Rural Roads and Local Agro-Firm Development: Evidence from India

Gowthami Venkateswaran, University of Illinois, Urbana-Champaign (gv4@illinois.edu)
Hemant Pullabhotla, Deakin University, Melbourne (h.pullabhotla@deakin.edu.au)
Kathy Baylis, University of California, Santa Barbara (baylis@ucsb.edu)

Selected Paper prepared for presentation at the 2023 Agricultural & Applied Economics Association Annual Meeting, Washington DC; July 23-25, 2023

Rural Roads and Local Agro-Firm Development: Evidence from India

Gowthami Venkateswaran[1]*, Hemant Pullabhotla*[2] *and Kathy Baylis*[3]

Abstract

Most of the extreme poor live in rural areas with agriculture as their major source of income. Poor infrastructure and high transaction costs severely limit markets and constrain economic development in these rural agricultural areas. By exploiting the timing of India's flagship rural road construction program, this study estimates the effects of improved transportation connectivity on the growth of agricultural industries. Results suggest that rural road construction encourages the entry of new agricultural firms. Furthermore, we find that the largest positive impacts on the growth of agricultural firms are observed in villages that have closer proximity to towns. This implies that while there is an overall positive effect, well-connected villages exhibit the most favorable outcomes.

Keywords: Transportation, Agricultural Firms, Rural Roads

[1] Department of Agriculture and Consumer Economics, University of Illinois, Urbana-Champaign (email: gv4@illinois.edu)
[2] Department of Economics, Deakin University, Melbourne (email: h.pullabhotla@deakin.edu.au)
[3] Department of Geography, University of California, Santa Barbara (email: baylis@ucsb.edu)

Rural areas are home to the majority of the world's extreme poor, with agriculture being their primary source of income (Sachs, 2014). However, due to inadequate infrastructure, smallholder farmers in these areas face significant transportation challenges and costs. The lack of access to agricultural markets exacerbates the last mile challenges in rural areas, making the delivery of technology and inputs to farmers expensive and challenging. Studies show that connectivity issues significantly hinder development and discourage agro-retailers from operating in these regions leading to restricted input access for farmers (Duflo et al., 2011; Duflo et al., 2008). Moreover, limited market access and the monopoly power of intermediaries contribute to low output prices, further burdening farmers in remote areas (Chatterjee, 2019). Beyond the constraints on input access and low output prices, farmers also encounter challenges in accessing other vital agricultural sectors that offer services such as processing, post-harvest handling and storage facilities. Hence, this study aims to investigate whether the improved transportation can effectively enhance the presence and growth of agricultural industries, including firms engaged in various agricultural support activities, processing, storage and input sales.

In this paper, we study the causal relationship between rural road rural road constructions and the growth of agricultural industries. We use village-level variation in the rollout of the world's largest rural road construction program, the *Pradhan Mantri Gram Sadak Yojana* (PMGSY), in combination with multiple rounds of census data on village-level firms and enterprises. To understand the underlying mechanisms driving the increase or decrease in firm numbers, we estimate heterogeneity effects based on village location, aiming to separate differential impacts on well-connected and remote villages. Additionally, to examine differences between the growth of agricultural and non-agricultural firms, we provide estimates for non-agricultural firms using our main specification.

We find that rural roads have substantial positive effects on the number of local agricultural firms. Following construction of PMGSY rural roads, we find an 8.6% increase in the number of agricultural firms in villages that received a road between 2001 and 2013. The estimates are precise and statistically significant. We also estimate that connecting a village with a new paved road causes a 17.3 % increase in non-agricultural firms on average. Results are robust to multiple specifications and sample definitions.

Several studies have explored the impact of rural roads and reduced transportation costs on agricultural-related outcomes. Using cross sectional survey data, Bell & Van Dillen (2014) find that PMGSY roads in Orissa, India increased paddy prices received by farmers by 5 percent. Khandker et al. (2009) find that newly placed rural roads in Bangladesh induced higher aggregate crop yields, lower input and transportation costs and higher output prices. Ali (2010) finds that a feeder road rehabilitation project in Bangladesh increased the adoption of modern inputs among farmers. Iimi et al., (2015) find that paved rural roads and bridge construction in Brazil reduced the distance from the farmers and their nearest populated areas. However, little has been studied about how improved infrastructure can influence the location of agro-dealers, agricultural processors and the location of other agro-firms critical to modern agricultural supply chains.

Increased entry of agro-firms can impact agricultural productivity through two primary pathways. Firstly, increased access to input and processing industries could address the problem of high input prices and low output prices faced by farmers. Secondly, economic geography literature suggests that agglomeration economics[4] play a critical role in the spatial concentration and development of economic activity. Local markets play a critical role in shifting production structurers to higher valuer activities and also promoting economic development (Van de Walle & Mu , 2011). In an agricultural context, this implies that the entry of a few grain storage industries may pave way for the establishment of grain processing industries such as rice mills, input stores etc., enabling farmers to enhance productivity. Hence, increased access to agricultural industries can contribute to farmer welfare in multiple dimensions.

In 2000, the Government of India launched the world's largest rural road program to fund new feeder roads to villages that did not previously have paved roads. Despite such a large investment, the impact of these roads on the rural economy, particularly on those that depend on agriculture, is not well known. Existing evidence on the agricultural outcomes of these new roads is mixed. Aggarwal (2018) finds a 2% and 3% increase in the use of hybrid seeds and fertilizers respectively among food crop cultivators using a district-level analysis. However, in a village-level analysis, Asher and Novosad (2020) find that four years after PMGSY, the main effect of the new roads is a reduction of the share of agricultural workers, with no significant changes in

[4] Agglomeration economics can be defined as an economy where large number of industries and services exist in close proximity to on another resulting in benefits from cost reductions and gains in efficiency.

agricultural yields. Analyzing the same program with a sample of remote household data, Shamdasani (2021) also finds large gains in agricultural outcomes including a significant increase in the diversification of crop portfolio, adoption of input technologies among villages where access to the non-agricultural sector remains limited despite road connectivity.

