on infrastructure intervention within the context of developing countries. Furthermore, in contrast to previous work that looked at medium and large firms, we focus on the effects on microenterprises, which in our context employs close to three-quarters of the work force.[6] Finally, we are able to shed light on one plausible channel through which the infrastructure grants improved the performance of microenterprises.

Given RSVY's focus on infrastructure, our study also relates to the growing literature that seeks to establish causal links of infrastructure on rural economic development in poor countries. More specifically, it directly complements the findings that investments in core public goods such as rural road networks and electrification can generate massive changes in rural economies.[7] In India, Asher and Novosad (2017); Adukia et al. (2017); Aggarwal (2018); Donaldson (2018) find that better rural access can positively influence welfare, rural market integration, and education and occupational choice. Furthermore, increased electrification has also been documented to improve industrialization (Rud, 2012), firm performance (Abeberese (2017); Allcott et al. (2016)) and household welfare (Chakravorty et al., 2014; Dinkelman, 2011),[8] and the human development index (Lipscomb et al., 2013). Cross-country analyses also indicate that electricity outage can be a significant constraint to production, especially to small firms. Alby et al. (2013) show that there is a significantly lower share of

[5] 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., forthcoming), the French ZFUs (Mayer et al., 2012; 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 welfare (Wang, 2013; Lu et al., 2015; 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.

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

[7] There is also evidence that other infrastructure investments such as dams (Duflo and Pande, 2007) and fast Internet (Hjort and Poulsen, 2018) increased agricultural production, and employment respectively.

[8] In contrast, there is recent evidence that rural electrification may not generate large gains for rural households (Burlig and Preonas, 2016; Lee et al., 2016). For example, Burlig and Preonas (2016) study the Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY), a national rural electrification scheme launched in 2005 to expand electricity access in over 400,000 rural Indian villages, and do not find evidence of large effects across measures of economic outcomes. Note that, our paper studies an earlier scheme (RSVY) launched in 2003-04, and we look at effects on microenterprises in 2005-06.

small firms in electricity-intensive sectors in high-outage countries. Hardy and McCasland (2017), support this finding with evidence on Ghanaian garment-making small firms, and find that blackouts lead to significant declines in revenues, work hours, wages, and profits.

Finally, our paper relates to another strand of literature - the economics of microenterprises. Despite the important role played by microenterprises in developing economies, the existing evidence on them is scant, partially due to the lack of data. Existing 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 (McKenzie, 2017; Karlan and Zinman, 2009; Karlan et al., 2014, 2015; Banerjee et al., 2015; Banerjee, 2013; de Mel et al., 2008; Rotemberg, 2017) and managerial constraints (Cole et al., 2011; Bruhn and Zia, 2013; Drexler et al., 2014; Bloom and van Reenen, 2007; Bruhn et al., 2018) on firm performance. We contribute to this body of research by analyzing the effect of relaxing a different production constraint; the improvement of infrastructure environment in which small firms operate. We are also able to provide evidence on negative spillovers in nearby areas as workers migrated in to the treated districts after the RSVY-infrastructure grants. Our results suggest that infrastructure investments (especially electricity) is important for the growth of microenterprises.

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

2 The Policy

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, 2003b). RSVY covered a total of 147 backward districts, out of approximately 600 districts in the country. Under the policy's guidelines, each district was entitled to receive unconditional cash transfer amounts of 450,000,000 Rupees (approximately $7 million USD) over the course of 3 fiscal years beginning 2003-04. The proposed transfer mechanism was equal payments of 150,000,000 Rupees, i.e. one-third of the total fund, per year. Figure 1A graphically details the recipient locations, broken down by 2 separate groups: (i) 115 regular districts that were selected specifically based on a transparent assignment mechanism discussed in the next subsection, and (ii) 32 left-wing districts affected by Naxalite violence, that were automatically included.

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, 2003a). 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 fund was invested in infrastructural interventions, including rural connectivity, electrification, agricultural and irrigation improvement projects (Program Evaluation Organization, 2010).

2.1 Assignment Mechanism

Unlike most place-based programs that are subject to endogenous 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 (Planning Commission, 2003a). In the second step, each state government, in accordance with the assigned quota, chose the specific districts to allocate the RSVY development 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) agricultural wage rate (1996-1997); and (iii) percentage of low-caste population in the district - Scheduled Castes/ Scheduled Tribes (1991) (Planning Commission, 2003b). The Backwardness Index ranked a total of 447 districts in the 17 major states with available data for all three parameters above. Perfect compliance with the proposed selection would imply that the most backward districts, based on their relative backwardness scores in each state would be selected for RSVY. In addition to the above algorithm, the government had a separate list of 32 districts that were heavily affected by Maoist/Naxalite violence. These districts were automatically selected into the RSVY program.

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

3 Empirical Methodology

3.1 Reconstruction of Backwardness Score Index

Since RSVY selection process followed a transparent score-based rule, it is feasible to evaluate the effect 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. We show that by reconstructing the central government's selection guideline, we can generate an instrumental variable to address the potential endogeneity concerns related to the actual allocation of the RSVY funds across districts. 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. As discussed earlier, 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, 2003a). We perfectly reconstruct the composite score for each district in our sample. By ranking the district scores relative to other districts within each state, we 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) the districts' score distance to the state-specific cutoff, which we refer to as the "standardized distance score".

From (i) we obtain the full list of districts that should have been granted RSVY funding if there had been perfect compliance with the central government's guidelines. This list includes all districts with state-specific backwardness scores below their state's cutoff. Compared to the list of districts that should have received funding, we find some non-compliance, in the sense that some districts that were not supposed to receive the RSVY grants did in fact receive them. To address endogeneity concerns with the policy's actual selection, we utilize the reconstructed selection as an instrument. With regards to our instrumented selection, we exclude any district potentially subject to endogenous assignment status. More explicitly, the "Backwardness" Index ranking data is available for 447 districts for the 17 major states in India.[10] In our sample of 147 districts that actually received the RSVY grants, 19 (12.9%) belong to states with missing ranking data. To the extent that the actual RSVY assignments to these 19 districts are endogenous i.e. they were funded without having Backwardness Index information, we remove them from our estimation sample. Quantitatively, our estimates should thereby provide a lower-bound of the actual impact of RSVY. Among the districts with available ranking data, we further drop the 32 districts affected by left-wing extremist violence, as their selection was not based on the backwardness index. This leaves us with 115 districts that actually received the grants.

Out of 115 RSVY districts, 96 had available ranking data. The assignment algorithm had a prediction accuracy of 80.2% and we correctly predicted 77 of the 96 districts that received RSVY (and had backwardness ranking/scores). Our prediction accuracy is distinctly different from a random draw of districts from the pool (21.48%),[11] and provides credence to our approach.

We use the standardized distance score as the running variable for our RD design. Formally, the standardized 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 standardized distance score of 0. That is:

$$z_{ds} = x_{ds} - x_{ds}^k \tag{1}$$

The district's state-specific standardized 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

[10] 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.

[11] Randomly drawing 96 districts from the pool of 447 index-available districts results in a prediction accuracy of 21.48%.

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 which study the impacts of NREGA – an employment guarantee program implemented in later years (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 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 discrete running variable in the RDD 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) using local polynomial regression. For example, Lee and Card (2008) show that local polynomial regression may not be feasible if the running variable only takes on a moderate number of distinct values on at least one side of the threshold. Kolesár and Rothe (2018) also show that the common practice of using confidence intervals based on clustered standard errors with a discrete running variable usually has poor coverage properties and affects inference. 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 standardized 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 standardized 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.

To fix ideas, we next motivate an empirical discussion on a hypothetical setting where the districts' program assignment was perfectly implemented. In such a "clean" setting with perfect compliance, RSVY would have been assigned to the most backward districts with non-missing data according to the "Backwardness" Index. Under the identifying assumption that the expected level of the districts' outcome variables is continuous in the index in the absence of RSVY, we would estimate the Local Average Treatment Effect of RSVY using a Sharp RD. We would regress our outcome variables on an indicator variable for RSVY treated districts and a polynomial function in the index ranking. The regression coefficient on the indicator variable would provide a consistent estimate of the effect of RSVY on districts around the cutoff value.

However, the implementation of RSVY deviated from this clean setting. Specifically, the expected treatment assignment (i.e. based on districts' backwardness ranking) and actual program receipt did not completely coincide. Hence, we use a Fuzzy RD design to estimate the causal effects of the policy change.

Another important identifying assumption is that districts should not have been able to manipulate their treatment status. This implies that states and districts should not have been able to take purposeful actions in ways that would have influenced their RSVY assignment. In our setting, it is highly unlikely that states or districts were able to manipulate the Backwardness Index. The index was constructed based on historical information that predated RSVY. The Planning Commission used data from the early to mid-1990s to derive the composite under-development index. This limits the possibility of districts strategically misreporting information.

The first stage of our Fuzzy RDD approach 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 (standardized 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.

3.2 Empirical Design

Formally, in our empirical analysis we use both the parametric RD functional form as suggested by Imbens and Lemieux (2008), as well as the local polynomial (non-parametric) approach. In each regression, we report the RD coefficients which capture both the Intent-to-Treat (ITT) and Treatment-on-the-Treated (TOT) effects of the policy. Specifically, the parametric regression takes the following form:

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

where the subscripts refer to a firm-level observation i, in district d, in state s, in year t. Thus, y_{idst} is the firm-level outcome variables of interest (profit, employment, etc.), $RSVY_{ds}$

is an indicator representing actual treatment status, that equals one, if the district was selected to receive RSVY. z_{ds} is the constructed standardized score distance discussed in the previous section, which serves as the running variable in our RD design. $\delta(z_{ds})$ is a polynomial function of the score variable that allows for both linear and quadratic 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 (in log values): population, population share of Schedule Caste/Schedule Tribe (SC/ST, the socially disadvantaged groups), prevalence of important and representative public facilities such as education, medical, postal, and banking services. Vector X^2_d further includes relevant district's time-invariant covariates (in log values): area, boundary perimeter, elevation, and distance to the nearest metropolitan cities.[12] Additionally, all regressions control for a vector of firm-specific exogenous characteristics X_{isdt}, which helps control for potential imbalance at the firm-level. These covariates include the microenterprises' physical operating structure (inside or outside of the household, and whether with fixed premises or not), and owner's gender as well as highest education level. Also, since cut-offs are state-specific, we control for π_s, the state fixed effects, in all specifications. Finally, ε_{idst} is a stochastic error term clustered at the district-level.

As discussed, estimating equation (2) would likely produce biased estimates of the policy effect,[13] since actual treatment $RSVY_{ds}$ is an endogenous regressor. We therefore use the predicted treatment indicator $\mathbf{1}\{z_{ds} \leq 0\}$. This binary instrumental 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:

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

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 at the discontinuity are perfectly comparable, β_1 represents the Local Average Intent-to-Treat ("ITT") effect of the policy. To estimate the Treatment-on-the-Treated ("TOT") effects, we run instrumental variable regressions, instrumenting $RSVY_{ds}$ in equation (2) with $\mathbf{1}\{z_{ds} \leq 0\}$.

Furthermore, we test for the sensitivity of our RD estimates by reporting the corresponding

[12]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. The results are consistent with including either one of the two measures.

[13]Specifically, the estimated coefficient α_1 would represent the treatment effect under a "clean" setting with perfect compliance.

local polynomial effects of RSVY. To do so, we estimate the RD coefficients of interest using non-parametric 2SLS procedure where the bandwidths are calculated following two data-driven bandwidth selection algorithms, including the mean-square-error-optimal ("MSE") and coverage-error-rate-optimal ("CER") approaches. MSE-optimal is a popular bandwidth selection method which was first introduced by Imbens and Kalyanaraman (2012) who developed the first-generation (henceforth IK), plug-in rule leading to the identification of an objective neighborhood selector tailored for RD local-linear regression point estimator. The IK method was extended and generalized to local polynomial point estimators by Calonico et al. (2014), which, relative to the original IK bandwidth selector, is demonstrably superior in terms of finite and large sample properties (Cattaneo and Vazquez-Bare, 2016). Finally, we provide another robustness check to bandwidth selection methods by reporting results obtained under CER-optimal bandwidth selection algorithm as proposed by Calonico et al. (2018).[14]

3.3 Validating the identification assumptions

Treatment assignment at the threshold is only "as good as random" when the polynomial function of the running variable is a smooth, or continuous function. In essence, districts must not be able to perfectly manipulate their relative backwardness scores so as to perfectly 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. Regardless, we visually check for the potential existence of treatment status manipulation by looking at Figure 3. This figure plots the distribution of districts over their standardized distance score measure.[15] If there was strategic manipulation, we should have seen visual evidence of "bunching" in the density of the assignment variable at the treatment cutoff. Figure 3 shows no such bunching and the kernel density function of the standardized distance scores is smooth around the threshold.

