Journal of Development Economics 152 (2021) 102686

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

# Rural road infrastructure & agricultural production: Evidence from India[☆]

Yogita Shamdasani

*National University of Singapore, Department of Economics, 1 Arts Link, 117570, Singapore*

ARTICLE INFO

*JEL classification:*
Q12
J43
O13
O18

*Keywords:*
Infrastructure
Agricultural productivity
Technology adoption
Labor markets

ABSTRACT

A large share of the world's poor live in remote regions, where high costs associated with limited connectivity constrain economic activity. This paper estimates the effects of improvements in infrastructure under a large rural road-building program on production decisions in agriculture. Remote households that gain access to program roads diversify their crop portfolio, adopt modern agricultural technologies and increase hired-labor use. Supporting evidence suggests that program roads increase the mobility of agricultural workers by integrating village labor markets across space, in turn enabling the adoption of labor-intensive production practices. These findings highlight the importance of last-mile connectivity in remote areas across the developing world.

## 1. Introduction

Important innovations such as high-yielding varieties (HYVs) have significantly increased agricultural yields across the world in recent decades.[1] These gains in yields, however, have been unequal, with developing economies substantially lagging behind top producers. For example, recent FAO estimates suggest that paddy yields in India are less than half of that in the United States. Numerous factors drive these cross-country differences, including limited take up of productive technologies such as HYVs by rural households in developing economies. This has subsequently given rise to interventions in recent years aimed at increasing adoption of such technologies – fertilizer subsidies, rainfall index insurance and mobile-based agricultural advice, to name a few – with mixed success.

In this paper, I examine the importance of poor transport infrastructure as an alternative factor impeding take up. Earlier works have discussed the potential role of infrastructure — for example, Suri (2011) proposed that farmers face heterogeneous net returns to productive technologies, such that even those with high estimated gross returns may fail to adopt due to high costs attributed to poor infrastructure. I empirically test the validity of this view. Specifically, I estimate the impacts of improving access to rural road infrastructure on agricultural households in remote villages – for whom the costs of adoption are likely to be very high – by studying a national road-building program in rural India.

At the turn of the century, approximately 40% of India's rural villages lacked any single all-weather road connectivity. In response, the Central government launched a national road-building program specifically targeted at increasing connectivity in these isolated villages. The goal was to equip villages with small paved roads so as to enable access to the closest market center, which serves as a rural business hub. Once built, program roads would provide households with connectivity to the broader rural road network, and subsequently, to surrounding villages, agricultural markets and towns.

Using a combination of program administrative records and a comprehensive household panel that is well-suited to study agricultural choices, I study the impact of increased access to rural roads on production decisions in agriculture. I look at two types of connections made under the program. First, I exploit variation in the timing of road construction in program-eligible villages. Timing of road construction in an eligible village is determined by its rank relative to that of other eligible villages in the state, with ranks generated using village population size. As such, within a state, a larger eligible village receives a program road before a smaller eligible village. Second, I take advantage of the fact that in some cases, program-ineligible villages

☆ I am deeply grateful to Emily Breza, Supreet Kaur and Eric Verhoogen for feedback and encouragement. I thank Jenny Aker, Dan Berkowitz, Seema Jayachandran, Ajin Lee, Maggie Liu, Suresh Naidu, Dan O'Flaherty, Kiki Pop-Eleches, James Roumasset, Adam Storeygard and participants of the Columbia Development Colloquium, numerous seminars and conferences for helpful discussions. I thank Andrew Foster and NCAER for sharing the REDS secure data files with identifiers. Arnesh Chowdhury and Santosh Kumar Sahoo provided excellent research assistance in the field. All errors are my own.

*E-mail address:* yogita@nus.edu.sg.

[1] 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.

https://doi.org/10.1016/j.jdeveco.2021.102686
Received 21 February 2020; Received in revised form 29 April 2021; Accepted 14 May 2021
Available online 21 May 2021
0304-3878/

are exposed to new program roads as they are in the proximity of eligible villages, and program roads are built connecting them. Using a difference-in-differences framework, I exploit variation in timing of road construction, comparing the evolution of outcomes for households in villages that receive a program road to households in villages that have yet to receive a program road.