The conflicting findings regarding agricultural yields despite studying the same program may be attributed, at least in part, to differences in the granularity of data used across the studies. Shamdasani (2021) uses data from 221 villages with around 4000 households, whereas Asher and Novosad (2020) rely on coarser remotely sensed data such as NDVI and EVI, for village-level agricultural production. The variation in yield outcomes suggests the presence of treatment heterogeneity, where the effects of improved roads may vary across different contexts. Furthermore, other studies have found contrasting outcomes. For example, Chinese highways lead to a decrease in local GDP for rural areas that were newly connected to more productive urban centers (Faber, 2014). Ghani et al. (2015) show that the upgradation of a central highway network in India instigated manufacturing activity disproportionately along the road network. Therefore, while some studies find a positive impact from rural roads, it is possible that rural roads do not lead to major economic changes in rural areas other than facilitating reallocation of labor from agriculture to more productive jobs.

The benefits of infrastructural projects are often hard to measure due to the endogenous placement of roads that relate to economic, political or social factors. The comparisons of villages with and without roads will be biased if the roads are targeted to wealthier villages or poorer villages. The rule laid by the government for receiving roads was based on the population of villages; the implementation rule was supposed to target villages with thresholds of 1000 people first, followed by villages with more than 500 people. However, by working closely with the National Rural Roads Development Agent, Asher and Novosad (2020) identified that most states except a few did not follow the population threshold prioritization rules as given by the national guidelines of PMGSY. To overcome any potential bias arising from non-random treatment allocation, we exploit the timing of program roll-out across villages, estimating a model with village and time fixed effects. Village fixed effects control for all unobserved factors at the village-level that may have influenced the timing of the treatment among villages. Time fixed effects control for time variant shocks as well as constant differences between villages.

Due to challenges in interpreting the coefficient from two-way fixed effects models when treatment effects are heterogenous, we use the framework from Callaway & Sant'Anna (2020) to retrieve disaggregated causal parameters called group-time average treatment effect, i.e., the average treatment effect for village group *g* at time *t*, where a group is defined by the time period when villages first receive treatment. We also explore heterogeneity in treatment effects on sample villages in terms of connectivity to nearby towns and the percentage of agricultural lands they hold.

This paper highlights the potential of road connectivity projects to positively impact agricultural firms by reducing transportation costs, improved market access and increasing incentives for retailers to establish closer connections with farmers. The paper also presents an important potential mechanism that could explain the positive agricultural outcomes in previous studies on this PMGSY project. While further data and research are needed to estimate the effectiveness of this mechanism, our findings suggest that improved roads can play an important role in enhancing agricultural outcomes.

The paper is organized as follows. Section I presents the background and context of the road infrastructure program studied in this paper. Section II provides details on data and Section III describes the empirical strategy and model specification. Section IV presents results; Section V explores the mechanism and heterogeneity among villages. Section VI concludes.

I. Context of the Road Construction Program

The *Pradhan Mantri Gram Sadak Yojana* was launched in 2000 by the Central Government of India to provide all-weather roads to villages that are unconnected across India. At that time, around 330,000 of India's 600,000 rural villages consisted of dirt or fair-weather roads with poor drainage systems that were susceptible to flooding during monsoon rains. The objective of the program was to provide all-weather roads by constructing paved roads in eligible villages. For the village to be eligible, it had to be at least 500 meters away from an all-weather road or another village that has an all-weather road and needed to have a population of 500 (and 1000 in some cases). The threshold for people in tribal areas, deserts and mountainous regions was 250. The population for all villages were determined using the 2001 census.

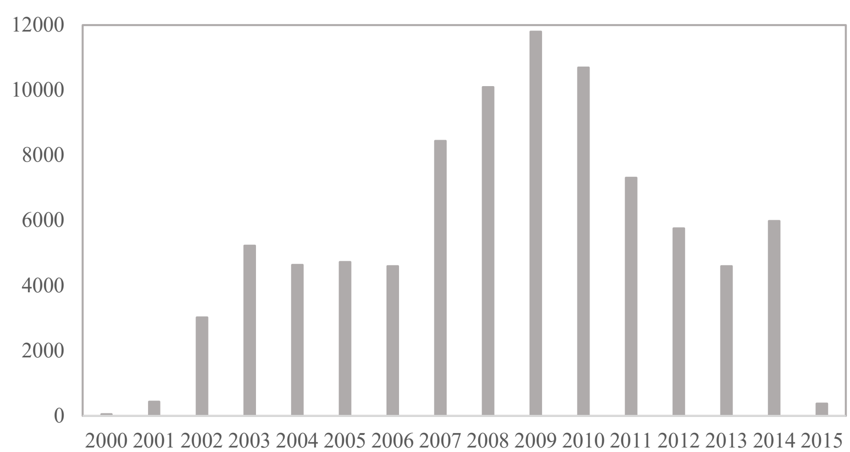
By 2013, around 96000 villages received newly paved roads under the program. Constructed roads were usually single lane roads that either connected villages with each other or to the nearest road of a higher category or other main roads. Construction was prioritized based on population categories. Villages with more than 1000 people were to be connected first, followed by those villages with 500 to 1000 people, followed by villages with 250-500 (in specific areas). These rules were applied on a state-by-state basis where if states had all larger villages connected, then they could proceed to smaller villages. A few states used only the 500 people threshold instead of first implementing the project in villages with 1000 people. Some states did not comply with the rule-based program at all. Asher and Novosad (2020) identified the complying states based on meetings with the National Rural Roads Development Agency. The states that followed the rules were: Chhattisgarh (500, 1000), Gujarat (500), Madhya Pradesh (500, 1000), Maharashtra (500), Orissa (500) and Rajasthan (500). If the state is denoted by (500, 1000), this means that the villages with 1000 people were connected first, followed by the villages with 500 people. Other criteria were used to determine the eligibility for treatment. Smaller villages could be connected earlier if they laid on the least-cost path among larger villages. Because of the population threshold, early treated villages inclined to have larger populations than later-treated villages. Though the project was federally funded, roads were implemented under the respective individual states. The project was funded by support from central government, loans from Asian Development Bank, World Bank and taxes of 0.75 INR (Indian Rupee) placed on diesel fuel. For more details on this program, see Asher and Novosad (2020).