Another potential threat to identification would be the presence of contemporaneous public program/intervention with the same development focus that also differentially affected the outcome variables in the RSVY sample of districts employed in our analysis. To the best of our knowledge, no such program existed during this time. The RSVY program was the first national public infrastructure development initiative that the Government of India

[14]This more recent method addresses certain weaknesses with inference properties under the MSE-optimal neighborhood by introducing a revised algorithm to obtain robust bias-corrected confidence intervals for RD point estimates.

[15]This validation exercise is similar to the McCrary (2008) density test for potential manipulation of the running variable.

introduced, that adopted a transparent assignment formula on the basis of a 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 was, in fact, the successor of RSVY, 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, and also did not assign treatment to the same districts as in our empirical estimation. Therefore, it is safe to conclude that these two programs do not contaminate our results at least in the short run (i.e. two years after the policy).

3.4 Difference-in-differences approach to study spillover effects

In the last empirical section of the paper, we provide evidence that addresses the potential geographical spillover effects of RSVY infrastructure cash grants. Place-based policies often induce economic spillovers (both negative and positive) to nearby regions after they are introduced. To the extent that RSVY provided a boost to infrastructural development in the backward districts that potentially relaxed firms' production constraint, one could expect economic relocation from nearby districts into the treated areas. Furthermore, migration would be particularly relevant to informal workers, who very often do not have stable formal job opportunities. It is also likely that these workers are induced to move because the greater economic prospects triggered by RSVY outweighs the cost of relocation. We specifically address geographical spillovers by estimating two separate difference-in-differences regressions, comparing employment outcomes among RSVY-untreated districts based on their proximity to the treated areas:

$$y_{idt} = \delta_1(post_t * neighbor_d) + \pi_d + \lambda_t + \gamma(X_{idt}) + \epsilon_{idt} \tag{4}$$

and

$$y_{idt} = \delta_2(post_t * inverseDistance_d) + \pi_d + \lambda_t + \gamma(X_{idt}) + \epsilon_{idt} \tag{5}$$

Equation (4) illustrates a standard DID regression where

$$post_t = \begin{cases} 1, & \text{if year} \geq 2004 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$neighbor_d = \begin{cases} 1, & \text{if the district shares a border with an RSVY district} \\ 0, & \text{if the district is located further away.} \end{cases}$$

Note that the single regressors $neighbor_d$ and $post_t$ are omitted due to the inclusion of district and year fixed effects, π_d and λ_t. Similar to equation (3), vector X_{idt} represents firm-specific characteristics, and ϵ_{idt} represents firm-specific idiosyncratic error terms clustered at the district level. The estimated coefficient of interest is δ_1, which shows the differential effect of RSVY on neighboring districts compared to those further away.

If there were indeed spillover effects of RSVY, we would expect such spillovers to be differentially stronger in neighboring districts of RSVY-treated districts. One potential concern with the use of $neighbor_d$ as a measure for nearness is the unbalanced sample sizes for treated (RSVY-neighboring districts) and control groups (RSVY-non-neighboring districts).[16] We tackle this arbitrariness by estimating equation (5), by replacing $neighbor_d$ with $inverseDistance_d$, where distance is measured by the shortest possible distance between a particular non-RSVY district to an RSVY-treated district (in '00 kilometers).[17] The coefficient δ_2, thus measures the differential impact of RSVY on a district located one hundred kilometers closer to the treated region relative to its counterpart.

4 Data and Variables Formation

We use several data sources for the analysis. For the main firm-level outcomes on profits, sales revenues and hiring activities, we use information from rounds 56 (2000-01) and 62 (2005-06) of the National Sample Survey - Manufacturing Enterprises Schedule "NSS - Sch. 2.2"). We control for observable baseline and time-invariant covariates at the district-level by utilizing information from the 2001 Population Census and GIS-processed shapefiles for the country. Because we study the impact of RSVY on manufacturing enterprise performance within the rural sector in India, we only keep rural observations in our analysis for all data sources. Finally, we proxy for the overall district's infrastructure environment with a measure of night-time light intensity processed from NASA's satellite transmitted data. We discuss each of these sources below.

[16] There are over 300 RSVY neighbors, whereas there are only 88 non-neighbors.

[17] As a robustness exercise, we also vary this nearness measure by experimenting with the averages of the shortest possible distances to the nearest two, three, four, and five RSVY districts. In terms of the migration context, this exercise allows for the possibility that informal workers do not have to migrate to the nearest RSVY for better economic opportunities. In addition, we also experiment with another proximity measure of centroid-to-centroid distance between districts. All results remain robust.

4.1 National Sample Surveys (NSS)

The NSS - Sch. 2.2 is the most comprehensive nationally representative survey in India that provides detailed information on manufacturing microenterprises' business activities and performance.[18] Only small, "unorganized" firms with less than 10 workers and use electrical power, or 20 workers which do not operate with electrical power, are included in this survey.[19] It is worth noting that micro firms meeting these employment criteria account for nearly 80% of India manufacturing employment (Nataraj, 2011), although are often overlooked from more established, high-frequency surveys[20] due to their small scale.[21]

Sampled firms are asked questions regarding their cash flows and operating activities such as operating revenues (including total value of production outputs), expenditures (including total value of production inputs), employment, wage, sources of capital, as well as various types of investments in assets. Quantitative questions are often addressed 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. Given that RSVY was introduced in the fiscal year 2003-04, information from round 62 (2005-06) perfectly 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.[22]

For our analysis, we calculate an enterprise's gross monthly profit by deducting from its total monthly revenues any monthly production expenditures. Firms' total revenues consists of all cash inflows collected from sales of completed and semi-completed outputs. Total production cost includes operating expenditures (total value of physical inputs plus other operating expenses such as warehousing or sub-contracting), cost of labor (total wage), as well as capital (payables on loan interesting and other financing activities). To accurately capture the effect of the policy on a firm's performance, we restrict our sample to perennial businesses.[23] We also consider information on firm-level employment. To test for the underlying mechanisms, we particularly utilize the subjective infrastructure-related question that ask enterprises whether (they) experienced power cut during production in the last year? We

[18] The Economic Census conducted by the Ministry of Statistics and Program Implementation (MoSPI) also surveys small firms, however provides only basic business information, making it insufficient for our study.

[19] 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.

[20] For example, the Annual Survey of Industries (ASI) includes only medium and large firms - those who registered under the 1948 Factories Act, or the Bidi & Cigar Workers Act.

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

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

[23] Seasonal businesses only account for less than one percent of the observations in our sample.

also define an indicator variable for "electricity use", if the firm incurred positive expenses on electricity-related charges.

To study the effects of the policy on migration outcomes, we use data from the National Sample Survey (Schedule 10.2) employment-unemployment survey round 64 (2007-08). The survey elicits information about the last usual place of residence for the household members. We define a migrant as one whose last usual place of residence was another district in the same state.[24]

4.2 Population Census (2001) & Geographic data

Since RSVY was directed at districts, our empirical design additionally accounts for various observable district-level baseline characteristics.[25] We use the 2001 Population Census (pre-RSVY) to construct 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 education, medical, postal, and banking facilities.

In addition, it is important to also control for district's geographic characteristics.[26] 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),[27] and distance (in kilometers) to the nearest metropolitan cities.

4.3 Night-time Light Intensity

Besides documenting the reduced-form effect of infrastructure development grants on microenterprise 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

[24]In India, migration in general has been shown to be low (Munshi and Rosenzweig, 2016), but migration across districts within a state is still significantly higher than migration across state borders (Kone et al., 2018).

[25]These are variables included in vectors X^1_{dt-1} and X^2_d in equation (3). Under the RD identifying assumption, districts located barely to the left and right of the assignment threshold should be "identical". Even though we will visually show that on aggregate, the observable district's characteristics are smooth functions around the cutoff, to be conservative, we still control for these variables as covariates in our regressions.

[26]One can argue that location of districts matter for economic performance of enterprises following an improvement in infrastructure, such as roads. For instance, better roads and connectivity would differentially improve economic communication for districts located near a metropolitan area, rather than those further away.

[27]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).

the policy on the overall progress of infrastructural environment in the treated districts.[28] 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 which provide information on certain public goods. However, they are conducted only decennially, and thus cannot 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.[29]

Figure A1 in the Appendix graphically illustrates the use of night light as a proxy for infrastructure and economic development in India. Light intensity peaks in metropolitan areas where the level of development is high. Regions of low development, including most of the backward areas in our analysis, were almost entirely unlit in 2003. Although they are relatively darker compared to the developed areas ten years later, there is a visible improvement. For our empirical analysis, we further process the raw GIS digital light raster to obtain taluk-level[30] population-weighted light intensity.[31] By design, we assign more weight to light intensity in populated areas where the majority of infrastructure development would take place. Unlit segments that also had low levels of inhabitation respectively receive lesser weight on aggregate.

The use of night-light as a proxy for economic and infrastructure activities has become popular among economists. Its pioneering use in the economics literature was first introduced by

[28] This can be thought of as the first stage in an instrumental variable regression.

[29] 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.

[30] 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.

[31] 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.

Henderson et al. (2012). Recently, a growing body of economic research has started to adopt night light measures for analysis (Alesina et al., 2016; Chen and Nordhaus, 2011; Hodler and Raschky, 2014; Klomp, 2016; Shenoy, 2018). There is currently 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 of public-service provisions such as electricity. Burlig and Preonas (2016) also uses changes in nighttime brightness as an indicator of electrification under RGGVY, a national rural electrification scheme in India.

4.4 Summary Statistics

Table 1 presents the summary statistics of the main outcome and control variables, separately at the firm-level (Panel A) and district-level (Panel B). Our analysis relies on several econometric techniques to select optimal RD bandwidths for each of the outcome variables. Thus, the data-driven optimal bandwidths vary for each of the variables in question. For the purposes of summary statistics, we choose a common bandwidth to present an overview of the outcome and control variables, that consists of all districts located 0.05 score distance around the RD assignment threshold ($|z| \leq 0.05$). For the majority of the variables we study, this range lies between the MSE-optimal and CER-optimal bandwidths, and thus serves as a representative RD window frame. There are 184 districts located within this restrictive window, with over thirteen thousands rural enterprises surveyed.

On average, in 2005-06, a microenterprise in our sample employed 2.44 workers, used around 87 thousand Rs. (approx. USD 1,200) of fixed capital, and earned Rs. 3,790 in monthly profit (USD 55). Firms' operating scales also vary significantly, with a standard deviation of employment and monthly profit being 2.98 workers and 12,532 Rs respectively. For this reason, we will report the regression results in both changes in magnitude and as a percentage of a standard deviation. More than a tenth of the firms are young and started operating in the last three years (with the standard deviation of 32%). Additionally, most of the microenterprises are owned by low-education workers – the majority of firm owners only completed middle school or below.[32] Electricity use in our sample is low with only 42 percent of firms reporting positive electricity-related expenses. Besides the low prevalence of electricity usage, the quality of electricity provision in rural India has been shown to be another major obstacle for firms' production (Abeberese, 2017; Allcott et al., 2016). 18 percent

[32]The categorical measure of owner's education ranks from 1 (lowest – "not literate) to 11 (highest – "post graduate or above") with 5 being "middle school or below".

of firms in the sample reported to have experienced a power cut in the year of the survey.

The high probability of experiencing production constraints due to insufficient infrastructure provision corresponds with the low luminosity levels of night-light emission, which is coded at about 2.8 overall in our sample of districts. Many districts in the sample are located in remote and mountainous regions, with average elevation of 235 meters above the sea-level. An average district's population in the sample is over 1.7 million.[33] Even though the RD estimate is supposed to be informative about the sub-population of individuals at the discontinuity (Lee and Lemieux, 2010), the large population around the threshold provides some generalizable conclusions.