I find that some households move out of agriculture in response to new road connectivity under the program. This exit from agriculture is concentrated among households within commuting distances to towns, where jobs in non-agriculture are plentiful. In contrast, I find no significant movement out of agriculture for households in villages further away from towns. This suggests that access to urban centers remains severely limited in these remote villages *in spite of* infrastructure provision.

In my subsequent analysis, I focus on understanding how road connectivity impacts agricultural outcomes for these isolated households, where movement out of agriculture is not first order. The existing literature on rural roads, which I describe in detail below, has largely focused on effects arising from changes in access to urban centers following improvements in road connectivity. Little is known about impacts of improvements in road connectivity on more remote populations that are severely constrained in leveraging opportunities in non-agriculture.

I find that households in remote villages that gain improvements in rural road connectivity diversify their crop portfolio – in addition to cultivating traditional cereal crops, households begin cultivating higher return, non-cereal hybrid crops that are more labor- and capital-intensive in production. The degree of diversification is substantial – I find a 17 percentage point increase in the share of total land cultivated under non-cereals. I also find that households significantly increase take up of complementary modern inputs and improved technologies such as HYV seeds, chemical fertilizers and irrigation, and they intensify hiring of agricultural workers. Finally, I observe commercialization of farm output, with a 15 percentage point increase in households selling any crops. Heterogeneity analysis reveals that these effects are largely concentrated among small-scale cultivators, implying that technology adoption was most likely constrained for this sub-population in the presence of poor infrastructure.

By connecting village labor markets across space, road infrastructure effectively increases the labor supply from which households can draw from. This in turn might enable households to switch to labor-intensive production technologies and intensify agricultural hiring as needed. I provide several pieces of evidence which suggest that the mobility of agricultural laborers across connected labor markets is an important channel. First, when I restrict my analysis to program-ineligible villages that were exposed to program roads, I find that the effects persist. Given that a majority of these villages had pre-existing road infrastructure and were already connected to a market center at baseline, a new road linkage to an unconnected village is unlikely to significantly alter access to input and output markets. Instead, such a linkage plausibly increases access to agricultural laborers from this newly connected village labor market. Second, I find that the increase in hiring of agricultural laborers is concentrated in villages with relatively low labor densities at baseline. This suggests an inflow of workers from nearby surrounding villages, facilitated by improvements in road connectivity. Third, I find a 35 percent increase in bicycle ownership among households who gain access to improved road connectivity, consistent with households investing in assets that allow for greater mobility in the presence of better infrastructure. This channel is consistent with prior works that have discussed the role of labor availability in influencing agricultural practices (Hicks and Johnson, 1979; Feder et al., 1985). I also consider several alternative channels such as changes in access to credit, increased information flows, and social learning; I find weak evidence in support of these channels.

This paper adds to a large literature that seeks to understand barriers to productive investments in agriculture, which include limited access to information and learning, credit constraints, aversion to risk, high costs and behavioral biases.[2] This has given rise to numerous attempts in recent years to relax these barriers through privately-provided interventions such as agricultural advising services (Cole and Fernando, 2021), cash grants and rainfall index insurance (Karlan et al., 2014; Cole et al., 2017), fertilizer subsidies (Duflo et al., 2011) and flood-tolerant crop varieties (Emerick et al., 2016), with mixed success. This paper is the first to document how improvements in rural road infrastructure can induce households to adopt modern inputs and technologies in agriculture. In particular, my findings suggest that small-scale cultivators may fail to adopt high-return, labor-intensive production technologies due to potentially high transaction costs in the casual daily labor market generated by poor infrastructure.[3] The effect sizes I find are economically significant, suggesting that infrastructure investments have the potential to significantly transform remote agrarian economies. Additionally, in contrast to the privately provided micro interventions described above, I study a massive government-administered program, which suggests a feasible policy instrument that can be implemented and scaled up in settings where connectivity is severely limited.