II. Data

To estimate the changes in the agricultural market sector, the study requires data on the number of agricultural-allied industries at the village level around the time of the road program launch in 2001. For this, we use data from the Indian Economic Census (EC) surveys held in the years 1990, 1998, 2005 and 2013. The survey covers all economic establishments in India excluding only those involved in crop production, public administration and defense. The survey provides information on firm-activity, location of firm, number of people employed and a few other firm characteristics. The outcome variable used in the study is the number of agricultural firms. So we include establishments coded with agricultural nomenclature such as "Support activities to agriculture and post-harvest crop activities", "Dairy Industries", "Processing and preserving fruits and vegetables", "Wholesale of agricultural raw materials and live animals", "Manufacture of basic chemicals, fertilizer and nitrogen compounds, plastics and synthetic rubber in primary forms". By including these specific types of agricultural firms, the study aims to capture the diverse range of activities and sectors that contribute to the agricultural industry.

To identify treated villages and completion dates of roads, we use the dataset provided by Asher and Novosad (2020) in their paper. These details are also available on the official PMGSY website (http://omms.nic.in) from which the authors retrieved information until January 2015. Next, we use a dataset that combines multiple rounds of population and economic census named "Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG)" to correctly match villages across all four years used in this study despite boundary changes that occurred throughout the years (Asher et al., 2021).

Figure 1: Number of Villages that received new roads through PMGSY



Notes: This figure shows the staggered rollout of treatment among sample villages used in the study. The study period is restricted to 2013 due to data constraints.

Table 1 shows summary statistics of villages at baseline. There is a notable difference between agricultural firms and total firms. On average, a village has close to 2 agricultural firms, while the total number of firms reaches approximately 32.

Table 1: Summary Statistics at Baseline

	Mean	SD
Population (2001 Population census)	1539.2	1754.6
Number of households	281.9	335.2
Number of Primary Schools	1.4	1.2
Number of Colleges	0.0	0.1
Area of Village (Hectares)	546.0	1011.3
Agricultural Firms (1990 Economic census)	1.97	5.07
Total Firms (1990 Economic census)	31.48	65.42

Notes: The table shows means and standard deviations of village-level variables at baseline in the sample of villages that matched in all the datasets used in the study.

III. Empirical Strategy

Two-way fixed effects

Our main empirical specification is a panel fixed effects regression that exploits the timing of road placement within villages that were eligible to receive a road under the PMGSY program to estimate the impact of rural roads on the development of local agricultural firms. Under the road program settings, eligible villages received treatment at different points in time. The panel estimation exploits variation in the year of receiving a road connection. The following equation defines the two-way fixed effects estimator:

$$y_{st} = \alpha + \beta\ ROAD_{st} + \mu_s + \lambda_t + \varepsilon_{st} \qquad (1)$$

where y_{st} denotes the outcome variable, the number of agricultural-allied firms in village s at time t; $ROAD_{st}$ is an indicator variable for treatment by year t, $ROAD_{st} = 1[t \geq \text{year of receiving road}]$. μ_s and λ_t are village and year fixed effects and ε_{st} is an unobserved error term and is clustered at the village level to account for serial correlation in the outcome variable. The β coefficient estimates the impact of a new road connection on the change in the number of agricultural industries and firms at the village level. β describe outcomes relative to the period before any road construction began. All villages eligible for the program are included in the study. As a robustness check, we also run the analysis limiting the sample of villages to those that received a road at some point under the program between 2000 and 2013 i.e., all villages are treated in 2013 to provide evidence that the comparison between villages are not biased. We use the later-treated village group as a control for the earlier-treated villages. The identification assumption is that the number of agricultural firms would have followed the same path over time in villages that received a paved road after removing village and time fixed effects if the PMGSY program did not have happened. The village fixed effects control for all time-invariant observables among villages that received the program during different years and villages that did not receive the road. No additional controls are included because village-level fixed effects account for all static village characteristics. We also present robustness checks using specifications that include state-year fixed effects. State-year fixed effects control for systematic differences and enrollment across states.

Event Study

This methodology estimates the impact of the road program which is implemented only in certain villages and in different periods, i.e. staggered treatment. The villages in which the road program is not implemented or is yet to be implemented are used as counterfactuals. We look at the variation in outcomes around the occurrence of the road program compared with a baseline reference period. The model to specify an event study in this study setting is:

$$y_{st} = \alpha + \sum_{j=2}^{3} \beta_j(\text{ Lead } j)_{st} + \sum_{k=0}^{1} \gamma_k (Lag\ k)_{st} + \mu_s + \lambda_t + \varepsilon_{st} \quad (2)$$

where y denotes the number of agricultural firms in village s at time t, μ_s and λ_t are village and year fixed effects. ε_{st} is an unobserved error term. The lags and leads variables are defined as follows:

$$(\text{ Lead } j)_{st} = 1[t = \text{ Road }_s - j] \text{ for } j \in \{1,2,3\}$$
$$(\text{ Lag } k)_{st} = 1[t = \text{ Road }_s + k] \text{ for } k \in \{0,1\}$$

Hence, leads and lags are binary variables indicating that a given village was a given number of periods away from the treatment in the respective time period. From the above specification, the baseline omitted case is the first lead where j=1. This time period represents the period just before the treatment takes effect. This model clusters standard error at the village level.

IV. Results

A. Average Impacts on Agricultural Firms and Non-Agricultural Firms

Table 2 displays the point estimates of lag and lead coefficients by equation (2) along with their 95% confidence intervals, in relation to the time from receiving treatment. Column 1 and 2 show the number of agricultural and non-agricultural firms respectively as a function of the number of years before and after the implementation of the road program, while controlling for village and year fixed effects. The relative time coefficients from the one period before treatment are omitted as the reference point.

Previous studies that use PMGSY data often restrict their sample of villages to those that had no road before the program in 2001 and had a road completed during the study period (Adukia et al., 2020; Asher et al., 2020). In doing so, the identification rests on the assumption that any systematic changes observed in the villages are caused only by the road program. In our analysis, we also test our results with a similar subsample of villages for our study period. We include a spatial fixed effects specification to address spatial heterogeneity by clustering the standard error at sub-district and district levels, equivalent to a spatial error model with block weights at sub-district and district levels.[5]

[5] Due to the large sample size of around 800,000 observations, we were unable to compute the weight matrix required for spatial error models.

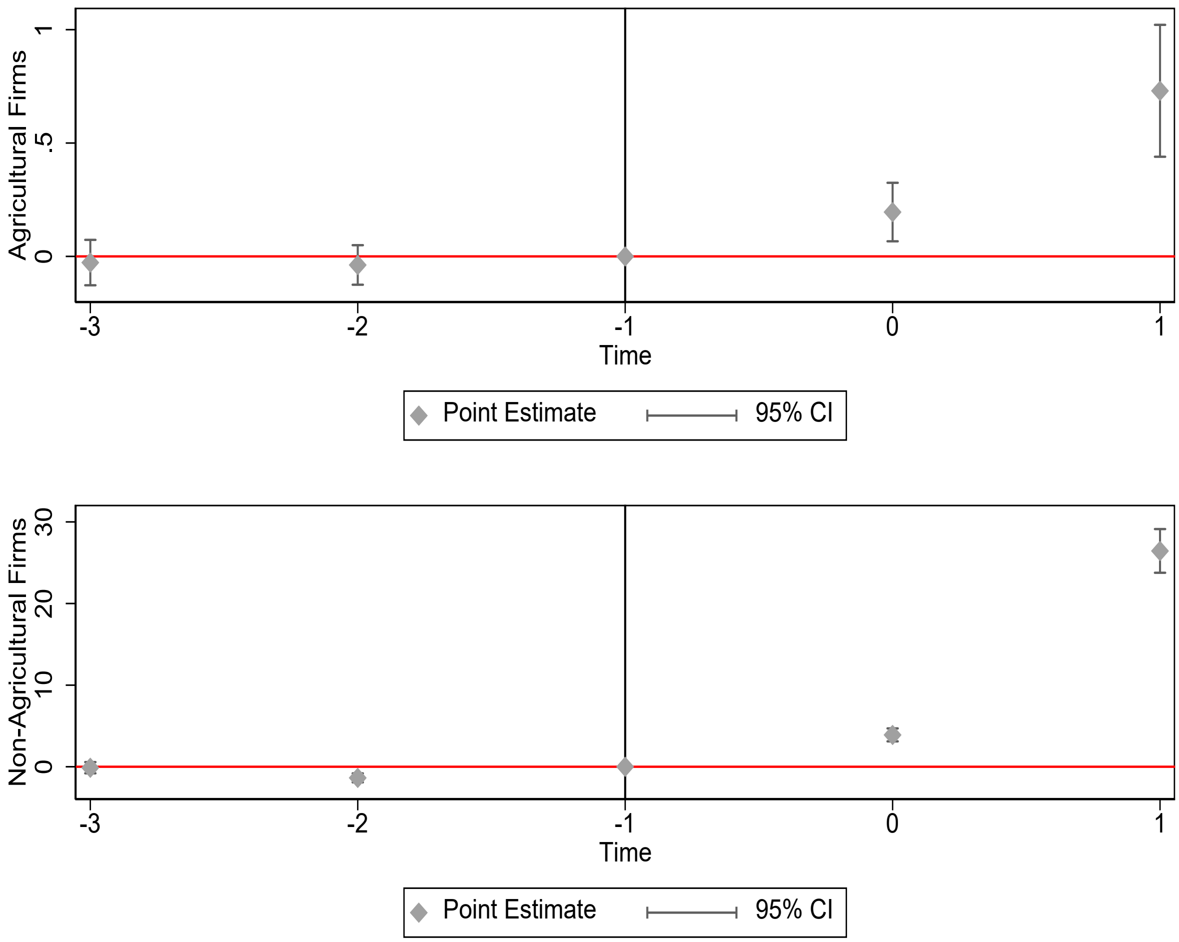
Table 2: Coefficients Estimates from Event Study

	(1) Agricultural Firms	(2) Non-Agricultural Firms
Lead 3	-0.0271	-0.118
	(0.0511)	(0.350)
Lead 2	-0.0377	-1.363^{***}
	(0.0445)	(0.280)
Lag 0	0.196^{***}	3.903^{***}
	(0.0659)	(0.406)
Lag 1	0.730^{***}	26.45^{***}
	(0.148)	(1.366)
Constant	3.097^{***}	44.84^{***}
	(0.00860)	(0.0529)
Observations	822268	822268
R square	0.445	0.762

Notes: This table displays point estimates of lag and lead coefficients from equation (2). The omitted time period is one period before receiving treatment. Column 1 reports estimates for Agricultural Firms and Column 2 reports estimates for Non-Agricultural Firms. The lag coefficients are close to zero and non-significant suggesting parallel trends. Standard errors are clustered at village level. Standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Figure 2 plots the coefficients derived from Table 2. The plots indicate that the number of agricultural firms correspond to the timing of the rural road construction. All lead coefficients are statistically insignificant in column 1, satisfying the pre-trend assumption. From the event study model, it is first evident that there is no change in the outcome variable before the treatment of roads. Second, the number of firms marginally increase in both the periods after road construction.

