Rural Road Infrastructure & Agricultural Production: Evidence from India*

Yogita Shamdasani†

December 23, 2016

Job Market Paper

Click here for the latest version

Abstract

This paper estimates the role of improvements in transport infrastructure on households' production decisions in agriculture. The Central Government of India launched a large-scale rural road-building program in 2000, targeting villages that lacked any single all-weather connectivity. Strict guidelines governed eligibility and timing of program road provision. I exploit the precise timing of road construction as a source of exogenous variation in connectivity using a household-level panel in a difference-in-differences framework. I find that households who gain access to improved rural road infrastructure diversify their crop portfolio – they begin to cultivate higher return, non-cereal hybrid crops. Households also increase take up of complementary productive inputs and intensify labor hiring. I find that households subsequently enter into the sales of farm output, indicating a transition from subsistence to market-oriented farming. Evidence from a field survey suggests that these effects operate through an increase in mobility of agricultural workers across connected village labor markets. These findings emphasize the substantial barrier to productive investments in agriculture generated by poor rural road connectivity that hampers the integration of labor markets across space.

*I am deeply grateful to Emily Breza, Supreet Kaur and Eric Verhoogen for feedback and encouragement. I thank Suresh Naidu, Kiki Pop-Eleches and participants of the Columbia Development Colloquium for helpful discussions. Arnesh Chowdhury and Santosh Kumar Sahoo provided excellent research assistance in the field. All errors are my own.

†Department of Economics, Columbia University. Email: yjs2112@columbia.edu.

1 Introduction

The provision of transport infrastructure has gained prominence in economic development policy in recent decades, with substantial investments led by both local governments and international organizations.[1] Consequently, this has given rise to a growing literature that seeks to understand the causal effects of transportation infrastructure on economic outcomes.[2] A majority of these works have focused on the effects of large-scale infrastructure systems - highways and railroads that connect distant markets across regions - on the spatial distribution of economic activity and on aggregate macro outcomes such as GDP growth. On the other hand, the effects of small-scale infrastructure systems such as rural roads, which are essential for spurring local growth and development, are less well understood.

In this paper, I exploit a natural experiment in a large-scale rural road building program to estimate the causal effects of improvements in rural connectivity on households' production decisions in agriculture. Using a combination of program administrative records and a comprehensive panel data set that is well-suited to study agricultural choices at the household level, I find that improved rural road infrastructure led to diversification of crop portfolios, take-up of productive agricultural inputs and commercialization of farm output. Further, I present suggestive evidence using a survey I conduct in the field for these effects operating through an increase in mobility of agricultural workers across connected labor markets.

Understanding the effects of improvements in rural connectivity on agricultural production is first order given that agriculture remains to be the primary source of income for rural households. Despite important innovations in recent decades that have led to significant increases in agricultural

[1]For example, the Indian government pledged USD 33 billion to transportation infrastructure investments in the recent 2016-2017 Union Budget, and the World Bank allocated 14% of total lending to transportation infrastructure projects in 2016.

[2]Estimating the causal effects of infrastructure poses an empirical challenge as the placement of transport infrastructure is non-random and often driven by political and economic factors. For example, Burgess et al. (2015) find strong evidence of ethnic favoritism in the provision of paved roads across the post-independence period in Kenya; districts that shared the same ethnicity as the president had double the amount of expenditure on roads, and five times the length of paved roads. The literature addresses this issue of endogenous placement by exploiting geographical variation in the placement of large-scale infrastructure such as highway and railroad systems along trunk routes. For example, Banerjee et al. (2012) exploit the fact that railroad networks built in the 19th and 20th century in China connected historical cities. Faber (2014) uses a straight-line instrument based on the construction of least cost path spanning tree networks to address non-random placement of trunk highways in China. Donaldson (forthcoming) argues that military motives trumped economic arguments in the design and construction of railroad networks by the British government in colonial India.

productivity, such as the introduction of high-yielding varieties during the Green Revolution,[3] traditional practices still dominate and take up of improved technologies by rural households is far from universal. While poor transport infrastructure has been highlighted as a potentially important barrier to the adoption of improved technologies in agriculture (Suri, 2011), there is no empirical evidence to date that documents how households' production decisions in agriculture respond to improvements in rural road infrastructure.[4]

The program that I examine - India's Prime Minister's Rural Road Building Program - targeted rural, *unconnected* villages i.e. villages that lacked any single, all-weather connectivity. Under the program, a small hard-topped road was constructed with the goal of enabling access to the closest market center,[5] which served as a rural business hub. This was achieved through the construction of a road that linked the village to either the closest village with an all-weather road, the closest all-weather road, the block headquarters or directly to the market center. Program roads thus enabled connectivity to the broader rural road network, and subsequently, to surrounding villages, agricultural markets and towns. Further, program roads were equipped with the necessary cross-drainage structures so as to remain operable during all weathers; this was especially relevant as the peak season in agriculture coincides with the onset of monsoon rains, and inadequate drainage often led to water-logged, and subsequently, inoperable roads.

The order in which roads were built under this program was determined by an unconnected village's population relative to that of other unconnected villages in the state. In every state, a priority ranking was generated among unconnected villages based on village population size, with larger villages ranked higher. Program roads were then constructed following this priority ranking. As such, within a state, a larger unconnected village received a program road before a smaller unconnected village. I exploit this exogenous variation in timing of road construction under the program in a difference-in-differences framework, comparing the evolution of outcomes for households in villages that received a program road to households in villages that had yet to receive a program

[3] These new varieties were adopted in tandem with an expansion of irrigation infrastructure, modernization of management techniques, as well as increased utilization of synthetic fertilizers and pesticides.

[4] The existing literature has examined the effects of rural roads on the movement of workers *out* of agriculture (Asher and Novosad, 2016), on agricultural land values (Jacoby, 2000; Shrestha, 2015; Gonzalez-Navarro and Quintana-Domeque, 2016) and on market prices of local crops (Casaburi et al., 2013).

[5] Market centers were identified by the program as "centers of activities for marketing agricultural produce and inputs, servicing of agricultural implements, health, higher education, postal, banking services etc." (PMGSY Operations Manual)

road.

I also take advantage of the fact that in some cases, program-ineligible villages were *incidentally* treated as they were in the proximity of unconnected villages, and a new road was built connecting them. Despite being ineligible for the program, these villages benefitted from an increase in connectivity to the broader rural road network. I construct the relevant control group for these incidentally treated villages, which consists of program-ineligible villages with unconnected villages in their proximity that had yet to be treated under the program, in ArcGIS using a database spanning the universe of geocoded villages in India. Estimating this spillover effect on incidental villages has numerous important advantages: first, it abstracts from program rules and subsequently, is not susceptible to any bias that could arise from deviations in program rules; second, it allows me to go beyond estimating the effects of the program on the targeted program-eligible population; and third, it can be used to rule out numerous channels that could be underlying the effects on agricultural production.

Prior to estimating the effects on agricultural production, I first account for potential changes in composition that may arise if improvements in road infrastructure induce the movement of workers out of agriculture. Asher and Novosad (2016) examine the same program and find that rural road construction led to a movement of households out of agriculture into wage labor, with effects being strongest in villages close to major cities, suggesting the importance of access to external labor markets. I document a similar movement of households out of agriculture in my sample, and I show that this effect is driven entirely by households with access to the non-agricultural sector proxied for by distance to the closest town. Subsequently, in order to obtain unbiased estimates of the effects of improved rural road infrastructure on households' production decisions in agriculture, I focus my analysis on households in remote villages far from towns, where I observe no movement out of agriculture.

I find that households in remote villages who gain access to improved rural road connectivity diversify their crop portfolio - in addition to cultivating staple cereal grains, they begin cultivating higher return, non-cereal hybrid crops that are typically sold in markets. I find a 55% increase in the share of total land cultivated under non-cereals. Households also significantly increase take up of productive agricultural technologies such as high-yielding variety seeds, chemical fertilizers and irrigation, and intensify hiring of agricultural workers. Lastly, I observe significant entry into sales of farm output, with a 24 percentage points increase in households selling high-yielding variety crops. I find significant heterogeneity by households' landholdings, with effects largely driven by small-scale cultivators.

I provide evidence in support of one channel underlying these effects: an increase in mobility of agricultural workers across connected village labor markets, which subsequently enabled households to engage in relative more input-intensive agricultural processes. The cultivation of commercial hybrid crops is both labor- and capital-intensive, relative to traditional cereals. I find a substantial 52% increase in casual agricultural hiring, consistent with improved infrastructure facilitating the movement of agricultural laborers across connected village labor markets. This increase in worker mobility is also reflected in a 35% increase in bicycle ownership among households who gain access to improved rural road connectivity. I provide evidence to rule out alternative channels such as a reduction in input prices and an increase in access to input and output markets. Finally, I provide suggestive evidence from a field survey conducted across 18 remote villages in 3 districts in rural Odisha, India, that suggests that the reliability of agricultural labor is a big concern among cultivators.

This paper adds to a large literature that seeks to understand barriers to productive investments in agriculture.[6] Low take up of technologies that increase agricultural productivity has been attributed to high costs, credit constraints, limited access to information, aversion to risk, and behavioral biases, to name a few.[7] This has given rise to recent micro studies that attempt to relax these barriers through privately-provided interventions such as fertilizer subsidies (Duflo et al., 2011), cash grants and rainfall index insurance (Cole et al., 2013; Karlan et al., 2014), and agricultural advice services (Cole and Fernando, 2016), with mixed success. While the role of infrastructure as a barrier in take up has been discussed in this literature - Suri (2011) attributes the high costs of

[6] See Feder et al. (1985) for a survey of barriers to technology adoption in agriculture in low-income countries, and Foster and Rosenzweig (2010) for a more recent survey on barriers to technology adoption generalized to a range of technologies and settings.

[7] Suri (2011) demonstrates that low rates of adoption of technologies such as hybrid maize are correlated with high costs of acquiring the technologies due to poor infrastructure, and Porteous (2016) shows using a counterfactual estimation that agricultural technology adoption only increases farmers' income when trade costs are low. Foster and Rosenzweig (1995) document that information frictions serve as a barrier in adoption of hybrid seeds, and explore the role of learning by doing and learning from others in alleviating this barrier. Munshi (2004), Bandiera and Rasul (2006) and Conley and Udry (2010) further explore the role of social learning in the diffusion of new technologies.

acquiring new technologies to high transport costs that accrue due to poor infrastructure - there exists no empirical evidence to date that estimates the causal effects of transport infrastructure on subsequent productive investments. I add to this literature by documenting that improved rural road connectivity leads to significant take up of improved agricultural technologies, with these effects operating through increased integration of village labor markets across space. In doing so, I examine a unique setting - a large-scale intervention that is publicly provided by the Indian government - which contrasts with the privately provided micro interventions described above.

This paper also contributes to a growing empirical literature that estimates the causal effects of transportation infrastructure investments on economic outcomes. A bulk of these studies have evaluated large-scale transport infrastructure systems that reduce transport costs of goods and travel times of individuals across regions, focusing on aggregate outcomes such as GDP levels, real income, and trade volumes.[8] The transport infrastructure that I examine in this paper is a much smaller system of local roads that link villages to other nearby villages or to the local road network. Such a system would be expected to generate effects on *local* economic growth and development, in contrast to the macro effects that are observed with large-scale infrastructure systems. I study India's Prime Minister's Rural Road Building Program as a natural experiment to provide empirical evidence on the effects of improved rural road connectivity on household-level production decisions.

Finally, this paper makes several contributions to a smaller literature on local intra-regional roads. First, I use a household-level panel that is well-suited to studying agricultural outcomes in order to examine how households' production decisions in agriculture evolve over time in response to improved infrastructure. Previous studies examining the impacts of local infrastructure provision have used cross-sectional or aggregated panel data and have focused on outcomes such as land values (Jacoby, 2000; Shrestha, 2015; Gonzalez-Navarro and Quintana-Domeque, 2016), market prices of crops (Casaburi et al., 2013), wage labor market participation (Asher and Novosad, 2016), em-

[8]For example, Banerjee et al. (2012) estimates the effect of highways in China on per capita GDP levels and growth; Allen and Atkin (2016) examine the effects of expanding India's highway network on trade costs and subsequently, farmers' revenue volatility and portfolio choice; Jedwab and Moradi (2016) estimate the effect of colonial railroads in Ghana on the spatial distribution and aggregate level of economic activity; Donaldson (forthcoming) estimates the impact of railroads in India on agricultural trade costs, interregional price gaps, trade volumes and real income; and Donaldson and Hornbeck (2016) estimate the effect of the expansion of railroads in the US on agricultural land values.

ployment in the manufacturing sector (Gertler et al., 2016) and farmland cultivated under hybrids at the district level (Aggarwal, 2015). Second, I explore distributional effects across the population, which bears important implications for policy. Third, I conduct a survey among cultivators in rural Odisha, India in order to provide suggestive evidence for a labor mobility channel underlying the main effects.