This paper also contributes to a growing literature that estimates the causal effects of transportation infrastructure on economic outcomes. Estimating the causal impacts of infrastructure poses an important empirical challenge as placement is non-random and often driven by political and economic factors.[4] The literature has made progress in addressing these endogeneity concerns using identification strategies that typically exploit geographical constraints or historical events. A bulk of these studies have evaluated large-scale infrastructure systems that reduce transport costs of goods and travel times of individuals across distant regions, focusing on aggregate outcomes such as GDP levels, real income, and trade volumes.[5] 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

[2] For example, Foster and Rosenzweig (1995) document that information frictions serve as a barrier in the adoption of HYV 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. 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 (2019) shows using a counterfactual estimation that agricultural technology adoption only increases farmers' income when trade costs are low. 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.

[3] Rural labor markets in developing countries are characterized by substantial frictions (Singh et al., 1986; Benjamin, 1992; Fafchamps, 1993; Bryan et al., 2014; LaFave and Thomas, 2016; Kaur, 2019) that are consequential for labor market functioning (Breza et al., 2021). Recent work by Jones et al. (2020) illustrates that labor and land market failures constrain farmers' adoption of a profitable agricultural technology — hillside irrigation — in Rwanda.

[4] For example, Burgess et al. (2015) find 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.

[5] For example, Banerjee et al. (2021) estimate 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 (2018) 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.

a system generates effects on local economic growth and development, in contrast to the macro effects observed with large-scale infrastructure systems.

This paper makes several contributions to the literature on small-scale, local infrastructure.[6] It complements concurrent work by Aggarwal (2018) and Asher and Novosad (2020), who examine impacts of the same road-building program on economic outcomes. Aggarwal (2018) aggregates program exposure to the district-level and finds an increase in the area cultivated under hybrids in districts with more program roads. I further improve upon the identification by examining exposure to program roads at the village-level, which alleviates concerns about endogenous road placement. I similarly find that program road connectivity leads to increased usage of hybrid seeds as well as other complementary technologies in agriculture. Asher and Novosad (2020) find that program roads led to a reduction in the share of workers in agriculture, and no major changes in agricultural production inside the village. I similarly find that improvements in road connectivity induce movement of workers out of agriculture, and I show that these effects are driven by villages with access to the non-agricultural sector. Further, I find large gains in agriculture for households in villages where access to the non-agricultural sector remains limited in spite of road connectivity. At first glance, this result appears to be contradictory to their findings. However, it can be easily reconciled by understanding differences in the study sample and in the granularity of data used across the two studies — I focus on households in remote villages, and I use household-level data that allows me to observe changes in agricultural decisions at a highly disaggregated level. While I similarly find no change in mechanization among households who gain road connectivity, I do find that remote households begin to diversify their crop portfolio through the cultivation of non-cereals — pulses, in particular. This nuanced result would not be picked up by the outcome variable used in Asher and Novosad (2020), which measures the cultivation of non-cereals/non-pulses as the top three major crops aggregated at the village level. At the same time, my findings importantly demonstrate that road connectivity in remote villages does lead to diversification in agriculture, which has been documented in the literature as an essential pre-condition for local economic development (Bustos et al., 2016).

Further, I build on the findings in these two papers in several ways. First, I focus on remote villages, where access to urban centers remains severely limited even after new roads are constructed. Prior work has largely described treatment effects such as the movement out of agriculture as operating through increased rural–urban market integration. These effects are not first-order for remote populations. Instead, I find that isolated households experience large gains in agriculture, with effects operating through greater *rural-to-rural* labor market integration. Second, I estimate the spillover effects of program roads to other villages in the road network by considering incidental connections made under the program. Prior work has focused solely on impacts on targeted program beneficiaries. I find large effects for ineligible households that gain road connectivity under the program, which highlights the value of last-mile connectivity for isolated households living in remote regions. Third, I use rich household-level data that spans various dimensions of agricultural production that are typically hard to observe, such as hired labor use and crop sales. This allows me to provide a more complete picture of how agricultural production is changing as households gain improved road connectivity, and also allows me to explore potential channels underlying the reduced-form effects. I find suggestive evidence that new roads facilitate the mobility of agricultural workers across connected rural labor markets, which in turn enables households to switch to labor-intensive production technologies.