Figure 2: Event Study Estimates of Impact of Rural Roads on Agricultural and Non-Agricultural Firms



Notes: This figure shows coefficient estimates from a panel event study regression from equation (2) for agricultural and non-agricultural firms. Firms are regressed on a set of indicator variables that indicate the time from receiving treatment, including village and year fixed effects. Time 0 is the first period in which a road connection was present when firm data was collected. Time period t = -1 is omitted. 95% confidence intervals are displayed for each estimate. Standard errors are clustered at the village level.

To test these results with two-way fixed effects, we show estimates from equation (1) in Table 3. Table 3 shows that the estimated average effect on the number of agricultural firms is robust to a range of empirical specifications and sample definitions. Our main estimate in

Column 1 shows that villages had on average an increase of 8.6 % after receiving a completed road from the project.

Table 3: Difference-in-Differences Estimates of Impact of Rural Roads on Agricultural Firms

Agricultural Firms	(1)	(2)	(3)	(4)	(5)
New Road	0.270***	0.382***	0.270**	0.270	0.199*
	(0.0630)	(0.0652)	(0.104)	(0.165)	(0.116)
Constant	3.098***	3.090***	3.098***	3.098***	3.481***
	(0.00463)	(0.00480)	(0.00765)	(0.0121)	(0.0330)
Village F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	No	Yes	Yes	Yes
State-year F.E.	No	Yes	No	No	No
Standard Error	Village	Village	Sub-District	District	Village
Observations	822268	822268	822268	822268	211704
Dep var mean	3.118	3.118	3.118	3.118	3.538
R^2	0.445	0.452	0.445	0.445	0.487

Notes: The table reports panel estimates of the effect of rural road placement on village-level agricultural firms, estimated with equation (1). The estimates are similar to those in Table 2 with some modifications. Column 1 shows results from a village and year fixed effects model. Column 2 shows results from a state-year fixed effects model. Columns 3 and 4 are similar to Column 1 except that the standard errors are clustered at the sub-district and district level respectively. Column 5 restricts the samples to villages that received a road between the years 2000 to 2013. All specifications have a balanced panel. Columns 1, 2 and 5 have standard errors clustered at the village level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Table 4 displays the estimates from equation (1) for non-agricultural firms. In table 3 and 4, column 1 uses village and year fixed effects. In column 2, we use a model with state-year fixed effects. To account for spatial errors, we cluster standard errors at the district and subdistrict level in columns 3 and 4 respectively for a spatial fixed effects model. In column 5, the model is the same as column 1 but the sample is restricted to the villages that received a road at some point between 2000 and 2013 and use later-treated villages groups as control for villages that received treatment earlier. We test the empirical specification with this restricted sample to ensure that the villages that received new roads and those that did not are comparable. Asher, Garg and Novosad (2020) restrict their main sample to villages that received treatment at some point in their study that measures the impact of the same rural roads program as this study on forest cover. Their identification uses the assumption that within the set of villages that received treatment, there are no other systematic changes that are not caused by roads specific to villages in the years where roads were built. Our main results are comparable to the results estimated with the restricted sample.

Table 4: Difference-in-Differences Estimates of Impact of Rural Roads on Non-Agricultural Firms

Non-Agricultural Firms	(1)	(2)	(3)	(4)	(5)
New Road	7.875***	10.83***	7.875***	7.875***	3.311***
	(0.493)	(0.494)	(1.652)	(2.181)	(0.718)
Constant	44.84***	44.62***	44.84***	44.84***	52.16***
	(0.0363)	(0.0364)	(0.122)	(0.160)	(0.205)
Village F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	No	Yes	Yes	Yes
State-year F.E.	No	Yes	No	No	No
Standard Error	Village	Village	Sub-District	District	Village
Observations	822268	822268	822268	822268	211704
Dep var mean	45.41	45.41	45.41	45.41	53.10
R^2	0.761	0.775	0.761	0.761	0.770

Notes: The table reports panel estimates of the effect of rural road placement on village-level non-agricultural firms, estimated with equation (1). The estimates are similar to those in Table 2 with some modifications. Column 1 shows results from a village and year fixed effects model. Column 2 shows results from a state-year fixed effects model. Columns 3 and 4 are similar to Column 1 except that the standard errors are clustered at the sub-district and district level respectively. Column 5 restricts the samples to villages that received a road between the years 2000 to 2013. All specifications have a balanced panel. Columns 1, 2 and 5 have standard errors clustered at the village level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

B. Treatment Effect by Length of Exposure

In this section, we examine the potential variation in treatment among villages based on length of exposure to the road program. It is plausible that villages that receive treatment earlier may experience larger treatment effects compared to the later-treated villages. The status-quo for estimating treatment effect with multiple time periods and variation in treatment timing is to use a two-way fixed effects linear regression model along with leads and lags. However, recent and emerging literature point out challenges in interpreting the coefficient from two-way fixed effects modes when treatment effects are heterogenous (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2020; Athey & Imbens, 2018). To address this issue and ensure unbiased results in our main equation (1), we employ the framework developed by Callaway and Sant'Anna (2020). This framework focuses on the disaggregated causal parameter known as the group-time average treatment effect. This parameter represents the average treatment effect for a specific group *g* at a given time *t*, where a group is defined by the time period when units first receive treatment. In the standard two period setup, these parameters reduce to the average treatment effect on the treated (ATT).

Table 5: Group-Time Average Treatment Effects

Group	Time	ATT	SE	95% Confidence bands	
All	-	0.2942	0.0651	0.1666	0.4218 *
2005	1998	0.0659	0.0659	-0.1054	0.2372
2005	2005	0.5269	0.1063	0.2456	0.8081 *
2005	2013	0.8072	0.1669	0.3659	1.2485 *
2013	1998	0.0363	0.0339	-0.0534	0.1260
2013	2005	0.0548	0.0536	-0.0870	0.1966
2013	2013	0.0451	0.0816	-0.1708	0.2610

Notes: The table reports group-time average treatment effects with weights proportional to group size. The column titled Group represents the year by which the sample received treatment; column titled Time represents the year. For instance, Group 2005 in Time 2013 has a longer period of exposure to the roads when compared to Group 2013 in Time 2013.

Table 5 provides estimates from a weighted average of all group-time average treatment effects, with weights proportional to the size of each group. This type of aggregation avoids the negative weights issue that two-way fixed effects typically suffer from. Our results from the weighted average closely align with those obtained from the main model specification using two-way fixed effects. Table 5 also includes estimates of group-time average treatment effects in the column labeled ATT and the corresponding bootstrapped-based standard errors are presented in the SE column. The Group and Time columns provide information on the respective groups and time periods. Table 5 also reports pseudo group-time average treatment effects when $t < g$ i.e. pre-treatment periods for group g. Callaway and Sant'Anna (2020) state that these pre-treatment estimates can be used as a pre-test for the parallel trends assumption as long as the no-anticipation assumption holds.

Figure 3 shows separate plots for each group identified in Table 5. The first panel, labeled "Group 2005" shows estimates for the group of villages that received treatment by the end of 2005. Similarly, the "Group 2013" panel marks villages that received a road by the end of 2013. Given that Group 2005 are the early-treated villages, they have been exposed to the road infrastructure for a larger duration than the Group 2013. In each time period, the figure contains estimates of group-time average treatment effects for both groups, along with a confidence interval. The treatment impacts of roads on villages that received the road by the end of 2013 are not significant. However, the estimates of Group 2005 are statistically significant at the 10% level, suggesting that the treatment effects primarily come from early-treated villages.

Figure 3: Group-time average treatment effects

Notes: Panel Group 2005 show group-time average treatment estimates for the group of villages that received treatment by the end of 2005. Panel Group 2013 show estimates for villages that received treatment by the end of 2013 but had not received treatment by 2005. Group 2005 are exposed to the road for a longer period than the Group 2013. The treatment impacts for villages that received the road by the end of 2013 are not significant. However, the estimates of Group 2005 are statistically significant.

Although Table 5 reports treatment estimates with simple aggregation using weighted average of all group-time average treatment effects, simple aggregation tends to overweight the effect of early-treated villages because there are more of these observations during post-treatment periods. To address this potential bias, an alternative approach is to aggregate group-time effects into an event study plot. In the event study plot, the group-time average treatment effects are averaged into average treatment effects at different lengths of exposure to treatment instead of overweighting the early-treated samples.

Table 6 shows the overall average treatment effects based on dynamic aggregation. Under dynamic effects, a length of exposure equal to 0 indicates the average effect of receiving treatment across all groups at the time they first receive the treatment (instantaneous treatment effect). A length of exposure of -1 the effect one period after groups first received treatment.

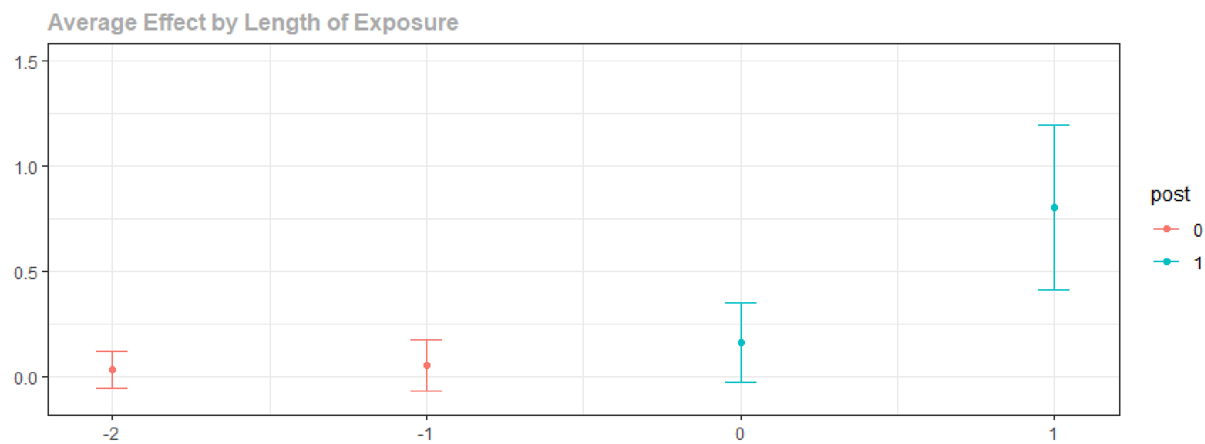
Table 6: Average treatment effects based on dynamic aggregation

Event Time	ATT	SE	95% Confidence bands	
Overall ATT				
-	0.4858	0.0852	0.3189	0.6527 *
Dynamic Effects				
-2	0.0363	0.0352	-0.0535	0.1261
-1	0.0576	0.0473	-0.0235	0.3523
0	0.1644	0.0737	-0.0235	0.3523
1	0.8072	0.1536	0.4153	1.1990 *

Notes: Overall ATT is the overall average treatment effect based on dynamic aggregation. Under dynamic effects, 0 provides the average effect of receiving treatment across groups in the time period they first receive the treatment. -1 denotes one period before villages first received treatment and so on.

Figure 4 plots the estimates from Table 6. From the estimates, parallel trends hold in pre-treatment periods. Additionally, the impact of receiving roads becomes more pronounced with a longer length of exposure. This suggests that the longer villages have been exposed to the newly constructed roads, the greater the positive effects on the number of agricultural firms.

Figure 4: Average Effect by Length of Exposure



Notes: In this figure, the x-axis denotes the length of exposure to new roads. If x-axis is 0, it represents the time period when the village first received the road. -1 corresponds to one time period before villages first received a road. Parallel trends hold in pre-treatment periods and the effect of participating in the treatment increases with time of exposure to treatment.

V. Heterogeneity

A. Remoteness of village location

In this section, we examine the impact of village location and connectivity on the growth of firms. We recognize the possibility that even if a village receives a new road, its effects may not be significant if the village is situated in a remote area with limited connections to other villages and towns. Proximity to larger villages or towns may significantly reduce transaction costs, but transportation costs may not decrease with a rural road connection if the village lacks sufficient connectivity to other towns and major roads. To test this hypothesis empirically, we use the distance of each village from its nearest town as a measure of remoteness. Based on the distance variable, we categorize villages into different groups: 0-10km, 10-20km, 20-30km, 30-40km and 40-50kms to the nearest town. The omitted category consists of villages that are farthest from their nearest towns i.e., more than 50kms away from their nearest town.

Table 7 shows the heterogenous treatment effects estimated with the main model in equation (1). Columns 1 and 2 show results for the number of Agricultural Firms. Columns 3 and 4 show results for the number of Non-Agricultural Firms. Columns 1 and 3 present the main specification, without interaction terms. Columns 2 and 4 include interaction terms between the treatment indicator for receiving new roads and the indicator variable for the distance from the village to the nearest town. The inclusion of interaction terms allows to examine if the impact of road construction varies across distance groups, providing insights into the role of remoteness as a channel through which new roads affect firm growth.

The estimated interaction effects are consistent with the fact that villages need to be well-connected to towns in order to experience increased agricultural firms as a result of road construction. Road construction has larger effects on both agricultural and non-agricultural industries in villages that are closer to towns. The largest effects of PMGSY are in villages that are within 10km of a town. As the distance from a town increases, the impact of the program decreases.

Table 7: Impact of new rural roads on Firms by distance to major town

	Agricultural Firms		Non-Agricultural Firms	
	(1)	(2)	(3)	(4)
New Road	0.270***	-0.292**	7.875***	2.091
	(0.0630)	(0.147)	(0.493)	(1.445)
New Road × (0 – 10 km)		1.072***		9.059***
		(0.202)		(1.781)
New Road × (10 – 20 km)		0.577***		6.578***
		(0.177)		(1.648)
New Road × (20 – 30 km)		0.443**		4.997***
		(0.191)		(1.788)
New Road × (30 – 40 km)		0.413*		5.682***
		(0.225)		(2.074)
New Road × (40 – 50 km)		0.315		2.448
		(0.241)		(2.290)
Village F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Panel sample	Balanced	Balanced	Balanced	Balanced
Observations	822268	822268	822268	822268
Dep var mean	3.118	3.118	45.41	45.41
R square	0.445	0.445	0.761	0.761

Notes: This table presents the heterogenous treatment estimates of agricultural and non-agricultural firms based on distance to the nearest town from each of the village. The omitted group is the sample of villages that are more than 50km away from their respective nearest towns. Columns 1 and 3 repeat the main specification from equation (1). Columns 2 and 4 include interaction terms between the treatment indicator and the distance indicator variable. The effect of the road program decreases as the distance to the nearest town increases. Standard error is clustered at village level in all columns. Standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

B. Percentage of Agricultural Land

To assess the impact of road construction on villages with varying levels of agricultural focus, we calculate the cropping intensity of the sample villages by dividing the land use for cropping by the total village area. Using this information, we create an indicator variable to identify villages with less than 50% of their land used for agriculture. The omitted category represents the villages with more than 50% of their land used for agriculture.

Table 8 shows the estimated heterogenous treatment effects of cropland areas estimated with the main model in equation (1). Columns 1 and 2 report results for the number of Agricultural Firms. Columns 3 and 4 show results for the number of Non-Agricultural Firms. Columns 1 and 3 present the main specification, without interaction terms. Columns 2 and 4 include the interaction terms between the treatment indicator for receiving new roads and the indicator variable for the percent of land used for agriculture. The findings indicate that the effects are smaller for villages with a smaller share of agricultural land. This is true for both agricultural and non-agricultural firms, implying that villages with a lower proportion of land may not experience significant growth in other types of firms as well.

Table 8: Impact of new rural roads on Firms by cropland area

	Agricultural Firms		Non-Agricultural Firms	
	(1)	(2)	(3)	(4)
New Road	0.270^{***}	0.553^{***}	7.875^{***}	10.94^{***}
	(0.0630)	(0.0922)	(0.493)	(0.720)
New Road × Less than 50% of cropland		-0.534^{***}		-5.790^{***}
		(0.115)		(0.941)
Constant	3.098^{***}	3.098^{***}	44.84^{***}	44.84^{***}
	(0.00463)	(0.00463)	(0.0363)	(0.0363)
Observations	822268	822268	822268	822268
Dep var mean	3.118	3.118	45.41	45.41
R square	0.445	0.445	0.761	0.761

Notes: This table reports the heterogenous treatment estimates of agricultural and non-agricultural firms based on the percentage of cropland. The omitted group is the sample of villages that have more than 50% of their area devoted to croplands. Columns 1 and 3 repeat the main specification from equation (1). Columns 2 and 4 include interaction terms between the treatment indicator and the percentage of cropland indicator variable. The effect of the road program decreases as the cropland area decreases. Standard error is clustered at village level in all columns. Standard errors in parentheses. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

VI. Conclusion

Investing in infrastructure in remote villages to improve connectivity and agricultural income has been a key focus for governments worldwide. Rural areas often face challenges due to incomplete and fragmented agricultural input and processing markets hindering technology adoption and agricultural production. While improved transportation programs, such as all-weather roads, are expected to facilitate agricultural production through reduced costs, empirical evidence on the impacts of roads particularly on agricultural firm development is limited.

The anticipated effect of reduced transportation costs on the number of firms in rural areas is not straightforward. On one hand, decreased transportation costs may lead to a decrease in the number of agricultural firms as it becomes cheaper to sell to locations further away. On the other hand, it could result in an agglomeration of firms in rural areas. Considering that agricultural firms in our study area are typically small scale mom-and-pop stores, the possibility of a decrease in the number of firms is lower. Additionally, new roads can create new job opportunities allowing low productivity labor to switch to higher productivity jobs within the village and opening agricultural-related markets, thus leading to overall productivity gains.

We use India's flagship rural road construction program between 2001 and 2013 where high quality feeder roads were built in around 96000 villages. These roads connected villages with nearby villages and bigger roads, potentially changing the incentives for input-retailers and agro-processors in terms of transportation costs. Our study focuses on whether this program facilitates the growth of agricultural firms. We find that the program increased the entry of new agricultural firms. Further, we find heterogeneous treatment effects based on the remoteness of villages where the largest positive effects on agricultural firms are on villages that are within 10km of the nearest town. Treatment effects diminish as the villages get further away from their respective nearest towns. We also find that effects are significantly larger for villages where more than 50% of the area is used for agriculture. This paper also highlights that treatment effects vary based on the length of exposure to the roads. Earlier-treated villages have higher treatment effects than later-treated villages in the study sample.

While various studies have identified positive agricultural outcomes resulting from new roads, the underlying mechanism behind these outcomes have not been empirically explored.

The creation of new local agricultural markets and an integrated supply chain are potential channels through which these outcomes might be affected. For instance, Shamdasani (2021) explores access to agricultural markets as a potential mechanism that might explain the positive results in technology adoption and crop diversification. However, the study uses prices of high yielding seed varieties and fertilizers as a proxy for access to agricultural markets, without finding statistically significant impacts of rural roads on these prices. Our study, on the other hand, utilizes the actual number of agricultural firms, providing a more accurate measure of increased market access. Findings from this paper suggest that improved access to local agro-firms is a significant factor driving the positive impacts of road programs on agricultural outcomes. However, it is important to note that our study does not empirically examine the specific mechanism of new agricultural firms facilitating agricultural outcomes. Further research should focus on empirically testing the role of new markets and other potential mechanisms to better understand the underlying channels through which improved transportation infrastructure affects agricultural outcomes.

In summary, our study contributes to the existing literature by providing causal evidence of the relationship between improved roads and agricultural firm growth in rural areas. This result highlights the critical importance of infrastructure investments in fostering entrepreneurship in rural areas, enhancing market access and driving economic growth within the agricultural sector.

References

Adukia, A., Asher, S., & Novosad, P. (2020). Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. *American Economic Journal: Applied Economics*, *12*(1), 348–376. https://doi.org/10.1257/app.20180036

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, *133*, 375–395. https://doi.org/https://doi.org/10.1016/j.jdeveco.2018.01.004

Ali, R. (2010). *Impact of Rural Road Improvement on High Yield Variety Technology Adoption: Evidence from Bangladesh* (Issue 95255). https://doi.org/10.22004/ag.econ.95255

Asher, S., Garg, T., & Novosad, P. (2020). The Ecological Impact of Transportation Infrastructure. *The Economic Journal*, *130*(629), 1173–1199. https://doi.org/10.1093/ej/ueaa013

Asher, S., Lunt, T., Matsuura, R., & Novosad, P. (2021). *Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform*.

Asher, S., & Novosad, P. (2020). Rural Roads and Local Economic Development. *American Economic Review*, *110*(3), 797–823. https://doi.org/10.1257/aer.20180268

Athey, S., & Imbens, G. W. (2018). *Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption* (Issue 24963). https://ideas.repec.org/p/nbr/nberwo/24963.html

Bell, C., & Van Dillen, S. (2014). How does India's rural roads program affect the grassroots? Findings from a survey in upland Orissa. *Land Economics*, *90*(2), 372–394. https://doi.org/10.3368/le.90.2.372

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 1–37. https://doi.org/10.1016/j.jeconom.2020.12.001

Chatterjee, S. (2019). Market Power and Spatial Competition in Rural India. *Mimeo*, *January*. http://www.princeton.edu/

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with

Heterogeneous Treatment Effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

Duflo, E., Kremer, M., & Robinson, J. (2008). How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. *American Economic Review*, *98*(2), 482–488. https://doi.org/10.1257/aer.98.2.482

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review*, *101*(6), 2350–2390. https://doi.org/10.1257/aer.101.6.2350

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, *81*(3), 1046–1070. https://doi.org/10.1093/restud/rdu010

Ghani, E., Goswami, A. G., & Kerr, W. R. (2015). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*, *126*(591), 317–357. https://doi.org/10.1111/ecoj.12207

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. https://doi.org/https://doi.org/10.1016/j.jeconom.2021.03.014

Iimi, A., Lancelot, E. R., Manelici, I., & Ogita, S. (2015). *Social and economic impacts of rural road improvements in the state of Tocantins, Brazil* (Issue 7249). https://ideas.repec.org/p/wbk/wbrwps/7249.html

Khandker, S. R., Bakht, Z., & Koolwal, G. B. (2009). The poverty impact of Rural roads: Evidence from Bangladesh. *Economic Development and Cultural Change*, *57*(4), 685–722. https://doi.org/10.1086/598765

Sachs, J. (2014). Ending Extreme Poverty. In *The Age of Sustainable Development*. http://www.worldbank.org/en/publication/global-monitoring-report/report-card/twin-goals/ending-extreme-poverty

Shamdasani, Y. (2021). Rural road infrastructure & agricultural production: Evidence from India. *Journal of Development Economics*, *152*, 102686. https://doi.org/https://doi.org/10.1016/j.jdeveco.2021.102686

Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.09.006

van de Walle, D., & Mu, R. (2011). Rural Roads and Local Market Development in Vietnam. *The Journal of Development Studies*, *47*, 709–734. https://doi.org/10.1080/00220381003599436