5 Empirical Results

In this section, we first present empirical evidence on the effects of RSVY on rural enterprises' profits, employment, revenues, and fixed capital. Then, we discuss results on a potential channel driving our main results. Finally, we focus on migration, young firms, and spillovers from nearby areas. For all our outcome variables, we test for the robustness of our estimates across five alternative specifications. The five alternative specifications are as discussed in section 3: (i) linear intent to treat, (ii) quadratic intent to treat, (iii) linear treatment on the treated, (iv) quadratic treatment on the treated, and (v) local polynomial. Furthermore, for all our main outcome variables, we report results across two sets of optimal bandwidth selection procedures (MSE-optimal bandwidth, and CER-optimal bandwidth). In the Appendix tables, we conduct robustness tests for our main results.

5.1 Firm-level results

In Table 2, we look at the impact of RSVY on the profits of enterprises. We use NSS round 62 for the analysis, and this provides us with information on a microenterprise's performance two years after the program's introduction. We find a sizable and statistically significant increase in rural manufacturing enterprise's monthly gross profit, across columns 1 through 5 (in both panels A and B), in treated districts relative to those in the control districts. In terms of the intent-to-treat effect of RSVY, we find a robust increase in gross monthly profit between 23.7%-34% across different RD functional specifications (linear (column 1) and quadratic (column 2)) and optimal bandwidth selection techniques (MSE-optimal (Panel A)

[33] With only the rural sector, the total population still amounts to over 100 million for our restricted sample.

and CER-optimal (Panel B)). This amounts to a positive increase of 898-1,289 Rs above the average monthly profit for a microenterprise in our sample, or equivalently a change of 0.072-0.103 of a standard deviation. Given the imperfect compliance in RSVY selection, we also report the treatment-on-the-treated coefficients in columns 3, and 4. Depending on the RD specifications and optimal bandwidths, we find a statistically significant increase in profits between 38.8%-59.5% for treated firms. This translates to 0.12 to 0.18 of a standard deviation. In column 5, we also look at the effects of RSVY using a local polynomial. We find that average profit increased discontinuously at the RSVY eligibility threshold by between 28.8% (0.087 standard deviation) and 45.7% (0.138 of a standard deviation). Taken altogether, it is evident that the introduction of RSVY generated significant profits for treated firms. Since profits could be negative, we also use two separate profit measures – in levels and log-modulus transformation in Appendix Table A1, and find robust results.

In terms of production inputs, in Table 3, we look at the effects of RSVY on firm-level employment. Across panels A and B, and across columns 1 through 5, we find increases in firm-level employment in treated districts. Firms in treated districts increased employment by 0.325 to 0.56 workers in panel A, and between 0.285 to 0.8 workers in panel B. This is a sizable increase in the number of workers compared to a sample mean of 2.44 workers.

Figure 4 presents the graphical representations of the results from Tables 2 and 3. Each scattered point in the graph represents bin-averaged values of log profits and levels of employment after partialling out the state fixed effects and the district-level covariates used in equation (3).[34] To be consistent with the summary statistics section, we continue to use the representative bandwidth of 0.05 ($|z| \leq 0.05$). The graphs plot linear and quadratic fitted curves on both sides of the cutoff and the corresponding 95% confidence interval bands. Panel A1 visually shows that there was no discontinuous jump in profits in the baseline period (2000-01), whereas Panel A2 shows a discontinuous jump in profits in the treated districts after RSVY was introduced (2005-06). Similarly, Panel B1 shows that there was no discontinuous jump in employment in the baseline period (2000-01), but a discontinuous jump in employment (Panel B2) in the treated districts after RSVY was introduced (2005-06).

In Table 4, we further look at the effects of RSVY on firm-level revenues (panel A) and fixed capital (panel B). Consistent with an increase in firm-level profits, we find a statistically significant increase in monthly revenues for firms in the treated districts. Even under our most conservative RD estimates (i.e. the intent-to-treat effects with an MSE-optimal bandwidth), we find an average increase of 39.2% above the mean, or 10,622 Rs. (approx. USD150), for microenterprises in RSVY treated districts. This effect is equivalent to an increase of 0.049 standard deviations in revenues. This evidence, combined with the results

[34] 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).

for profits and employment, suggest that microenterprises were, on average, able to expand and generate higher revenues and profits, shortly after the introduction of RSVY. Next, we look at firms' fixed capital investment. According to Panel B in Table 4, we find no statistically significant changes in fixed capital at the firm-level for the treated districts relative to control districts. This is not surprising given that we are looking at the short-term effects of RSVY, whereas firms undertake capital investment in the medium to longer term. The corresponding graphical representation is shown in Figure 5. Panels A1, and B1 show the effects of RSVY on revenues and capital in the baseline period whereas Panels A2, and B2 show the effects after RSVY.

To summarize, our results suggest that in the short-term, firms in districts that received RSVY infrastructure grants had increases in revenues, employment, and profits. These increases could be related to the relaxation of production constraints for microenterprises due to improvements in infrastructure. We look at the mechanisms driving these results next.

5.2 Mechanisms

The RSVY cash grants were fundamentally injected to foster infrastructural development in the backward districts. Therefore, one would expect that a main driver of the reduced-from policy effects on micro-enterprises is through direct improvements in the overall infrastructural environment in treated districts. First, we proxy for districts' level of infrastructure development by night-time light luminosity. We then estimate the RD coefficients using both MSE-optimal and CER-optimal bandwidths 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 nightlight outcome 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 A4, where due to space constraints we only show the coefficients between 2001 and 2010. We find that night-lights started to grow differentially faster 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 2008-09. 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 ones under the new policy.

Next, we look at whether firms benefited from these overall infrastructure improvements following RSVY. To answer this, we look at how enterprises responded to infrastructure-related questions that were asked to them in the NSS survey. Specifically, we focus on two measures. First, for each firm, we define an indicator variable for electricity use if they incurred a positive expenditure on electricity-related charges. Second, we focus on firms' responses to the question on the nature of problems faced, if any, during the reference year with respect to power cuts. Table 5 provides suggestive evidence that the probability of electricity use went up and the likelihood of experiencing power cuts went down for the treated firms. Across different specifications, firms operating in the treated districts reported a significant increase in electricity use by 8.7% to 23.5%, and a reduction in problems related to power cuts by between 5% and 15%. Although RSVY may have relaxed other infrastructure-related constraints for firms, our findings indicate that improvements in electrification was a major channel leading to higher firm performance. Our results therefore support existing studies on the importance of electricity provision for firm growth (Reinikka and Svensson (2002), Allcott et al. (2016), Abeberese (2017)).

5.3 New Firms and Migration

RSVY grants increased overall infrastructure (proxied by nighttime lights) in the treated districts and reduced electricity related constraints for firms. These positive effects may have resulted in new firm creation. We use the NSS Sch 2.2 (2005-06) to define a new firm if it had operated for less than 3 years. Although imperfect, this is the only measure of firm age available in our data set. Since our data covers a period of around 1.5-2 years after RSVY came in to effect, we believe that the regression estimates for new firms may be an overestimate. In Table 6, panel A, we look at the effects of RSVY on the probability that a firm reported as being new. We find an 8%-17% increase in the likelihood of a firm reporting as being new in treated districts compared to control districts. However, the local polynomial regressions in column 5 (panels A1 and A2) show statistically insignificant results for new firms.

We next look at whether RSVY districts had an increase in migrants as compared to control districts. For this analysis, we use NSS Sch 10.2 (2007-08), and define a migrant as one whose last usual place of residence was the same state but another district. This definition is similar to the one used in (Chaurey, 2017). In Table 6, panel B, we find across our various specifications, that treated states had around 1.6%-3.5% more migrants than control districts. This is a novel result in the Indian context where previous work has found that migration rates are very low ((Munshi and Rosenzweig, 2016)).

Taken together, these results suggest that RSVY increased new firms and migrants in the treated districts compared to control districts. However, these increases in migrants and new firms may have been the result of a relocation in economic activity between treated and control districts. We explore these spillovers next.

5.4 Spillovers

To analyze spillovers in economic activity, we remove our treated districts from the sample, and run difference-in-difference regressions comparing outcomes in districts closer to RSVY treated districts to those that are further away. We use two measures of distance to define our "spillover" group. First, we use neighbors of districts that received the RSVY districts as the spillover group and second, we use an inverse distance measure from an RSVY treated district. We use these regressions to look at spillovers in employment (using NSS data for 2001 and 2005) and firm count (using the Economic Census for 1998 and 2005). In Table 7, columns 1 and 2, we find that employment declined differentially by 4.8% in neighboring districts as compared to those further away, after the policy relative to before RSVY. In columns 3 and 4, we use the inverse distance measure interacted with a indicator variable for the period after RSVY was in place. We find that a district that was 100 kms closer to a RSVY district saw a decline of employment of 0.049% compared to those further away, after RSVY relative to before the policy. However, for the firm counts regressions in Table 7, we find no evidence of spillovers across columns 1 through 4. To summarize, we find that districts closer to RSVY districts saw a decline in firm-level employment compared to those further away, but there were no such differences in the number of firms between districts. These negative spillovers on employment in nearby districts provides additional support to the increase in migrants in the treated districts discussed previously. However, the absence of negative spillovers in terms of firm counts seems to suggest that the results on new firms in the treated districts were not due to relocation of firms from nearby areas to the treated areas, but rather due to actual firm births.

5.5 Robustness Tests

Having discussed the results in detail, we perform two falsification tests, and show that policy effect becomes statistically indistinguishable from zero under counterfactual events. We show that the policy had no effect on districts that did not receive RSVY grants or in districts before RSVY was implemented. First, we run the RD regressions with baseline data from NSS Round 56 (2000-01) survey, that is 3 years before RSVY was implemented. In Appendix Table A2, we show the results for this analysis. We find no significant effects of the policy on our main outcomes of interest before RSVY was implemented. In the second test, we replicate our regressions 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. In Appendix Table A3, we find no statistically significant effects at these hypothetical cutoffs. Apart from these tests, in Appendix Figures A2 and A3, we also graphically show that there are no discontinuous jumps at the cutoff for district-level observable characteristics such as population, SC/ST population, agricultural output, agricultural wages, area, elevation, and the distance to the closest city. 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 place-based infrastructural development scheme (RSVY) on the performance of rural microenterprises in India's backward districts. We exploit RSVY's unique characteristics, including the program's focus on improving the regional infrastructure development, and its transparent treatment-selection mechanism. Utilizing the district's backwardness score index, we reconstruct state-specific score cutoffs based on detailed official guidelines, which allows us to estimate the effects using a Fuzzy Regression-Discontinuity Design. By comparing firm-level outcomes in districts around the RD thresholds, we find significant treatment effects of RSVY on firm's profits, revenues, and employment. Our empirical exercise also sheds light on a potential mechanism underlying the effect on firm's outcomes. We show evidence that RSVY cash grants directly improved treated districts' infrastructural development, with night-light luminosity, and this improvement is in turn is realized by firms, who report significantly lower likelihood of power cuts and an increase in electricity use. We also find an increase in in-migration and new firms in the treated districts. Some of these positive effects in the treated districts may be due to negative spillovers in the nearby areas. We find that districts closer to RSVY treated districts saw a reduction in employment at the firm-level.

While most of the literature on place-based policies has focused on provision of financial incentives such as tax or land subsidies, our paper highlights the effectiveness of an infrastructural development scheme. Finally, this paper adds to the growing body of research focusing on micro-enterprises, and provides new evidence that investments in infrastructure can lead to large gains for them.

A limitation of our analysis is that we are only able to look at the short-term effects of infrastructure grants. Furthermore, we are only able to provide suggestive evidence on one potential channel through which the gains were realized by firms (electrification). 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, Ama Baafra, "Electricity Cost and Firm Performance: Evidence from India," *Review of Economics and Statistics*, 2017, *99*, 839–852.

Adukia, Anjali, Sam Asher, and Paul Novosad, "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction," *Becker Friedman Institute for Research in Economics Working Paper*, 2017.