The body of the paper proceeds as follows. Section 2 discusses the Indian setting and provides institutional detail on the road-building program. Section 3 describes the data and presents descriptive statistics. Section 4 details the empirical strategy. Section 5 presents the main estimation results. Section 6 discusses potential channels and presents survey evidence. Section 7 presents robustness checks, and Section 8 concludes.

2 Setting

This section describes the Indian context and provides institutional detail on the road-building program. Program rules generated exogenous variation in the timing of road construction, which I exploit to overcome the challenge of endogenous road placement. Program administration utilized two key features of a village - baseline connectivity status and total population size - in determining eligibility for a program road as well as the precise timing of program road construction.

In 2000, an estimated 330,000 of India's 825,000 rural villages lacked any all-weather road access.[9] Existing road infrastructure in these villages consisted of dirt or fair-weather roads - often filled with potholes and equipped with poor drainage systems - rendering them prone to water logging once the monsoon rains set in. This translated to approximately 300 million people living in villages characterized by low spatial mobility, with little/no motorized traffic volume and the movement of workers and goods mostly done on foot, by bicycle, or using hand-held carts.

In response to this lack of connectivity, the Central Government of India launched the *Pradhan Mantri Gram Sadak Yojana*, or the Prime Minister's Rural Road Building Program, hereafter

[9]To put this in a global context, the Rural Access Index developed at World Bank estimates that over one billion people, 98% of them in developing countries, live more than 2 kilometers away from the nearest all-weather road. Improving access and mobility through infrastructure thus remains a key priority in economic development policy not only in India, but across many other developing nations.

PMGSY, in December 2000, with the primary objective of providing single, all-weather connectivity in targeted villages. PMGSY was launched as a 100% Centrally-sponsored scheme, with the Central Government providing full funding and State Governments managing project implementation. Figure 1 illustrates the scale at which the program was implemented. Over the first decade of the program, an average of 22,000 kilometers of roads were built annually. With 480,000 kilometers of rural roads built to date,[10] PMGSY has doubled the size of the existing paved road network in rural India.

PMGSY envisaged the provision of *single, all-weather*[11] connectivity in all rural villages across India.[12] Any village with a pre-existing all-weather road within 500 meters of its boundaries was classified as connected, and subsequently, program-ineligible. Only unconnected villages - villages located at least 500 meters away from an all-weather road or from another village with an all-weather road - were deemed eligible for program roads. Program roads were constructed so as to enable access to the closest market center, identified by the program as the "center of activities for marketing agricultural produce and inputs, servicing of agricultural implements, health, higher education, postal, banking services etc."

PMGSY was specific about the categories of rural roads that were eligible for construction under the program and the possible linkages that could be generated by these roads. All program roads had to fall under one of the two lowest categories of rural roads: they could be either Village Roads (VR) that connected villages with each other or to the nearest road of a higher category, or Other District Roads (ODR) that connected villages to other main roads, the block headquarters, or the market center. All higher category rural roads (Major District Roads, State Highways and National Highways) were strictly excluded under the program. Program roads thus linked unconnected villages to either the closest village with an all-weather road, the closest all-weather road, the market center or the block headquarters. (PMGSY Manual for the Preparation of District Rural Road Plans)

[10] September 2016 figures. The program is still ongoing, and is currently in Phase XI of implementation. The program has an ambitious goal of completing road construction in the remaining 65,000 unconnected villages by 2019 (India Budget 2016).

[11] By definition, an all-weather road is a road that is equipped with the necessary cross-drainage structures that allow it to be operable during all weathers throughout the year. (PMGSY Manual for the Preparation of District Rural Road Plans)

[12] Rural and urban boundaries are well-defined in the Indian Census.

Implementation of PMGSY proceeded as follows: every state was required to draw up a priority listing in which all unconnected villages within the state were ranked by village population size;[13] a village with a larger population ranked higher. This priority listing determined the order in which roads were to be constructed.[14] Based on the amount of funding made available each year, the State Level Standing Committee then shortlisted the planned road works from this priority listing.[15]

3 Data

This section describes the datasets I use in my empirical analysis. The data is a combination of a panel survey of rural households, the administrative PMGSY database, the Indian Population Census and a database containing the universe of geocoded natural villages in India. This highly disaggregated panel dataset allows me to estimate the causal effects of improved rural road infrastructure on production decisions in agriculture.

REDS Data. I use village- and household-level surveys from two rounds of the Rural Economic & Demographic Survey (REDS), administered by the National Council of Applied Economic Research (NCAER).[16] This nationally representative survey of rural households in India spans 242 villages spread across 100 districts in 17 major states, as mapped in Figure 2. I use the two most recent rounds - rounds 4 and 5, which correspond to years 1999 and 2006 respectively - to construct a balanced household-level panel. I restrict my sample to households that were observed in both periods.[17] Further, some household units split over time, resulting in redistribution of 1999 household

[13] As recorded in the 2001 Population Census

[14] The initial goal of PMGSY was for all unconnected villages with a population of 1000 persons and above to be covered by 2003 and for all unconnected villages with a population of 500 persons and above to be covered by 2007. For certain Hill States (North-East, Sikkim, Himachal Pradesh, Jammu & Kashmir, Uttaranchal), as well as designated Desert and Tribal areas, the program prioritized unconnected villages with a population of 250 persons and above. Given that the REDS dataset used in this analysis only covers the 17 major states, the 250 persons cutoff is not relevant to my empirical setting.

[15] In practice, program implementation first entailed having all districts create a District Rural Road Plan (DRRP) - a complete mapping of all existing and planned roads that would provide connectivity to every village within the district - along with an accompanying list of all unconnected villages. Once finalized by the District Panchayat (head), the DRRP and village listing were then submitted to the State Level Standing Committee. This plan constituted the extent of preparation of projects under PMGSY at the district-level. At the state level, a priority ranking of villages *across* all districts was then generated.

[16] I thank Andrew Foster and NCAER for sharing the REDS secure data files with village identifiers.

[17] There is attrition across each round of REDS. In order to maintain representativeness, households were added to the panel over time. In Section 7.4, I test for differential attrition across treatment and control villages.

members across multiple households in 2006. To maintain balance, I aggregate all 2006 households back to the original 1999 household unit level.[18] All observations are at the household-year level. Finally, given that NCAER over samples certain types of households, I weigh all observations using sampling weights provided by NCAER.

I use the following REDS survey modules for my analysis - households' landholdings, agricultural material inputs and labor use (hired and family), agricultural outputs, crop sales, crop revenues and profits, and households' labor supply. Table 1 summarizes key variables for all households in the balanced REDS panel. The average household size is 6. 79% of these households are Hindu, and 69% of households own some land. 59% of households are cultivating land in 1999; the average amount of land cultivated is 2.4 acres. 47% of households cultivate cereal crops, and a smaller 33% cultivate non-cereals. In terms of farm input investments, 25%, 27% and 56% of households purchase irrigation, high-yielding varieties (HYVs) and chemical fertilizers respectively. Approximately 30% of households engage in the casual agricultural and non-agricultural labor markets, working an average of 96 man-days in agriculture and 85 man-days in non-agriculture.

OMMS Data. I download administrative records of PMGSY from the Online Management & Monitoring System (OMMS), an online reporting system used by PMGSY administration to track and monitor road works.[19] The administrative records contain the following information at the village level: connectivity status and population size at baseline (which determine program eligibility), dates at which the program road was sanctioned and a work order was issued, costs associated with the program road, length of the program road as well as information on other villages connected by the program road.

Census of India. I download Primary Census Abstract data from the 1991, 2001 and 2011 Indian population censuses; I use this data to validate my parallel trends assumption in Section 7.1. This village-level data contains demographic variables such as population, gender and caste ratios, literacy rates, as well as a breakdown of the population across broad employment categories.

India Place Finder. I download the Hamlets database, which contains the universe of geocoded nat-

[18] I follow Munshi & Rosenzweig (2016), who perform a similar aggregation of households when creating a balanced panel using the 1982 and 1999 REDS rounds.

[19] This can be publicly accessed at http://omms.nic.in.

ural villages in India, along with latitude and longitude coordinates of the REDS villages from India Place Finder[20] to construct my control groups using ArcGIS software. I describe this construction in detail in Section 4.1.

I match across the multiple data sources outlined above using state, district and village-level identifiers.

4 Empirical Strategy

This section describes my empirical strategy. In Section 4.1, I detail the construction of treatment and control groups in my sample in order to estimate the causal effects of improved rural road connectivity on agricultural production. In Section 4.2, I specify the difference-in-differences framework. In Section 4.3, I account for compositional effects resulting from the movement of workers out of agriculture.

4.1 Defining Treatment & Control Groups

In constructing treatment and control groups, I exploit PMGSY rules outlined in Section 2 that governed the timing of road construction under the program. Road construction in an unconnected village was determined by its rank relative to that of other unconnected villages in the state, with ranks generated using village population size. As such, within a state, larger unconnected villages received program roads before smaller unconnected villages. The rollout of program roads based on this rank ordering generated exogenous variation in the timing of road construction.

Figure 3 summarizes the timeline. I observe households in Round 4 of the REDS survey conducted in 1999, prior to the start of PMGSY. The program launched in December 2000, and road construction commenced shortly after in 2001. I then observe households again in 2006, when Round 5 of the REDS survey was conducted. During this period, road construction was restricted to unconnected villages with a population greater than 500, hereafter referred to as program-eligible villages. By

[20] This can be publicly accessed at http://india.csis.u-tokyo.ac.jp/, courtesy of the Mizushima Laboratory, Department of Oriental History, Graduate School of Humanities and Sociology, The University of Tokyo.

2006, some of the larger program-eligible villages had received program roads, while the smaller program-eligible villages had not.

I exploit two types of connections made under the program. The first type of connection - the *direct* connection - exploits variation in the timing at which program-eligible villages were exposed to new roads under PMGSY. The second type of connection - the *incidental* connection - exploits the fact that some program-ineligible villages were also exposed to new roads under PMGSY. This occurred in cases where program-eligible villages received a new road under PMGSY, and the program road linked them to a nearby program-ineligible village with an all-weather road.[21] This nearby village thus gained an additional road linkage through the program, despite being classified as ineligible.

Exploiting the incidental connections made under the program has numerous important advantages. First, incidental connections abstract from program rules that determine eligibility and timing of road provision, which I exploit to estimate the effects of direct connections. Thus, any deviations in project implementation from program rules, for example, as a political favor, could bias estimates of the direct connections, but the incidental connections would not be susceptible to this bias. Second, the inclusion of incidental connections allows for the estimation of spillover effects in addition to direct effects of PMGSY roads. Despite having a pre-existing all-weather road, there can be substantial gains in agriculture from an expansion of the rural road network. For example, additional road connectivity to a previously unconnected village can potentially lead to: (i) access to new markets to hire farm labor; (ii) access to new markets to sell farm output; (iii) access to traders who travel to villages to procure farm output; and (iv) access to extension workers who travel to villages to promote/sell farm inputs. Third, it informs the potential channels underlying the effects on agricultural production as it allows me to rule out several classes of channels that might be at play.

To identify treatment and control villages under the direct connection, I match the REDS villages to OMMS administrative records so as to determine program eligibility and timing of road construction in program-eligible villages. Program-eligible villages that receive PMGSY roads prior to 2006 are

[21] As detailed in Section 2, roads built under the program linked an eligible village to the closest: (i) village with an all-weather road, (ii) all-weather road, (iii) market center or (iv) block headquarters.

classified as treated and program-eligible villages that receive PMGSY roads after 2006 are classified as control.

To identify treatment and control villages under the incidental connection, I need to be able to spatially track the REDS villages as well as identify other Census villages in their proximity.[22] First, I geo-reference all the REDS villages by downloading longitude and latitude coordinates from India Place Finder. Second, I identify all Census villages within a five kilometer radius[23] of each program-ineligible REDS village using a database containing the universe of geocoded villages in India in ArcGIS software.[24] Third, I match these Census villages to the OMMS database in order to determine program eligibility as well as timing of road construction. Program-ineligible villages with a program-eligible village in its proximity that receives a PMGSY road prior to 2006 are classified as treated.

Once all treatment villages under the incidental connection are identified, the next step is to construct the relevant counterfactual control group. Using all program-ineligible villages that did not receive this incidental treatment would appear to be a natural control group; however, this group could easily violate the parallel trends assumption. For example, a village within a densely connected area where all villages are connected could possibly trend differentially from a village within a more sparsely connected area, where some villages are connected, but there exists other unconnected, eligible villages in their proximity. Instead, I construct a control group for the incidentally treated that is analogous to the control group for the directly treated. This group consists of all program-ineligible villages with program-eligible villages in their proximity that receive PMGSY roads after 2006.