The body of the paper proceeds as follows. Section 2 discusses the Indian setting and the road-building program. Section 3 describes a conceptual framework that outlines how improvements in rural road connectivity may impact agricultural production. Section 4 describes the data and presents descriptive statistics. Section 5 details the empirical strategy, while Section 6 presents the results. Section 7 discusses potential channels that can explain the reduced-form results and provides some supplementary survey evidence from the field. Section 8 concludes.

## 2. Setting

In 2000, an estimated 330,000 of India's 825,000 rural villages lacked any all-weather road access. 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 monsoon rains set in. This translated to approximately 300 million people living in villages characterized by low spatial mobility, with limited access to basic economic and social services.

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 PMGSY, in December 2000. The primary objective was to provide single, all-weather connectivity to targeted villages through the construction of paved roads equipped with cross-drainage structures that allowed them to be operable in all weathers (PMGSY, 2001).[7] To be program-eligible, villages had to be unconnected, i.e. at least 500 meters away from an all-weather road or another village with an all-weather road.

Roads constructed under PMGSY were typically small, single lane roads[8] that belonged to one of two lowest categories of rural roads: (1) "village roads" that connected villages with each other or to the nearest road of a higher category; or (2) "other district roads" that connected villages to other main roads, the block headquarters, or the market center.[9] 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.

Given the dependency on agriculture in this setting, local road networks play an important role in facilitating travel to nearby markets, villages and towns for agricultural input purchases, farm output sales, as well as transactions in the casual daily labor market. Enabling access to the closest market center – identified by the program as centers of activities for marketing agricultural inputs and outputs, as well as servicing of agricultural tools – was thus an important goal under PMGSY.

[6] Previous studies examining the impacts of road provision in developing countries have used cross-sectional or aggregated panel data to exploit non-random construction and/or rehabilitation of rural roads. This includes work examining impacts on agricultural land values (Jacoby, 2000; Shrestha, 2020), acreage of local variety and HYV rice (Ali, 2011); market prices of local crops (Khandker et al., 2009; Casaburi et al., 2013), development of local markets (Mu and Van de Walle, 2011), non-farm employment (Gertler et al., 2016; Gibson and Olivia, 2010), poverty and consumption growth (Gibson and Rozelle, 2003; Dercon et al., 2009). In concurrent work, Brooks and Donovan (2020) evaluate the impact of integrating rural villages and urban markets through the construction of small footbridges. They find that footbridges eliminated the uncertainty of access to labor markets outside the village, subsequently resulting in increased labor market earnings.

[7] The unit of targeting under PMGSY is the habitation, a sub-village entity defined as a population cluster living in a location that does not change over time. A village consists of one or more habitations. The unit of analysis in this paper is the village; thus, for ease of exposition, I will use the terms habitation and village interchangeably.

[8] In the PMGSY rural roads manual, the suggested carriageway width is 3.7 m, "similar to that of a single lane carriageway of 'other district roads'", and the estimated road length is 4 kilometers.

[9] All higher category rural roads (major district roads, state highways and national highways) were strictly excluded.

PMGSY was launched as a Centrally-sponsored scheme, with the central government providing full funding and state governments managing project implementation, which proceeded as follows. Each state generated a list of all unconnected villages within the state, with villages ranked in descending order by population size (as recorded in the 2001 Population Census). Based on the amount of funding made available each year, the State Level Standing Committee then shortlisted the planned road works from this list.[10] 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. (PMGSY, 2005, 2012) 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. I exploit this variation in timing of road construction by village size in my empirical strategy in order to overcome the challenge of endogenous road placement.

Over the first decade of the program, an average of 22,000 kilometers of PMGSY roads were built annually. With 576,000 kilometers of roads built to date,[11] the program has more than doubled the size of the existing paved road network in rural India.