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

Alby, Philippe, Jean-Jacques Dethier, and Stephane Straub, "Firms Operating under Electricity Constraints in Developing Countries," *The World Bank Economic Review*, 2013, *27*, 109–132.

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

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou, "Ethnic Inequality," *Journal of Political Economy*, 2016, *124* (2), 428–488.

Allcott, Hunt, Allan Collard-Wexler, and Stephen O'Connell, "How Do Electricity Shortages Affect Industry? Evidence from India," *American Economic Review*, 2016, *106*, 587–624.

Asher, Sam and Paul Novosad, "Rural Roads and Structural Transformation," *Working Paper*, 2017.

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

_, Dean Karlan, and Jonathan Zinman, "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, 2015, *7* (1), 2015.

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

Becker, Sascha O., Peter H. Egger, and Maximilian von Ehrlich, "Going NUTS: The effect of EU Structural Funds on regional performance," *Journal of Public Economics*, 2010, *94* (9), 578 – 590.

_ , _ , and _ , "Too much of a good thing? On the growth effects of the EU's regional policy," *European Economic Review*, 2012, *56* (4), 648 – 668.

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

Bloom, Nicholas and John van Reenen, "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

Bondonio, D. and R.T. Greenbaum, "Do local tax incentives affect economic growth? What mean impacts miss in the analysis of enterprise one policies," *Regional Science of Urban Economics*, 2007, *37* (1), 121–136.

Bronzini, R. and G. de Blasio, "Evaluating the Impact of Investment Incentives: The Case of Italy's law 488/1992," *Journal of Urban Economics*, 2006, *60* (2), 327–349.

Bruhn, Miriam and Bilal Zia, "Stimulating managerial capital in emerging markets: the impact of business training for young entrepreneurs," *Journal of Development Effectiveness*, 2013, *5* (2), 232–266.

_ , Dean Karlan, and Antoinette Schoar, "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico," *Journal of Political Economy*, 2018, *126* (2), 635–687.

Burlig, Fiona and Louis Preonas, "Out of the Darkness and Into the Light? Development Effects of Rural Electrification," *Working Paper*, 2016.

Busso, M., J. Gregory, and P. Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 2013, *103* (2), 897–947.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 2018, *0* (0), 1–13.

_ , _ , and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82*, 2295–2326.

Cattaneo, Matias and Gonzalo Vazquez-Bare, "The Choice of Neighborhood in Regression Discontinuity Designs," *Observational Studies*, 2016, *2*, 134–146.

Chakravorty, Ujjayant, Martino Pellib, and Beyza Ural Marchand, "Does the Quality of Electricity Matter? Evidence from Rural India," *Journal of Economic Behavior & Organization*, 2014, *107*, 228–247.

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

Chen, Xi and William D. Nordhaus, "Using Luminosity Data as a Proxy for Economic Statistics," *Proceedings of the National Academy of Sciences*, 2011, *108* (21), 8589–8594.

Cheng, Y., "Place-based Policies in a Development Context - Evidence from China," *Working Paper, UC Berkeley*, 2014.

Cole, Shawn, Thomas Sampson, and Bilal Zia, "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets," *Journal of Finance*, 2011, *66*, 1933–1967.

Criscuolo, C., R. Martin, H. Overman, and J. van Reenen, "Some Causal Effects of an Industrial Policy," *American Economic Review*, forthcoming.

de Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.

Dinkelman, Taryn, "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *American Economic Review*, 2011, *101* (7), 3078–3108.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, April 2018, *108* (4-5), 899–934.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, "Keeping It Simple: Financial Literacy and Rules of Thumb," *American Economic Journal: Applied Economics*, 2014, *6* (2), 1–31.

Duflo, Esther and Rohini Pande, "Dams," *The Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Givord, P., R. Rathelot, and P. Sillard, "Place-based Tax Exemptions and Displacement Effects: An Evaluation of the Zones Franches Urbaines Program," *Regional Science Urban Economics*, 2013, *43* (1), 151–163.

Glaeser, E. and J. Gottlieb, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, (1), 155–239.

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

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

Hardy, Morgan and Jamie McCasland, "Light off, Light on: The Effects of Electricity Shortages on Small Firms," *Working Paper*, 2017.

Hari, Siddharth and Kalyani Raghunathan, "Providing more than just Employment: Evidence from NREGA in India?," *Working Paper*, 2017.

Hasan, Rana, Yi Jiang, and Radine Michelle Rafols, "Place-Based Preferential Tax Policy and Its Spatial Effects: Evidence from India's Program on Industrially Backward Districts," *Working Paper*, 2017.

Henderson, Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 2012.

Hjort, Jonas and Jonas Poulsen, "The Arrival of Fast Internet and Employment in Africa," *NBER Working Paper No. 23582*, 2018.

Hodler, Roland and P.A. Raschky, "Regional Favoritism," *Quarterly Journal of Economics*, 2014, *129.*

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

_ and Peter J. Klenow, "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 2014, *129*, 1035–1084.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 2012, *79* (3), 933–959.

Imbens, Guido W. and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 2008, *142* (2), 615 – 635. The regression discontinuity design: Theory and applications.

Karlan, Dean and Jonathan Zinman, "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts," *Review of Financial Studies*, 2009, *23* (1), 433–464.

_ , Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics*, 2014, *129* (2), 597–652.

_ , Ryan Knight, and Christopher Udry, "Consulting and Capital Experiments with Microenterprise Tailors in Ghana," *Journal of Economic Behavior & Organization*, 2015, *118* (281-302).

Kline, P. and E. Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 2014, (129), 275–331.

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

Kolesár, Michal and Christopher Rothe, "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 2018, *108* (8), 2277–2304.

Kone, Zovanga L, Maggie Y Liu, Aaditya Mattoo, Caglar Ozden, and Siddharth Sharma, "Internal borders and migration in India," *journal of Economic Geography*, 2018, *18* (4), 729–759.

Lee, David S. and David Card, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 2008, (142), 655–674.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram, "Experimental Evidence on the Demand for and Costs of Rural Electrification," Working Paper 22292, National Bureau of Economic Research May 2016.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham, "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil," *American Economic Journal: Applied Economics*, April 2013, *5* (2), 200–231.

Lu, Y., J. Wang, and L. Zhu, "Place-based Policies, Creation, and Displacement: Evidence from China's Economic Zone Program," *Working Paper*, 2015.

Mayer, T., F. Mayneris, and L. Py, "The Impact of Urban Enterprise Zones on Establishments' Location Decisions: Evidence from French ZFU," *mimeo, Sciences-Po*, 2012.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

McKenzie, David, "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition," *American Economic Review*, August 2017, *107* (8), 2278–2307.

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

Munshi, Kaivan and Mark Rosenzweig, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, January 2016, *106* (1), 46–98.

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

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

Neumark, David and Helen Simpson, "Chapter 18 - Place-Based Policies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 1197 – 1287.

Planning Commission, "Backward Districts Initiative - Rashtriya Sam VikasYojana - The Scheme and Guidelines for Preparation of District Plans," Technical Report, Government of India 2003.

_ , "Report of the Task Force: Identification of Districts for Wage and Self Employment Program," Technical Report, Government of India 2003.

Program Evaluation Organization, "Evaluation Study on Rashtriya Sam Vikas Yojana (RSVY)," Technical Report, Government of India 2010.

Reinikka, Ritva and Jakob Svensson, "Coping with Poor Public Capital," *Journal of Development Economics*, 2002, *69* (1), 51–69.

Rotemberg, Martin, "Equilibrium Effects of Firm Subsidies," *Working Paper*, 2017.

Rud, Juan Pablo, "Electricity Provision and industrial Development: Evidence from India," *Journal of Development Economics*, 2012, *97* (2), 352–367.

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

Wang, J., "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities," *Journal of Development Economics*, 2013, (101), 133147.

World Bank, "World Development Report 1994: Infrastructure for Development," Technical Report, World Bank 1994.

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

_ and Gaurav Khanna, "Guns and Butter? Fighting Violence with the Promise of Development," *Journal of Development Economics*, 2017.

Results

Table 1: Summary Statistics (using representative bandwidth $|z| \leq 0.05$)

	Observations	Mean	SD	Source
Panel A: Firm-Level variables				
1. Firm-related Outcomes ('000 Rs.):				
Employment (labor count)	13,333	2.441	2.976	NSS62 -Sch. 2.2
Revenue	13,333	27.097	215.906	NSS62 -Sch. 2.2
Profit	13,333	3.790	12.532	NSS62 -Sch. 2.2
Capital	13,333	86.844	461.083	NSS62 -Sch. 2.2
2. Mechanism Measures:				
Share of Microenterprises using Electricity (%)	13,333	0.420	0.494	NSS62 -Sch. 2.2
Probability of Experiencing Power Cut (%)	13,333	0.178	0.382	NSS62 -Sch. 2.2
Inception within 3 Years (%)	13,333	0.112	0.316	NSS62 -Sch. 2.2
Internal (within-State) Migration (%)	183,343	0.0875	0.283	NSS64 - Migration
3. Covariate Measures:				
Owner's Education Level (1-11)	13,333	4.074	2.083	NSS62 -Sch. 2.2
Firm Location/Structure (1-6)	13,333	1.640	1.172	NSS62 -Sch. 2.2
Panel B: District-Level variables				
1. Infrastructure Development:				
Night-light density (proxy)	184	2.802	3.647	NASA/NOAA
2. Geographic Characteristics:				
Area (km sq.)	184	5413.640	3758.439	GIS
Boundary Length (km)		544.991	287.048	GIS
Elevation (m)	184	234.692	210.076	GIS
Distance to nearest city (km)	184	116.375	60.692	GIS
Average distance to nearest 5 cities (km)	184	221.157	86.464	GIS
3. Socio-Demographic Characteristics:				
2001 Population ('000)	184	1710.501	1052.908	DC 2001
Share of SC/ST Population (% 1991)	184	0.273	0.125	PC 2003
Output per Agricultural Worker (Rs. 1990-93)	184	6263.364	4802.480	PC 2003
Agricultural Wage Rate (Rs. 1996-97)	184	33.783	9.409	PC 2003
4. Infrastructural Facilities (Count):				
Education	184	1484.321	1384.488	DC 2001
Medical	184	2028.902	1739.301	DC 2001
Postal	184	292.946	193.507	DC 2001
Banking	184	2352.772	1944.161	DC 2001
5. RD Running Variables:				
Backwardness Composite Score	184	0.329	0.069	PC 2003
Distance to Cutoffs (z)	184	0.008	0.023	PC 2003

Note: This table shows summary statistics for the main outcomes and control variables. The sample of analysis includes all firms operating in the districts with the standardized Backwardness Index Scores (z) within 0.05 point from the cutoff, i.e. $|z| \leq 0.05$ (optimal bandwidth). Sources: 1. NSS62 -Sch. 2.2: National Sample Survey, Round 62 (2005-06) - Manufacturing Enterprises Schedule; 2. "GIS": Geographic Information System - data constructed using ArcGIS software; 3. PC 2001: Population Census 2001; 4. NSS64 - Migration: National Sample Survey, Round 64 (2006-07) - Migration Module.

Table 2: RSVY Impact on Microenterprises' Profits (in logs)

	(1)	(2)	(3)	(4)	(5)
	Linear ITT	Quadratic ITT	Linear TOT	Quadratic TOT	Local Polynomial
Panel A: MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.241**	0.237**	0.393**	0.388**	0.288**
S.E.	(0.108)	(0.103)	(0.200)	(0.193)	(0.148)
Observations	13,352	13,352	13,352	13,352	13,352
Panel B: CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.340***	0.337***	0.595***	0.593***	0.457**
S.E.	(0.0997)	(0.0978)	(0.210)	(0.213)	(0.184)
Observations	11,233	11,233	11,233	11,233	11,233
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows the impact on microenterprise profits (dependent variable: log of monthly profit). Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Panel A consists of estimates using data-driven MSE-optimal bandwidth selection procedure as shown in Calonico et al. (2014) and Imbens and Kalyanaraman (2012). Panel B consists of estimates using instead the CER-optimal bandwidth selection procedure as suggested by Calonico et al. (2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table 3: RSVY Impact on Microenterprises' Employment

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
Panel A: MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.325*	0.325*	0.526*	0.527*	0.560*
S.E.	(0.172)	(0.171)	(0.282)	(0.281)	(0.468)
Observations	13,230	13,230	13,230	13,230	13,230
Panel B: CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.285*	0.299*	0.503*	0.534*	0.800**
S.E.	(0.170)	(0.169)	(0.296)	(0.299)	(0.349)
Observations	10,942	10,942	10,942	10,942	10,942
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows the impact on microenterprises' hiring activity (dependent variable: total employment). Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Panel A consists of estimates using data-driven MSE-optimal bandwidth selection procedure as shown in Calonico et al. (2014) and Imbens and Kalyanaraman (2012). Panel B consists of estimates using instead the CER-optimal bandwidth selection procedure as suggested by Calonico et al. (2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table 4: RSVY Impact on Microenterprises' Revenues & Capital

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
[Panel A] Dependent Variable: Revenues					
[A1] MSE - optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.392***	0.392***	0.653***	0.653***	0.394**
S.E.	(0.112)	(0.111)	(0.225)	(0.223)	(0.161)
Observations	12,537	12,537	12,537	12,537	12,538
[A2] CER - optimal bandwidth (CCF 2017)					
RD Estimate	0.713***	0.704***	1.402***	1.384***	0.668**
S.E.	(0.129)	(0.126)	(0.448)	(0.432)	(0.220)
Observations	9,008	9,008	9,008	9,008	9,008
[Panel B] Dependent Variable: Capital					
[B1] MSE - optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.578	0.615	1.102	1.185	0.587
S.E.	(0.426)	(0.421)	(0.783)	(0.779)	(0.441)
Observations	9,661	9,661	9,661	9,661	9,662
[B2] CER - optimal bandwidth (CCF 2017)					
RD Estimate	-0.00768	0.155	-0.0162	0.330	0.666
S.E.	(0.679)	(0.548)	(1.423)	(1.139)	(0.712)
Observations	7,839	7,839	7,839	7,839	7,839
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows the impact on microenterprises' monthly revenues, and capital. Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Panel A1 and B1 show estimates using MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). Panel A2 and B2 show estimates using CER-optimal bandwidth selection procedure (Calonico et al., 2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table 5: Mechanism – Electricity-related Channel

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
[Panel A] Dependent Variable: Probability that Firm used Electricity in Production					
[A1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.0869**	0.0881**	0.158**	0.160**	0.0710
S.E.	(0.0413)	(0.0411)	(0.0744)	(0.0746)	(0.0628)
Observations	10,254	10,254	10,254	10,254	10,254
[A2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.109*	0.110**	0.231*	0.235*	0.083
S.E.	(0.0551)	(0.0536)	(0.126)	(0.126)	(0.0727)
Observations	7,905	7,905	7,905	7,905	7,905
[Panel B] Dependent Variable: Probability that Firm experienced Power Cut					
[B1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	-0.0503*	-0.0503*	-0.0836*	-0.0838*	-0.087***
S.E.	(0.0270)	(0.0271)	(0.0482)	(0.0485)	(0.033)
Observations	12,338	12,338	12,338	12,338	12,339
[B2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	-0.0638*	-0.0637*	-0.125	-0.125	-0.151***
S.E.	(0.0380)	(0.0381)	(0.0860)	(0.0866)	(0.043)
Observations	8,941	8,941	8,941	8,941	8,942
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table reports RD estimates for two electricity-related measures including probability of electricity usage (Panel A) and probability that firm reported to had experienced power cut in production. Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Panel A1 and B1 show estimates using MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). Panel A2 and B2 show estimates using CER-optimal bandwidth selection procedure (Calonico et al., 2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table 6: Firm Inception and Migration

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
[Panel A] Dependent Variable: Firm Inception					
[A1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.0808**	0.0810**	0.147**	0.173*	0.02466
S.E.	(0.0362)	(0.0363)	(0.0635)	(0.103)	(0.086)
Observations	10,254	10,254	10,254	10,254	10,254
[A2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.0818*	0.0778*	0.147**	0.166*	0.003
S.E.	(0.0492)	(0.0446)	(0.0637)	(0.0986)	(0.005)
Observations	7,905	7,905	7,905	7,905	7,905
Source: NSS 62-Schedule 2.2					
[Panel B] Dependent Variable: Migration					
[B1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.0194***	0.0188***	0.0346**	0.0347**	0.024**
S.E.	(0.0067)	(0.0067)	(0.0148)	(0.0143)	(0.0106)
Observations	148,047	148,047	148,047	148,047	148,047
[B2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.0159**	0.0156**	0.0319*	0.0314*	0.020*
S.E.	(0.0076)	(0.0075)	(0.0171)	(0.0170)	(0.0118)
Observations	125,230	125,230	125,230	125,230	125,230
Source: NSS 64-Employment/Unemployment					
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table reports RD estimates on microenterprises' inception probability (Panel A), and internal migration (Panel B). Panel A employs firm-level inception data from the NSS 62, Schedule 2.2. Panel B employs immigration data at the individual level from the NSS Round 64 – Employment/Unemployment Module. Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Sub-panels A1 and B1 consist of estimates using data-driven MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). Sub-panels A2 and B2 consist of estimates using the CER-optimal bandwidth selection procedure (Calonico et al., 2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

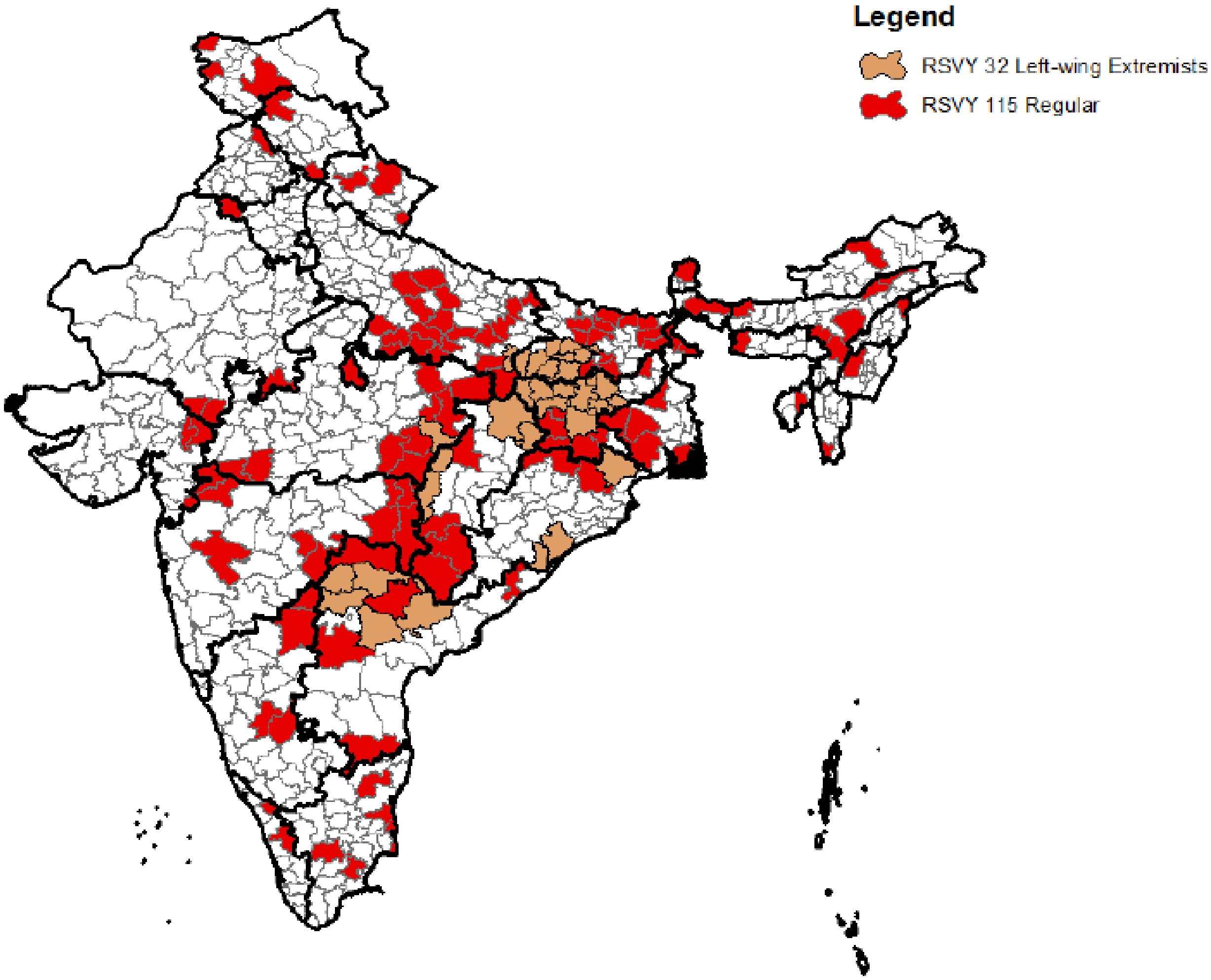
Table 7: Spillover Effects of RSVY

	Neighbor Interaction		Inverse Distance Interaction	
	(1)	(2)	(3)	(4)
Employment (log)				
DID Estimates	-0.0485**	-0.0494**	-0.0049***	-0.0047***
S.E.	(0.0232)	(0.0223)	(0.0014)	(0.0013)
Observations	212,013	212,013	212,013	212,013
District Fixed Effects	Yes	Yes	Yes	Yes
Firm-Specific Controls	No	Yes	No	Yes
Source: NSS - Schedule 2.2 (2001-2005)				
Firm Count (log)				
DID Estimates	-0.0091	0.0075	-0.0025	-0.0015
S.E.	(0.0456)	(0.0430)	(0.0022)	(0.0018)
Observations	47,153	47,153	47,153	47,153
District Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	No	Yes	No	Yes
Source: Economic Census (1998-2005)				

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

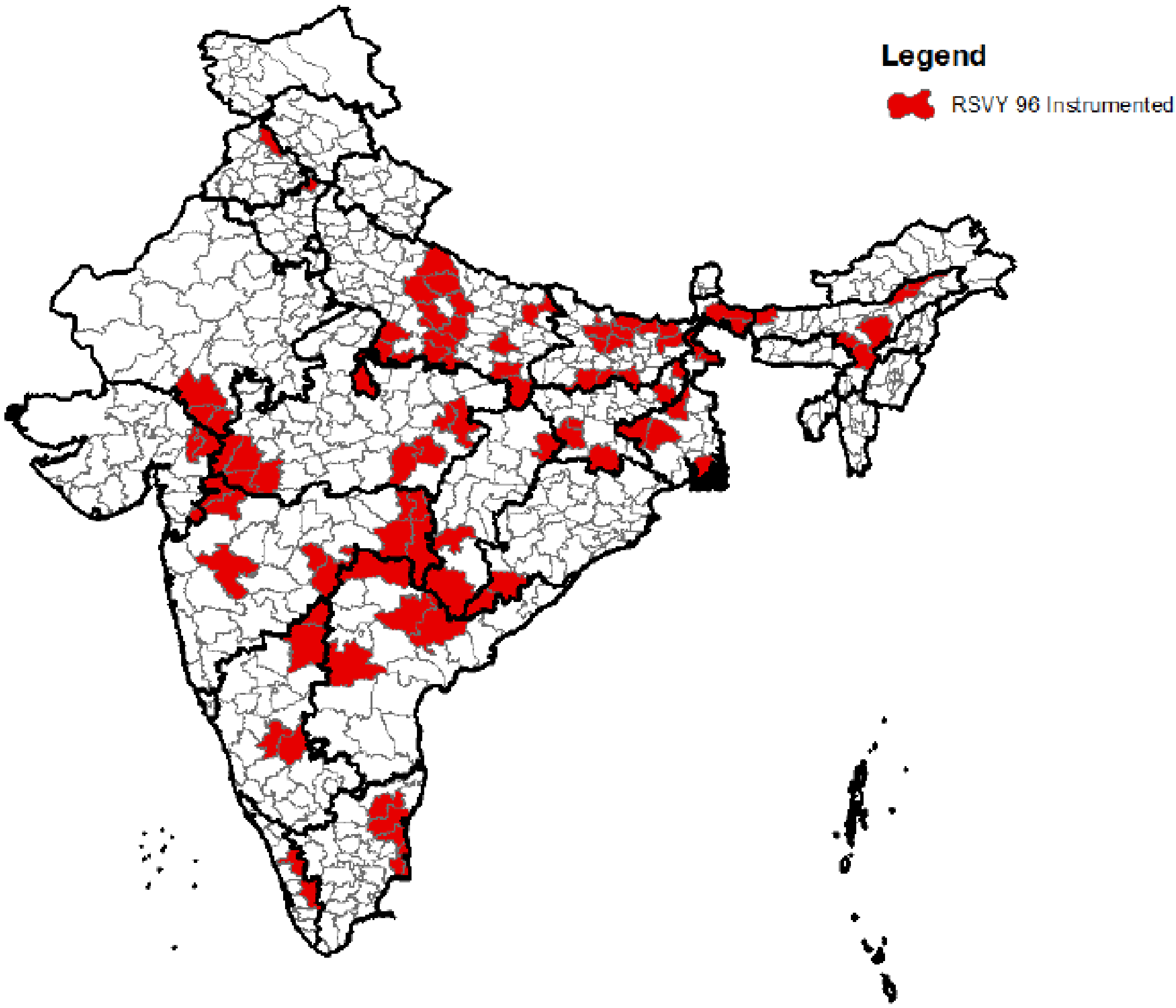
Note: This table shows spillover effects of employment and total microenterprises to non-treated neighboring districts (column 1 and 2) and to non-treated districts locating closer to RSVY regions using inverse distance (column 3 and 4). All regressions employ difference-in-differences approach. For employment variables, we use Round 56 (2000-01) and 61 (2005-06) of the NSS–Schedule 2.2. For the count of microenterprises, we use the Economic Censuses 1998 (pre-RSVY) and 2005 (post-RSVY). All columns control for district fixed effects. Columns (2) and (4) further control for Industry-specific Fixed Effects. Robust standard errors are clustered at the district level.

Figure 1A: 147 RSVY Recipient Districts



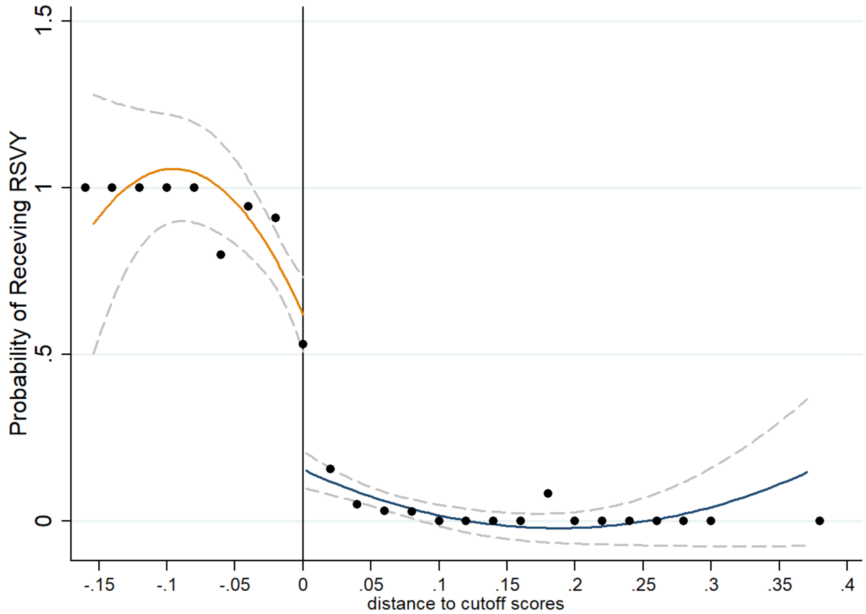
Note: The graph depicts the 147 districts that were chosen to receive RSVY program. Out of the 147 districts, 115 were selected based on the Central Government's assignment mechanism using the "Backwardness" Ranking. 32 other districts that were affected by Naxalite movement (left-wing extremists) were automatically included in the list, bypassing the assignment mechanism. Thick lines represent state boundaries. Thin lines represent district boundaries.

Figure 1B: 96 Instrumented Districts



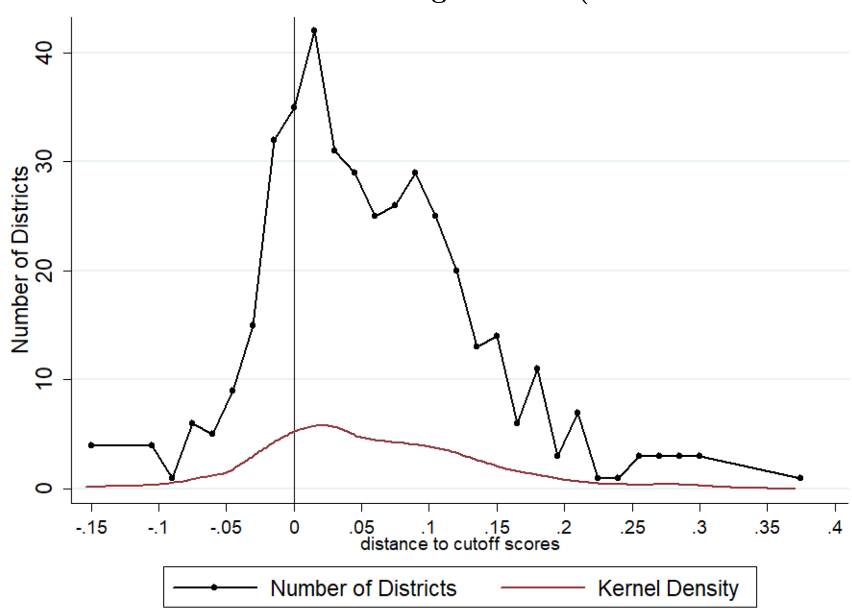
Note: The graph depicts the 96 districts selected as instruments for the actual assignment in the first stage of the 2SLS analysis (Fuzzy Regression Discontinuity Design). Selection criteria are discussed in section 4. Thick lines represent state boundaries. Thin lines represent district boundaries.

Figure 2: Discontinuity in Treatment Probability (First Stage)



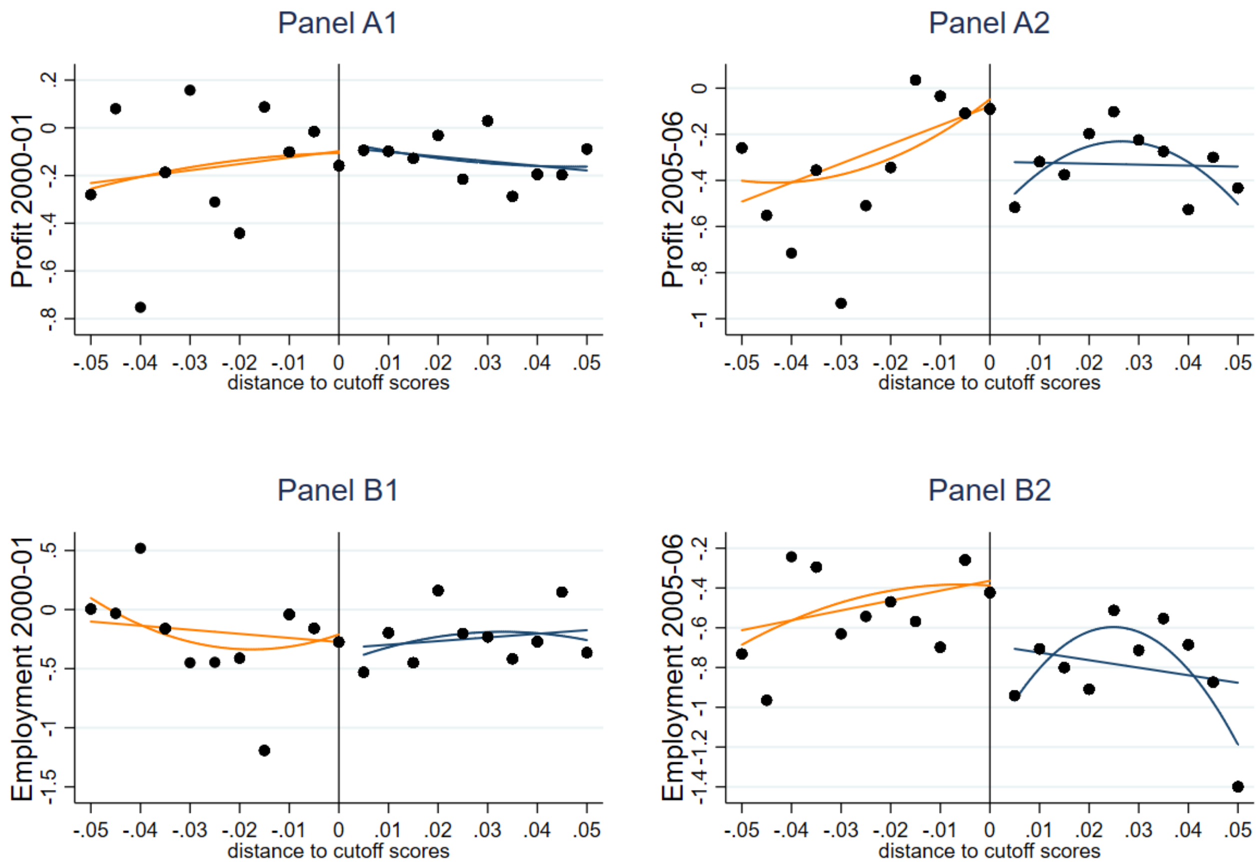
Note: The graph shows cutoff discontinuity on the probability of a district receiving RSVY treatment over the RD running variable (district's standardized distance scores from the cutoff). Quadratic fitted curves on each side of the cutoff as well as 95% confidence interval bands are also included.

Figure 3: Distribution of Districts over Running Variable (Check for Threat of Manipulation)



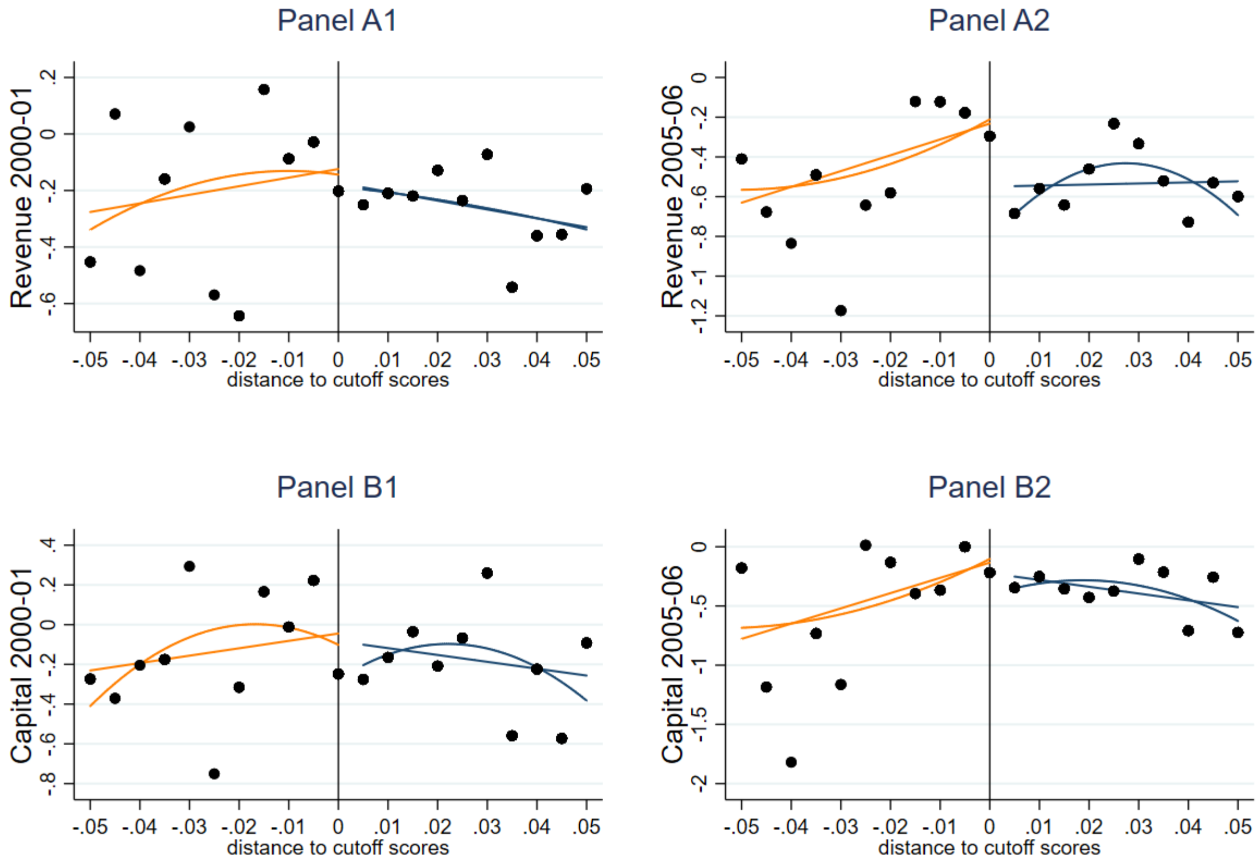
Note: The graph shows the distribution of districts over the RD running variable (district's standardized distance scores from the cutoff).

Figure 4: RD Estimate of RSVY Impact –Discontinuity in Profits & Employment



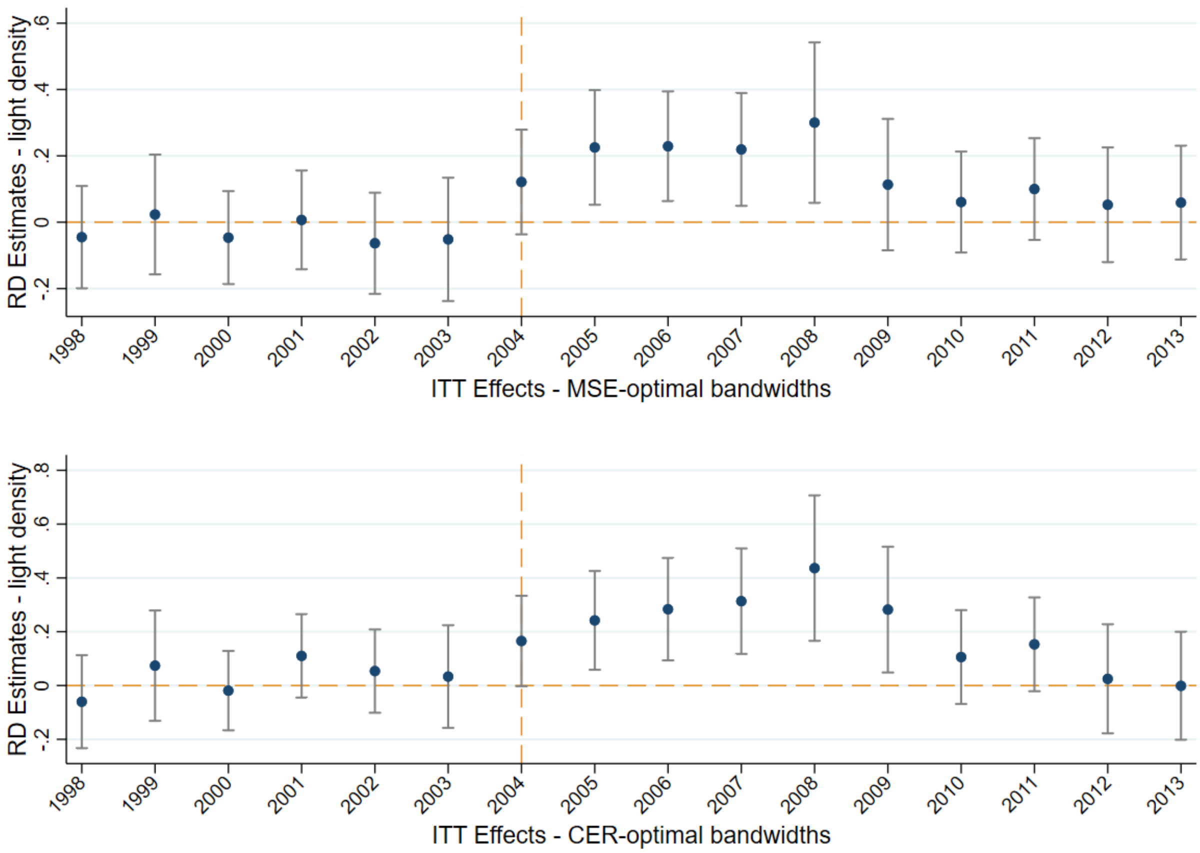
Note: The graph shows treatment cutoff discontinuity in profits (in log) and total employment of firms operated in the districts with the standardized Backwardness Index Scores (z) within the restricted 0.05 point from the cutoff, i.e. $|z| \leq 0.05$. To be consistent with the regressions, each scatter point represents the bin-average of firm's profit and employment, residualized from State fixed effects and districts' baseline socio-demographic, and geographic characteristics. 1. Socio-demographic controls include log values of district's total population as well as the historical under-development measures used to construct the Backwardness Index: share of SC/ST population, historical agricultural wage rate, and historical per-capita agricultural output; 2. Geographic controls include log values of district area (km sq.), boundary perimeter (km); average elevation (m); and distance to nearest city (km). In all graphs, both linear and quadratic fitted curves are presented.

Figure 5: RD Estimate of RSVY Impact –Discontinuities in Revenue & Capital



Note: The graph shows treatment cutoff discontinuity in important measures of microenterprise's operating activities, including firm's capital and total monthly revenues. Sample includes all firms operated in the districts with the standardized Backwardness Index Scores (z) within the restricted 0.05 point from the cutoff, i.e. $|z| \leq 0.05$. To be consistent with the regressions, each scatter point represents the bin-average of firm's profit and employment, residualized from state fixed effects and districts' baseline socio-demographic, and geographic characteristics. 1. Socio-demographic controls include log values of district's total population as well as the historical under-development measures used to construct the Backwardness Index: share of SC/ST population, historical agricultural wage rate, and historical per-capita agricultural output; 2. Geographic controls include log values of district area (km sq.), boundary perimeter (km); average elevation (m); and distance to nearest city (km). In all graphs, both linear and quadratic fitted curves are presented.

Figure 6: RD Estimate of RSVY Impact - Mechanism Tests



Note: The graph plots RD estimates (i.e. equation (3)'s β_1) 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. Both graphs show yearly RD estimates for the intent-to-treat effects. The upper panel employs MSE-optimal bandwidths, and lower panel employs CER-optimal bandwidths. All regressions include taluk-level per capita light density observations and control for state fixed effects, district's baseline (2000-01) socio-demographic, and geographic characteristics. 1. Socio-demographic controls include log values of total household, population, share of SC/ST population, and total number of education, medical, postal, and banking facilities; 2. Geographic controls include log values of district area (km sq.), average elevation (m); distance to nearest city; and average distance to the nearest 5 cities (population $\geq$ 500 thousands). Robust standard errors are clustered at the district level.

Appendix

Table A1: RSVY impact on Profits – Robustness to Different Measures of Profits

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
[Panel A] Dependent Variable: Profit (in Levels)					
[A1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	405.7*	406.8*	662.7*	665.2*	187.8
S.E.	(211.8)	(212.1)	(359.2)	(361.1)	(258.1)
Observations	13,386	13,386	13,386	13,386	13,386
[A2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	569.4***	585.0***	995.0***	1,029***	470.9
S.E.	(193.8)	(194.6)	(344.1)	(351.3)	(370.7)
Observations	11,259	11,259	11,259	11,259	11,259
[Panel B] Dependent Variable: Profit (Log-modulus Transformation)					
[B1] MSE-optimal bandwidth (CCT 2015 & IK 2012)					
RD Estimate	0.264**	0.260**	0.432**	0.425**	0.315*
S.E.	(0.114)	(0.108)	(0.215)	(0.207)	(0.163)
Observations	13,386	13,386	13,386	13,386	13,386
[B2] CER-optimal bandwidth (CCF 2017)					
RD Estimate	0.356***	0.353***	0.621***	0.621***	0.471**
S.E.	(0.102)	(0.1000)	(0.217)	(0.220)	(0.189)
Observations	11,259	11,259	11,259	11,259	11,259
State FE	Yes	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows the RSVY impact on microenterprise profits, with profits either measured in levels (Panel A) and log-modulus transformed (Panel B), as opposed to the natural-log transformation reported in Table 1. Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Panel A1 and B1 show estimates using MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). Panel A2 and B2 show estimates using CER-optimal bandwidth selection procedure (Calonico et al., 2018). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table A2: Placebo Test 1 – Counterfactual Treatment Effect on Pre-determined Variables (2000-01)

	(1) Linear ITT	(2) Quadratic ITT	(3) Linear TOT	(4) Quadratic TOT	(5) Local Polynomial
Profit (in log)					
RD Estimate	-0.149	-0.137	-0.323	-0.296	0.061
S.E.	(0.110)	(0.108)	(0.272)	(0.258)	(0.125)
Observations	27,548	27,548	27,548	27,548	27,548
Capital (in log)					
RD Estimate	-0.123	-0.0977	-0.123	-0.0977	0.197
S.E.	(0.268)	(0.265)	(0.268)	(0.265)	(0.126)
Observations	27,391	27,391	27,391	27,391	27,391
Employment					
RD Estimate	-0.0300	-0.0418	-0.0651	-0.0899	-0.122
S.E.	(0.0717)	(0.0680)	(0.153)	(0.144)	(0.296)
Observations	27,079	27,079	27,079	27,079	27,080
Revenue (in log)					
RD Estimate	-0.151	-0.141	-0.330	-0.304	-0.0899
S.E.	(0.140)	(0.139)	(0.335)	(0.324)	(0.695)
Observations	27,391	27,391	27,391	27,391	27,391
Used Electricity (%)					
RD Estimate	-0.0251	-0.0252	-0.0603	-0.0604	-0.0810
S.E.	(0.0532)	(0.0532)	(0.134)	(0.134)	(0.1023)
Observations	24,211	24,211	24,211	24,211	24,211
Power Cut (%)					
RD Estimate	-0.0666	-0.0617	-0.256	-0.245	0.692
S.E.	(0.0649)	(0.0639)	(0.342)	(0.347)	(0.789)
Observations	20,189	20,189	20,189	20,189	20,189

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows a placebo test for counterfactual RSVY impact on outcome variables previously studied. Regressions use predetermined dependent variables collected in the baseline period (2000-01). Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) and (4) repeat the approach, showing the treatment-on-treated effects. Column (5) documents the RD effect using local polynomial (non-parametric) estimation. Estimates are reported using MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table A3: Placebo Test 2 – Counterfactual Treatment Effect with Hypothetical RD Threshold

	(1) Linear ITT	(2) Quadratic ITT	(3) Local Polynomial
Profit (in log)			
RD Estimate	0.639	0.485	0.023
S.E.	(0.519)	(0.476)	(0.326)
Observations	2,605	2,605	2,605
Capital (in log)			
RD Estimate	0.489	0.420	-0.778
S.E.	(0.574)	(0.554)	(0.647)
Observations	3,005	3,005	3,005
Employment			
RD Estimate	0.605	0.693	0.392
S.E.	(0.721)	(0.537)	(0.316)
Observations	1,609	1,434	1,435
Revenue (in log)			
RD Estimate	0.886	0.677	0.0698
S.E.	(0.596)	(0.542)	(0.438)
Observations	2,613	2,613	2,613
Used Electricity (%)			
RD Estimate	0.0383	0.0451	-0.0417
S.E.	(0.152)	(0.153)	(0.189)
Observations	3,573	3,573	3,573
PowerCut (%)			
RD Estimate	-0.304	-0.478	-0.529
S.E.	(0.308)	(0.303)	(0.404)
Observations	2,145	2,145	2,145
Inception (%)			
RD Estimate	0.105	0.110	0.045
S.E.	(0.182)	(0.181)	(0.038)
Observations	3,005	3,005	3,005

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: This table shows a placebo test for counterfactual RSVY impact, using hypothetical cutoff in district eligibility (threshold is hypothetically moved to $z = 0.01$). Column (1) and (2) in each panel refer to the intent-to-treat RD estimates, measured under both the 1st and 2nd order polynomial specifications. Column (3) documents the RD effect with local polynomial (non-parametric) estimation. Estimates are reported using MSE-optimal bandwidth selection procedure (Calonico et al., 2014; Imbens and Kalyanaraman, 2012). All regressions control additionally for 1. state fixed effects; 2. district's under-development parameters which are components of the backwardness index score: percentage of SC/ST population, agricultural wage rate, and per-capita agricultural output; 3. district's geographic characteristics: District area (km sq); perimeter (km); average elevation (m); average distance to nearest 5 cities ($\geq$ 500k population); and 4. individual firm's characteristics: ownership status, owner's education level, and establishment's location status. Robust standard errors are clustered at the district level.

Table A4: **Regression-Discontinuity Estimates for Night-light Density**

	(1) First Order	(2) Second Order	(3) First Order	(4) Second Order
RD Estimates - Light 2001	0.00690	0.00979	0.110	0.119
	(0.0901)	(0.0941)	(0.0938)	(0.0948)
Observations	863	863	692	692
RD Estimates - Light 2002	-0.0636	-0.0640	0.0536	0.0510
	(0.0923)	(0.0940)	(0.0936)	(0.0942)
Observations	829	829	654	654
RD Estimates - Light 2003	-0.0517	-0.0476	0.0333	0.0464
	(0.112)	(0.114)	(0.115)	(0.117)
Observations	835	835	668	668
RD Estimates - Light 2004	0.121	0.122	0.166	0.170*
	(0.0955)	(0.0984)	(0.102)	(0.102)
Observations	864	864	681	681
RD Estimates - Light 2005	0.225**	0.230**	0.242**	0.242**
	(0.105)	(0.105)	(0.111)	(0.110)
Observations	704	704	567	567
RD Estimates - Light 2006	0.229**	0.233**	0.284**	0.282**
	(0.0999)	(0.102)	(0.115)	(0.115)
Observations	693	693	546	546
RD Estimates - Light 2007	0.219**	0.216**	0.314***	0.319***
	(0.103)	(0.102)	(0.118)	(0.117)
Observations	668	668	530	530
RD Estimates - Light 2008	0.300**	0.288*	0.436***	0.422**
	(0.146)	(0.148)	(0.163)	(0.163)
Observations	692	692	545	545
RD Estimates - Light 2009	0.113	0.107	0.282**	0.292**
	(0.120)	(0.120)	(0.141)	(0.143)
Observations	667	667	498	498
RD Estimates - Light 2010	0.0608	0.0586	0.106	0.114
	(0.0920)	(0.0918)	(0.105)	(0.104)
Observations	655	655	499	499
State FEs	Yes	Yes	Yes	Yes
District Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

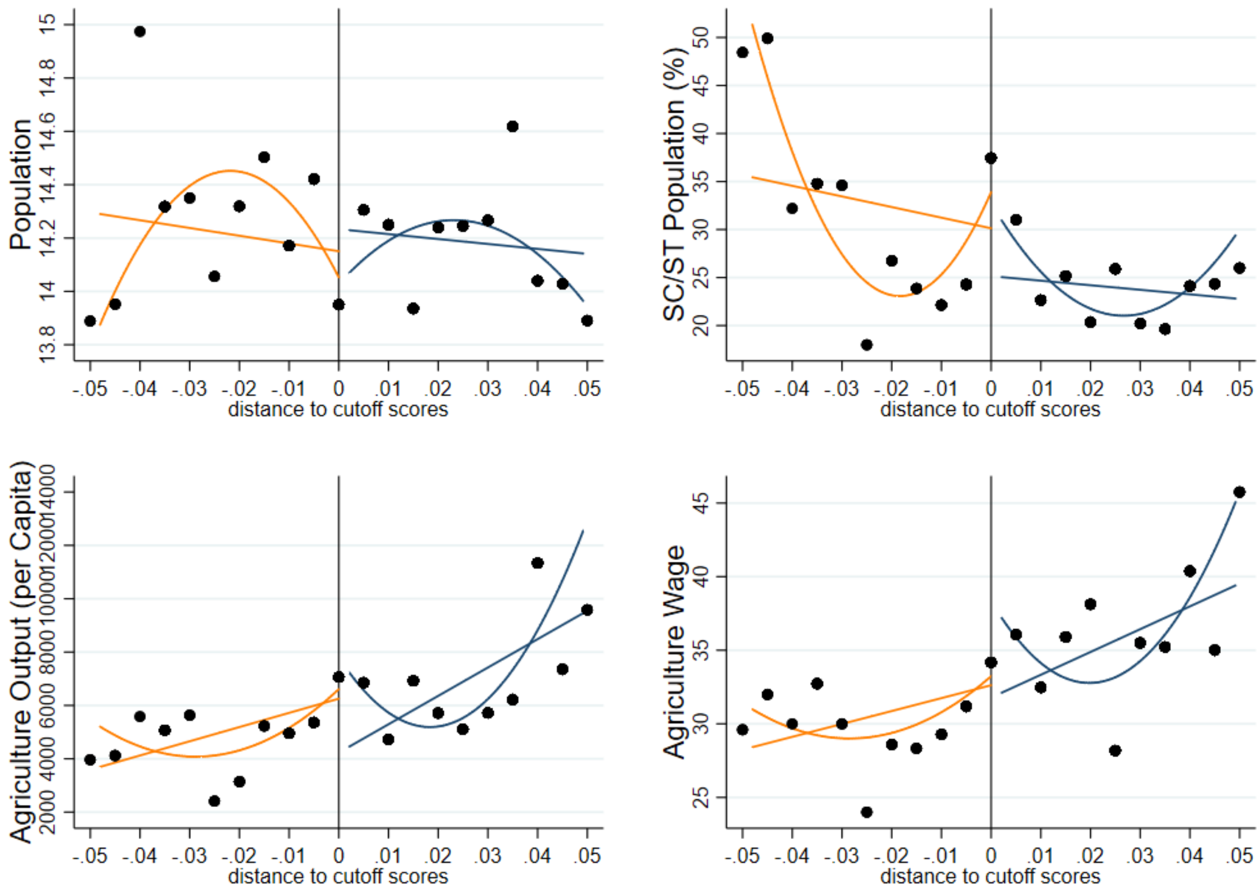
Note: The table shows RD estimates for night-light density, which corresponds with Figure 6. Observations include all taluks (sub-district administrative unit) in the districts with standardized Backwardness Index Scores (z) within the optimal bandwidths computed using data-driven MSE-optimal bandwidth selection procedure (Columns 1 and 2) (Calonico et al., 2014; Imbens and Kalyanaraman, 2012), and using CER-optimal bandwidth selection procedure (Columns 3 and 4) (Calonico et al., 2018).

Figure A1: Night Light Intensity Comparison: 2003 versus 2013

Night Light Intensity 2003 **Night Light Intensity 2013**

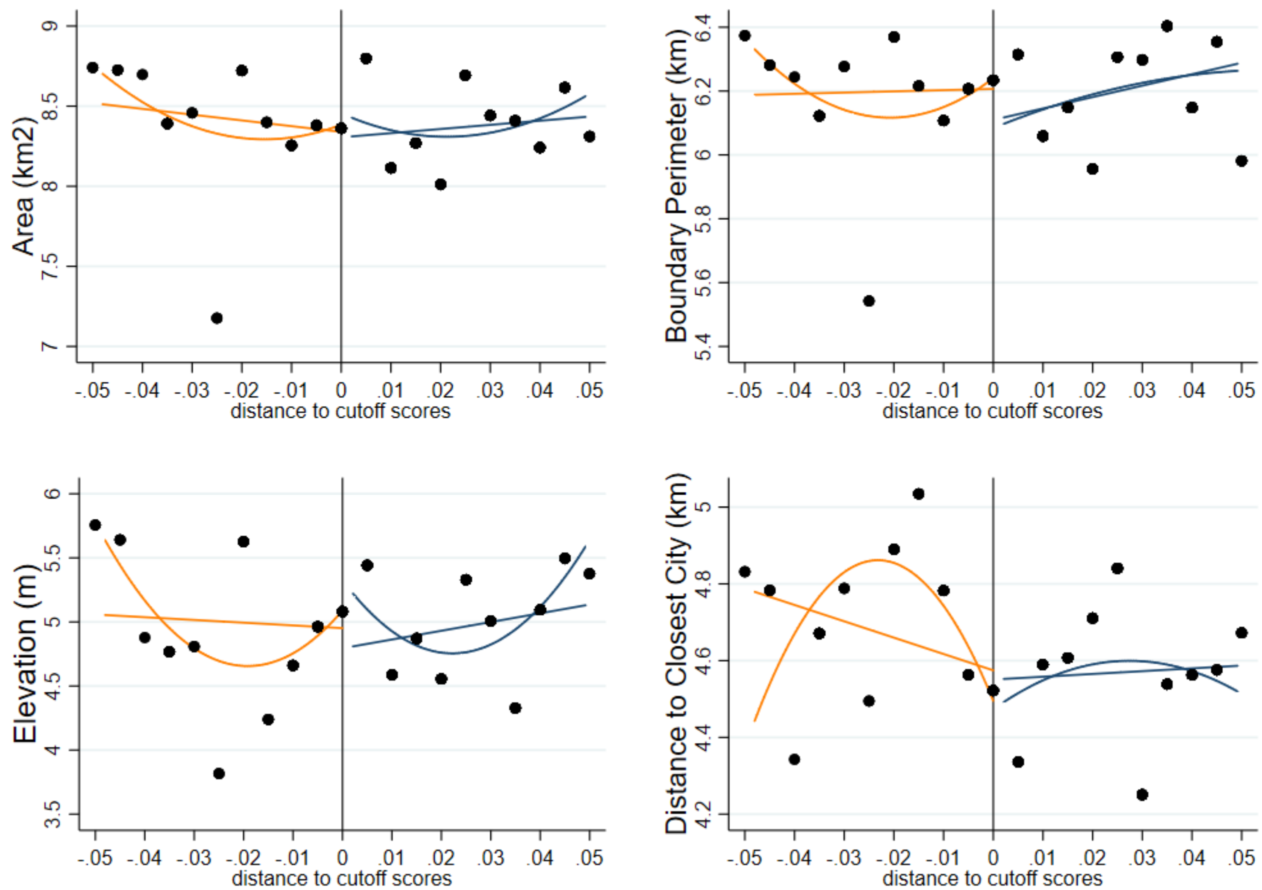
Note: The graph visually illustrates the use of night-light luminosity as a proxy for economic development for entire India between 2003 (left) and 2013 (right). Nightlight luminosity is obtained from satellite imagery of the earth at night, recording light output at the 30 arc-second level, equivalent to approximately 1 square kilometer at the equator. Each light intensity pixel stores a coded digital number as an integer between 0 (no light) and 63 (top-coded, brightest level).

Figure A2: RD Estimate of RSVY Impact – Robustness Check for Discontinuities in Districts' Socio-demographic & Physical Facilities



Note: The graph shows smooth linear and quadratic fitted functions in district-level baseline socio-demographic at the cutoffs. Observations include all districts with the standardized Backwardness Index Scores (z) within 0.05 point from the cutoff, i.e. $|z| \leq 0.05$. Each scatter point represents the bin-average of all firm's outcomes in log values. Socio-demographic characteristics include log of district's population, percentage of SC/ST population in the district, historical agricultural wage rate, and historical per-capita agricultural output.

Figure A3: RD Estimate of RSVY Impact – Robustness Check for Discontinuities in Districts' Geographic characteristics



Note: The graph shows smooth linear and quadratic fitted functions in district-level geographic variables at the cutoffs. Observations include all districts with the standardized Backwardness Index Scores (z) within 0.05 point from the cutoff, i.e. $|z| \leq 0.05$. Each scatter point represents the bin-average of all firm's outcomes in log values. Geographic characteristics include log values of district area (km sq.), average elevation (m); average length of boundary (m); distance to nearest city (km). Both linear and quadratic fits are presented.