4.2 Empirical Specification

I estimate the effects of improved rural road infrastructure on a set of household-level outcomes in a difference-in-differences framework, where I compare the evolution of outcomes for households

[22] I am unable to directly observe the local road network, so I employ this method in identifying treatment and control villages.

[23] The mean PMGSY road length is 4.2 km. Results are robust to different specifications for this radius, and are available upon request.

[24] I thank Jeremiah Trinidad-Christensen at the Columbia Digital Social Science Center for sharing the Census State and District Boundaries shape files, as well as for his help with ArcGIS.

in villages that have received a program road to households in villages that have yet to receive a program road.

Direct Effect. The most basic estimating equation restricts the sample to program-eligible villages and takes the form:

$$y_{ivst} = \beta \, DirectTreat_v * Post + \rho_i + \gamma_{st} + \delta \, \mathbf{X_v} * Post + \epsilon_{ivst} \qquad (1)$$

where y_{ivst} are outcome variables for household i in village v and state s at time t; $DirectTreat_v$ is an indicator that equals 1 if program-eligible village v receives a program road before 2006; $Post$ is an indicator for the year 2006; and β is the coefficient of interest. Any time-invariant or household characteristics are absorbed by the household fixed effects ρ_i, and any annual shocks that are common across villages in a state are captured by the state-year fixed effects, γ_{st}. Any trends correlated with village population are picked up by the baseline number of households in village v and its square $\mathbf{X_v}$ interacted with Post. Observations are weighted using 1999 sample weights. To allow for serial correlation of ϵ_{ivst} within villages over time, I adjust the standard errors by clustering at the village level.

In restricting the sample to program-eligible villages, I drop observations from all program-ineligible villages, which are useful in estimating state-year fixed effects. To improve statistical power, I augment Equation 1 to include this group:

$$y_{ivst} = \beta \, DirectTreat_v * Post + \alpha \, Ineligible_v * Post + \rho_i + \gamma_{st} + \delta \, \mathbf{X_v} * Post + \epsilon_{ivst} \qquad (2)$$

$Ineligible_v$ is an indicator that equals 1 if village v was deemed ineligible for a program road. The coefficient of interest β is still estimated off the sample of program-eligible villages, as the $Ineligible_v * Post$ interaction term absorbs changes in outcomes for the program-ineligible villages in the post period. I use Equation 2 as my main specification when estimating the direct effect of road construction; results are robust to restricting the analysis to program-eligible villages in Equation 1.

Interpreting the coefficient of interest β as the causal effect of access to improved rural road infrastructure relies on the assumption that within a state, the construction of program roads is not correlated with time-varying village characteristics that affect outcomes through channels other than the program road. I test the validity of this assumption in Section 7.

Table 2 presents baseline village-level characteristics of all program-eligible villages, by treatment status. I look at characteristics such as village size, proximity to key infrastructure, aggregate agricultural activity, harvest prices as well as prevailing wages in the casual labor market. Consistent with the program rules, I observe that the mean population size is greater in Treatment villages than Control villages, though this difference is not significant (p-value of difference 0.677). Treatment villages are also relatively further away from key infrastructure such as bus stops, banks, wholesale markets and towns; these differences are not significant. Agricultural wages appear to be significantly higher in treatment villages at baseline.

Incidental Effect. The most basic estimating equation restricts the sample to the incidentally treated and control villages, which I hereafter refer to as the incidental villages, and takes the form:

$$y_{ivst} = \beta\; IncidentalTreat_v * Post + \rho_i + \gamma_{st} + \delta\; \mathbf{X_v} * Post + \epsilon_{ivst} \tag{3}$$

$IncidentalTreat_v$ is an indicator that equals 1 if program-ineligible village v receives a program road before 2006. To improve statistical power, I augment Equation 3 to include all non-incidental villages which help in the estimation of state-year fixed effects. The estimating equation takes the form:

$$y_{ivst} = \beta\; IncidentalTreat_v * Post + \alpha\; NonInc_v * Post + \rho_i + \gamma_{st} + \delta\; \mathbf{X_v} * Post + \epsilon_{ivst} \tag{4}$$

$NonInc_v$ is an indicator that equals 1 if village v was classified as non-incidental. The coefficient of interest β is still estimated off the sample of incidental villages as the $NonInc_v * Post$ interaction term absorbs changes in outcomes for the non-incidental villages in the post period. I use Equation 4 as my main specification when estimating the incidental effect of road construction; results are robust to restricting the analysis to incidental villages in Equation 3.

Table 3 presents baseline village-level characteristics of all incidental villages, by treatment status. I look at the same set of outcomes as in Table 2; none of these outcomes are significantly different across Treatment and Control villages.

Pooled Effect. To maximize power, I examine the pooled effect of treatment across the direct and incidental connections for several key outcomes. The estimating equation takes the form:

$$y_{ivst} = \beta\, Treat_v * Post + \eta\, Ineligible_v * Post + \mu\, NonInc_v * Post + \rho_i + \gamma_{st} + \delta\, \mathbf{X_v} * Post + \epsilon_{ivst} \tag{5}$$

$Treat$ is an indicator that equals 1 if village v receives a program road before 2006.

4.3 Accounting for Movement out of Agriculture

In order to obtain unbiased estimates of the effects of improved rural road infrastructure on households' production decisions in agriculture, I have to first account for compositional effects as road infrastructure may induce the movement of workers out of agriculture. Asher and Novosad (2016) find that PMGSY led to a 10 pp reduction in the share of households in agriculture, with an equivalent increase in participation in wage labor. As such, simply conditioning on households within a village that remain in agriculture could lead to biased estimates as there may be selection in the types of households that move out of agriculture.

To account for this potential movement, I use a proxy for households' access to employment opportunities in the non-agricultural sector. Within village boundaries, non-agricultural opportunities are limited. Workers often commute to nearby factories or towns[25] when seeking non-agricultural employment in low-skilled manufacturing or in construction. I thus use the distance between villages and their closest town to construct a proxy for access to the non-agricultural sector. Figure 4 plots the distribution of distances between each village in the REDS sample and the nearest town at baseline; the median distance is 10 kilometers.

I define my proxy for access to employment in non-agriculture to be a binary indicator that takes the value 1 when a village is within 10 kilometers of the nearest town. This cutoff is conservative in the Indian context where daily commuting distances are low. Appendix Figure A1 summarizes information collected in the 2011 India Census on commuting patterns among non-agricultural workers in rural India. 38% of workers have no commute to work, 22% commute by foot, and 13% commute by bicycle. Further, for over 80% of workers, the daily commuting distance is less than

[25] A Census town is defined as: (i) Population exceeding 5000 and (ii) At least 75% of the male working population employed outside the agricultural sector.

10 kilometers. Using a 10 kilometer cutoff as a proxy for access to employment in non-agriculture is thus reasonable in this setting. My results are robust to alternative distance cutoffs; I discuss this further in Section 7.3.

I fully interact the variables in Equations 2, 4 and 5 with this binary proxy measure. The estimating equations for the direct, incidental and pooled effects respectively take the form:

$$\begin{aligned} y_{ivst} = \beta_1\ DirectTreat_v * Post + \beta_2\ DirectTreat_v * Post * Close_v + \alpha_1\ Ineligible_v * Post \\ + \alpha_2\ Ineligible_v * Post * Close_v + \rho_i + \gamma_{st} + \delta\ \mathbf{X_v} * Post + \epsilon_{ivst} \end{aligned} \quad (6)$$

$$\begin{aligned} y_{ivst} = \beta_1\ IncidentalTreat_v * Post + \beta_2\ IncidentalTreat_v * Post * Close_v + \alpha_1\ NonInc_v * Post \\ + \alpha_2\ NonInc_v * Post * Close_v + \rho_i + \gamma_{st} + \delta\ \mathbf{X_v} * Post + \epsilon_{ivst} \end{aligned} \quad (7)$$

$$\begin{aligned} y_{ivst} = \beta_1\ Treat_v * Post + \beta_2\ Treat_v * Post * Close_v + \eta_1\ Ineligible_v * Post + \mu_1\ NonInc_v * Post \\ + \eta_2\ Ineligible_v * Post * Close_v + \mu_2\ NonInc_v * Post * Close_v + \rho_i + \gamma_{st} + \delta\ \mathbf{X_v} * Post + \epsilon_{ivst} \end{aligned} \quad (8)$$

$Close_v$ is an indicator that equals 1 if village v is within 10 kilometers of the nearest town at baseline. $\beta_1 + \beta_2$ captures the effect of access to improved road infrastructure for households that are in close proximity to a town, and subsequently, to employment opportunities in the non-agricultural sector, while β_1 captures the effect of improved access to road infrastructure for households that are further away from towns.

5 Results

This section presents the empirical results. In Section 5.1, I validate the use of program rules in obtaining causal estimates by establishing that program rules were followed. In Section 5.2, I show that the provision of improved rural road infrastructure led to crop diversification, take up of productive agricultural inputs and commercialization of farm output. In Section 5.3, I explore heterogeneous effects across the population.

5.1 Compliance with Program Rules

This subsection provides evidence for compliance of program implementation with the stated program rules. Column 1 of Table 4 shows that villages that were deemed unconnected by program administration were 55% less likely to report having any type of road in the village at baseline.[26] Next, Column 2 shows that for villages that were deemed eligible for a program road i.e. unconnected at baseline with a village population greater than 500, there was a 38.8 percentage points increase in the probability of having a program road in their village by 2006, significant at the 1% level. This establishes that program rules were enforced, validating the use of these rules in obtaining causal estimates of improved rural road infrastructure on agricultural outcomes.

5.2 Reduced Form Results

This subsection presents my main results on production decisions in agriculture. I begin by first estimating the effects of improved rural road infrastructure on households' activity status. As discussed in Section 4.3, I have to account for the potential movement of workers across sectors before estimating the effects on households' production decisions in agriculture. To do this, I look at two outcomes that reflect households' activity status in Table 5: an indicator for whether the household cultivates any agricultural land; and the total number of man-days worked by all members of the household in the casual labor market for non-agriculture.

Panel A of Table 5 summarizes the pooled treatment effects of improved rural road infrastructure on households' activity status. Among households treated by the program, there is a 10 percentage points decline in cultivation as shown in Column (1), though this effect is not significant.[27] The F-test p-value of 0.019 in Column (2) suggests that this decline is driven entirely by households in villages close to towns: there is a 33.8 percentage points reduction - equivalent to a 57% decline, relative to the baseline mean - in households' engagement in cultivation. This decline is accompanied by an increase of 44 man-days worked in non-agriculture for households in villages close to towns,

[26]The program was very particular in the type of road it constructed - they had to be *hard-topped, all-weather* roads. This implied that villages that with dirt/fair-weather roads at baseline were considered unconnected, for the purposes of the program.

[27]This is similar in magnitude to effect sizes in Asher & Novosad (2016); they find a 10 percentage points reduction in the share of households reporting cultivation as their primary income source.

as shown in Column (4). Note that for households further away from towns, there is no significant exit from cultivation and no change in the number of man-days worked in the casual labor market for non-agriculture.

Panels B and C of Table 5 summarize the treatment effects for the direct and incidental connections respectively. In both panels, a similar pattern of results emerge for households in villages close to towns - a reduction in households' engagement in agriculture and an increase in man-days worked in non-agriculture. The results in Table 5 demonstrate that households who gain access to increased rural road connectivity exit from agriculture, and this effect entirely driven by households with access to employment in the non-agricultural sector. Given this significant movement of households out of the agricultural sector in villages close to towns, I focus only on households further away from towns when analyzing the effects of improved road infrastructure on agricultural outcomes. To maximize power, I use the pooled Equation 8 for the remainder of my analysis, and I restrict attention to the coefficient β_1, the treatment effect on households further away from towns.

I now move to estimating the effects of improved road infrastructure on households' crop choices. Households often cultivate staple cereal grains such as paddy, wheat and maize for subsistence use, and cultivate commercial non-cereal crops such as fruits, vegetables and fiber crops for sale in the market. This is evident among households in my REDS sample at baseline - only 0.3% of cultivators with no observed market activity grow non-cereal crops, while 64.2% of cultivators engaged in some market activity grow non-cereal crops. Examining crop choice is thus indicative of households' intentions to retain their farm output for home consumption or to sell their farm output in nearby product markets.

In the REDS agricultural modules, I observe for every season and for every crop grown by the household the amount of land that is cultivated as well as the use of high-yielding varieties. I aggregate this data to the household level in constructing the following outcomes: indicators for whether the household cultivates any non-cereal and cereal crops; the share of land that is cultivated under non-cereals; and indicators for the use of high-yielding varieties, broken down into two categories - non-cereal and cereal.

Table 6 summarizes the pooled treatment effects of improved rural road infrastructure on households' crop choice. First, there is a significant extensive margin response in non-cereal cultivation - I find a 25.7 percentage points increase in households cultivating any non-cereals in Column (1). I observe a non-significant decline in households cultivating any cereals in Column (2). Second, there also appears to be a significant intensive margin response - conditioning on households cultivating land in both periods in Column (4), I find a 55% increase in the share of land cultivated under non-cereals, relative to the baseline mean. Third, there is substantial take up of high-yielding varieties for non-cereal crops, significant at the 1% level, in Column (5). While take up of high-yielding varieties is also positive for cereal crops in Column (6), this coefficient is not significant.[28]

The results in Table 6 demonstrate that households who gain improved road connectivity begin cultivating hybrid non-cereal crops, and at the same time, continue to cultivate staple cereal grains on some portion of their land. This shift towards cultivation of commercial, non-cereal crops is suggestive of households' intentions to market their farm output; I empirically examine households' market activity in Table 8.

Next, I examine the effects of improved road infrastructure on households' usage of agricultural inputs. In particular, I focus on decisions to invest in productivity-enhancing agricultural technologies as well as farm labor use. To do this, I construct the following outcomes at the household level: indicators for the purchase of high-yielding variety seeds, irrigation (pumps, storage tanks, water etc.), organic manure and chemical fertilizers; and the number of labor-days utilized per cultivated acre, broken down into 3 categories - total, hired and family.

Table 7 summarizes the pooled treatment effects of improved road infrastructure on households' farm input decisions. There is a significant increase in households' take up of material inputs such high-yielding variety seeds, irrigation, manure and fertilizer, as seen in Columns (1)-(4). Further, there is a substantial increase in farm labor use, with a 46% increase in the total number of labor days invested per acre of cultivated land, significant at the 5% level in Column (5). This is driven by a 52% increase in the number of labor-days hired per acre of cultivated land in Column (6). This increase is accompanied by an increase in the number of family labor-days employed per acre of cultivated land in Column (7), though this effect is not significant. Households are not substituting

[28] Panels A and B of Appendix Table A1 summarize the treatment effects of improved rural road infrastructure on households' crop choice for the direct and incidental connections respectively.

between hired labor and family labor, but instead, increasing investments in both categories.[29] Given strong complementarities between high-yielding variety seeds and other inputs of production[30] (Duflo and Pande, 2007; Foster and Rosenzweig, 2010; Suri, 2011), this result is consistent with households increasingly cultivating high-yielding variety crops and subsequently investing in the complementary material and labor inputs required to cultivate these crops successfully.

The results in Table 7 are robust to alternative definitions of the farm input variables, as shown in Appendix Table A3, where I use the log of total expenditures for each of the inputs. Appendix Table A4 illustrates that these results are also robust to restricting the sample to villages further away from towns (Panel A), to eligible and incidental villages (Panel B), and to the eligible and incidental households in villages further away from towns (Panel C).

It is worth noting that the effect sizes that I find - a 26pp (77%) increase in cultivation of non-cereals, a 24pp (91%) increase in HYV use, and a 25pp (100%) increase in irrigation use - are larger than effect sizes found in other micro studies. For example, Cole et al. (2013) find a 12% increase in planting of higher-return/higher-risk cash crops among Indian farmers provided with rainfall index insurance. Duflo et al. (2011) find a 47-60% increase in fertilizer adoption among Kenyan farmers offered free fertilizer delivery early in the season. Cole and Fernando (2016) find a 60% increase in irrigation expenditures among Indian farmers provided with mobile-phone based agricultural advice.

Finally, I examine the effects of improved road infrastructure on households' product market activity. To do this, I construct the following outcome variables: an indicator for whether the household engages in any crop sales; an indicator for whether the household engages in the sales of high-yield variety crops; log sales revenues; log empirical profits; and log imputed profits, where I impute the cost of family labor valued at the prevailing market wage at the village level. I find a 16 percentage points, or 30% increase, in entry into sales of farm output among households that received a program road in Column (1) of Table 8. Further, there is a 24 percentage points increase in households

[29] Panels A and B of Appendix Table A2 summarize the treatment effects of improved rural road infrastructure on households' farm input decisions for the direct and incidental connections respectively.

[30] Previous work has documented that high-yielding varieties are more sensitive to modern inputs such as chemical fertilizers, as well as traditional inputs such as water. The sowing of high-yielding variety seeds, the application of chemical fertilizers and manure, and the management of irrigation systems are all labor-intensive processes. Further, as yields are higher, the adoption of high-yielding varieties is also associated with an increased demand for labor during harvests.

selling high-yielding variety crops, significant at the 10% level, in Column (2). While imprecisely estimated, evidence presented in Columns (3)-(5) suggests that households subsequently experienced gains in farm sales revenue and profits. Improved rural road infrastructure thus led households to transition from subsistence to market-oriented, commercial farming.

5.3 Heterogeneous Results

This subsection explores heterogeneous effects of improved road infrastructure. This exercise is helpful in understanding how gains from connectivity are distributed across households and has potentially important implications for policy. Given that the adoption of agricultural technologies is associated with high upfront fixed costs and scale effects, I separately test for treatment effects among households in different subgroups of cultivated landholdings.

Table 9 reports the treatment effects of improved road infrastructure on households' crop choice and usage of agricultural inputs separately for two subgroups: non and below median cultivators in Panel A, and above median cultivators in Panel B. I construct these groups using the within-district gross cultivated land distribution at baseline.

I find that the large positive average treatment effects described above in Section 5.2 are concentrated among small-scale (below median) cultivators. First, in Column (1) of Table 9, I find a significant extensive margin response in non-cereal cultivation for below median cultivators in Panel A; this response is not significant for above median cultivators in Panel B. Second, in Column (2), I find substantial take up of non-cereal high-yielding varieties for below median cultivators in Panel A, significant at the 5% level. Take up of high-yielding varieties is also positive for above median cultivators in Panel B, though this coefficient is not significant. Third, in Columns (3)-(7), I find a significant increase in investments in material inputs and farm labor, concentrated among below median cultivators in Panel A. This increase in farm labor is driven by a significant increase in the number of labor-days hired per acre of cultivated land, as seen in Column (8). These findings suggest that small-scale cultivators were more likely to be constrained by the high costs associated with making productive investments in agriculture in the presence of poor infrastructure.

6 Potential Channels

This section presents evidence that supports a channel through which improved rural road infrastructure led to crop diversification and take up of agricultural technologies: an increase in mobility of agricultural workers across connected labor markets. Next, it rules out alternative channels such as a decline in input prices and an increase in access to farm input/output markets. Finally, it presents survey evidence which suggests that agricultural labor reliability is an important concern among cultivators.

As described in Section 5.2, the cultivation of commercial hybrid crops is labor- and capital-intensive relative to traditional crops, and there exists strong complementarities between hybrid seeds and other factors of production such as material inputs and labor. Thus, for a household to successfully cultivate these crops, it requires timely access to these complementary inputs to production. In Column (6) of Table 7, I find that improved rural road infrastructure led households to increase the intensity of farm labor use, particularly that of hired labor - there is a substantial 52% increase in the number of labor-days hired per acre of cultivated land. This is consistent with the idea that improved rural road infrastructure increased the mobility of agricultural workers by connecting village labor markets across space. As a result, households were able to intensify agricultural hiring to support farm operations as needed.

To further explore this channel of increased labor mobility, I examine the reduced form effect of improved rural road infrastructure on households' ownership of bicycles and scooters, two important modes of transport in this setting.[31] I construct indicators that equal 1 if the household reports owning a bicycle and scooter respectively. In Column (1) of Table 10, I find a 35% increase in bicycle ownership among households that received a program road, significant at the 5% level. I find a non-significant increase in scooter ownership in Column (2). This is consistent with the idea that households increasingly invest in assets that allow for greater mobility in the presence of improved rural road infrastructure.

There are several alternative channels that could also explain my results. While I am unable to

[31] Car ownership is very low in this setting - only 0.1% of households own a car at baseline. Rural residents commute short distances and predominantly rely on travel by foot or by two-wheelers such as bicycles and scooters. This is consistent with evidence presented in Section 4.3 on commuting patterns among rural Indian workers.

rule out all possible alternatives, I consider the following three: (1) a decline in farm input prices; (2) an increase in access to input markets; and (3) an increase in access to output markets where farm output is sold.

To rule out channel (1) outlined above, I estimate the effects of improved rural road connectivity on the imputed price of high-yielding variety seeds and the casual daily agricultural cash wage. I find no significant changes in the price of hybrid seeds and agricultural labor in Columns (3) and (4) of Table 10 respectively, suggesting that a reduction in farm input prices is not a key channel driving the effects.[32]

To rule out channels (2) and (3) outlined above, I exploit the incidental connections described in Section 4.1. Given that incidental villages were already connected at baseline, any changes in access to input and output markets through a new road linkage to an unconnected village would be second order. As shown in Appendix Table A1 and A2, the effects on crop choice and usage of farm inputs are persistent when restricting to the incidental villages, suggesting that at least for this set of villages, a change in access to input and output markets is not a key channel driving the effects. Instead, effects are likely to be driven by an increase in the reliability of agricultural labor drawn from connected labor markets within commuting distance of a village.

To provide suggestive evidence on the role of labor reliability, I surveyed 114 male cultivators in 18 rural villages across 3 districts in Odisha, India. All villages surveyed have all-weather road access and are greater than 10km from the closest town. The cultivators in my survey sample are similar on observables to cultivators in the REDS sample at baseline; for example, the average amount of cultivated land in my sample is 2.52 acres (2.39 acres in REDS sample), and 53% cultivate non-cereal crops (56% in REDS sample).

Despite having road connectivity, 42% of cultivators express facing difficulties in finding enough hired laborers to work on their land in the past agricultural cycle, with a majority attributing this to a high demand for workers at the same time. In contrast, only 5% and 7% of cultivators express having difficulty procuring agricultural inputs and selling farm output respectively. This highlights that the reliability of agricultural labor remains to be a concern, even among cultivators with access

[32] This result on agricultural wages is consistent with the presence of nominal wage rigidity in markets for casual daily agricultural labor, which has been documented in this setting (Kaur, 2015).

to a road network.

I present cultivators with a series of questions to better understand their hiring practices. 92% of cultivators report hiring laborers from inside the village first before hiring laborers from outside the village. When asked why they choose to hire from inside the village first, 56% report knowing laborers inside the village better, while 37% report that laborers from inside the village are more reliable in terms of showing up to work. Further, when cultivators do hire laborers from outside the village, only 32% report having to pay a higher wage in cash, while 82% report having to pay more in kind. This increase in in-kind transfers could potentially explain why the casual daily agricultural cash wage does not appear to respond in Table 10.

Next, I present a hypothetical scenario where cultivators can choose to hire workers from two equidistant villages - one with a hard-topped road and one with a dirt road. 76% of cultivators report a strict preference for hiring from the village with a hard-topped road, even though 99% report that workers from both villages would be equally hardworking. This suggests that improved road infrastructure can play a role in alleviating this labor reliability constraint, even in villages that have existing road connectivity at baseline.

I present a second hypothetical scenario where I describe the construction of a new road that connects the village to a nearby village that was not easily accessible previously by road, and elicit responses on how cultivation practices would be affected. 43% of cultivators report that it would be easier for them to find workers to hire for work on their land and 29% report that they would be able to hire workers more frequently, while only 8% report that it would be less expensive to hire labor.

Finally, I present a third hypothetical scenario where I describe a deterioration in the quality of the road leading up to the village, and elicit responses on how a series of agricultural activities would be affected. 34%, 35% and 26% report that purchasing and transporting of inputs, traveling to nearby markets to sell output, and traders traveling in to the village to buy outputs respectively would definitely be a concern. This suggests that poor road quality may affect access to input and output markets in villages with no road connectivity at baseline, for example, by deterring traders traveling in to the village to purchase output. However, as noted previously, these effects would not be first order in villages with road connectivity at baseline, unless there exist complementarities e.g. traders only travel in to the village if there is a large enough network of villages to buy from.

7 Robustness Checks

This section details several robustness checks. First, I test the validity of the underlying parallel trends assumption in my empirical strategy by looking at trends in outcomes using data from the Indian Census as well as by conducting a placebo test. I find no evidence of violations of this assumption. Second, I show that my results are not sensitive to alternate specifications of my proxy for access to non-agriculture. Third, I test for differential attrition and splitting of households in my constructed panel by treatment status, I find no significance differences. These robustness checks validate the empirical strategy and the sample that I use in estimating the causal impact of improved infrastructure on agricultural outcomes.

7.1 Pre-Trends

The difference-in-differences specification in my setting relies on the assumption that in the absence of PMGSY, the evolution of outcomes in villages that received roads in earlier years would have parallel trends with that of villages that received roads in later years; I empirically test the validity of this assumption by looking at trends in outcomes using data from the Indian Population Census.[33] In particular, I use three rounds of the decennial census (1991, 2001 and 2011) to examine trends in three outcomes - total village population, share of village population that is literate and the total number of non-workers. Given that PMGSY was launched in December 2000 and road construction only commenced in 2001, I treat 1991 and 2001 as pre-PMGSY years. Next, I classify all villages that received PMGSY roads between 2001-2010 as treated. One limitation of this exercise is that the data only contains village demographic variables and a breakdown of the village population across broad employment categories, so I am unable to look specifically at agricultural outcomes.[34]

[33] There exists several earlier rounds of REDS which I do not use in this analysis. The surveys in Round 3 were conducted in 1982, 17 years prior to Round 4. Given the long gap between the two rounds, the earlier rounds of REDS do not appear to be ideal for pre-trend analysis.

[34] The Indian Government does conduct a separate Agricultural Census and Farm Inputs Survey, but this data is aggregated to the block level, rendering it unsuitable for the purposes of this analysis.

Figure 5 plots the yearly means by treatment status for program-eligible villages in Panel A, and for incidental villages in Panel B. There are two features worth highlighting. First, the treated villages are larger than the control villages in Panel A; this is consistent with program rules that dictated larger unconnected villages were to receive program roads before smaller unconnected villages. Second, in all of the outcomes in both panels, the parallel trends assumption appears to hold prior to 2011.

7.2 Placebo Test

To further validate my identifying assumption, I carry out a placebo test, where $Treat$ is an indicator that equals 1 if the village received a program road between 2007-2010. I drop all villages that received a program road prior to 2007; the relevant control group in this set up thus consists of villages that received the program roads after 2010. If villages that were treated earlier by the program were on differential trends from those that were treated later by the program, I would expect to see significant effects for this placebo group. Appendix Table A7 shows the results from this placebo test on a subset of outcomes described in 5.2; none of the coefficients are significant at the standard levels.

7.3 Alternate Specifications of Proxy for Access to Non-Agriculture

To validate the use of a binary proxy in my empirical strategy, I test the sensitivity of my results to alternate specifications of this proxy. Throughout my analysis, the proxy I use for households' access to employment opportunities in the non-agricultural sector is a binary indicator that takes the value 1 when a village is within 10 kilometers of the nearest town at baseline. I verify that my results are robust to this proxy specification by repeating my analysis using two alternative distance specifications: 5 and 15 kilometers.

Appendix Table A5 estimates the effects of improved road infrastructure on households' activity status, using a 5 km cutoff in Columns (1) and (2), and a 15 km cutoff in Columns (3) and (4). I find that the results are highly consistent with my findings in Table 5. Among directly treated households in villages within 5 km of towns, I find a significant decline in households' engagement in cultivation (F-test p-value of 0.017), accompanied by an increase in man-days worked in the casual labor market for non-agriculture (F-test p-value of 0.021). These effect sizes are larger in magnitude than the effects I find with the 10 km cutoff in Table 5 - this is unsurprising and lends further support to the spatial concentration of non-agricultural opportunities closer to towns. When I move to the 15 km cutoff, the coefficients are imprecisely estimated but the sign of the effects point in the right direction.

Appendix Table A6 estimates the effects of improved road infrastructure on households' crop choices and farm input decisions using a 5 km cutoff in Panel A, and a 15 km cutoff in Panel B. As before, I find that the results are highly consistent with my findings in Tables 6 and 7.

7.4 REDS Sample Selection

There is some attrition between each round of the REDS survey. To maintain representativeness in the sample, new households in each village were added every round. Further, household units split over time. In constructing my balanced panel, I restrict my sample to households that appeared in both Rounds 4 and 5, and aggregate them back to the original household unit level, as detailed in Section 3. To ensure that the provision of rural roads under PMGSY did not generate selection in this sample, I test for differential attrition and splitting of households, by treatment status. Appendix Table A8 shows that there was no differential attrition or splitting; none of the coefficients are significant at the standard levels.

8 Discussion

In this paper, I provide causal evidence for the effects of improved rural road infrastructure on households' production decisions in agriculture. Using a panel of rural Indian households, I find that the provision of rural roads led to crop diversification - households begin cultivating commercial non-cereal hybrid crops in addition to staple cereal grains. It also led to the modernization of cultivation practices through the adoption of improved technologies, an increase in labor hiring and commercialization of farm output. I provide suggestive evidence using a survey in the field for these effects operating through an increase in agricultural labor mobility across connected village labor markets.

My findings suggest that poor rural road infrastructure which hampers the integration of labor markets across space can serve as a substantial barrier to productive investments in agriculture. There is a strong policy interest in both the provision of infrastructure and the adoption of agricultural technologies as means towards sustainable poverty reduction - in the 2016-2017 Budget, the Central Government of India allocated USD 14 billion to the roads sector and USD 11 billion to fertilizer subsidies for rural Indian farmers.[35] This suggests that there are potential gains that could arise from coupling infrastructure projects with other commonly used policy instruments such as fertilizer and irrigation subsidies. Further, given heterogeneous effects of improved rural road connectivity across space, my findings highlight the need to think about spatial targeting when designing such policy interventions.

My findings also demonstrate that the gains that accrue to agricultural households extend beyond the households targeted by the program, which suggests the importance of incorporating network effects in the design of transport policy. Finally, my findings can be used to inform an ongoing policy debate on whether costly infrastructure investments can be justified by the benefits that accrue to agricultural households.

[35] The Central Government also committed USD 7.4 billion to a flagship irrigation scheme, the *Pradhan Mantri Krishi Sinchai Yojana*, in 2015.

References

Aggarwal, S. (2015). Do Rural Roads Create Pathways out of Poverty? Evidence from India. *Working Paper.*

Allen, T. and Atkin, D. (2016). Volatility, Insurance, and the Gains from Trade. *Working Paper.*

Asher, S. and Novosad, P. (2016). Market Access and Structural Transformation: Evidence from Rural Roads in India. *Working Paper.*

Bandiera, O. and Rasul, I. (2006). Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*, 116(514):869–902.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the Road: Transportation Infrastructure and Economic Development. *NBER Working Paper.*

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padro i Miquel, G. (2015). The Value of Democracy: Evidence from Road Building in Kenya. *The American Economic Review*, 105(6):1817–1851.

Casaburi, L., Glennerster, R., and Suri, T. (2013). Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone. *Working Paper.*

Cole, S. and Fernando, N. (2016). Mobile'izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability. *Working Paper.*

Cole, S. A., Giné, X., and Vickery, J. I. (2013). How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment. *World Bank Policy Research Working Paper*, (6546).

Conley, T. G. and Udry, C. R. (2010). Learning About a New Technology: Pineapple in Ghana. *The American Economic Review*, 100(1):35–69.

Donaldson, D. (forthcoming). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review.*

Donaldson, D. and Hornbeck, R. (2016). Railroads and American Economic Growth: A "Market Access" Approach. *Quarterly Journal of Economics.*

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *The American Economic Review*, 101(6):2350–2390.

Duflo, E. and Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2):601–646.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*.

Feder, G., Just, R. E., and Zilberman, D. (1985). Adoption of Agricultural Innovations in Developing Countries: A Survey. *Economic Development and Cultural Change*, 33(2):255–298.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, pages 1176–1209.

Foster, A. D. and Rosenzweig, M. R. (2010). Microeconomics of Technology Adoption. *Annual Review of Economics*, 2:395–424.

Gertler, P. J., Gonzalez-Navarro, M., Gracner, T., and Rothenberg, A. D. (2016). Road Quality, Local Economic Activity, and Welfare: Evidence from Indonesia's Highways. *Working Paper*.

Gonzalez-Navarro, M. and Quintana-Domeque, C. (2016). Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects. *Review of Economics and Statistics*, 98(2):254–267.

Jacoby, H. G. (2000). Access to Markets and the Benefits of Rural Roads. *The Economic Journal*, 110(465):713–737.

Jedwab, R. and Moradi, A. (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *Review of Economics and Statistics*, 98(2):268–284.

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

Kaur, S. (2015). Nominal Wage Rigidity in Village Labor Markets. *NBER Working Paper*.

Munshi, K. (2004). Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics*, 73(1):185–213.

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

PMGSY (2001a). Manual for the Preparation of District Rural Road Plans. http://pmgsy.nic.in/pmg61.asp.

PMGSY (2001b). Operations Manual. http://pmgsy.nic.in/opmn1.htm.

Porteous, O. (2016). High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade. *Working Paper*.

Shrestha, S. A. (2015). The Effects of Rural Roads on Farmland Values: Evidence from the Topography-based Road Network in Nepal. *Working Paper*.

Suri, T. (2011). Selection and Comparative Advantage in Technology Adoption. *Econometrica*, 79(1):159–209.

Figure 1: Total Length of Roads Built under PMGSY

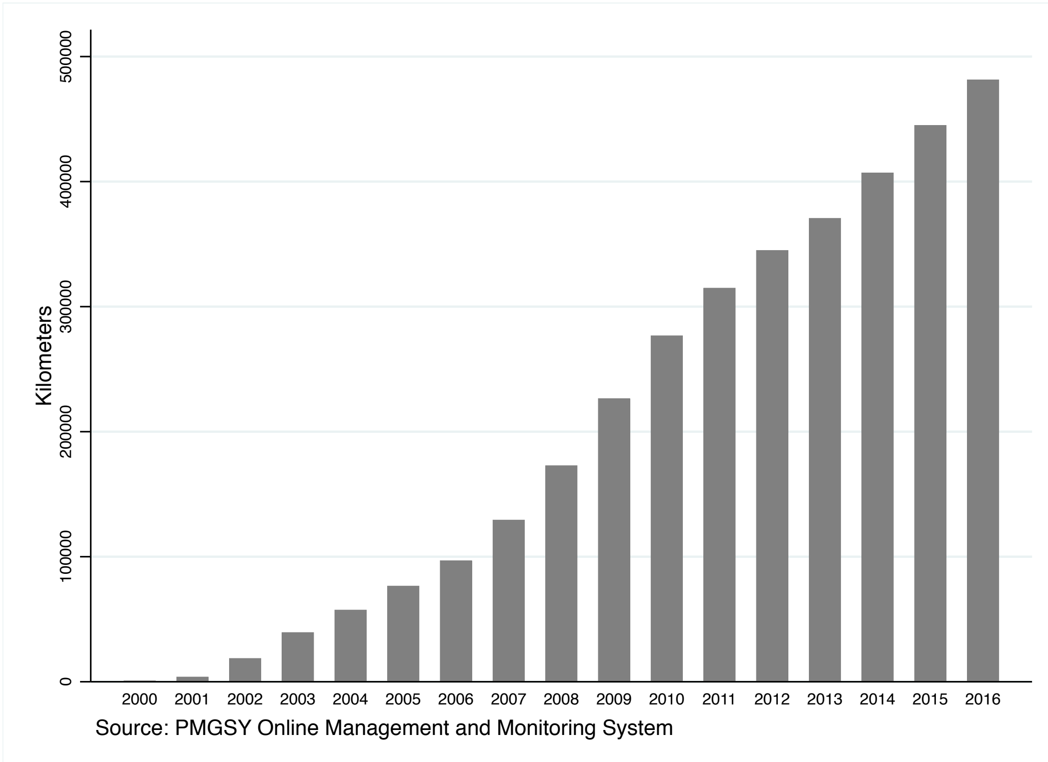


Figure 2: Spatial distribution of villages in *Rural, Economic and Demographic Survey*

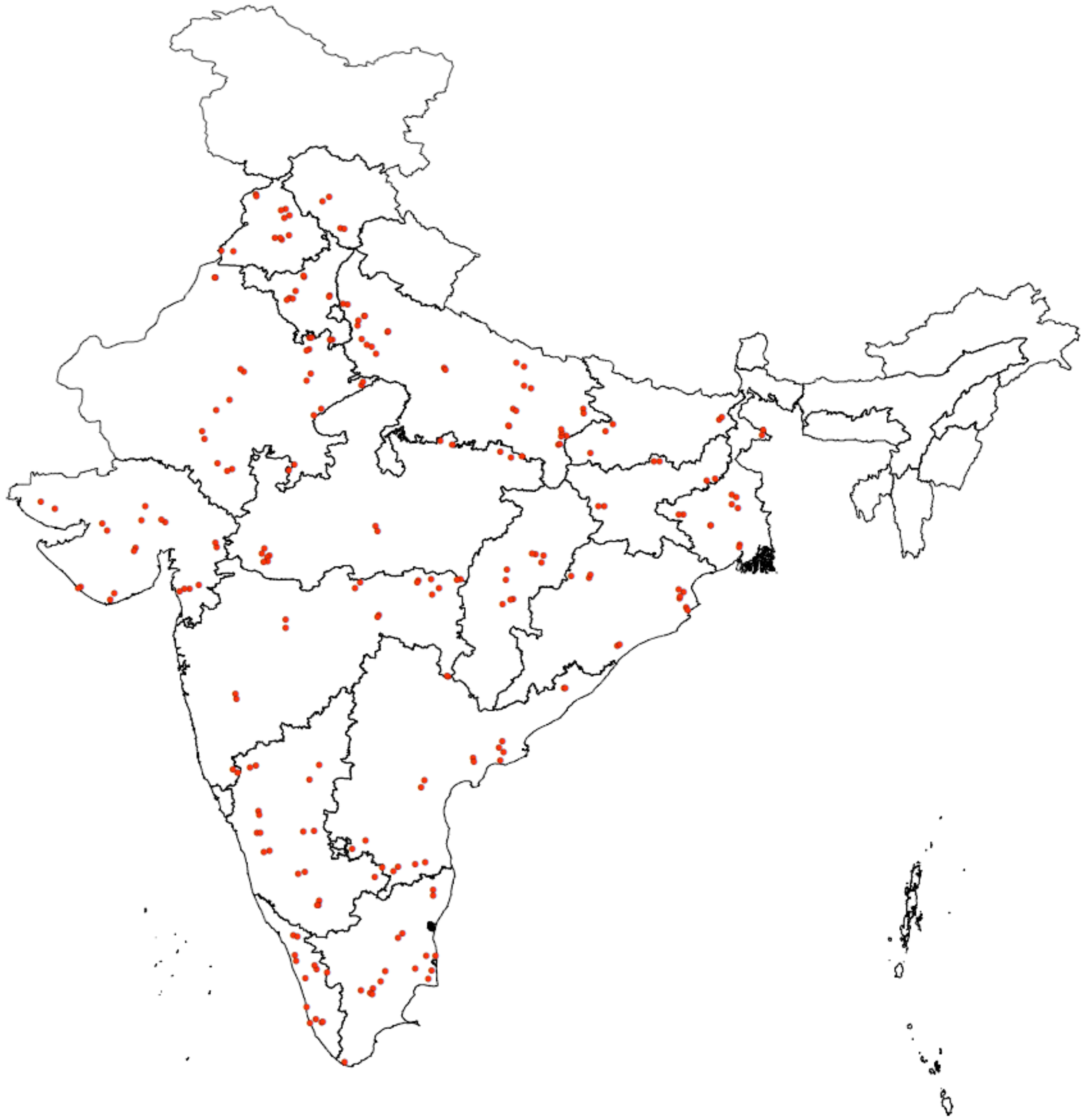


Figure 3: Timeline

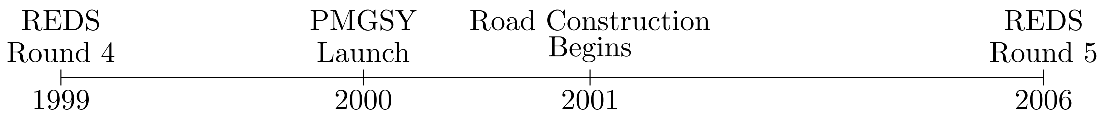
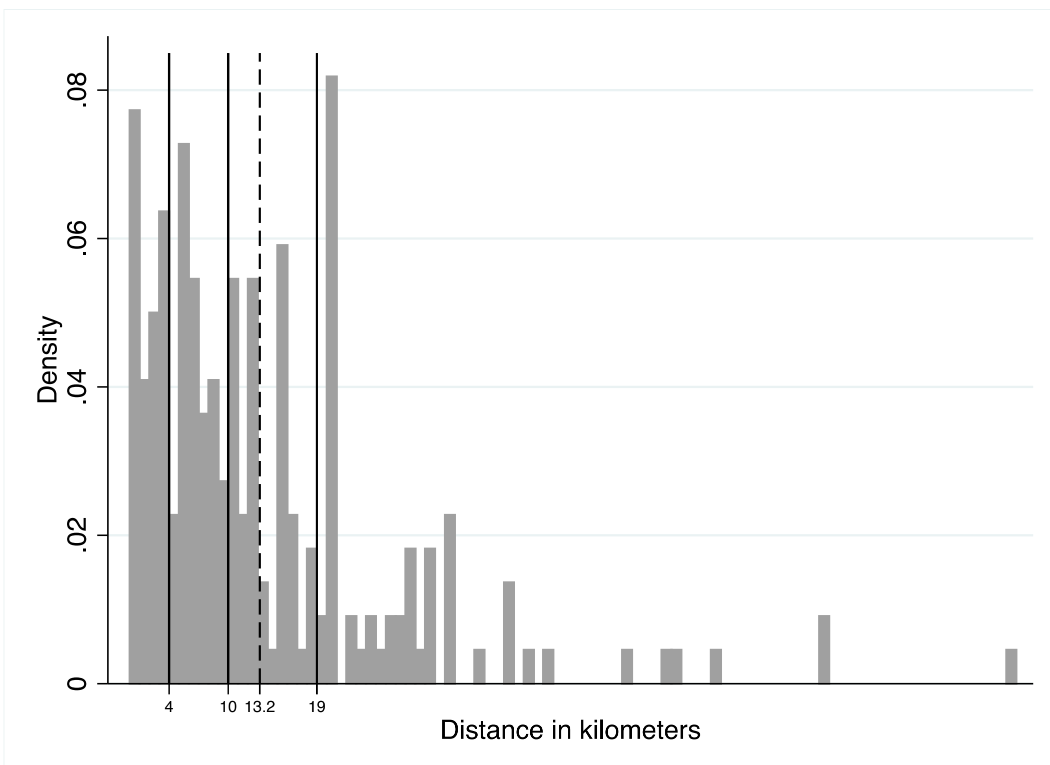


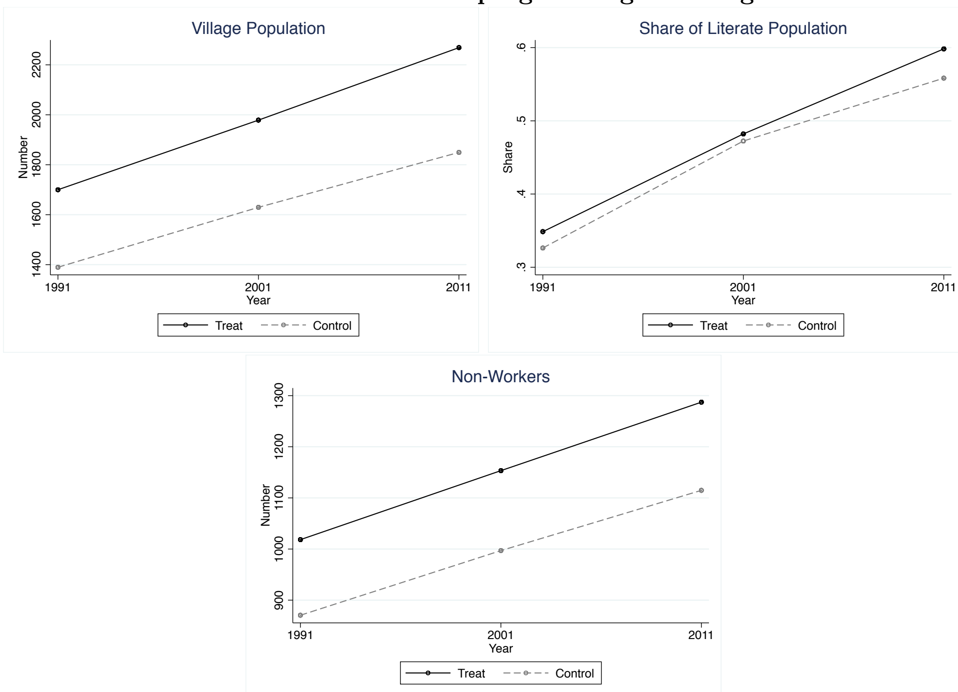
Figure 4: Distance between Village and Closest Census Town in 1999



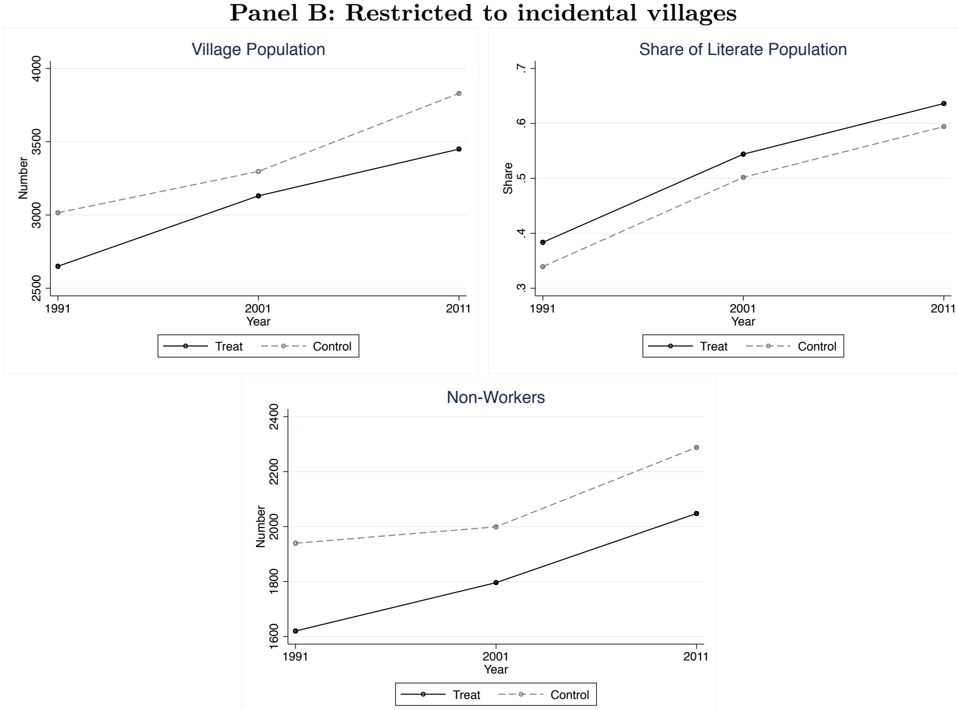
Notes: The distance between a REDS village and its closest town is recorded in the REDS 1999 village survey. The 25th, 50th and 75th percentiles of this distribution are 4, 10, and 19 kilometers respectively. The mean distance is 13.2 kilometers.

Figure 5: Pre-Trends using Indian Census Outcomes

Panel A: Restricted to program-eligible villages



Panel B: Restricted to incidental villages



Notes: Means constructed using village-level data from the 1991, 2001 & 2011 Indian Census Primary Census Abstracts. All villages that received program roads under PMGSY prior to 2010 are classified as treated.

Table 1: Baseline Summary Statistics of REDS Households

Demographics	
Number of Household Members	6.03
	(3.60)
Hindu	0.79
	(0.41)
SC/ST	0.18
	(0.38)
OBC	0.33
	(0.47)
Engagement in Agriculture	
Owns Land	0.69
	(0.46)
Sharecrops Land	0.02
	(0.13)
Cultivates Land	0.59
	(0.49)
Gross Amount of Land Cultivated (Acres)	2.39
	(4.53)
Cultivates Cereal Crops	0.47
	(0.50)
Cultivates Non-Cereal Crops	0.33
	(0.47)
Input Use: Irrigation	0.25
	(0.43)
Input Use: Chemical Fertilizers	0.56
	(0.50)
Input Use: High-Yielding Varieties	0.27
	(0.44)
Engagement in Casual Labor Markets	
Works in Casual Agriculture	0.29
	(0.45)
Days Worked in Casual Agriculture	96.52
	(181.12)
Works in Casual Non-Agriculture	0.30
	(0.46)
Days Worked in Casual Non-Agriculture	84.80
	(189.93)

Notes: Outcomes are measured in the 1999 REDS household survey. Table presents household-weighted baseline means, with standard deviations reported in parentheses. N = 4246

Table 2: Baseline Summary Statistics for Eligible Villages

	Control (1)	Treatment (2)	P-Value (3)
Village Population	1512.6	1640.8	0.677
	(947.1)	(967.6)	
Number of Households	369.4	313.1	0.586
	(381.5)	(189.9)	
Distance to Bus Stop	2.8	4.4	0.215
	(2.8)	(5.6)	
Distance to Bank	4.8	6.9	0.312
	(4.8)	(9.1)	
Distance to Weekly Market	4.7	5.4	0.692
	(4.6)	(8.1)	
Distance to Block HQ	12.8	16.2	0.315
	(10.1)	(11.4)	
Distance to Town	16.3	22.9	0.278
	(15.2)	(24.7)	
Gross Area Cultivated (Acres)	964.3	981.1	0.948
	(854.1)	(753.0)	
Average Harvest Price of Paddy (Rs./quintile)	532.2	560.9	0.467
	(85.0)	(102.5)	
Average Harvest Price of Wheat (Rs./quintile)	481.7	434.8	0.460
	(184.7)	(139.0)	
Agricultural Wage, Male (Rs.)	42.9	52.2	0.021
	(8.5)	(16.8)	
Agricultural Wage, Female (Rs.)	32.5	39.1	0.042
	(8.6)	(11.8)	
Non-Agricultural Wage, Male (Rs.)	76.3	90.3	0.033
	(15.0)	(26.8)	
Non-Agricultural Wage, Female (Rs.)	58.3	49.0	0.621
	(70.7)	(14.8)	

Notes: Restricted to program-eligible villages at baseline. Means and standard deviations (in parentheses) are shown in columns (1) and (2). Column (3) displays the p-values of the comparison of means across the Treatment and Control villages. N=42

Table 3: Baseline Summary Statistics for Incidental Villages

	Control (1)	Treatment (2)	P-Value (3)
Village Population	3598.8	2573.3	0.471
	(4825.7)	(3255.7)	
Number of Households	694.6	489.1	0.495
	(1040.4)	(617.7)	
Distance to Bus Stop	5.2	2.8	0.369
	(9.5)	(3.9)	
Distance to Bank	5.3	5.2	0.953
	(5.8)	(6.1)	
Distance to Weekly Market	6.5	10.2	0.386
	(7.4)	(22.2)	
Distance to Block HQ	21.9	20.2	0.804
	(23.0)	(18.2)	
Distance to Town	16.4	10.8	0.290
	(18.4)	(11.0)	
Gross Area Cultivated (Acres)	1159.4	1460.5	0.483
	(1200.8)	(1561.8)	
Average Harvest Price of Paddy (Rs./quintile)	549.5	542.0	0.823
	(88.7)	(85.9)	
Average Harvest Price of Wheat (Rs./quintile)	478.2	503.4	0.689
	(167.0)	(94.3)	
Agricultural Wage, Male (Rs.)	51.0	56.1	0.247
	(12.9)	(15.1)	
Agricultural Wage, Female (Rs.)	39.3	41.2	0.603
	(12.0)	(10.1)	
Non-Agricultural Wage, Male (Rs.)	90.5	94.6	0.712
	(30.2)	(44.0)	
Non-Agricultural Wage, Female (Rs.)	49.7	55.9	0.368
	(19.5)	(24.7)	

Notes: Restricted to program-ineligible villages with a neighboring eligible village at baseline. Means and standard deviations (in parentheses) are shown in columns (1) and (2). Column (3) displays the p-values of the comparison of means across the Treatment and Control villages. N=47

Table 4: Compliance with Program Rules

	Any Road in Village at Baseline	PMGSY Road in Village by 2006
	(1)	(2)
Unconnected at Baseline	-0.409***	
	(0.0764)	
Eligible for Program		0.388***
		(0.0840)
State FE	No	Yes
Dependent Variable Mean	0.747	0.127
Observations	221	221

Notes: The dependent variable in Column (1) is an indicator variable that equals 1 if there is any (fair-weather or all-weather) road in the village at baseline. This variable is recorded in the 1999 REDS village survey. The dependent variable in Column (2) is an indicator variable that equals 1 if there is a PMGSY road in the village by 2006. This variable is constructed using information downloaded from OMMS, the PMGSY administrative database. A village is classified by program administration as unconnected at baseline if it is located at least 500m away from an all-weather road or from another village with an all-weather road. A village is classified as program-eligible in the study period if it is unconnected at baseline and has a village population greater than 500. The regression in Column (2) includes state fixed effects. Robust standard errors are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Activity Status

	Engaged in Cultivation		Casual Labor Days Worked in Non-Agriculture	
Panel A: Pooled Effect	(1)	(2)	(3)	(4)
Treat x Post	-0.099	0.005	13.124	-2.915
	(0.093)	(0.099)	(30.675)	(32.449)
Treat x Post x Close to Town		-0.343**		47.358
		(0.163)		(53.919)
F-test p-value: Sum of coefficients		0.019		0.335
Panel B: Direct Effect				
Treat x Post	-0.115	0.032	43.827	-4.005
	(0.120)	(0.086)	(41.284)	(43.352)
Treat x Post x Close to Town		-0.393**		101.132
		(0.193)		(61.590)
F-test p-value: Sum of coefficients		0.039		0.039
Panel C: Incidental Effect				
Treat x Post	-0.076	0.003	-9.425	-13.112
	(0.118)	(0.131)	(34.996)	(39.766)
Treat x Post x Close to Town		-0.167		18.652
		(0.224)		(68.030)
F-test p-value: Sum of coefficients		0.402		0.923
Household FE	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.592	0.592	84.799	84.799
Observations	8492	8492	8492	8492

Notes: Difference-in-differences regressions. I run the pooled regressions (Equations 5 and 8) in Panel A, the direct regressions (Equations 2 and 6) in Panel B, and the incidental regressions (Equations 4 and 7) in Panel C. Outcome in Columns (1)-(2) is an indicator that equals 1 if the household engages in any agricultural cultivation. Outcome in Columns (3)-(4) is the total number of person-days worked in the casual labor market for non-agriculture. Close to Town is an indicator that equals 1 for households within 10km of a town. The p-value from a F-test of the joint significance of the two coefficients is reported at the bottom of each panel. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Crop Diversification

	Indicator: Non-Cereal	Indicator: Cereal	Non-Cereal Area Share		High-Yielding Varieties: Non-Cereal	High-Yielding Varieties: Cereal
	(1)	(2)	(3)	(4)	(5)	(6)
Treat x Post	0.257**	-0.105	0.227**	0.186**	0.313***	0.191
	(0.102)	(0.107)	(0.090)	(0.086)	(0.098)	(0.124)
Household FE	Yes	Yes	No	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Village FE	No	No	Yes	No	No	No
Dependent Variable Mean	0.334	0.472	0.341	0.341	0.217	0.239
Observations	8492	8492	5443	4400	8492	8492

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcome in Columns (1)-(2) is an indicator that equals 1 if the household engages in cultivation of non-cereal and cereal crops, respectively. Outcome in Columns (3)-(4) is the share of gross land cultivated under non-cereal crops. Outcome in Columns (5) and (6) is an indicator that equals 1 if the household uses high-yielding varieties (HYVs) of non-cereal and cereal crops, respectively. Cereal crops include paddy, wheat, maize, jawar, bajra, ragi and barley; non-cereal crops include fruits and vegetables, pulses, oilseeds, fiber crops, sugarcane, spices, drugs and plantation crops. The regression in Column (4) is conditional on households that are cultivating some land in both periods. All regressions include state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Farm Inputs

	Material Inputs				Labor Days Per Acre		
	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat x Post	0.243**	0.249***	0.326***	0.196*	26.918**	17.613*	9.305
	(0.119)	(0.091)	(0.071)	(0.115)	(12.016)	(10.324)	(13.487)
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.267	0.248	0.167	0.562	58.955	34.103	24.852
Observations	8492	8492	8492	8492	4400	4400	4400

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcome in Columns (1)-(4) is an indicator that equals 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Outcome in Column (5) is the total number of labor-days per cultivated acre. Outcome in Column (6) is the total number of hired labor-days per cultivated acre. Outcome in Column (7) is the total number of family labor-days per cultivated acre. The regressions in Columns (5) - (7) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Market Activity

	Product Market		Revenues and Profits		
	Any Crop Sales	HYV Crop Sales	Sales Revenues	Empirical Profits	Imputed Profits
	(1)	(2)	(3)	(4)	(5)
Treat x Post	0.157**	0.238*	1.131	0.775	0.845
	(0.072)	(0.129)	(1.193)	(0.852)	(0.858)
Household FE	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.515	0.305	9.613	5.391	5.164
Observations	8492	8492	8492	8135	7929

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcome in Column (1) is an indicator that equals 1 if the household sells any crops. Outcome in Column (2) is an indicator that equals 1 if the household sells any HYV crops. Outcome in Column (3) is log(total sales revenues + 1) from crop production. Outcome in Column (4) is log(empirical profits + 1), and the outcome in Column (5) is log(imputed profits + 1), where the price of family labor is imputed using the prevailing market wage at the village level. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Heterogeneity by Landholdings

	Crop Diversification		Material Inputs				Labor Days Per Acre		
	Non-Cereal Indicator	HYV Non-Cereal	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
Panel A: Below Median Cultivators	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat x Post	0.239**	0.231**	0.234*	0.207**	0.330***	0.207	36.138**	25.635*	10.503
	(0.105)	(0.091)	(0.128)	(0.084)	(0.064)	(0.131)	(17.184)	(13.199)	(18.990)
Dependent Variable Mean	0.227	0.141	0.189	0.184	0.122	0.427	55.982	25.725	33.220
Observations	5910	5910	5910	5910	5910	5910	2270	2270	2270
Panel B: Above Median Cultivators									
Treat x Post	0.056	0.243	-0.080	0.094	0.101	-0.225	14.751	6.252	8.499
	(0.175)	(0.176)	(0.136)	(0.226)	(0.131)	(0.220)	(18.702)	(10.560)	(10.975)
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.668	0.455	0.513	0.447	0.310	0.983	40.206	27.449	12.757
Observations	2582	2582	2582	2582	2582	2582	2130	2130	2130

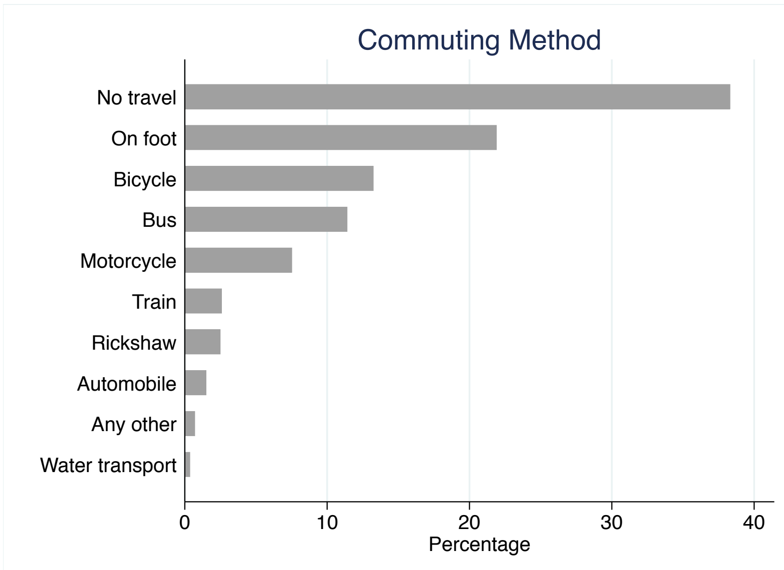
Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8 for baseline non-cultivators and below-median cultivators in Panel A, and above-median cultivators in Panel B. I restrict attention to β_1, the treatment effect on households further away from towns. I construct these groups using the within-district gross cultivated land distribution at baseline. Outcome in Column (1) is an indicator that equals 1 if the household engages in cultivation of non-cereal crops. Outcome in Column (2) is an indicator that equals 1 if the household uses high-yielding varieties (HYVs) of non-cereal crops. Outcomes in Columns (3)-(6) are indicators that equal 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Column (7) is the total number of labor-days per cultivated acre. Outcome in Column (8) is the total number of hired labor-days per cultivated acre. Outcome in Column (9) is the total number of family labor-days per cultivated acre. The regressions in Columns (7) -(9) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

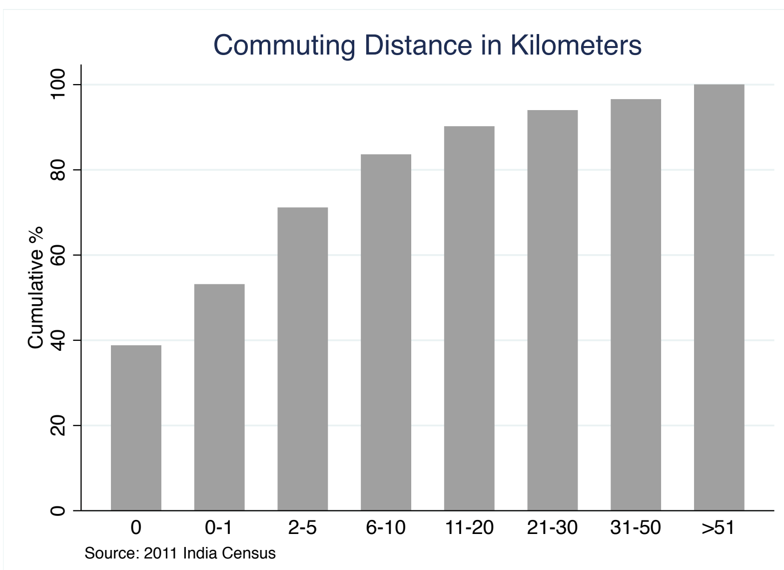
Table 10: Labor Mobility Channel

	Asset Ownership		Farm Input Prices	
	Bicycle	Scooter	HYV Seeds (Rs./Unit)	Agri Wage (Rs./Day)
	(1)	(2)	(3)	(4)
Treat x Post	0.157**	0.004	0.490	-2.629
	(0.076)	(0.005)	(8.152)	(6.388)
Dependent Variable Mean	0.449	0.018	23.697	42.836
Household FE	Yes	Yes	No	No
Village FE	No	No	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes
Observations	8492	8492	2443	4530

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcomes in Column (1)-(2) are indicators that equal 1 if the household owns a bicycle or scooter respectively. Outcome in Column (3) is the unit price of HYV seeds. Outcome in Column (4) is the daily agricultural cash wage. All regressions include state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: Commuting Patterns Among Rural Non-Agricultural Workers





Notes: These figures summarize information collected on commuting patterns among non-agricultural workers in rural India in the 2011 Census.

Table A1: Crop Diversification

	Non-Cereal			
	Indicator	Area Share		HYV
Panel A: Direct Effect	(1)	(2)	(3)	(4)
Direct Treat x Post	0.127	0.009	0.013	0.189
	(0.146)	(0.137)	(0.146)	(0.177)
Dependent Variable Mean	0.104	0.078	0.078	0.085
Panel B: Incidental Effect				
Incidental Treat x Post	0.395***	0.328***	0.263***	0.382***
	(0.105)	(0.067)	(0.069)	(0.098)
Dependent Variable Mean	0.423	0.255	0.255	0.187
Household FE	Yes	No	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes
Village FE	No	Yes	No	No
Observations	8492	5443	4400	8492

Notes: Difference-in-differences regressions. I run the direct regression in Equation 6 in Panel A, and the incidental regression in Equation 7 in Panel B. I restrict attention to β_1, the treatment effect on households further away from towns. Outcome in Column (1) is an indicator that equals 1 if the household engages in cultivation of non-cereal crops. Outcome in Columns (2)-(3) is the share of gross land cultivated under non-cereal crops. Outcome in Columns (4) is an indicator that equals 1 if the household uses high-yielding varieties (HYVs) of non-cereal crops. Non-cereal crops include fruits and vegetables, pulses, oilseeds, fiber crops, sugarcane, spices, drugs and plantation crops. The regression in Column (4) is conditional on households that are cultivating some land in both periods. All regressions include state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Farm Inputs

	Material Inputs				Labor Days Per Acre		
	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
Panel A: Direct Effect	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DirectTreat x Post	0.036	0.060	0.317***	-0.110	23.449	-13.731	37.180
	(0.092)	(0.150)	(0.094)	(0.126)	(25.057)	(9.537)	(22.760)
Dependent Variable Mean	0.135	0.258	0.046	0.424	50.984	22.630	28.354
Observations	8492	8492	8492	8492	4400	4400	4400
Panel B: Incidental Effect							
IncidentalTreat x Post	0.340**	0.332***	0.287***	0.395***	27.386**	35.429***	-8.042
	(0.155)	(0.104)	(0.097)	(0.146)	(12.209)	(12.988)	(7.437)
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.173	0.320	0.110	0.580	47.663	23.695	23.968
Observations	8492	8492	8492	8492	4400	4400	4400

Notes: Difference-in-differences regressions. I run the direct regression in Equation 6 in Panel A, and the incidental regression in Equation 7 in Panel B. I restrict attention to β_1, the treatment effect on households further away from towns. Outcome in Columns (1)-(4) is an indicator that equals 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Outcome in Column (5) is the total number of labor-days per cultivated acre. Outcome in Column (6) is the total number of hired labor-days per cultivated acre. Outcome in Column (7) is the total number of family labor-days per cultivated acre. The regressions in Columns (5) - (7) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Farm Inputs: Robustness to Alternate Definition of Outcome Variable

Log(1+ Expenditure in Rupees)	HYV Seeds	Irrigation	Manure	Fertilizer	Hired Labor
	(1)	(2)	(3)	(4)	(5)
Treat x Post	1.806*	1.575**	2.490***	1.612**	0.719
	(0.941)	(0.630)	(0.627)	(0.781)	(0.974)
Household FE	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	1.059	1.842	0.563	3.722	2.834
Observations	8492	8492	8492	8492	8492

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcome in Columns (1)-(5) is the log(total expenditures in Rupees +1) for high-yielding variety (HYV) seeds, irrigation, manure, chemical fertilizer and hired labor respectively. I impute the total expenditures for hired labor using the prevailing agricultural wage at the village level, and the reported number of days of hired labor. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Farm Inputs: Robustness to Sample Restrictions

	Material Inputs				Labor Days Per Acre		
	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
Panel A: Villages Far from Town	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat x Post	0.317***	0.325***	0.338***	0.132	30.798***	18.684*	12.113
	(0.113)	(0.092)	(0.070)	(0.106)	(10.795)	(10.780)	(11.391)
Observations	3516	3516	3516	3516	1954	1954	1954
Panel B: Eligible & Incidental Villages							
Treat x Post	0.149*	0.267***	0.327***	0.245**	32.588**	15.915	16.673
	(0.086)	(0.082)	(0.076)	(0.116)	(13.738)	(9.648)	(14.380)
Observations	3100	3100	3100	3100	1720	1720	1720
Panel C: Eligible & Incidental Villages Far from Town							
Treat x Post	0.141	0.334***	0.319***	0.203	31.456	3.452	28.004
	(0.086)	(0.114)	(0.105)	(0.138)	(22.164)	(9.662)	(20.379)
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1436	1436	1436	1436	858	858	858

Notes: Difference-in-differences regressions. Outcome in Columns (1)-(4) is an indicator that equals 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Outcome in Column (5) is the total number of labor-days per cultivated acre. Outcome in Column (6) is the total number of hired labor-days per cultivated acre. Outcome in Column (7) is the total number of family labor-days per cultivated acre. The regressions in Columns (5)-(7) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Activity Status: Robustness to Alternate Cutoffs for Close to Town Measure

Cutoff for Close to Town Measure:	Distance to Town < 5km		Distance to Town < 15km	
	Engaged in Cultivation	Casual Labor Days in Non-Agriculture	Engaged in Cultivation	Casual Labor Days in Non-Agriculture
Panel A: Direct Effect	(1)	(2)	(3)	(4)
Treat x Post	0.006	15.558	0.009	94.457
	(0.080)	(39.629)	(0.103)	(73.395)
Treat x Post x Close to Town	-0.423**	122.923*	-0.244	-86.435
	(0.178)	(68.521)	(0.177)	(88.959)
F-test p-value: Sum of coefficients	0.017	0.021	0.107	0.867
Panel B: Incidental Effect				
Treat x Post	0.005	-4.020	-0.023	0.013
	(0.132)	(39.494)	(0.124)	(42.811)
Treat x Post x Close to Town	-0.296	-34.372	-0.107	10.073
	(0.222)	(76.812)	(0.215)	(65.523)
F-test p-value: Sum of coefficients	0.142	0.588	0.504	0.847
Household FE	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes
Observations	8492	8492	8492	8492

Notes: Difference-in-differences regressions. I run the regression in Equation 6 in Panel A, and Equation 7 in Panel B. Outcome in Columns (1) and (3) is an indicator that equals 1 if the household engages in any agricultural cultivation. Outcome in Columns (2) and (4) is the total number of person-days worked in casual non-agriculture. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table A6: Crop Diversification & Farm Inputs: Robustness to Alternate Cutoffs for Close to Town Measure

	Crop Diversification		Material Inputs				Labor Days Per Acre		
	Non-Cereal Indicator	HYV Non-Cereal	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
Panel A: Distance to Town < 5 km	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat x Post	0.205**	0.275***	0.263***	0.206**	0.283***	0.187*	21.418*	12.624	8.793
	(0.081)	(0.089)	(0.098)	(0.088)	(0.070)	(0.103)	(11.669)	(9.272)	(12.519)
Dependent Variable Mean	0.334	0.217	0.267	0.248	0.167	0.562	58.955	34.103	24.852
Observations	8492	8492	8492	8492	8492	8492	4400	4400	4400
Panel B: Distance to Town < 15 km									
Treat x Post	0.172	0.338***	0.253*	0.276***	0.296***	0.182	14.166	16.483	-2.317
	(0.107)	(0.115)	(0.132)	(0.104)	(0.086)	(0.122)	(11.227)	(12.240)	(6.632)
Dependent Variable Mean	0.334	0.217	0.267	0.248	0.167	0.562	58.955	34.103	24.852
Observations	8492	8492	8492	8492	8492	8492	4400	4400	4400

Notes: Difference-in-differences regressions. I run the pooled regression in Equation 8 with $Close_v$ equals 1 if village v is within 5 kilometers of the nearest town at baseline in Panel A, and within 15 kilometers of the nearest town at baseline in Panel B. I restrict attention to β_1, the treatment effect on households further away from towns. Outcome in Column (1) is an indicator that equals 1 if the household engages in cultivation of non-cereal crops. Outcome in Column (2) is an indicator that equals 1 if the household uses high-yielding varieties (HYVs) of non-cereal crops. Outcomes in Columns (3)-(6) are indicators that equal 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Column (7) is the total number of labor-days per cultivated acre. Outcome in Column (8) is the total number of hired labor-days per cultivated acre. Outcome in Column (9) is the total number of family labor-days per cultivated acre. The regressions in Columns (7) -(9) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table A7: Placebo Test: Program Road Constructed between 2007-2010

	Engaged in Cultivation	HYV Seeds	Irrigation	Manure	Fertilizer
	(1)	(2)	(3)	(4)	(5)
Treat x Post	-0.085	0.001	-0.069	-0.048	0.041
	(0.107)	(0.089)	(0.086)	(0.080)	(0.211)
Treat x Post x Close to Town	0.022	-0.179	0.043	0.120	-0.084
	(0.211)	(0.215)	(0.270)	(0.102)	(0.278)
F-test p-value: Sum of coefficients	0.729	0.359	0.916	0.204	0.805
Household FE	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes
Observations	7958	7958	7958	7958	7958

Notes: Difference-in-differences regressions. I run the regression in Equation 6 with $Treat_v$ equals 1 if village v receives a program road between 2007 and 2010. I drop all households that received program roads before 2007. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Balance by Treatment Status: Robustness of Sample

	Household Attrits			Household Splits		
	Direct	Incidental	Pooled	Direct	Incidental	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.019	-0.075	-0.045	0.013	0.023	0.017
	(0.101)	(0.082)	(0.064)	(0.033)	(0.052)	(0.030)
Observations	1166	1468	2634	1166	1468	2634

Notes: Outcome in Columns (1)-(3) is an indicator that equals 1 if the household attrits between 1999 and 2006. Outcome in Columns (4)-(6) is an indicator that equals 1 if the household splits into multiple household units between 1999 and 2006. Columns (1) and (4) are restricted to households in program-eligible villages, Columns (2) and (5) are restricted to households in incidental villages, and Columns (3) and (6) are restricted to households in program-eligible and incidental villages. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Farm Inputs, Excluding Split Households

	Material Inputs				Labor Days Per Acre		
	HYV Seeds	Irrigation	Manure	Fertilizer	Total	Hired	Family
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat x Post	0.277**	0.294***	0.327***	0.176	24.113*	12.557	11.556
	(0.128)	(0.092)	(0.078)	(0.118)	(12.627)	(10.364)	(14.263)
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.204	0.323	0.101	0.572	59.725	34.200	25.525
Observations	6898	6898	6898	6898	3494	3494	3494

Notes: Difference-in-differences regressions. I drop all households that split between 1999 and 2006. I run the pooled regression in Equation 8, restricting attention to β_1, the treatment effect on households further away from towns. Outcome in Columns (1)-(4) is an indicator that equals 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Outcome in Column (5) is the total number of labor-days per cultivated acre. Outcome in Column (6) is the total number of hired labor-days per cultivated acre. Outcome in Column (7) is the total number of family labor-days per cultivated acre. The regressions in Columns (5) - (7) are conditional on households that are cultivating some land in both periods. All regressions include household and state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$