## 3. Conceptual framework

I describe a simple framework that outlines how changes in rural road connectivity may impact agricultural production. Improvements in access to rural roads lead to a reduction in transport costs of both goods and people. I first consider the impact of a reduction in transport costs of goods, specifically agricultural inputs and outputs. A reduction in the price of agricultural inputs such as hybrid seeds may lead to an increase in input use. However, access to cheaper inputs alone may not be sufficient to induce adoption; households also need access to output markets in order to make such input investments profitable. A reduction in the costs of bringing agricultural output to markets can induce households to shift into cultivation of market-oriented crops and increase input use, which subsequently leads to an increase in agricultural production. An increase in production can lead to an increase in agricultural labor demand, particularly if production shifts towards more labor-intensive crops and/or technologies such as hybrid seeds.

Next, I consider the impact of a reduction in transport costs of people. Lower transport costs for workers may lead to an increase in access to external employment opportunities. Households may reallocate labor to non-farm employment outside the village and subsequently reduce investments in agriculture. Further, if households that choose to reallocate to non-farm employment are positively selected – for example, if the relatively higher-skilled workers choose to reallocate – this can lead to an even larger reduction in agricultural production. Conversely, households may also benefit from increased farm employment outside the village boundaries.

Beyond these channels, there are numerous other ways in which improvements in access to rural roads can impact agricultural production. For example, it could increase access to credit, which in turn leads to an increase in productive investments in agriculture. It could also increase access to information on farming best practices through extension workers, which can directly impact agricultural production decisions.

Given the numerous channels described above, the overall impact of improvements in access to rural roads on agricultural production is ambiguous — it will depend on the magnitude of relative changes in transportation costs of goods and people, as well as households' proximity to employment opportunities outside the village.

## 4. Data

My main empirical analysis uses a panel of rural Indian households linked to PMGSY administrative records. In particular, I use the two most recent waves of the Rural Economic & Demographic Survey (REDS), a nationally representative survey administered by the National Council of Applied Economic Research (NCAER). The timing of the REDS is ideal for my analysis — I observe households in 1999, prior to the launch of PMGSY in December 2000, and again in 2006, several years after road construction began.[12] In constructing a balanced panel, I restrict my sample to households that were observed in both waves, and I aggregate all partitioned households back to the original 1999 household unit level.[13]

The REDS is extremely well-suited for studying agricultural outcomes at the household level — it contains detailed information on households' landholdings, crop choice, agricultural material inputs and labor use, agricultural outputs, crop sales, revenues and profits, and labor supply. In addition, the village-level REDS questionnaire includes demographic variables as well as information on existing road infrastructure and distance to various amenities.

Importantly, the restricted REDS data contains village-level identifiers that allow me to match the household panel to PMGSY administrative records. The Online Management & Monitoring System (OMMS) is an online reporting system used by program administration to track and monitor road works. I match the REDS villages to the OMMS database in order to observe village connectivity status and population size at baseline (which determine program eligibility), dates at which a program road was sanctioned and a work order was issued, costs associated with a program road, length of a program road and information on other villages connected by the program road. My matched sample consists of 4246 households from 221 villages.

I use three other datasets in this paper. I use the India Place Finder Hamlets database, which contains the universe of geocoded natural villages,[14] to construct treatment and control groups as described in Section 5. I also use remote sensing data on night time luminosity at the village level from the Socioeconomic High-resolution Rural–Urban Geographic Dataset on India (Henderson et al., 2011; Asher et al., 2019), along with the 1991, 2001 and 2011 Primary Census Abstracts (PCA), which contain key village-level demographic variables such as total population and worker counts across broad employment categories, to validate the parallel trends assumption underlying my empirical strategy.

[10] 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. The REDS dataset used in this analysis only covers the 17 major states, so the 250 persons cutoff is not relevant to my empirical setting. In practice, program implementation 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. In cases where there were multiple potential routes that could be constructed in an unconnected village, the District Panchayat (head) identified the most efficient route using weights assigned to different socio-economic services.

[11] January 2019 figures. Appendix Figure A1 illustrates the scale at which PMGSY was implemented over time.

[12] I also use the REDS 1982 wave (which precedes the wave in 1999) to validate the parallel trends assumption underlying my empirical strategy — I discuss this in Section 6.1.

[13] I follow Munshi and Rosenzweig (2016), who perform a similar aggregation with the 1982 and 1999 rounds of the REDS.

[14] 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.

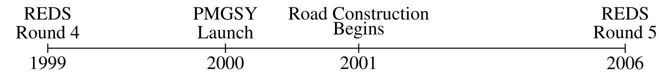


**Fig. 1. Timeline.** This figure presents a timeline of each round of REDS, as well as road construction under PMGSY.

## 5. Empirical strategy

### 5.1. Defining treatment and control groups

Fig. 1 summarizes the timeline over the study period. I observe households in 1999, prior to the start of PMGSY. The program launched in December 2000, and road construction commenced in 2001. I then observe households again in 2006.[15] During this period, road construction was restricted to unconnected villages with a population greater than 500, hereafter referred to as eligible villages.[16]

I study two types of connections made under the program. The first, which I refer to as a *direct* connection, exploits variation in the timing at which eligible villages were exposed to new program roads. The timing of road construction in an eligible village was determined by its rank relative to that of other eligible villages in the state, with ranks generated using village population size (PMGSY, 2005). By 2006, larger eligible villages had received program roads while smaller eligible villages had not. I classify program-eligible villages that receive PMGSY roads by 2006 as treated and program-eligible villages that receive PMGSY roads after 2006 as control.

The second, which I refer to as an *incidental* connection, exploits the fact that some ineligible villages were also exposed to new program roads. This occurred in cases where eligible villages received a new road under PMGSY, and the program road linked them to a nearby ineligible village. This nearby village thus gained a road linkage through the program, despite being classified as ineligible. I classify program-ineligible villages that are connected to a program-eligible village in their proximity through a PMGSY road constructed by 2006 as treated. While using all program-ineligible villages that did not receive an incidental treatment would appear to be a natural control group, this could easily violate the parallel trends assumption. For example, a village within a densely connected area where all neighboring villages are connected could possibly trend differentially from a village within a more sparsely connected area, where some neighboring villages are connected, but there exists other unconnected, eligible villages in their proximity. As such, 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 have yet to receive a PMGSY road by 2006.[17]

Incorporating incidental connections has numerous important advantages. First, it allows me to estimate the impacts of additional road connectivity, given that most incidental villages are already connected at baseline. 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 access to new markets to hire farm labor. It could also reduce transaction costs for extension workers and traders who travel to clusters of villages to promote/sell farm inputs and procure farm output respectively. Prior work on rural roads has largely focused on effects deriving from direct connections, ignoring potentially important spillovers to other villages in the network. Second, given that incidentally treated villages are program-ineligible, they abstract from program rules that govern the timing of road provision. Any deviations in project implementation from program rules, for example, as a political favor, could bias estimates of direct connections, but incidental connections would not be susceptible to this bias. Third, it can help inform the potential mechanisms underlying the effects of road infrastructure on agricultural production by allowing me to rule out several classes of channels that might be at play — I discuss this further in Section 7.

### 5.2. Difference-in-differences 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 in villages that have received a program road to households in villages that have yet to receive a program road.

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

$$y_{ivst} = \beta \ DirectTreat_v * Post + \rho_i + \gamma_{st} + \epsilon_{ivst} \tag{1}$$

$y_{ivst}$ is the outcome variable of interest 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 by 2006; $Post$ is an indicator for the survey wave in 2006; and $\beta$ is the coefficient of interest.[18] Any time-invariant household characteristics are absorbed by the household fixed effects $\rho_i$, and any annual shocks that are common across villages in a state are captured by the state-year fixed effects, $\gamma_{st}$. To allow for serial correlation of $\epsilon_{ivst}$ 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 Eq. (1) to include this group:

$$y_{ivst} = \alpha \ DirectTreat_v * Post + \zeta \ Ineligible_v * Post + \rho_i + \gamma_{st} + \epsilon_{ivst} \tag{2}$$

$Ineligible_v$ is an indicator that equals 1 if village $v$ was deemed ineligible for a program road. The coefficient of interest $\alpha$ 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. Interpreting the coefficient of interest $\alpha$ 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.

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

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

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

$$y_{ivst} = \alpha \ IncidentalTreat_v * Post + \phi \ NonInc_v * Post + \rho_i + \gamma_{st} + \epsilon_{ivst} \tag{4}$$

$NonInc_v$ is an indicator that equals 1 if village $v$ was classified as non-incidental. The coefficient of interest $\alpha$ is still estimated off the

[15] Household interviews for the REDS 2006 survey were canvassed over several years, starting from 2006 and ending in 2009. For ease of exposition, I will stick to NCAER's notation and refer to this wave as REDS 2006.

[16] Conversely, ineligible villages consist of villages that are already connected (i.e. within 500 meters of an all-weather road or another village with an all-weather road) and unconnected villages with a population less than 500.

[17] I provide details on the construction of incidental treatment and control groups in Appendix C.

[18] I use the year a work order was issued or a program road was sanctioned for a particular village as a proxy for the year of program road construction.

sample of incidental villages as the $NonInc_v * Post$ interaction term absorbs changes in outcomes for the non-incidental villages in the post period.

Out of the 221 villages in my matched sample, 37 are program-eligible villages and 48 are incidental villages. 15 program-eligible villages and 14 incidental villages receive program roads by 2006. The number of program-eligible treated and control households is 289 and 346 respectively, and the number of incidentally treated and control households is 291 and 594 respectively. Given that road construction under the program commenced in 2001, there is variation in the length of exposure to program roads for treated villages — on average, I observe treated households four years after their village receives a program road.

Table 1 presents baseline village-level characteristics by treatment status for all program-eligible villages (columns 1 and 2) and incidental villages (columns 4 and 5). For the program-eligible sample, consistent with the program rules, I observe that the mean total population is greater in treatment villages than control villages, though this difference is not significant ($p$-value of difference 0.304, column 3). The gross amount of cultivated land is higher in treatment villages relative to control villages ($p$-value of difference 0.062). Treatment villages are relatively further away from key infrastructure such as bus stops, banks and wholesale markets, and agricultural wages appear to be higher; though these differences are not significant. For the incidental sample, none of the outcomes are significantly different across treatment and control villages (column 6).

Appendix Table B1 presents baseline household-level characteristics by treatment status for program-eligible villages and incidental villages. Across both sub-samples, households' activity status, crop choice, farm input use and market activity prior to the start of the road-building program are similar in treatment and control villages, with a couple of exceptions. For the program- eligible sample, casual days worked in non-agriculture is lower and hybrid seed use is higher for treatment villages relative to control villages. For the incidental sample, casual days worked in non-agriculture is higher and irrigation and farm labor use are lower for treatment villages relative to control villages.

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:

$$y_{ivst} = \delta\ Treat_v * Post + \eta\ Ineligible_v * Post + \mu\ NonInc_v * Post + \rho_i + \gamma_{st} + \epsilon_{ivst} \quad (5)$$

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

### 5.3. Compositional effects

Road infrastructure may induce the movement of workers *out* of agriculture. In estimating the impact of improved road connectivity on households' production decisions in agriculture, simply conditioning on households within a village that remain in agriculture would bias estimates as there may be selection in the types of households that move out of agriculture. For example, if the least productive agricultural households in the village switch to the non-agricultural sector, this would lead to an over-estimate of the reduced-form impacts on agricultural production.

Within village boundaries, non-agricultural opportunities are limited. Workers who seek non-agricultural employment often commute to nearby factories or towns[19] for jobs in low-skilled manufacturing or construction. I construct a proxy for households' access to employment opportunities outside of agriculture using distance to the closest town at baseline. Fig. 2 plots the distribution of the distance between each REDS village and the nearest town in 1999 — the median is 10 kilometers.

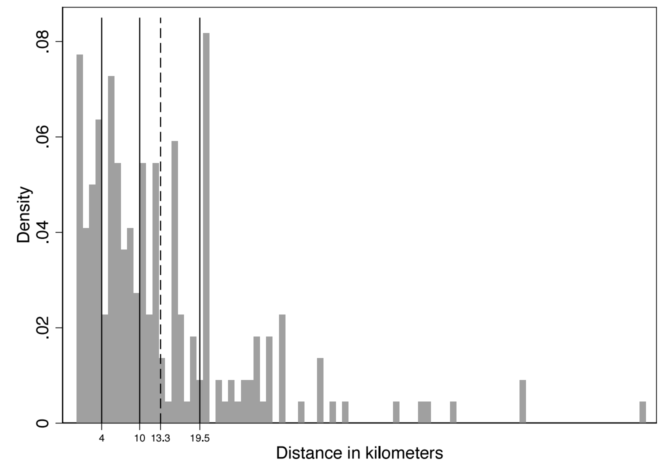


**Fig. 2. Distance between REDS village and closest Census town in 1999.** This figure plots the distance between a REDS village and its closest town, as recorded in the REDS 1999 village survey. The solid lines denote the 25th, 50th and 75th percentiles of this distribution, which are 4, 10, and 19.5 kilometers respectively. The dotted line denotes the mean (13.3 kilometers).

I define a proxy for access to employment in non-agriculture to be a binary indicator that takes the value 1 when a village is within the median distance of the closest town. Using this median cutoff to proxy for access to employment in non-agriculture is reasonable in this context — over 80 percent of rural non-agricultural workers report daily commutes of less than 10 kilometers in the 2011 Census.[20] I thus assume that labor markets are local due to limited labor mobility. I fully interact the variables in Eqs. (2), (4) and (5) with this binary measure.

The estimating equations for the direct, incidental and pooled effects respectively take the form:

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

$$\begin{aligned} y_{ivst} = {} & \theta_1\ IncidentalTreat_v * Post + \theta_2 \\ & IncidentalTreat_v * Post * Close_v + \phi_1\ NonInc_v * Post \\ & + \phi_2\ NonInc_v * Post * Close_v + \rho_i + \gamma_{st} + \epsilon_{ivst} \end{aligned} \quad (7)$$

$$\begin{aligned} y_{ivst} = {} & \delta_1\ Treat_v * Post + \delta_2\ Treat_v * Post * Close_v \\ & + \eta_1\ Ineligible_v * Post \\ & + \mu_1\ NonInc_v * Post + \eta_2\ Ineligible_v * Post * Close_v \\ & + \mu_2\ NonInc_v * Post * Close_v + \rho_i + \gamma_{st} + \epsilon_{ivst} \end{aligned} \quad (8)$$

$Close_v$ is an indicator that equals 1 if village $v$ is within 10 kilometers of the nearest town at baseline. $\delta_1 + \delta_2$ 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 $\delta_1$ captures the effect of improved access to road infrastructure for households that are further away from towns.

[19] A census town is defined as: (i) population exceeding 5000 and (ii) at least 75 percent of the male working population employed outside the agricultural sector.

[20] Appendix Figure A2 summarizes information collected in the 2011 India Census on commuting patterns among non-agricultural workers in rural India. 38 percent of workers have no commute to work, 22 percent commute by foot, and 13 percent commute by bicycle (panel A). Further, for over 80 percent of workers, the daily commuting distance is less than 10 kilometers (panel B).

**Table 1**
Characteristics of REDS villages at baseline.

| | Direct sample | | | Incidental sample | | |
|---|---|---|---|---|---|---|
| | Control (1) | Treatment (2) | *p*-value (3) | Control (4) | Treatment (5) | *p*-value (6) |
| Village Population | 1634.2 (945.7) | 1924.5 (1509.3) | 0.304 | 3598.8 (4825.7) | 2573.3 (3255.7) | 0.809 |
| Number of Households | 310.7 (200.1) | 339.5 (233.7) | 0.336 | 694.6 (1040.4) | 489.1 (617.7) | 0.632 |
| Distance to Bus Stop | 2.8 (2.2) | 4.4 (5.8) | 0.687 | 5.2 (9.5) | 2.8 (3.9) | 0.322 |
| