CSAE Working Paper WPS/2023-03

Lighting the Path Forward? The Impact of Rural Road Construction on Structural Transformation in India: New Evidence from the PMGSY Scheme and two complementary natural experiments

by

Thomas Kurian

‘

Centre for the Study of African Economies
Department of Economics · University of Oxford · Manor Road Building · Oxford OX1 3UQ
T: +44 (0)1865 271084 · F: +44 (0)1865 281447 · E: csae.enquiries@economics.ox.ac.uk · W: www.csae.ox.ac.uk

Abstract

1 billion people worldwide live over 2 km from a paved road. Consequently, I investigate medium-run impacts of rural road construction on structural transformation in India- identifying how responsive such benefits are based on a) external market conditions and b) in-village electrification. I leverage a regression discontinuity design and triple difference strategy, exploiting discontinuities in population-based eligibility and staggered rollout of the Indian PMGSY rural road program- which aimed to provided all-weather road (AWR) connectivity to 115, 000 villages nationwide. I combine the program with a unique natural experiment induced by the US fracking boom, which created a parallel agricultural commodity boom in the price of guar, a crop providing a necessary fracking input. I compare heterogeneous impacts of AWRs in villages with high and low-intensity exposure to the fracking boom, and separately investigate heterogeneity of roads by village electrification access, exploiting variable implementation intensity of the nationwide RGGVY electrification program. My results imply structural transformation benefits of AWRs are relatively unresponsive to village electrification, whereas external economic conditions can drastically influence these impacts. RD analysis showcases labor reallocation gains from AWRs were entirely concentrated in non-Boom villages- where roads caused a 12.1-7 percentage-point reduction in share of workers employed in agriculture, and 9.2-8 percentage-point increased share employed in non-agricultural manual labour. Conversely, AWRs caused significantly reduced (net zero) structural transformation benefits in boom villages. My findings are robust to multiple specification tests, varying electrification levels, and suggest substantial within-village heterogeneity, with largest discrepancies in new labor market entrants. A plausible mechanism is reduced out-migration impacts of AWRs in boom-villages. These results confirm theoretical predictions that local economic conditions can drastically influence the impact of infrastructure investments suggesting the need for effective spatial and temporal targeting.

Acknowledgements

I thank Jin Kim, for his invaluable guidance and support during the process of writing this paper. I thank Christopher Woodruff, Niclas Moneke, Pramila Krishnan for their useful comments in preparing the project. I thank Luke Milsom, Tom Carlowitz, and Julia Lawson-Johns for their advice on parts of this paper's methodology. I thank Faraz Usmani for his invaluable support and methodological advice throughout this project. I thank my family and friends for their personal support throughout.

1 Introduction

1 billion people worldwide live over 2 km from a Paved Road (World Bank, 2015)- 98% of whom live in developing countries. Poor availability of transportation infrastructure has been identified as a market friction preventing efficient allocation of labour from agriculture, which reports lower productivity than other sectors (Gollin 2014). The assumption that roads contribute to economic growth has spurred rapid investment in "last mile road construction" programs; aiming to expand road networks to remote regions.

However, whereas "first mile" infrastructure investments have generated large welfare improvements in low-income settings (networks in (Banerjee et al. (2020), railroads (Donaldson, 2018)), highways (Storeygard, 2016)), evidence in support of last-mile programs has been mixed. Evaluations of rural road programs in Bangladesh (Khandker 2019), Papua New Guinea (Rozelle 2003), Vietnam (Mu 2011) have observed gains to wages, productivity, and output prices. Conversely, Asher and Novosad (A-N, 2020), who conducted the first major evaluation of the PMGSY rural road program in India, observe far more muted effects[1]. A potential rationale is constraints in rural areas, like absence of complementary inputs and poor agglomeration economies (Rauch 2019) may preclude benefits despite productive infrastructure investments.

Substantial heterogeneity of rural roads' impacts worldwide presents an important empirical puzzle. Which factors influence heterogeneity in benefits of rural roads– either through augmenting (complementarities), or reducing (constraints) impacts of such investments on structural transformation? My paper addresses this question- by identifying how responsive benefits of rural roads on structural transformation are to a) external market conditions and b) in-village electrification.[2]

A challenge in causal identification of infrastructural impacts is endogenous placement, which can bias OLS estimates in unknown directions. I thus leverage a regression discontinuity and triple-difference strategy, exploiting discontinuities in population-criteria and staggered rollout of the Indian PMGSY rural road program- which pro-

[1] A parallel literature, studying the relationship between large scale expansion of electrification, and last-mile electrification expansion, has also uncovered heterogeneous impacts of rural electrification; for example, Kassem (2021) and Allcott (2016) find large productivity benefits from grid-scale electricity expansion in Indonesia and India. Lipscomb et al. (2013) and Dinkelman (2011) find large benefits of employment in rural electrification programs in Brazil and South Africa, whereas Burlig and Preonas (2016), Bernard and Torero (2015), and Lenz et al. (2017) have found relatively lackluster impacts on labor market outcomes in India and Sub-Saharan Africa

[2] In this sense, my results are partially related to an emerging 'big push' literature which aims to study the impact of road infrastructure investments in combination with external policy and other infrastructure shocks, rather than in isolation. (eg Moneke, 2020).

vided all-weather road (AWR) connectivity to 115, 000 villages from 2001-2015. I combine the program with a natural experiment (from Usmani 2020) induced by the US fracking boom, which, created a parallel commodity boom in the price of guar, a crop providing a necessary input to fracking. From 2007-2012, guar prices rose over 1, 000 percent triggered by increased US demand, leading to an exogenous agricultural boom in North-western India, where over 60% of guar is produced. Combining these natural experiments enables me to understand how causal impacts of rural roads vary with exogenous changes in external market conditions. I compare heterogenous impacts of rural roads in Boom and Non-Boom villages , and separately, by in-village electrification (utilising both direct and indirect electrification measures via incidental village exposure to the RGGVY national electrification program)

My results suggest benefits of PMGSY treatment are highly heterogeneous to external market conditions, with gains to structural transformation entirely concentrated in villages with low-intensity exposure to the fracking Boom. My work builds on A-N (2020), who identified the program increased rates of structural transformation, as reflected by labor reallocation away from agriculture, but had limited benefits on other socioeconomic variables. Findings in Non-Boom villages are consistent with previous analysis. RD analysis showcases rural roads caused a 12.1-12.7 percentage-point reduction in share of village households employed in agriculture, and a 9.2-9.8 percentage-point increase in share employed in manual labour. In contrast, I find villages with high-exposure to the fracking-boom saw an economically substantial reduction in structural transformation gains from road construction- amounting to net-zero labor reallocation. Regressions are robust to multiple bandwidth, kernel functions, electrification measures, various specification tests, and placebo RD estimation. Furthermore, I find evidence of within-village heterogeneity - with largest discrepancies observed for young men and new entrants to the labour market.

I find relative reductions in temporary out-migration in PMGSY recipient boom-villages provides the most plausible causal mechanism ; I attribute this to the agricultural boom increasing opportunity costs of leaving agriculture, reducing incentives for temporary out-migration. I reject alternative explanations including a) variation in non-agricultural village firm creation b) permanent migration, c) agricultural productivity and also find evidence suggesting results are unlikely to be driven by d) regional variation between Boom and Non-Boom villages.

Conversely, I fail to observe similar evidence of heterogeneity of structural transformation benefits of rural roads by village electrification level (utilising both direct measures of village power availability, and indirect proxy measures of village exposure to electrification). My findings imply structural transformation benefits of rural roads are not particularly responsive to village electrification, whereas external economic conditions can drastically influence such investments' impact. These insights could be relevant to policymakers in targeting of infrastructure investments to maximize labor reallocation.

I make three contributions to past literature. Firstly, I am the first author to study interactions of the Fracking Boom in Northwestern India and the PMGSY roads program. Originally, Usmani et al (2020) finds evidence Boom villages exposed to rural electrification saw higher rates of structural transformation than non-Boom villages. The authors find electrification led to six percentage-point SR and 50 percentage-point LR increase in non-agricultural employment in villages exposed to the commodity boom, but do not observe evidence the policy benefited non-Boom villages—attributing these findings to complementary growth in industrial-scale non-agricultural processing which benefited from combinations of increased global demand for guar gum and local electrification. In contrast, I find Boom villages observed significantly lower (net zero) benefits of structural transformation from rural roads.

Secondly, my results provide insight to Usmani (2020)'s findings that infrastructure investment's benefits are tied to local economic conditions. Here, my results imply complementarities between varying infrastructure programs (rural roads and rural electrification) and local economic conditions may differ substantially, generating diverging outcomes for structural transformation– even in the same localities. This implies precise short-run benefits of rural road investments may be counter-intuitively reduced by external demand shocks if these shocks reduce the incentive for households to switch away from agriculture.

Finally, my research contributes to literature in treatment effects heterogeneity. If policy benefits differ substantially observables, then ATE estimation may overestimate welfare gains to specific sub-populations. (Meager, 2015). Limited research has investigated heterogeneous treatment impacts of the PMGSY program (an exception is Lawson-Johns, 2021, who studies heterogeneity of PMGSY effects by village flood exposure). Here, I study two alternative sources of treatment heterogeneity- external market conditions and in-village electrification.

The rest of this paper is organized as follows: (2) Research Question (3) Background Information (4) Theoretical Framework (5) Data (6) Methodology and Empirical Strategy (7) Results (8) Mechanisms (9) Conclusion.

2 Research Question

I investigate two key questions:

1. *How are causal impacts of rural roads on structural transformation affected (augmented or reduced) by the presence of i) complementary external market conditions to agriculture or ii) in-village electrification*
2. *What mechanisms drive this process?*

3 Background Information

3.1 The Pradhan Mantri Gram Sadak Yojana (PMGSY) rural road construction program

India has historically been constrained by high costs of road construction, large geographic size and population, and rapid degradation due to unpredictable weather conditions which have impeded efforts to connect every village to road networks (Adukia, 2020).

In response to such challenges, the Indian Government launched the PMGSY rural road program in 2000. The program aimed to provide AWRs to every unconnected village in India. The program focused on providing new feeder roads to localities without existing paved road, whilst additionally upgrading pre-existing roads (A-N 2020).

Under PMGSY, national guidelines determined AWR prioritization. Guidelines initially targeted construction of paved roads to villages with populations above 1000 by 2003, population above 500 by 2007, population above 250 by 2009.[3] Although rules were dictated by the federal government, responsibility for administration of village-level road allocation was delegated to states. Population guidelines did not not entirely determine treatment status. States could utilise other rules in determining allocation including measures of local economic importance like presence of weekly markets (A-N 2020). Given program rules, Adukia (2020) highlights early-treated villages generally had larger populations but were not systematically different from

[3] The unit of targeting was technically the habitation- an administrative unit smaller than a village. As highlighted by Burlig and Preonas (2020), a village typically consists of between one to three habitations. In this paper, I focus on villages as the unit of analysis as (i) limited economic information at the habitation level exists (ii) many villages consist of a single habitation and (iii) many habitations were ultimately pooled to the village level for the purposes of the program (Adukia, 2020).

later-receiving states in other characteristics.

By 2015, over 185000 villages had paved roads upgraded or built by PMGSY at a cost of over $37 billion (A-N 2020) and median road length of 4.4 km (Adukia, 2020).

3.2 Fracking, guar, and the US Shale Gas boom

Hydraulic fracturing (fracking) is an industrial process which involves injecting a combination of sand, water, and chemicals at high pressure to generate fractures in sedimentary rock formations. Technological advances have led to a rapid increase in utilisation of fracking in fossil fuel extraction. The share of natural gas and oil production in "fracked" vs conventional wells rose from 7-67% and 2-50% from 2000-2014 in the US (Usmani (2020)).

A key input to fracking is guar gum, a thickening agent which reduces fluid requirements for a given job. (Fetter et al, 2018). (Elsner and Hoelzer, 2016), highlights up to 50% of fracking production is dependent on guar gum. Guar Gum is derived from guar, a drought-resistant crop primarily grown in the Northwestern Indian desert states of Rajasthan, Gujarat, and Haryana (Kuravadi 2013). The National Rainfed Area Authority (2014) states India accounts for 80% of global guar production; India also produces the majority of guar gum relevant for international trade. Usmani (2020) finds production of guar is decentralised and grown by marginal farmers throughout the North-western region.

From 2009-2014, prices of guar gum, a key commodity in the hydraulic fracturing process, increased by almost 1, 000 % due to surge in demand from the US shale exploration industry. This commodity boom induced a parallel augmentation in both agricultural production of guar and Indian exports of guar-gum, which rose from $241 million to $3.9 billion. (Usmani, 2020), highlights India's share of global guar-gum exports increased by 3.5x over the same period.

3.3 Rural Electrification and the Rajiv Gandhi Grameen Vidyutikaran (RGGVY) national electrification program

The RGGVY program was introduced in 2005 with dual mandate to (i) connect 100, 000 unelectrified rural villages to the grid (ii) more intensively electrify infrastructure 300, 000 "under-electrified" villages, across 27 Indian states. The program included both installation of electricity infrastructure and connecting unelectrified households.

The program was funded under in two waves (the first funded in 2005-8, the second in 2008-11). The program also utilised population thresholds; only villages with population over 300 (2001 Census) were prioritized for electrification. Burlig and Preonas (2016) exploited both methodologies to identify causal impacts of RGGVY- finding the program increased village electrification rates utilising both methods.

These methods are unsuitable for my design as a) key outcomes are from 2012, when most districts had received electrification, and b) all villages in my RD sample have populations above the 300 cut-off. I instead leverage variation in nationwide implementation intensity of RGGVY as an indirect measure of electrification exposure (conditional on district fixed effects to eliminate baseline between-district variation .

4 Theoretical Framework

4.1 A (Modified) Model of Labor Allocation across sectors with Heterogeneous Treatment Effects

My theoretical framework builds upon the A-N model of Labor Allocation Across Sectors (A-N 2020) with additions of a) constraints and complementarities and b) worker heterogeneity. These additions generate heterogeneous between-village benefits of road construction based on a) external market conditions or b) in-village electrification, and within-village variations in labor allocation based on worker characteristics.

I consider a three-sector economy consisting of Village and External Market. The Village Economy is subdivided into Village Agriculture, and Village Non-Agriculture. The External Economy is entirely Non-Agriculture. The village is geographically isolated from the External Economy, and in absence of road, the economy is modelled as one of Total Autarky (High Transportation Costs, precluding exchange of goods and services with External Market). With rural roads, I assume complete integration- with prices and wage equalisation for goods and input markets.

Labour is the only input to production. I denote Value of Production in Village Agriculture and Village Non-Agriculture as Y_{av}, Y_{nv}.I define the following production functions:

$$Y_{av} = A_{ap} ln(L_{av})$$

$$Y_{nv} = A_{np} ln(L_{nv})$$

The total labor endowment of the village economy is 1. Allocations L_{av}, L_{nv}, L_{ne}refer to labor input allocations in the respective sector. Thus, $(L_{av} + L_{nv} + L_{ne} = 1)$.

A_{ap}, A_{np} are parameters for the Agriculture and Non-Agriculture Sector, containing both productivity and the prices of outputs and inputs. p is an indicator variable for whether a village is isolated or connected via a paved road.

In the model, following A-N, I assume paved roads increase revenue productivity of labor in the village agricultural and non-agricultural sector [4].

Hence, we have potential values of p ($p = 0, 1$)- taking the value 1 if a village has paved road and 0 otherwise.

Without roads, villages are isolated from the external labor market ($L\_ne = 0$). Assuming friction-less internal markets and homogeneous labour, workers should move between village sectors until the marginal productivity of labour(MPL) and wage is equalised between village agriculture and non-agriculture.

Formally , the MPL in sector i, $MPL_i = \frac{A_{ip}}{L_i}$. In-village, in autarky:

$$MPL_{ap} = MPL_{np}$$

$$\frac{A_{ap}}{L_{av}} = \frac{A_{np}}{L_{nv}}$$
$$\frac{A_{ap}}{L_{av}} = \frac{A_{np}}{(1 - L_{av})}$$

$$A_{ap}(1 - L_{av}) = A_{np}L_{av}$$

$$A_{ap} = L_{av}(A_{np} + A_{ap})$$

Recalling $p = 0$ in both sectors:

$$L_{av} = \frac{A_{a0}}{(A_{n0} + A_{a0})}$$

Taking the productivity parameter A_{a0} (agricultural productivity) as the numeraire, and A_{N0} as the relative (revenue) productivity of the non-agricultural versus

[4]For example, by expanding access to critical inputs like fertiliser (agriculture) , manufacturing equipment (non-agriculture), lowering the cost of inputs (net of transportation costs), and increasing the price of outputs in the village economy (net of transportation costs)

agricultural sector when p = 0:

$$L_{av} = \frac{1}{(A_{N0} + 1)} = L_{av}^{0} \tag{1}$$

The village is assumed a price-taker to the external market wage, ω_e. Roads enable villagers to access the external labour market with elimination of transportation costs-meaning villagers should migrate between sectors until MPL in the village agriculture and non-agriculture sector is equal to ω_e

In-village, with roads:

$$\frac{A_{ap}}{L_{av}} = \frac{A_{a1}}{L_{av}} \tag{2}$$

$$\omega_e = \frac{A_{np}}{L_{nv}} = \frac{A_{n1}}{L_{nv}} \tag{3}$$

From (3):

$$L_{av} = \frac{A_{a1}}{\omega_e} = L_{av}^{1} \tag{4}$$

In autarky, the share of village agricultural workers in is $L_{av}^{0} = \frac{1}{(1+A_{N0})}$, driven by the relative productivity of village non-agriculture vs. agriculture in the absence of roads. With roads, the share of village agricultural workers is $L_{av}^{1} = \frac{A_{a1}}{\omega_e}$, driven by the productivity of village agriculture with paved road access and the external wage rate.

The basic A-N model predicts structural transformation (the movement of workers from agriculture to non-agriculture) should occur if:

$$L_{av}^{1} < L_{av}^{0} \tag{5}$$

$$\frac{A_{a1}}{\omega_e} < \frac{1}{(1 + A_{N0})} \tag{6}$$

$$A_{a1} < \frac{\omega_e}{(1 + A_{N0})} \tag{7}$$

I now augment the basic A-N model with two key additions:

4.1.1 Complementarities and Constraints

Firstly, I add village heterogeneity in shock terms, defined as complementarities and constraints, which may influence the opportunity cost of agriculture vs. non-agriculture. In simpler terms, these factors either increase or reduce the effectiveness of rural roads in promoting structural transformation. For example, rural roads may only be effective in supporting labor reallocation in the presence of high village human capital; in this case, high human capital levels would be a complement to non-agricultural growth.

I define Z_{cti}, which defines a vector of complementary shocks experienced by a village v in community c at time c in sector i (agriculture, a or non-agriculture, n) which influences the impact of rural roads.

4.1.2 Worker Heterogeneity

Secondly, consider there is worker heterogeneity. A simple addition to the A-N model is to allow the external wage ω_e, and A_{a1} relative productivity gain in agriculture to differ worker skill and specialisation. I add individual subscripts i to generate ω_{ei} and A_{a1i}.

4.2 Augmented A-N Model and Predictions

I rewrite the basic A-N model as follows:

$$A_{a1i} + Z_{cta} < \frac{\omega_{ei}}{(1 + A_{N0})} + Z_{ctn}$$

Rural roads should generate structural transformation for a worker if the worker-specific external wage rate, plus the village-specific complementary shock to non-agriculture Z_{ctn} at time t is is high relative to:

(a) Potential (worker-specific) agricultural productivity gains in presence of roads (A_{a1i}),

(b) the village-specific complementary shock to agriculture Z_{ctn} at time t

(c) relative non-agricultural productivity in absence of roads (which defines baseline labor share of workers in village non-agriculture)

The trade-off between vectors Z_{cta} and Z_{ctn} captures village-specific factors affecting opportunity costs of staying vs. leaving agriculture.

4.2.1 Prediction 1: Roads and External Market Conditions

The Boom could have multiple interactive impacts on rural roads' structural transformation benefits. The boom induced a large surge in prices in the agricultural sector (due to a boom in cultivation of guar in guar-intensive districts)- leading to positive income effects for farmers and increasing Z_{cta}. Simultaneously, Usmani (2020) highlights, the Boom increased prices for non-agricultural firms associated with the trade and manufacturing of guar-gum- increasing parameters Z_{ctn}. This implies effects on structural transformation are theoretically ambiguous.

However, provided the Boom induced a larger magnitude of agricultural vs. non-agricultural sector growth, the model predicts the Boom should lead to a lower rate of structural transformation.

4.2.2 Prediction 2: Roads and Village Electrification

Electrification could increase productivity of agricultural or non-agricultural firms in combination with road construction, increasing Z_{cta} and Z_{ctn} parameters. Assuming Electrification is more beneficial to non-agriculture than agricultural production when paired with road construction, however, the model predicts electrification should lead to higher rates of structural transformation.

4.2.3 Prediction 3: Within-Village Labour Allocation Decisions

The model predicts workers with lowest agricultural productivity gains (opportunity cost of leaving agriculture, A_{a1i}) and highest external wages (ω_{ei}) should see the largest structural transformation gains from rural road by being most likely to change sectors.

5 Data

I combine four main sources of data. Firstly, I use high-resolution information on village labour force composition from the SHRUG. Secondly, I use multiple waves of the Economic Census (1990, 1998, 2005, and 2013) to obtain detailed information on non-agricultural firm creation, size, sectoral composition, and total employment. Thirdly, following Usmani (2020), I use reports from the Government of India and United States to identify India's main guar-growing districts. Finally, I obtain data on village electrification through both direct measures of reported village power availability and additional information on village exposure to the RGGVY national electrification program (2005-2012).

5.1 Data Description and Key Variables

5.1.1 Socioeconomic High Resolution Urban Geographic Platform for India (SHRUG)

My paper's core dataset comes from the SHRUG (A-N) , an open-access repository covering India's 500,000 villages with common geographic identifiers. The dataset includes village demographic data from 1991, 2001, 2011 Population Censuses, data from the 2002 BPL (Below Poverty Line) Census, and aggregated household data from the 2011-12 Socioeconomic and Caste Census (SECC)- which provides information on structural transformation

The SHRUG provides village-level program data on PMGSY AWR construction. I employ spatial and temporal variation in PMGSY treatment in analysis. I define a village as having received a rural road if an AWR was constructed in the year prior to the relevant survey year .

5.1.2 Firm Data- 3rd-6th Economic Census (EC) of India

Firm data comes from the 3rd-6th waves of the Indian EC (1990, 1998, 2005, and 2013). The EC contains information on every non-farm establishment in the country. This includes formal and informal private-sector firms and public sector establishments. I collapse raw EC establishment-level information to the SHRUG-village level to create a village-EC panel utilised in difference-in-difference analyses. Firm data includes counts of firms, employment, firm composition (sectoral, size, and social distribution of firm ownership).

Various EC rounds used different NIC industry classification systems. The 1990 and 1998 census use the 1987 NIC, the 2005 EC uses the 2004 NIC, and the 2013 EC uses the 2008 NIC. I use government concordance tables to create consistent sectoral groupings of firm entry at the village-level.

Data on agricultural firms is incomplete in the EC, excluding cultivation and inconsistently listed across EC rounds. I exclude agricultural firms across all waves for DD/triple-difference analyses. Finally, different waves of the Economic Census utilise different inclusion criteria with the 2013 EC excluding National Administration and Defence, and Compulsory Social Security enterprise. I exclude these firms from earlier rounds of the EC to obtain consistent data on village employment.

5.1.3 Fracking Data

I obtain data on districts with greater exposure to the US Fracking Boom ('boom districts') from Usmani (2020)- who denotes such districts as those recognised as intensive in guar-cultivation at baseline according to at leat two official Government reports.[5]

5.1.4 Rural Electrification Data

I utilise multiple data sources of village electrification. My primary (indirect) metric exploits nationwide rollout of the RGGVY national electrification program- considering PMGSY rural road recipient villages in high ('electrified') and low-intensity ('unelectrified') RGGVY districts. Electrification for RGGVY also expanded at the village, rather than district level- but data from the DDUGJY monitoring website is inconsistent and missing in over 40% of cases. For this reason, I obtained web-scraped data on district-level implementation of the RGGVY program from state-level reports. [6]

For my analysis, I focus on 10th plan RGGVY districts. I define high-intensity 10th plan ('electrified') RGGVY districts as those where 60% of the villages in a district benefited from electrification in either the intensive or extensive margin under RGGVY, following Burlig and Preonas (2020); I compare these districts to

[5] 2001 Census Districts defined as Guar-Intensive (Boom Districts) by Usmani (2020) are: Alwar, Barmer, Bhilwara, Bikaner, Churu, Dausa, Hanumangarh, Jaipur, Jaisalmer, Jhunjhunu, Jodhpur, Nagaur, Pali, Sikar, Sri Ganganagar in Rajasthan; Bhiwani, Gurgaon, Mahendragarh, Rewari in Haryana; Ahmedabad, Banasantha, Kutch, Mehsana, Sabarkantha, Vadodara in Gujarat

[6] These are ntitled 'Report C-Physical Financial Progress of RGGVY Projects Under Implementation These are available on the website of the Deendayal Uphayaya Gram Jyoti Yojana (DDUGY) program- into which RGGVY was ultimately subsumed, at http://www.ddugjy.gov.in/.

a pooled sample of low-RGGVY implementation intensity 10th plan districts, and non-RGGVY ('unelectrified') districts in my RD design. Conditional on inclusion of controls for baseline electrification status in 2001, and district-cut-off fixed-effects, this measure highlights varying incidental village exposure to increases in electrification concurrent with PMGSY.

I complement this strategy with direct measures of village power availability– recorded in 1991, 2001, and 2011 Population Census rounds. The 1991 and 2001 rounds contain only binary metrics of village power whereas the 2011 round contains detailed figures on mean hours of village electricity. Hourly data is missing for previous population rounds and in the 2011 Population Census, for approximately 20% of villages in my RD sample. For RD analysis I define high power villages as those with 8 hours of mean summer hours of commercial power availability, which represents the 75th Percentile in my RD sample

This direct measure is presumably less exogeneous to dependent outcomes , but overall a stronger proxy for actual village electrification.

5.2 Summary Statistics

I investigate the interaction between rural roads and village exposure to a) the fracking boom in North-western India and b) in-village electrification. Balance Tables 1 and 2 show baseline balance tables of my RD sample (N= 11, 431) for Boom and Non-Boom villages and Electrified vs Non-Electrified Villages. The top section includes variables from the 2001 Population Census, middle section variables from the 2002 BPL Census, and bottom section variables from the 1998 EC.

The RD sample shows Non-Boom villages report significantly higher access to medical services, proximity to Census towns, literate share, and Non-Boom villages show significantly higher availability of bus services, share of irrigated land, share of households earning $\mathring{4}$ USD (significantly higher in Boom villages). Similar differences are observed between Electrified vs non-electrified district-villages. Variation in covariates justifies inclusions of baseline controls across specifications.

Balance Table 1

Balance Table (Boom and Non-Boom District) - Main Sample (N= 11, 431)

	(Mean Non-Boom)	(Mean Boom)	(Difference in Means)	(N)
Number of Primary Schools in village (2001)	1.010	1.011	0.002	11,431
	(0.344)	(0.325)	(0.010)	
Number Of Maternity And Child Welfare Centres (2001)	0.015	0.002	-0.013***	11,431
	(0.130)	(0.040)	(0.004)	
Medical Centre Available (Y/N) (2001)	0.169	0.113	-0.056***	11,431
	(0.375)	(0.317)	(0.011)	
Electricity Available (Y/N) (2001)	0.428	0.401	-0.027*	11,431
	(0.495)	(0.490)	(0.015)	
Bus Services Available (Y/N)(2001)	0.134	0.280	0.146***	11,431
	(0.341)	(0.449)	(0.011)	
Banking Facilities Available (Y/N) (2001)	0.007	0.006	-0.001	11,431
	(0.081)	(0.075)	(0.002)	
Distance to Closest Census Town (km) (2001)	26.046	32.805	6.760***	11,431
	(21.866)	(25.202)	(0.667)	
Land irrigated share (2001)	0.273	0.351	0.078***	11,431
	(0.290)	(0.327)	(0.009)	
Ln Land Area (Hectares) (2001)	5.677	5.962	0.285***	11,431
	(0.734)	(0.798)	(0.022)	
Literate share (2001)	0.457	0.445	-0.012***	11,431
	(0.156)	(0.137)	(0.005)	
Scheduled Caste Population Share (2001)	0.472	0.288	-0.183***	11,431
	(0.326)	(0.297)	(0.010)	
Share of Households With Land (2002)	0.718	0.882	0.164***	11,431
	(0.234)	(0.153)	(0.007)	
Share of Households with Subsistence Agriculture as Primary Income Source (2002)	0.428	0.535	0.107***	11,431
	(0.267)	(0.241)	(0.008)	
Share of Households Earning (> 4 USD per month) (2002)	0.744	0.857	0.112***	11,431
	(0.281)	(0.222)	(0.008)	
Share of Establishments with Access to Finance (1998)	0.059	0.040	-0.019***	11,431
	(0.153)	(0.135)	(0.005)	
Share of Establishments Formally Registered (1998)	0.005	0.002	-0.002*	11,431
	(0.043)	(0.034)	(0.001)	
Share of Establishments using Fuel/Electricity as main energy source (1998)	0.131	0.146	0.016***	11,431
	(0.198)	(0.231)	(0.006)	
Share of Population Employed in Non-Agriculture Village Establishments	0.034	0.018	-0.017***	11, 431
	(0.067)	(0.036)	(0.002)	
	(0.070)	(0.060)	(0.001)	
Observations	10,183	1,248	11,431	

Balance Table 2

Balance Table (Electrified and Non-Electrified District) - Main Sample (N= 11, 431)

	(Mean Non-Electrified)	(Mean Electrified)	(Difference in Means)	(N)
Number of Primary Schools in village (2001)	1.011	1.007	-0.004	11,431
	(0.362)	(0.257)	(0.008)	
Number Of Maternity And Child Welfare Centres (2001)	0.016	0.006	-0.010***	11,431
	(0.131)	(0.092)	(0.003)	
Medical Centre Available (Y/N) (2001)	0.171	0.134	-0.038***	11,431
	(0.377)	(0.340)	(0.008)	
Electricity Available (Y/N) (2001)	0.404	0.500	0.096***	11,431
	(0.491)	(0.500)	(0.011)	
Bus Services Available (Y/N)(2001)	0.151	0.146	-0.005	11,431
	(0.358)	(0.353)	(0.008)	
Banking Facilities Available (Y/N) (2001)	0.007	0.006	-0.001	11,431
	(0.082)	(0.078)	(0.002)	
Distance to Closest Census Town (km) (2001)	27.410	24.519	-2.891***	11,431
	(23.312)	(18.297)	(0.507)	
Land irrigated share (2001)	0.264	0.341	0.077***	11,431
	(0.293)	(0.299)	(0.007)	
Ln Land Area (Hectares) (2001)	5.690	5.775	0.085***	11,431
	(0.772)	(0.641)	(0.017)	
Literate share (2001)	0.456	0.455	-0.001	11,431
	(0.156)	(0.149)	(0.004)	
Scheduled Caste Population Share (2001)	0.460	0.421	-0.039***	11,431
	(0.330)	(0.318)	(0.007)	
Share of Households With Land (2002)	0.729	0.760	0.031***	11,431
	(0.234)	(0.225)	(0.005)	
Share of Households with Subsistence Agriculture as Primary Income Source (2002)	0.434	0.460	0.026***	11,431
	(0.267)	(0.265)	(0.006)	
Share of Households Earning (> 4 USD per month) (2002)	0.750	0.780	0.030***	11,431
	(0.279)	(0.272)	(0.006)	
Share of Establishments with Access to Finance (1998)	0.055	0.063	0.008**	11,431
	(0.148)	(0.161)	(0.003)	
Share of Establishments Formally Registered (1998)	0.005	0.003	-0.001	11,431
	(0.043)	(0.039)	(0.001)	
Share of Establishments using Fuel/Electricity as main energy source (1998)	0.126	0.155	0.029***	11,431
	(0.197)	(0.217)	(0.005)	
Share of Population Employed in Non-Agriculture Village Establishments (1998	0.029	0.026	-0.002	11,431
	(0.065)	(0.047)	(0.001)	
Observations	8,955	2,476	11,431	

6 Methodology and Empirical Strategy

6.1 Overview

This paper investigates **heterogeneous treatment impacts** of rural roads on structural transformation in India based on a) external market conditions and b) in-village electrification. A pre and post comparison of outcomes in PMGSY-recipient villages in boom or electrified districts would likely yield a biased estimates of causal impacts of AWRs in presence of complementary external market conditions or in-village electrification- as this regression lacks a suitable "non electrified" or "non-boom" counterfactual and ignores confounding variation over time. Cross-sectional comparison of PMGSY-recipient villages in electrified vs. unelectrified or boom vs. non-boom districts would also generate unreliable estimates. Electrified districts may also be more isolated from technology at baseline, and most boom districts are located in Rajasthan, one of India's poorest states.

I thus utilise a modified regression discontinuity design and triple-difference approach.

6.2 Regression Discontinuity Design and the Heterogeneous Local Average Treatment Effect (HLATE)

My RD strategy leverages discontinuities in population-based eligibility for PMGSY combined with interactions with heterogeneous i) exposure to complementary or adverse external market conditions and ii) in-village electrification– to estimate the heterogeneous Local Average Treatment Effect (HLATE- (Becker 2013)

Under PMGSY, villages were eligible for AWRs if their population exceeded 250, 500 or 1000 (with thresholds relaxed nationwide over time). Compliance to guidelines was imperfect, necessitating a fuzzy RD approach. A-N (2020) previously worked closely with the National Government to identify states complying with guidelines. Few villages in complier states between 250-500 population received roads by 2012, the year of key outcomes; I limit my sample to villages in complier states with populations close to 500 and 1000).[7] Further, following (A-N), I pool villages utilising the same optimal bandwidth on either side of the 500 or 1000 population cut-off to maximise power.

[7] Complier states identified by A-N are: Gujarat, Rajasthan, Orissa, Chhattisgarh, Madhya Pradesh, Maharashtra

The standard (fuzzy) RD approach typically only allows heterogeneity of treatment effects based on variables which also influence treatment status - generating an interpretation of a Local Average Treatment Effect (LATE) for a complier sub-population- villages who would have received PMGSY treatment due to being above the population threshold but not otherwise.

Following Becker (2013), Lawson-Johns (2021), I employ a modified RD approach to estimate a heterogeneous local average treatment effect (HLATE)- a **LATE conditional on observable covariates that may not influence treatment status** (here, external market conditions and in-village electrification).

I define R_i as a treatment dummy indicating ($R_i = 1$) whether a village received road treatment (0 otherwise), P_i as a forcing variable (2001 Population, with cut-off P_0). Y_i is a dependent outcome of interest which is influenced by R_i, and a vector, Z_i of additional interaction terms $Z_1...Z_m$.

Notably, these interaction variables do not influence treatment status but may augment or reduce the impact of AWRs (in my example, external market conditions or in-village electrification).

I aim to identify the Heterogeneous Local Average Treatment Effect (HLATE), defined in Becker (2013) as:

$$HLATE(P_i = p_0, z_i) = HLATE(p_0, z_i) = E[Y_1v|p_0, z_i] - E[Y_{0v}|p_0, z_i]$$

Y1v and Y0v refer to potential outcomes of village v with and without treatment.

Under Fuzzy RD, treatment assignment is non-random, but probability of treatment status varies discontinuously with the forcing variable, Pi:

$$(\mathrm{P}(\mathrm{R}_i = 1|\ \mathrm{P}_i) = \begin{cases} g_1(P_i), & \text{if } (P_i \geq P_0) \\ g_0(P_i), & \text{if } (P_i < P_0) \end{cases}$$

Where $0 < g_1(P_i) – g_0(P_i) \leq 1$

(Becker 2013) highlights the HLATE may vary widely for a single LATE based on observable covariates. Past investigations into AWR may underestimate or overestimate benefits for specific subpopulations due to failure to consider heterogeneous village treatment effects a) external market conditions and b) in-village electrification.

I estimate a fuzzy RD specification to estimate the HLATE of PMGSY AWRs and its interaction with i) external market conditions or ii) in-village electrification. RD estimation of the HLATE relies upon five assumptions.

Assumptions

1. Continuity of potential outcomes across the RD Threshold RD analysis assumes continuity across the RD-threshold for all observable and unobservable characteristics which might correlate with my outcome variable. This assumption is fundamentally untestable, however in support of it, Figure 1 plots a parallel RD plot of illustrating the continuity assumption is satisfied as pre-PMGSY village-level covariates appear to be smooth across normalised population cut-off.

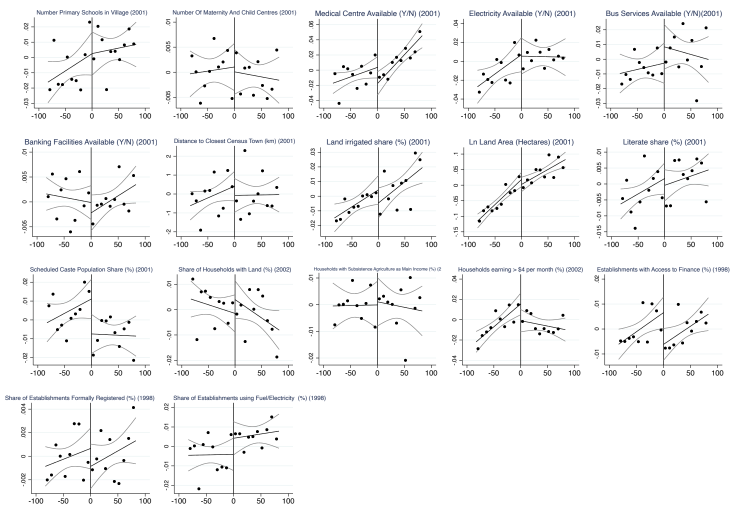

Figure 1: Parallel RD Plot of Baseline Covariates

2. No Perfect Manipulation of Running Variable at Threshold As in Lee and Lemieux (2013), even if manipulation occurs, provided villages are unable to precisely manipulate the assignment variable, a direct consequence of this fact is the threshold can be considered randomized as though from a RCT. This assumption almost certainly holds, as, I use population estimates from the 2001 census, independent to PMGSY. Figure 2 plots a histogram of the distribution of 2001 village population in my sample- indicating no evidence of bunching at the 500 and 1000 cut-off points. I also employ the RD manipulation procedure developed by Cattaneo (2018) to test for bunching at the cut-off. I obtain a robust variance-corrected test statistic of -0.9379 (P-Value 0.3483) against null hypothesis of no manipulation confirming this assumption is valid. The Cattaneo Manipulation testing plot of normalised village population is shown in Figure 3.

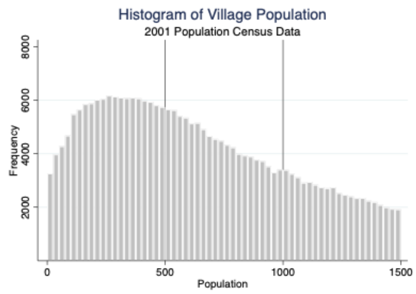

Figure 2: Histogram of 2001 Population

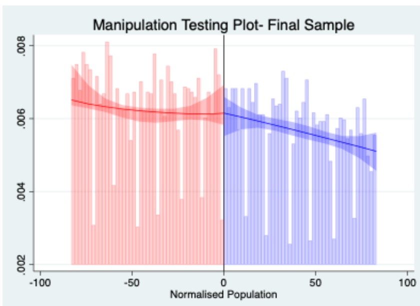

Figure 3: Cattaneo Manipulation Testing Plot

3. Discontinuous Jump in Treatment Probability around Threshold

Figure 4 shows the First Stage share of villages which received a new road under PMGSY by 2011 in my sample , across a range of population bandwidths including the I-K Optimal Bandwidth. The Probability of a New Road by 2011 increases by 21% across the normalised population cut-off, indicating the assumption is satisfied (Figure 5 shows this graphically).

	(1) 84- IK Optimal	(2) 60	(3) 70	(4) 80	(5) 90	(6) 100
t	.205***	.212***	.209***	.206***	.203***	.201***
	(.029)	(.031)	(.03)	(.029)	(.029)	(.028)
Observations	11431	8291	9661	11036	11431	11431
R-squared	.308	.315	.311	.309	.306	.305

Figure 4: First Stage Regressions

4. Random Assignment of Interaction Variables, Z_i , conditional on X_i

Here, this assumption implies, conditional on the population running variable, villages in Boom or electrified RGGVY districts, or villages with high power availability do not differ in unobservable dimensions from the counterfactual in ways which are also correlated with the outcome variable.

It is unlikely this assumption is directly satisfied in absolute terms- e.g., Boom villages may be inherently more productive or have stronger institutions to facili-

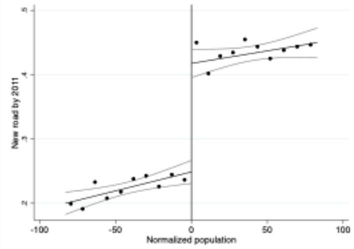

Figure 5: Probability of New Road by 2011

tate structural transformation. I thus include controls for district-cut-off FEs in my RD regression. Conditional on district-cut-off FEs , the treatment parameter identifies the additional benefit of village co-location in Boom and electrified districts. In specifications using direct village-power availability, I supplement district-cut-off FE and baseline electrification controls with additional controls for village endline power availability to capture remaining potential endogeneity of unobservable village-characteristic s influencing occupational choice.

5. Continuity of Interaction Variables, Z_i, at Threshold

To detect genuine variation in interaction variables, the interaction term, Z_i, must not be discontinuous at the threshold. I check this assumption by plotting graphs of Mean Hours of Commercial Power Availability in 2011 and probability of village co-location in Boom or Electrified district against normalised village population in 2001. I do not observe discontinuity around the threshold showing this condition is satisfied.

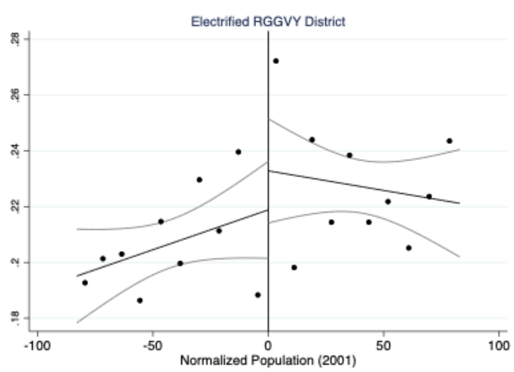

Figure 6: RGGVY Electrified District Bin Scatterplot

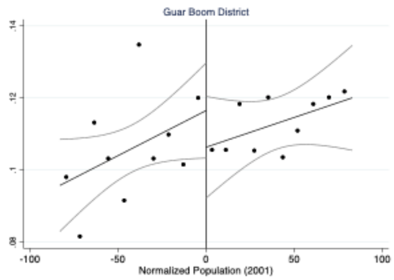

Figure 7: BOOM District Bin Scatterplot

My RD strategy follows recommendations of Gelman and Imbens (2014) and Lee and Lemieux (2013). My specification uses a non-parametric local linear regression specification of the running variable (village population) on either side of the threshold cut-off. I use a triangular kernel which places greater weight to observations closer to the cut-off and restrict the sample population to villages within a narrow bandwidth of the population threshold. I follow Imbens and Kalyanaraman (2012) , and utilise an optimal bandwidth of 84.

Formally, I employ a two-stage-least-squares specification approach (IV-2SLS) as follows:

(i) PMGSY roads and Complementary Market Conditions (Boom or Non-Boom Districts)

(First stage)

$$\begin{aligned} R0AD_{v\ (t-1)} &= \lambda_0{}^* + \lambda_1{}^* T_v + \lambda_2{}^* (p - p_0) \\ &+ \lambda_3{}^* ((p - p_0) \times T_v) + \\ &+ \lambda_4{}^* \ T_v \times BOOM_{v_t} + X_{v_t} + D_d + \varepsilon_{v_t} \end{aligned} \tag{1}$$

$$\begin{aligned} ROAD_{v\ (t-1)} \times BOOM_{vt} &= \gamma_0{}^* + \gamma_1{}^* T_v + \gamma_2{}^* (p - p_0) + \\ &+ \gamma_3{}^* \ ((p - p_0) \times T_v) \\ &+ \gamma_4{}^* \ T_v \times BOOM_{v_t} + X_{v_t} + D_d + \varepsilon_{v_t} \end{aligned} \tag{2}$$

(Main stage)

$$\begin{aligned} Y_{v_t} &= \lambda_0 + \lambda_1 \widehat{R0AD}_{v\ (t-1)} + \lambda_2 \ (p - p_0) + \lambda_3 \ ((p - p_0) \times T_v) \\ &+ \lambda_4 \ R0AD_{v\ \widehat{(t-1) \times}} BOOM_{vt} + X_{vt} + D_{cd} + \varepsilon_{v_t} \end{aligned} \tag{3}$$

(ii) PMGSY roads and Electrification (Indirect Measure- Electrified vs non-electrified districts)

(First stage)

$$\begin{aligned} R0AD_{v\ (t-1)} &= \mu_0{}^* + \mu_1{}^* T_v + \mu_2{}^* (p - p_0) \\ &+ \mu_3{}^* ((p - p_0) \times T_v) + \\ &+ \mu_4{}^* \ T_v \times ELECTRIFIED_{vt} + X_{vt} + D_d + \varepsilon_{vt} \end{aligned} \tag{4}$$

$$\begin{aligned} ROAD_{v\ (t-1)} \times ELECTRIFIED_{vt} &= \pi_0{}^* + \pi_1{}^* T_v + \pi_2{}^* (p - p_0) + \\ &+ \pi_3{}^* \ ((p - p_0) \times T_v) \\ &+ \pi_4{}^* \ T_v \times ELECTRIFIED_{vt} + X_{vt} + D_d + \varepsilon_{vt} \end{aligned} \tag{5}$$

(Main stage)

$$\begin{aligned} Y_{v_t} &= \mu_0 + \mu_1 \widehat{R0AD}_{v\ (t-1)} + \mu_2\ (p-p_0) + \ \mu_3 \quad ((p-p_0)\times T_v) \\ &+ \ \mu_4 \ \ R0AD_{v\ (t-1)} \times \widehat{ELECTRIFIED}_{vt} \ + X_{vt} \ + D_{cd} + \ \varepsilon_{vt} \end{aligned} \tag{6}$$

(iii) PMGSY roads and Electrification (Direct Measure- High vs Low Village Power Availability)

(First stage)

$$\begin{aligned} R0AD_{v\ (t-1)} &= \tau_0\ ^* + \tau_1\ ^*T_v + \tau_2\ ^*(p-p_0\ ) \\ &+ \tau_3\ ^*((p-p_0)\ \times T_v) + \\ &+ \tau_4\ ^* \ \ T_v \times HIGHPOWER_{vt} + \tau_5\ ^*HIGHPOWER_{vt} + X_{vt} + D_d + \varepsilon_{vt} \end{aligned} \tag{7}$$

$$\begin{aligned} ROAD_{v\ (t-1)} \times HIGHPOWER_{vt} \ &= \kappa_0\ ^* + \kappa_1\ ^*T_v + \kappa_2\ ^*(p-p_0\ ) + \\ &+ \kappa_3\ ^* \ \ ((p-p_0)\ \ \times T_v) \\ &+ \ \kappa_4\ ^* \ \ T_v \times HIGHPOWER_{vt} + X_{vt} \\ &+ \kappa_5\ ^*HIGHPOWER_{vt} + D_d + \varepsilon_{vt} \end{aligned} \tag{8}$$

(Main stage)

$$\begin{aligned} Y_{v_t} &= \tau_0 + \tau_1 \widehat{R0AD}_{v\ (t-1)} + \tau_2\ (p-p_0) + \ \tau_3 \quad ((p-p_0)\times T_v) \\ &+ \ \tau_4 \ \ R0AD_{v\ (t-1)} \times \widehat{HIGHPOWER}_{vt} \\ &+ \ \tau_5 \ \ HIGHPOWER_{vt} + D_d + \varepsilon_{vt} \end{aligned} \tag{9}$$

T_v is a binary variable equal to 1 if a village has a population greater than the relevant population-based threshold according to the 2001 Census. $ROAD_(v(t-1))$ is a binary variable equal to 1 if a village received electricity at the year prior to the survey round of interest. $(ROAD)_v,(t-1))$ refers to the part of the take up of PMGSY AWRs in a given village which is explained by exogenous variation in the village population eligibility cut-off point.

$Y_{vd}$is a village-level economic outcome measured at time t for village v located in district d. *BOOM* in equations (i) refers to whether a village is co-located in a guar-growing Boom District at time t, *ELECTRIFIED* in equations (ii) refers to whether a village is located in an electrified RGGVY district, and *HIGHPOWER* in equation (iii) refers to whether a village has high power availability, according to the 2011 census. My RD model includes controls for district cut-off fixed effects (D_{cd}) controlling for all time-invariant district characteristics around a given population bandwidth, and a vector of baseline village-level characteristics[8] . In model (iii), I add additional controls for HIGHPOWER to absorb potential endogeneity between actual village power availability and village unobservables also influencing Y_{vd}. To overcome omitted variable bias, due to likely correlation between village-level power availability and baseline share of village population employed in non-agriculture, all my electrification specifications include controls for village non-agricultural employment share from the 1998 Economic Census- in addition to varying levels of (district cut-off, district, and state-level fixed effects), baseline binary controls for village electrification in 2001, and both interaction terms for my treatment variable of interest *ROADxELECTRIFIED* or *ROADxHIGHPOWER*) and level controls (*HIGHPOWER*) in specifications using village power availability as my electrification measure.

Provided the HLATE assumptions are satisfied, λ_1, μ_1 , τ_1, respectively represent the HLATE (heterogeneous local average treatment effect) of AWRs in villages in non-boom, unelectrified districts or villages with low (commercial) power availability. λ_4, μ_4, τ_4, are my key parameters, which represent the additional relative impact of AWRs in villages affected by the guar boom (λ_4), in electrified districts (μ_4), or in villages with high power availability (τ_4)respectively. Conditional on inclusion of district-cutoff FEs (and additional controls for endline power availability in 2011 in relevant specifications), these parameters highlight the extent to which the impacts of AWRs are augmented or reduced by a) complementary economic conditions (here, generated by the exogenous guar boom) and b) complementary access to electrification.

[8]These include indicators for village level primary school, medical centre, banking and bus facilities access, electrification availability, literacy rate, share of residents belonging to a scheduled caste or scheduled tribe, distance in Km from the closest census town, share of irrigated land, share log of total agricultural land area from the 2001 Population Census. From the 2002 BPL census, I include controls for the share of households owning agricultural land, the share of households earning over 4 USD per month in cash, the share of households employed in subsistence agriculture. From the 1998 Economic Census, I include controls for the share of village non-farm establishments with access to finance outside of self-funding, the share of village non-farm establishments with formal registration, the share of village non-farm establishments using fuel or electricity as a source of energy, and share of village population employed in (village) non-farm establishments (utilising data from the 1998 EC). As highlighted by (Imbens and Kalyanaraman, 2012), covariates and fixed-effects are not strictly necessary in a standard fuzzy RD regression, but can improve the precision of RD estimaton.

My RD sample is restricted to i) villages within a 84-person bandwidth of the population threshold (500 or 1000) according to the 2001 Population Census ii) located in complier states to PMGSY population-based guidelines (A-N 2020) iii) without paved road access in 2001 (receiving AWRs on the extensive margin) and iv) matched across all rounds of the Population Census, SECC, and PMGSY data. My final RD sample includes N= 11 431 villages across six Indian states[9].

6.3 Triple-difference Strategy

My RD Design generates causal interpretation of a HLATE local to a subset of villages within narrow population bandwidth of AWR population thresholds, in villages within complier states, for the complier village subpopulation who received AWRs due to being above the threshold, and would not have done so otherwise.

To investigate firm-level mechanisms driving structural transformation patterns, I extend analysis with a Triple-difference Specification. I exploit the staggered nationwide rollout of the PMGSY rural road construction Indian program across Indian states, comparing early-treated villages with later and never-treated villages, using four waves of village-level panel data from the Indian EC.[10] My triple-difference specification investigates interactive effects of AWRs in:

(a) villages located in Boom and Non-Boom districts

(b) villages with and without power availability in the closest (1991, 2001, 2011) Population census [11]

I use a modified triple difference design to allow for a) variation in treatment timing and b) time and village fixed effects

[9]I test robustness of my RD estimates to different bandwidths and kernel weighting functions in my robustness section

[10]See Roth et al . for an overview of modern DiD methodologies used in my paper

[11]For consistency, I use a binary indicator for Power Availability in triple-difference design, as hourly village power is not available in Population Census rounds prior to 2011. I do not utilise the RGGVY program in this part of my research design as the program was not present prior to 2005, when early-treated PMGSY recipient villages received AWRs. Furthermore I believe I am better able to alleviate potential endogeneity of village power availability in my triple-difference design through the inclusion of village fixed effects, which controls for all time-invariant village heterogeneity. NB: due to insignificant findings in RD analysis, results of Power x Road construction for triple-difference analysis are provided in Appendix only. RD results for both interactions are presented in my main body.

My main triple-difference specification(s) are:

$$Y_{vt} = \alpha_0 + \alpha_1 ROAD_{i(t-1)} + \alpha_2 ROAD_{i(t-1)} \times BOOM_{vt} + \alpha_3 BOOM_{vt} + X'\beta + d_t + v_v + e_{vt}$$

$$Y_{vt} = \beta_0 + \beta_1 ROAD_{i(t-1)} + \beta_2 ROAD_{i(t-1)} \times POWER_{vt} + \beta_3 POWER_{vt} + X'\beta + d_t + v_v + e_{vt}$$

where Y_{vt} is the firm-level economic outcome for village v in year t, T_{it} is an indicator equal to 1 if a village is treated by the year prior to Economic Census Round t (remaining 1 for subsequent rounds), d_t refers to time level fixed effects. v_v refers to village-level fixed effects Finally, $X'b$ is a vector of time-varying village-level covariates imputed from the 1991, 2001, and 2011 Population Census rounds.[12]. α_2 and β_2 are the key triple-difference parameters, the additional impact of roads on village-firm outcome Y in village v in Boom vs Non-Boom villages (villages with vs. without power availability) at round t.

A basic DD specification assumes in absence of treatment, treated units would have followed parallel trends as the control (later or never-treated) villages. The triple-differences specification relies on weaker assumptions than this. My identifying assumption is, in absence of being co-located in boom districts or with village power access, such villages would have followed parallel trends as PMGSY treated villages absent complementarities or constraints. Identification would be challenged only by a) a persistent shock only affecting AWR-recipient boom villages b) a shock affecting villages with both AWRs and power availability rather than only one.

In spite of this, I present graphical evidence of pre-treatment trends (from the period 1990-2013) separately for never-treated, early-treated, and later-treated AWR-recipient villages for two key dependent outcomes- number of non-agricultural firms and non-agricultural employment (Figure 8. The figures confirm pre-treatment, untreated villages follow approximately parallel trends- suggesting the weaker triple difference assumption is also likely to hold.

[12]In my sample this includes : village population and number of households, literate share, scheduled tribe/caste share, number of primary schools in a village, access to power and paved road (independent of PMGSY construction), imputed from 1991, 2001, and 2011 rounds of the Indian Population Census.

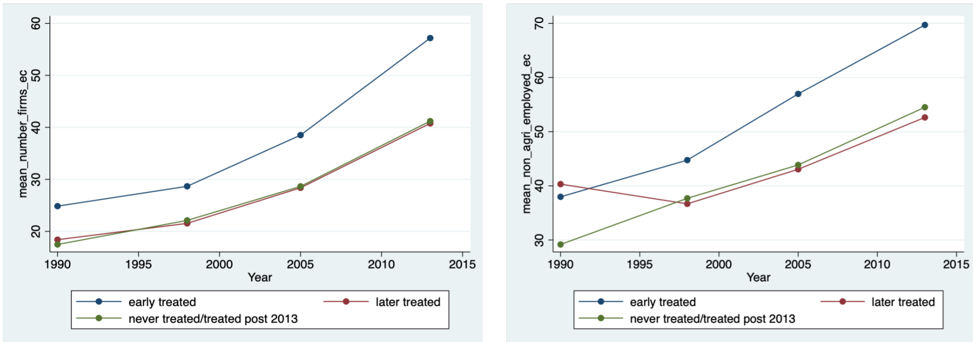

Figure 8: Parallel Trends Plot

I also robustify analysis for potential violation of parallel trends by comparing results in models excluding and including state-specific and state-district specific trends.

My triple-difference sample comprises all villages (a) without paved road in 2001 (b) merged to all data (PMGSY, Population Census, at least 3 waves of EC data) (c) with 2001 population greater than 500 (exceeding the lowest PMGSY population threshold in RD analysis (d) in states with at least N¿ 500 observations. My final sample consists of N= 74, 062 independent villages across 13 states[13]. The sample includes 6, 419 'early-treated' (treated 2000-2005), 22, 836 'late-treated' (treated 2006-2013), and 44, 807 'never-treated' (never treated or treated after the 2013 EC) villages.

[13](Andhra Pradesh, Assam, Bihar, Chhatisgarh, Gujarat, Jharkhand, Karnataka , Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Uttar Pradesh, Uttarakhand.

7 Results

7.1 Main RD Results

I first present RD results on impact of AWRs on (i) short-to-medium-run labor market outcomes (2011 Population Census) and (ii) occupational choice and structural transformation (2012 SECC). I explore how impacts of AWRs are affected by a) village co-location in Boom districts and b) village electrification, and examine whether such impacts exhibit iii) within-village heterogeneity.

7.1.1 Short-Run Labour Market Outcomes

I firstly compare interactive effects SR labour market outcomes from the 2011 Population Census and external market conditions Table 1A presents impacts of AWRs on short-term labour force participation- in absolute figures and as share of 2011 village population. Columns (1) and (4) present results with my main RD specification (including district-cut-off FEs) for two short-term labour market outcomes – Log(employment) and Employment Rate. My main specification implies in non-Boom villages, AWRs caused approximate 9.7 percentage-point decrease in total employment and 3.8 percentage-point reduction in total village employment rate. Both coefficients are weakly significant at the 10% level. Columns (2-3) and (5-6) show results are consistent with lower levels of fixed-effects, although more imprecisely estimated, with neither set of coefficients statistically significant at conventional levels.

These figures imply AWRs generated either negligible effects to a small decrease in extensive margin labour force participation in Boom and Non-Boom villages. In specification 1 and 4) the reported decrease in 2011 Employment Rate is lower in magnitude than the decrease in total employment A potential mechanism is AWRs caused increases in permanent out-migration generating a simultaneous decline in population size. Even when controlling for population changes (4) implies AWRs caused a decrease in village employment rate. A rationale is AWRs increased access to the external labour market, caused systematic attrition of existing employed village citizens to more productive opportunities (compositional labour force changes), and increased internal competition in Boom and Non-Boom villages.

The second row of table IA illustrates my key parameter: Road×BOOM, which captures additional impacts of AWRs on total village employment in Boom vs. Non-Boom villages. I do not observe significant relative variation in total employment in Boom villages. Although I cannot rule out differential rates of entry and exit of workers leaving employment unchanged, these findings indicate AWRs did not lead households to shift extensive margin labour force participation differentially to those in non-Boom villages.

I next consider the interaction of AWRs and village electrification on short-term labour market outcomes from the 2011 Population Census[14] . Table 2A illustrates results on SR Labour Market Outcomes from the 2011 Population Census using village level power availability (HIGHPOWER) as the electrification measure. In my primary specification incorporating district-cut-off FEs, I observe AWRs caused an approximate 11.0 percentage-point reduction in total employment in villages with low power availability, and a simultaneous (smaller) 4.2 percentage-point reduction in village employment rate. Similar to findings in Boom villages, I do not observe evidence villages with High Power Availability (Road $\times$HIGHPOWER) observed differential SR trends in total employment or employment rate relative to villages with low power availability. This result is robust across all specifications in Table 2A alongside an indirect measured of electrification in Table 2B- where I do not observe evidence villages in high RGGVY implementation districts, conditional on district-cut-off FEs saw larger or smaller impacts of AWRs on labour force participation on the extensive margin (Road $\times$ ELECTRIFIED) relative to unelectrified districts. The consistency of null results utilizing direct measures of village electrification implies my results are not unduly biased by sample attrition.

[14]I include controls for 1998 Employment Share in all electrification measures due to likely omitted variable bias between baseline cultivation share and village electrification

Figure 1: SR Labour Market Outcomes – ROAD X BOOM

Table 1A- Short Term Labour Market Outcomes – 2011 Population Census- (ROAD x Boom)

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln (employed) 2011	Ln (employed) 2011	Ln (employed) 2011	Employment Rate 2011	Employment Rate 2011	Employment Rate 2011
ROAD	-.097*	-.116	-.098	-.039*	-.038	-.021
	(.057)	(.073)	(.066)	(.023)	(.025)	(.022)
ROAD x BOOM	.065	.087	-.047	.026	.026	-.018
	(.053)	(.066)	(.047)	(.022)	(.022)	(.024)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.606	.309	.229	.280	.276	.106
Fixed Effects	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 2: SR Labour Market Outcomes – ROAD X HIGH POWER, ROAD x ELECTRIFIED)

Table 2A- Short Term Labour Market Outcomes – 2011 Population Census- (ROAD x High Power)
(HIGH POWER = Mean Village Daily Hours of Commercial Power ≥ 8 Hours)

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln (employed) 2011	Ln (employed) 2011	Ln (employed) 2011	Employment Rate 2011	Employment Rate 2011	Employment Rate 2011
ROAD	-.110**	-.129*	-.154**	-.042*	-.041*	-.04*
	(.05)	(.068)	(.07)	(.021)	(.021)	(.022)
ROAD x HIGH POWER	.117	.133	.106	.036	.038	.039
	(.076)	(.089)	(.097)	(.03)	(.03)	(.032)
HIGH POWER	-.037	-.038	-.017	-.013	-.013	-.012
	(.026)	(.029)	(.034)	(.01)	(.01)	(.012)
Observations	9195	9195	9195	9195	9195	9195
R-squared	.604	.312	.21	.277	.273	.096
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 2B- Short Term Labour Market Outcomes – 2011 Population Census- (ROAD x ELECTRIFIED)
(ELECTRIFIED = Over 60% of Villages in District received Electrification under RGGVY program)

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln (employed) 2011	Ln (employed) 2011	Ln (employed) 2011	Employment Rate 2011	Employment Rate 2011	Employment Rate 2011
ROAD	-.099	-.094	-.127**	-.038	-.036	-.063***
	(.062)	(.08)	(.059)	(.025)	(.024)	(.024)
ROAD x ELECTRIFIED	.042	-.022	-.005	.012	.008	.043
	(.054)	(.07)	(.07)	(.023)	(.022)	(.027)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.606	.316	.213	.280	.277	.074
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

7.1.2 Occupational Choice and Structural Transformation

I now explore occupational choice from the 2012 SECC. Table 3A presents RD results (Road $\times$BOOM) for impacts of AWRs on household occupational choice- separately for the share of households employed in agriculture (1-3) and (non-agricultural) manual labour (4-6) . Table 3A highlights AWRs created a large labor reallocation away from cultivation that is entirely concentrated in Non-Boom villages. Table 3A , (1-2) illustrates in non-Boom villages, AWRs led to a 12.1 to 12.7 percentage-point decline in the share of agricultural households in models with Further evidence comes from (4-6) which illustrate AWRs simultaneously led to 8.6 (state FE) to 9.8 percentage-point increase in manual labour in non-boom villages. The robustness of findings implies AWRs augmented structural transformation rates in non-boom districts.

The second row in Table 3A shows Boom villages saw a significantly reduced relative structural transformation gains from AWRs. Columns (1) and (2) illustrate boom-villages saw a 16.9 to 17.3 percentage-point reduced decline in share of households with main occupation as cultivation. Similarly, (4) and (5) highlight relative to Boom villages, Non-Boom villages experienced 11.1 to 11.4 percentage-point smaller increase in share of households employed in manual labour. Noticeably, coefficient estimates together sum to an effect largely indistinguishable from zero in boom villages. Ultimately, these findings suggest whilst non-Boom villages saw substantial labor-reallocation, boom-villages saw a net zero gain in structural transformation from rural roads. These findings contrast with Usmani (2020), who demonstrated boom-villages saw a concurrent increase in structural transformation relative to non-Boom villages following the RGGVY rural electrification program – implying complementarities between infrastructure investments (AWRs and rural electrification) and identical external economic conditions may differ- leading to diverging structural transformation and labor re-allocation outcomes.

Table 3B presents occupational choice using an alternative configuration of Share of Households with the Main Source of Income as Cultivation (- (1-3)) and Manual Labour (4-6) . In contrast to Table 3A, I fail to observe effects significantly different from zero for both exit from cultivation (1-3) or entry to manual labour (4-6) in non-boom districts. Additionally, the only observed discrepancy (in the interaction term Road $\times$ Boom) is observed in column 6, figure 3B, which implies, relative to Non-Boom villages, Boom villages saw a 10.9 percentage-point reduction in share of households with manual labor as their main income source. However, these findings are only significant with a lower level of FEs which may fail to capture unobservable district-level variation. Table 2B implies gains to structural transformation from AWRs were largely captured by members of the household who were not the primary earners – like new labour market entrants in non-Boom villages. These findings support work by Kim & Topel (1995), who studied Korean industrialisation and found non-agricultural enterprise mainly hired new entrants to the labour market rather than former agricultural workers, leading to large societal structural transformation rates but limited household changes in occupational choice on the extensive margin (i.e. for existing farm workers).

Tables 4A-4D revisits this relationship using the interaction between AWR and village power availability (Road×HIGHPOWER, 4A & 4B) and village co-location in electrified districts (ROAD×ELECTRIFIED, 4C & 4D). Tables 4A and 4C examine my primary measures of structural transformation and Tables 4B and 4D utilize an alternative metric. Findings imply AWR led to labor reallocation away from agriculture towards non-agricultural manual labor in villages with low power availability, and in unelectrified districts. Figure 4A showcases villages with low power availability showed between a 9.6 (district-cut-off FE), 10.2 (district FE) to 10.8 (state FE) percentage-point reduction in share of households employed in agriculture, and a corresponding 8.3, 8.8, to 9.0 respective percentage-point increase in share employed in non-agricultural manual labour.

However, diverging from earlier results, I fail to observe evidence of differential rates of structural transformation in electrified villages or with high power availability relative to the counterfactual. Table 4A, reports relative to Low Power villages, villages with High Power saw a 4.0 percentage-point reduced decrease in share of households employed in agriculture from AWRs and a 5.1 percentage-point lower share of households employed in non-agricultural manual labour using a district-cut-off FEs specification. Coefficient estimates are insignificant across models featuring direct and indirect measures of electrification. These results imply, conditional on baseline availability,varying village electrification status did not impact labor reallocation gains of rural roads.

7.1.3 Within-Village Heterogeneity

My previous results illustrate average structural transformation benefits of AWRs were highly heterogeneous to external market conditions, but were not impacted by village electrification. To understand whether external market conditions created larger relative impacts on specific subgroups, Figure 5 (A & B) estimates my ROAD x

BOOM specification on share of village workers employed in agriculture separately by a) household landholding share and b) Male Age Bracket. Row 1 in both specifications illustrates that in Non-Boom Villages, AWRs led to the greatest impacts on structural transformation on households without land and for the youngest male demographic bracket. Row 2, illustrates that relative to non-Boom villages, Boom villages saw a significantly increased share of workers in agriculture from AWRs across all household land-sizes, but a significantly larger discrepancy vs. non-Boom villages only for males aged between 10 to 20.

These results are partially consistent with theoretical predictions that roads augmented structural transformation among sub-groups with lowest opportunity cost of leaving agriculture in non-Boom villages (young workers) and it is for these groups for which relative discrepancies in labor reallocation gains in Boom villages was largest.

Figure 3- Occupational Choice and Structural Transformation (ROAD x BOOM)

Table 3A- Share of Household Employed in Agriculture and Manual Labour (ROAD x BOOM)

	(1)	(2)	(3)	(4)	(5)	(6)
	AGRICULTURE	AGRICULTURE	AGRICULTURE	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR
ROAD	-.121**	-.127**	-.081	.094*	.100**	.094*
	(.054)	(.054)	(.057)	(.05)	(.049)	(.049)
ROAD x BOOM	.172***	.175***	-.015	-.111*	-.115*	-.112*
	(.064)	(.064)	(.05)	(.06)	(.06)	(.059)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.262	.256	.162	.245	.239	.245
Fixed Effects	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 3B- Share of Households with Main Source of Income as Cultivation and Share of Households with Main Source of Income as Manual Labour (ROAD x BOOM)

	(1)	(2)	(3)	(4)	(5)	(6)
	CULTIVATION	CULTIVATION	CULTIVATION	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR
ROAD	-.048	-.057	-.055	.005	.013	.034
	(.054)	(.056)	(.055)	(.054)	(.055)	(.056)
ROAD x BOOM	.062	.066	.048	-.026	-.03	-.082*
	(.077)	(.077)	(.05)	(.077)	(.077)	(.044)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.302	.296	.198	.272	.268	.172
Fixed Effects	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 4 -Occupational Choice and Structural Transformation (ROAD x HIGH POWER, ROAD x ELECTRIFIED)

Table 4A- Share of Household Employed in Agriculture and Manual Labour (ROAD x HIGH POWER)

(HIGH POWER = Mean Village Daily Hours of Commercial Power ≥ 8 Hours)

	(1)	(2)	(3)	(4)	(5)	(6)
	AGRICULTURE	AGRICULTURE	AGRICULTURE	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR
ROAD	-.096*	-.102*	-.108**	.083*	.088*	.09*
	(.052)	(.052)	(.054)	(.047)	(.047)	(.048)
ROAD x HIGH POWER	.04	.041	.045	-.051	-.052	-.041
	(.051)	(.051)	(.055)	(.051)	(.051)	(.056)
HIGH POWER	-.03	-.03	-.048**	.032*	.032*	.05**
	(.019)	(.019)	(.021)	(.019)	(.019)	(.021)
Observations	9195	9195	9195	9195	9195	9195
R-squared	.272	.267	.154	.249	.243	.138
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 4B- Share of Households with Main Source of Income as Cultivation and Share of Households with Main Source of Income as Manual Labour (ROAD x HIGH POWER)

(HIGH POWER = Mean Village Daily Hours of Commercial Power ≥ 8 Hours)

	(1)	(2)	(3)	(4)	(5)	(6)
	CULTIVATION	CULTIVATION	CULTIVATION	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR
ROAD	-.048	-.057	-.053	.016	.023	.021
	(.053)	(.054)	(.057)	(.053)	(.054)	(.056)
ROAD x HIGH POWER	.055	.061	.047	-.073	-.079	-.058
	(.069)	(.069)	(.073)	(.067)	(.067)	(.068)
HIGH POWER	-.024	-.026	-.043*	.027	.029	.047**
	(.025)	(.025)	(.026)	(.024)	(.024)	(.023)
Observations	9195	9195	9195	9195	9195	9195
R-squared	.300	.294	.197	.267	.263	.17
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses

**** p<.01, ** p<.05, * p<.1*

Table 4C- Share of Household Employed in Agriculture and Main Occupation as Manual Labour (ROAD x ELECTRIFIED)

(ELECTRIFIED = Over 60% of Villages in District received Electrification under RGGVY program)

	(1)	(2)	(3)	(4)	(5)	(6)
	AGRICULTURE	AGRICULTURE	AGRICULTURE	NON-AGRICULTURE MANUAL LABOUR	NON-AGRICULTURE MANUAL LABOUR	NON-AGRICULTURE MANUAL LABOUR
ROAD	-.117**	-.123**	-.05	.083*	.104**	.014
	(.058)	(.057)	(.074)	(.047)	(.051)	(.057)
ROAD x ELECTRIFIED	.078	.084	-.043	-.051	-.083	.024
	(.057)	(.057)	(.057)	(.051)	(.051)	(.06)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.274	.268	.177	.249	.250	.131
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses

**** p<.01, ** p<.05, * p<.1*

Table 4D- Share of Households with Main Source of Income as Cultivation and Share of Households with Main Source of Income as Manual Labour (ROAD x ELECTRIFIED)

(ELECTRIFIED = Over 60% of Villages in District received Electrification under RGGVY program)

	(1)	(2)	(3)	(4)	(5)	(6)
	CULTIVATION	CULTIVATION	CULTIVATION	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR	(NON-AGRI) MANUAL LABOUR
ROAD	-.047	-.056	.006	.002	.01	-.049
	(.062)	(.062)	(.059)	(.064)	(.064)	(.06)
ROAD x ELECTRIFIED	.044	.051	-.043	-.029	-.036	.053
	(.055)	(.055)	(.053)	(.055)	(.055)	(.055)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.303	.298	.201	.273	.269	.165
FE	DISTRICT-CUTOFF	DISTRICT	STATE	DISTRICT-CUTOFF	DISTRICT	STATE

Clustered Robust Standard errors are in parentheses

**** p<.01, ** p<.05, * p<.1*

Figure 5- Within Village Heterogeneity (ROAD x BOOM)

Table 5A- Within Village heterogeneity (Share of workers in Agriculture by Size of Household Landholding)

	(1) (Landless)	(2) (0-2 hectares)	(3) (2-4 hectares)	(4) (4+ hectares)
ROAD	-.141**	-.13**	-.115*	-.094
	(.059)	(.062)	(.062)	(.068)
ROAD x BOOM	.145**	.161**	.205***	.155**
	(.062)	(.069)	(.074)	(.071)
Observations	11100	10697	10379	9944
R-squared	.213	.173	.176	.202
Fixed Effects	District-Cutoff	District-Cutoff	District-Cutoff	District-Cutoff

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 5B- Within Village heterogeneity (Share of workers in Agriculture by Male Age Bracket)

	(1) (Landless)	(2) (0-2 hectares)	(3) (2-4 hectares)	(4) (4+ hectares)
ROAD	-.016*	.002	.007	-.006
	(.009)	(.009)	(.010)	(.011)
ROAD x BOOM	.033***	.001	-.015	-.004
	(.062)	(.069)	(.074)	(.071)
Observations	11100	10697	10379	9944
R-squared	.215	.178	.170	.201
Fixed Effects	District-Cutoff	District-Cutoff	District-Cutoff	District-Cutoff

Clustered Robust Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

8 Discussion and Plausible Mechanisms

I next discuss mechanisms behind my main finding: AWRs generated significantly reduced (net-zero) impact on structural transformation in Boom villages with benefits entirely accruing to non-Boom villages . I explore three primary channels: variation in (i) non-agricultural firm creation (ii) permanent migration (iii) temporary migration.[15]

–

8.1 Non-Agricultural Firm Creation

I conduct a Triple Difference analysis of non-agricultural firm creation using Indian EC data. My first firm channel is Non-Agricultural Enterprise Growth. Spatial heterogeneity in infrastructure has been shown (Martin & Rogers 1995) to guide firm-level investment decisions along extensive and intensive margins. Interactions between AWRs and variation in intensity of agricultural demand shocks may have reduced non-agricultural firm growth in Boom villages, explaining labor reallocation patterns.

8.1.1 Firm Channel 1: Firm Entry and Enterprise Growth

Addendum Figure 1 presents triple-difference analysis whose key parameter is ROADxBOOM, Figure 1A presents results for non-agricultural firm counts, Figure 1B for non-agricultural employment. Figure 1A illustrates in non-Boom villages, AWRs led to either an insignificant change or small decrease of 0.669 firms (column 2). Relative to non-Boom villages, the ROADxBOOM coefficient is insignificant in all specifications, implying no differential patterns of extensive margin firm entry. Figure 1B illustrates results for non-agricultural employment. The results for non-Boom villages are inconsistent implying roads generated between an approximate 4.3 percentage-point decrease to 3.2 percentage-point increase in employment (1). Notably, the coefficients are small in magnitude, illustrating the impact of AWRs on firm employment was economically insubstantial. The coefficient on ROADxBOOM is positively significant in all models except with state-district linear trends (model 4), rather than a negative sign, as would be required to support this channel. Results suggest Boom villages saw

[15] Due to insignificant findings relating electrification and occupational choice in the RD analysis, and the subject of my section focusing on mechanisms driving structural transformation benefits of roads, I do not discuss Power x Road triple-difference . These results are provided in Addendum Figure 5

up to an approximate 16.9 percentage-point relative increase in non-agricultural in-village employment. A rationale could be supply-side mechanisms- increased growth of non-agricultural firm employment complimentary to agricultural growth (Usmani, 2020) or demand-driven channels caused by BOOM-related income effects.

My findings reject (extensive or intensive margin) firm entry as the mechanism driving lower structural transformation from AWRs in BOOM villages; even if the insignificant model (4) result is correct, effect sizes are small whereas a large decrease in village employment would be required to explain this finding.

8.1.2 Firm Channel 2- Firm Sectoral and Size Composition

Addendum Figure 2A presents triple-difference results on firm counts (2A a)) and employment (2A b)) by firm size- subdivided into Own Account Enterprises (no hired employees), Non-Directory Establishments (2-5 employees) and Directory Establishments (6+ employees). Figure 2A illustrates in non-Boom villages, AWRs led to significant increase in OAEs and NDE numbers, and decline in DE numbers. Sign coefficients are consistent across specifications. Nevertheless, magnitude of effect size is small- amounting to .631-.831 increased OAEs number in Non-Boom villages and .278-.835 decreased DE count. The ROADxBOOM coefficient illustrates an inverse trend in Boom villages. Relative to Non-Boom villages, results imply Boom villages saw decrease in small OAEs and NDEs (columns 1-4) and an increase in DEs (column 5-6). As Indian firms follow an asymmetric size distribution (with a larger share of microenterprises than under normal distribution), the results imply a slight reduction of average firm size in non-Boom villages, and an increase in Boom villages. Figure 2A b) repeats analysis with employment by firm size. Notably, the patterns are much less clear with no significant ROADxBOOM coefficients in models including state-district trends. An interpretation is size growth was primarily driven by new larger firms in Boom villages, rather than intensive margin effects. Nevertheless, one would instead expect a decrease in firm size to justify higher structural transformation in Non-Boom villages, providing evidence against this channel.

Investigating patterns of sectoral firm composition, Figure 2B and 2C repeats analysis of firm count and employment by non-agricultural firm sector (manufacturing and services) Whilst I observe small relative decrease in count and relative increase in employment of services firms in BOOM-villages, implying growth in size of service firms- this does not support observed rates of labor reallocation between Boom and Non-Boom villages; I also fail to find effects on manufacturing firm entry.

8.1.3 Firm Channel 3- Social Grouping of Firm Ownership

Theory implies households with lowest opportunity cost of leaving agriculture may be more likely to shift sectors. Assuming individuals from disadvantaged castes/women face higher discrimination in external labour markets precluding out-migration, asymmetric relative growth favouring non-agricultural firms owned by these groups in Non-Boom villages could generate higher rates of structural transformation Testing intuition, Figures 2A and 2E disaggregate firm entry by owner sex and caste. Whilst finding no evidence of differential trends in male-owned firms, figure 2E presents limited evidence Boom villages saw relatively higher rates of growth in female-owned firms among OBC and General castes; this finding contradicts the channel's hypothesis.

8.1.4 Firm Conclusions

My analysis rejects non-agricultural village firm activity as a plausible mechanism driving lower relative labor reallocation gains in Boom villages.

To ensure variation between triple-difference and RD sample does not bias conclusions, Figure 1C shows RD results of firm entry using data from the 2013-4 EC. My findings are broadly consistent with DD/triple-difference analysis. I observe in Non-Boom villages, AWRs led to an approximate 29.4 percentage-point increase in non-agricultural employment- with no differential trend in boom-villages- confirming variation in village firm growth patterns cannot explain distinct structural transformation patterns between both subgroups.

8.2 Permanent Migration

Previous results imply variation in structural transformation between Boom and Non-Boom villages must be explained by out-village labour market variation, rather than in-village firm creation. This may be caused by either variations in relative AWR impacts on a) permanent migration or b) temporary migration. I am unable to directly test migration formally, but examine two proxy measures of permanent migration.

1. **Absolute Population discontinuities** - Firstly I show evidence of no divergences in discontinuities in village population between Boom and Non-Boom villages from AWRs using my original RD specification. Figure 3A presents findings, illustrating no evidence of significantly lower relative permanent migration rates in Boom villages.

2. **Village Age and Gender Distribution** – Secondly I examine evidence of divergences in village age distribution and share of population is male – between Boom and Non-Boom villages. Permanent population changes, driven by in and out-migration, should average to changes in village demographics. Figures 3B and 3C represent RD results on both total and male population divided in 10-year age bins. I observe a relative increase in village male-share aged 10-20 in Boom villages- implying AWRs led to reduced permanent out-migration among younger members of the labour force in Boom villages. However, estimated coefficient sizes are too small to explain observed structural transformation patterns. These findings imply variations in permanent migration, cannot fully explain discrepancies.

8.3 Conclusion: Temporary Migration

My evidence suggests RD results cannot be driven by a) variation in permanent migration or b) non-agricultural firm activity; thus, I identify diverging patterns of temporary out-migration between Boom and Non-Boom PMGSY-recipient villages as the most plausible mechanism. This finding supports theoretical predictions in Boom villages, the (predominately) Agricultural boom reduced the opportunity cost of staying in village agriculture, reducing patterns of temporary out-migration in combination with AWRs.

8.4 Other potential channels and counterarguments

8.4.1 Complementary Economic Conditions or Regional Effects?

One unaddressed possibility is findings are driven by local effects, rather than the specific boom-AWR interaction. All Boom villages are located in the Northwestern region, predominately Rajasthan – one of the poorest Indian states. Structural transformation AWR benefits could potentially driven by weaker external labour markets in these villages rather than the concurrent boom. I argue this is unlikely for three reasons.

Firstly, RD specifications feature district-cut-off FEs, eliminating mean district-specific unobservables. Assuming district characteristics may interact with AWRs in regionally distinctive ways, I also control for baseline village characteristics, including 1998 Employment Share of Village Non-Agriculture, and results are largely unchanged. Secondly, triple-difference analysis showed in some instances higher non-agricultural firm growth in Boom vs Non-Boom villages. If results were driven by weaker labor markets, it is unlikely this would manifest in higher relative in-village firm growth. Finally, Usmani (2020) finds rural electrification generated greater relative impacts on structural transformation in Boom villages. Given I observe inverse trends for AWRs, it is unlikely Boom districts could have sufficiently strong labor markets to facilitate higher labor reallocation in Usmani (2020), but sufficiently weak labor markets to drive reduced structural transformation here.

8.5 Further Robustness Checks

I finally examine further robustness of my key (RD) results with four additional tests, using share of workers in agriculture (2012 SECC) as the dependent outcome.[16]

1. Robustness of RD Specification to Varying Kernel Functions and Bandwidths

Addendum Figure 4A re-estimates ROAD x BOOM interaction results results on 'Share of Households Employed in Agriculture' utilising bandwidths 60, 80, 100 and both Triangular and Uniform kernel functions. Results are consistent across specifications

2. Robustness of RD specification to varying Power Level

In my initial specification, I assigned $\mathring{8}$ hours of Commercial Power Availability as a High Power indicator. I re-estimate my main (ROAD x POWER) equations in Addendum Figure 4B using a low (4 hours+), and high (10+ hours) measure of commercial power availability. The Road x POWER term is insignificant across all specifications

3. Robustness of RD specification by varying RGGVY District Implementation Intensity

Similarly, Addendum Figure 4C re-estimates my ROAD x ELECTRIFIED equation utilising a lower (40% villages electrified) and higher (80% villages electrified) measure of 10th plan district electrification under RGGVY. The ROAD x ELECTRIFIED term is insignificant across specifications. This implies the choice of electrification implementation intensity not bias RD results.

4. Placebo RD estimation

Finally, I estimate evidence of pre-treatment variation between Boom and Non-Boom villages by re-running the ROAD x BOOM regression on non-agricultural village employment from 1990 and 1998 (pre-treatment) EC waves (Addendum Figure 4D). I do not observe differences in either non-boom villages (the ROAD coefficient) or divergent patterns in Boom villages (the ROAD x BOOM coefficient). This suggests a) changes in village non-agricultural employment were driven by roads post-treatment and b) pre-treatment differences between Boom and Non-Boom villages cannot explain observed results.

[16] Results are provided in Addendum Figure 4

9 Conclusion

Leveraging a RD and triple-difference research design, I observe benefits to PMGSY recipient villages are highly heterogeneous to external market conditions, but fail to find similar evidence of heterogeneity to village electrification (utilising both direct and indirect measures of electrification). I observe structural transformation gains from AWRs were entirely observed by villages with low-intensity exposure to the US Fracking Boom. Conversely, Boom villages saw economically substantial and significantly reduced (net zero) gains from structural transformation caused by AWRs (a 16.9-17.3 percentage-point reduced decline in households employed in agriculture, and a 11.1-11.4 smaller percentage-point increase in households employed in non-agricultural manual labour). These findings are robust to multiple bandwidth, kernel functions, varying levels of (direct and indirect) electrification, and placebo RD estimation. Furthermore, results imply substantial within-village heterogeneity, with largest relative discrepancies observed for individuals with lowest opportunity cost of out-migration at baseline (young men, and new labour market entrants).

A plausible mechanism is substantially reduced temporary out-migration in PMGSY-recipient boom villages relative to non-Boom villages. I attribute this to income effects- with the (predominately) agricultural boom causing price increases of a key agricultural commodity, reducing incentives of agricultural workers to out-migrate in Boom-villages (reducing opportunity costs of leaving agriculture).

I reject other mechanisms including variations in a) non-agricultural firm creation and (size, sectoral, or social) composition b) permanent migration and argue this finding is unlikely to be driven by local variation between Boom and Non-Boom villages. Conversely, I fail to observe similar evidence of heterogeneity of structural transformation benefits of AWRs by village electrification level – implying conditional on baseline availability, road construction's impacts on structural transformation is not particularly responsive to village electrification. A policy implication is spatial and temporal targeting of road investments must consider complementary (local) economic conditions or provide sufficient counter-investment to ensure incentives correctly favour out-migration (such as migration subsidies – e.g. Bryan 2015).

The paper has four key limitations. Firstly, I do not have detailed information on firm inputs within the EC census- thus, am unable to infer details about whether rural roads led to enhanced productivity. Secondly, my primary measure of electrification in indirect exposure to the RGGVY program is imperfect; due to data limitations, I was unable to obtain information on village-level rollout of electrification in the program, and rely on both direct and indirect measures due to this fact. Thirdly, although in discussion I highlight my findings are unlikely to be driven by regional variation between Boom and Non-Boom districts- I am unable to test this mechanism directly. An extension to this paper could consider the usage of randomization inference as in Usmani (2020) as a robustness check on analysis.

Finally, there may be concerns about imperfect compliance to the guar treatment. The price increase was so substantial that farmers that did not grow guar initially may have decided to switch crops and plant guar following the price rise. If so, the line between treatment and control may not be so firm. A potential solution could be to use actual guar production as the main treatment variable. Pre-cultivation could be instead used as an instrument for guar intensity- and climate or geography could also provide useful exogeneous variation.

My paper presents four opportunities for future research. Firstly, I identify primarily short-to-medium run effects on structural transformation between Boom and Non-Boom villages. Assuming gains from labor reallocation accumulate over time, re-studying relative occupational choice patterns over a longer time horizon could highlight whether structural transformation patterns persisted even after market conditions subsided. Secondly, a promising research area would be examining whether benefits of large-scale infrastructure investment are also contingent on local economic conditions.

Thirdly, this paper has taken the perspective that structural transformation is economically desirable. Whilst this may be true, having higher priced exports is also beneficial. If the latter holds back the former, there may be a welfare trade-off. Perhaps welfare consequences of the fracking boom differ over time, helping guar-growing villages in the short run, but holding them bac in the longer run by locking them into agriculture. A more thorough examination of the aggregate welfare consequences of the fracking shock studied here could be a relevant area of future study.

Finally, I focus predominately on structural transformation, however, changing external economic conditions may have influenced other variables. Understanding how external economic conditions, local infrastructure investment, and outcomes like consumption, human capital investment, spillover effects in neighbouring cities, environmental degradation (Asher et al. 2018)- may be critical in pioneering opportunity-enhancing policies in rural low-income settings.

10 References

1. Adukia, A., Asher, S. and Novosad, P., 2017. Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. SSRN Electronic Journal,.

2. Allcott, H., Collard-Wexler, A. and O'Connell, S., 2014. How Do Electricity Shortages Affect Industry? Evidence from India. American Economic Review,.

3. Asher, S. and Novosad, P., 2020. Rural Roads and Local Economic Development. American Economic Review, 110(3), pp.797-823.

4. Asher, S., Garg, T. and Novosad, P., 2020. The Ecological Impact of Transportation Infrastructure. The Economic Journal, 130(629), pp.1173-1199.

5. Asher, S., Lunt, T., Matsuura, R. and Novosad, P., 2022. Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform.

6. Banerjee, A., Duflo, E., & Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2018637

7. Becker, S., Egger, P. and von Ehrlich, M., 2013. Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects. American Economic Journal: Economic Policy, 5(4), pp.29-77.

8. Bernard, T. and Torero, M., 2015. Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia. Economic Development and Cultural Change, 63(3), pp.459-484.

9. Bryan, G., Chowdhury, S. and Mobarak, A., 2013. Escaping Famine Through Seasonal Migration. SSRN Electronic Journal,.

10. Burlig, F. and Preonas, L., 2021. Out of the darkness and into the light? Development effects of rural electrification. Haas Energy Institute WP,.

11. Carlowitz, T., 2021. · Carlowitz, T. (2021). "The Effect of Rural Electrification on Firm Creation - New Evidence from Ghana. CSAE Working Paper Series, 10.

12. Dinkelman, T., 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. American Economic Review, 101(7), pp.3078-3108.

13. Donaldson, D. and Storeygard, A., 2016. The View from Above: Applications of Satellite Data in Economics. Journal of Economic Perspectives, 30(4), pp.171-198.

14. Donaldson, D., 2018. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. American Economic Review, 108(4-5), pp.899-934.

15. Fan, J., 1994. The Economic Analysis of Rural Industrialization of China and the Difference of Provincial Development Level. , 5, 398-407. Geographica Sinica, 5.

16. Gelman, A. and Imbens, G., 2018. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. Journal of Business amp; Economic Statistics, 37(3), pp.447-456.

17. Gibson, J. and Rozelle, S., 2003. Poverty and Access to Infrastructure in Papua New Guinea. SSRN Electronic Journal,.

18. Gollin, D., Lagakos, D. and Waugh, M., 2014. Agricultural Productivity Differences across Countries. American Economic Review, 104(5), pp.165-170.

19. Henderson, J., Storeygard, A. and Weil, D., 2012. Measuring Economic Growth from Outer Space. American Economic Review, 102(2), pp.994-1028.

20. Imbens, G. and Kalyanaraman, K., 2009. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. SSRN Electronic Journal,.

21. Imbens, G. and Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142(2), pp.615-635.

22. Kassem, D., 2021. Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia. CRC TR 224 Discussion Paper Series,.

23. Khandker, S., Samad, H., Ali, R. and Barnes, D., 2014. Who Benefits Most from Rural Electrification? Evidence in India. The Energy Journal, 35(2).

24. Kim, Dae-Il and Robert H Topel, "Labor Markets and Economic Growth: Lessons from Korea's Industrialization, 1970-1990," in Richard B. Freeman and Lawrence F. Katz, eds., Differences and Changes in Wage Structures, University of Chicago Press, 1995, chapter 7, pp. 227-264

25. Kothari, S., 2014. The Size Distribution of Manufacturing Plants and Development. SSRN Electronic Journal,.

26. Lawson-Johns, J., 2021. Roads to Resilience: Can rural road infrastructure improve household resilience to flood-shocks? A case study of India's PMGSY initiative. CSAE Working Paper Series, 10.

27. Lee, D. and Lemieux, T., 2010. Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48(2), pp.281-355.

28. Lenz, L., Munyehirwe, A., Peters, J. and Sievert, M., 2017. Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwanda's Electricity Access Roll-Out Program. World Development, 89, pp.88-110.

29. Lipscomb, M., Mobarak, A. and Barham, T., 2013. Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. American Economic Journal: Applied Economics, 5(2), pp.200-231.

30. Martin, P. and Rogers, C., 1995. Industrial location and public infrastructure. London: CEPR.

31. Meager, R., 2015. Understanding the Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of 7 Randomised Experiments. SSRN Electronic Journal,.

32. Moneke, N., n.d. an Big Push Infrastructure Unlock Development? Evidence from Ethiopia.

33. Mu, R. and van de Walle, D., 2011. Rural Roads and Local Market Development in Vietnam. Journal of Development Studies, 47(5), pp.709-734.

34. Roth, J., Sant'Anna, P., Bilinski,, A. and Poe, J., 2022. What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.

35. Usmani, F. and Fetter, T., 2020. Fracking, farmers, and rural electrification in India. Ruhr Economic Paper, 864.

11 Appendix

12 Addendum

ADDENDUM- Mechanism and Robustness Results

FIGURE 1-Triple Difference and Regression Discontinuity- Firm Entry

1A- Triple Difference- Number of Non-Agricultural Firms

	(1) Fixed Effects	(2) Fixed Effects	(3) Fixed Effects	(4) Fixed Effects
ROAD	.171	-.669**	.449	.155
	(.312)	(.311)	(.319)	(.321)
ROAD X BOOM	-1.03	-.902	-.974	-1.186
	(.771)	(.739)	(.742)	(.743)
BOOM	-1.511**	-.548	1.719***	2.072***
	(.658)	(.624)	(.648)	(.675)
N	277977	277977	277977	277977
R-squared	.054	.115	.152	.18
Year FE	YES	YES	YES	YES
Village FE	YES	YES	YES	YES
Controls	NO	YES	YES	YES
Linear Trends	NO	NO	STATE	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

1B - Triple Difference- ln(Non-Agricultural Employment)

	(1) Fixed Effects	(2) Fixed Effects	(3) Fixed Effects	(4) Fixed Effects
ROAD	-.043***	-.032***	.02***	.013*
	(.007)	(.007)	(.008)	(.008)
ROAD X BOOM	.164***	.169***	.069**	.033
	(.033)	(.033)	(.032)	(.033)
BOOM	.13***	.14***	.118***	.167***
	(.028)	(.028)	(.029)	(.031)
Observations	277977	277977	277977	277977
R-squared	.166	.172	.196	.225
Year FE	YES	YES	YES	YES
Village FE	YES	YES	YES	YES
Controls	NO	YES	YES	YES
Linear Trends	NO	NO	STATE	STATE-DISTRICT

1C- Regression Discontinuity Design- Non-Agricultural Employment (2013 Economic Census Data)

	(1) Ln (Non-Agri Employment)	(2) Ln (Non- Agri Employment)	(3) Non-Agri Employment Share (2011 Population)	(4) Non-Agri Employment Share (2011 Population)	(5) Non-Agri Employment Share (2011 Workers)	(6) Non-Agri Employment Share (2011 Workers)
ROAD	.294**	.321**	.011	.012	.02	.022
	(.147)	(.156)	(.011)	(.011)	(.016)	(.016)
ROAD x BOOM	-.128	-.141	.000	-.001	-.009	-.011
	(.18)	(.187)	(.014)	(.014)	(.022)	(.022)
Observations	10833	10833	10833	10833	10833	10833
R-squared	.24	.201	.152	.148	.245	.242
FE	District-Cut-off	District	District-Cut-off	District	District-Cut-off	District
EC98 Employment Share Control	YES	YES	YES	YES	YES	YES

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

FIGURE 2-Triple Difference- Firm Composition

FIGURE 2A Triple Difference- Firm Composition by Size

a) Firm Counts by Size

	(1) OAE	(2) OAE	(3) NDE (2-5)	(4) NDE (2-5)	(5) DE (6+)	(6) DE (6+)
ROAD	.631**	.831***	.227**	.159*	-.278***	-.835***
	(.3)	(.311)	(.092)	(.092)	(.097)	(.099)
ROAD X BOOM	-1.181*	-1.254*	-1.06***	-.732**	1.331***	.764***
	(.615)	(.66)	(.348)	(.3)	(.147)	(.156)
BOOM	-2.627***	6.305***	.382	.95***	1.666***	-5.141***
	(.512)	(.606)	(.314)	(.321)	(.119)	(.161)
N	277977	277977	277977	277977	277977	277977
R-squared	.175	.276	.068	.091	.191	.273
Year FE	YES	YES	YES	YES	YES	YES
Village FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Linear Trends	NO	STATE-DISTRICT	NO	STATE-DISTRICT	NO	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

b) Employment by Firm Size

	(1) OAE	(2) OAE	(3) NDE (2-5)	(4) NDE (2-5)	(5) DE (6+)	(6) DE (6+)
ROAD	.033***	.051***	.015*	.007	-.018**	-.012
	(.009)	(.009)	(.008)	(.008)	(.008)	(.008)
ROAD X BOOM	.071*	.028	.022	-.035	.12***	.02
	(.04)	(.041)	(.036)	(.037)	(.036)	(.039)
BOOM	.056*	.609***	-.007	.019	.34***	-.045
	(.034)	(.042)	(.03)	(.037)	(.03)	(.036)
N	277814	277814	277892	277892	277683	277683
R-squared	.508	.545	.189	.216	.237	.28
Year FE	YES	YES	YES	YES	YES	YES
Village FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Linear Trends	NO	STATE-DISTRICT	NO	STATE-DISTRICT	NO	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 2B- Triple Difference- Count and Employment of Manufacturing Firms

	(1) Firm Counts- Manufacturing	(2) Firm Counts- Manufacturing	(3) Employment- Manufacturing	(4) Employment- Manufacturing
ROAD	-.556***	-.249	-.062***	-.003
	(.184)	(.21)	(.01)	(.01)
ROAD X BOOM	.284	.266	.065	.013
	(.313)	(.345)	(.042)	(.044)
BOOM	-.494**	.572*	.176***	.164***
	(.243)	(.305)	(.035)	(.041)
N	277977	277977	277977	277977
R-squared	.034	.076	.012	.052
Year FE	YES	YES	YES	YES
Village FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Linear Trends	NO	STATE- DISTRICT	NO	STATE- DISTRICT

Figure 2C- Triple Difference- Count and Employment of Service Firms

	(1) Firm Counts- Services	(3) Firm Counts- Services	(4) Employment- Services	(6) Employment- Services
ROAD	-.109	.399*	-.019***	.012*
	(.21)	(.207)	(.007)	(.007)
ROAD X BOOM	-.778	-1.174**	.15***	.013
	(.541)	(.553)	(.031)	(.03)
BOOM	-.463	1.183**	.061**	.114***
	(.451)	(.501)	(.027)	(.029)
N	277977	277977	277977	277977
R-squared	.121	.187	.263	.309
Year FE	YES	YES	YES	YES
Village FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Linear Trends	NO	STATE- DISTRICT	NO	STATE- DISTRICT

Figure 2D- Triple Difference- Non-Agricultural Firms by Social Group of Owner (Male)

	(1) MALE SC/ST	(2) MALE SC/ST	(3) MALE OBC	(4) MALE OBC	(5) MALE OTHER	(6) MALE GENERAL
ROAD	2.888	2.888	-.111	-.111	1.172	1.172
	(2.324)	(2.324)	(2.572)	(2.572)	(1.255)	(1.255)
ROAD X BOOM	-1.343	-1.343	1.561	1.561	.03	.03
	(2.255)	(2.255)	(2.458)	(2.458)	(1.596)	(1.596)
BOOM	6.575***	3.107***	9.023***	-.479	1.035	7.248***
	(.904)	(.621)	(1.034)	(1.019)	(.707)	(.641)
N	211019	211019	211019	211019	211019	211019
R-squared	.022	.103	.044	.133	.022	.107
Year FE	YES	YES	YES	YES	YES	YES
Village FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Linear Trends	NO	STATE-DISTRICT	NO	STATE-DISTRICT	NO	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 2E- Triple Difference- Non-Agricultural Firms by Social Group of Owner (Female)

	(1) FEMALE SC/ST	(2) FEMALE SC/ST	(3) FEMALE OBC	(4) FEMALE OBC	(5) FEMALE OTHER	(6) FEMALE GENERAL
ROAD	.065	.065	-.806	-.806	-1.056	-1.056
	(.774)	(.774)	(.581)	(.581)	(.088)	(.707)
ROAD X BOOM	.197	.197	1.687**	1.687**	.722***	.991***
	(.719)	(.719)	(.748)	(.748)	(.279)	(.329)
BOOM	.825***	.633***	1.709***	.278	-1.68***	-.661**
	(.297)	(.242)	(.529)	(.605)	(.266)	(.326)
N	211019	211019	211019	211019	211019	211019
R-squared	.015	.058	.028	.077	.037	.100
Year FE	YES	YES	YES	YES	YES	YES
Village FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Linear Trends	NO	STATE-DISTRICT	NO	STATE-DISTRICT	NO	STATE-DISTRICT

Clustered Robust standard errors are in parentheses

FIGURE 3- MIGRATION TESTS

Figure 3A : RD Population Checks (2011 Population Census- Regression Discontinuity Design)

	(1) 2011 Population	(2) Log (2011 Population)
ROAD	-7.263	-.025
	(21.633)	(.031)
ROAD x BOOM	-7.918	.013
	(21.569)	(.03)
Observations	11431	11431
R-squared	.834	.786

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 3B : RD Population Checks (2011 Population Census- Regression Discontinuity Design (SECC share of population by age bin)

	(1) secc_age_share _11_20	(2) secc_age_share_ 21_30	(3) secc_age_share_ 31_40	(4) secc_age_share_ 41_50	(5) secc_age_share_ 51_60
ROAD	-.003	-.004	.000	-.003	.002
	(.006)	(.005)	(.005)	(.004)	(.003)
ROAD x BOOM	-.003	.002	.003	0	-.002
	(.006)	(.004)	(.004)	(.004)	(.003)
Observations	11431	11431	11431	11431	11431
R-squared	.224	.194	.267	.382	.399

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Figure 3C : RD Population Checks (SECC share of male population by age bin)

	(1) secc_male_shar e_1~20	(2) secc_male_share _2~30	(3) secc_male_share _3~40	(4) secc_male_share _4~50	(5) secc_male_share _5~60
ROAD	-.016*	.002	.007	-.006	.014
	(.009)	(.009)	(.01)	(.011)	(.014)
ROAD x BOOM	.033***	.001	-.015	-.004	.008
	(.008)	(.008)	(.011)	(.012)	(.017)
Observations	11431	11431	11431	11431	11431
R-squared	.122	.193	.092	.07	.056

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

FIGURE 4-FURTHER ROBUSTNESS CHECKS

FIGURE 4A: Robustness of Main RD Specification by Kernel Function and Bandwidth (ROAD x BOOM, Dependent Variable- Share of Workers employed in Agriculture, 2011-2 SECC)

(Name- Kernel-Bandwidth)
Dependent Variable- Share of Workers Employed in Agriculture)

	(1)	(2)	(3)	(4)	(5)	(6)
	Triangular- 60	Triangular- 80	Triangular- 100	Uniform- 60	Uniform- 80	Uniform- 100
ROAD	-.127**	-.124**	-.129**	-.127**	-.127**	-.126**
	(.055)	(.053)	(.052)	(.059)	(.052)	(.052)
ROAD x BOOM	.175***	.175***	.175***	.175***	.175***	.175***
	(.064)	(.065)	(.061)	(.064)	(.061)	(.062)
Observations	11431	11431	11431	11431	11431	11431
R-squared	.256	.256	.256	.256	.256	.256

FIGURE 4B: Robustness of Main RD Specification by Village Power indicator- HIGH POWER = >=8, 10, or 6 Hours of Mean Commercial Power Availability 2011 Census

(Name- Kernel-Bandwidth)
Dependent Variable- Share of Workers Employed in Agriculture)

	>= 8 hours	>= 10 hours	>= 4 hours
ROAD	-.096*	-.102*	-.108**
	(.052)	(.052)	(.054)
ROAD x HIGH POWER	.04	.041	.045
	(.051)	(.051)	(.055)
HIGH POWER	-.03	-.03	-.048**
	(.019)	(.019)	(.021)
Observations	9195	9195	9195
R-squared	.272	.267	.154
FE	DISTRICT-CUTOFF	DISTRICT	STATE

FIGURE 4C: Robustness of Main RD Specification by DISTRICT RGGVY ELECTRIFICATION- ELECTRIFIED = >=60, 40, or 80% share of villages electrified under RGGVY)

(Name- Kernel-Bandwidth)
Dependent Variable- Share of Workers Employed in Agriculture)

	>= 60%	>= 80%	>= 40%
ROAD	-.117**	-.125**	-.098**
	(.058)	(.058)	(.058)
ROAD x ELECTRIFIED	.078	.073	.071
	(.057)	(.057)	(.057)
Observations	11431	11431	11431
R-squared	.274	.278	.279

FIGURE 4D: Placebo RD Estimation (ln (Non-Agricultural Employment) 2013 (post-treatment), 1998, 1990 (pre-treatment) EC waves

	2013 Ln (Non-Agri Employment) 2013	**2013** Ln (Non- Agri Employment)	**1998** Ln (Non-Agri Employment)	**1998** Ln (Non- Agri Employment)	**1990** Ln (Non-Agri Employment)	**1990** Ln (Non- Agri Employment)
ROAD	.294**	.321**	.105	.094	.318	.321
	(.147)	(.156)	(.100)	(.102)	(.243)	(.287)
ROAD x BOOM	-.128	-.141	-.121	-.101	-.107	-.141
	(.18)	(.187)	(.180)	(.187)	(.180)	(.187)
Observations	10833	10833	10833	10833	10833	10833
R-squared	.24	.201	.24	.201	.24	.201
FE	District-Cut-off	District	District-Cut-off	District	District-Cut-off	District
EC98 Employment Share Control	YES	YES	YES	YES	YES	YES

Clustered Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

FIGURE 5: ROAD x POWER Triple Difference Results

Figure 5A- Non-Agricultural Firm Counts (ROAD x POWER, Triple Difference)

	(1) Pooled OLS	(2) Fixed Effects	(3) Fixed Effects	(4) Fixed Effects	(5) Fixed Effects
ROAD	.198***	.029***	.019**	.04***	.028***
	(.011)	(.009)	(.009)	(.01)	(.009)
ROAD X POWER	-.141***	-.1***	-.089***	-.029**	-.022**
	(.014)	(.012)	(.011)	(.011)	(.011)
POWER	.329***	.039***	.042***	.038***	.028***
	(.007)	(.006)	(.006)	(.006)	(.006)
N	277977	277977	277977	277977	277977
R-squared	.082	.162	.169	.195	.223
Year FE	YES	YES	YES	YES	YES
Village FE	NO	YES	YES	YES	YES
Controls	NO	NO	YES	YES	YES
Linear Trends	NO	NO	NO	STATE	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1*

Figure 5B- Non-Agricultural Firm Employment (ROAD x POWER, Triple Difference)

	(1) Pooled OLS	(2) Fixed Effects	(3) Fixed Effects	(4) Fixed Effects	(5) Fixed Effects
ROAD	.205***	.034***	.024**	.041***	.026***
	(.011)	(.009)	(.01)	(.01)	(.01)
ROAD X POWER	-.151***	-.109***	-.098***	-.035***	-.024**
	(.014)	(.012)	(.012)	(.012)	(.011)
POWER	.321***	.037***	.041***	.034***	.025***
	(.007)	(.006)	(.006)	(.006)	(.006)
N	277977	277977	277977	277977	277977
R-squared	.083	.165	.172	.196	.225
Year FE	YES	YES	YES	YES	YES
Village FE	NO	YES	YES	YES	YES
Controls	NO	NO	YES	YES	YES
Linear Trends	NO	NO	NO	STATE	STATE-DISTRICT

Clustered Robust standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1*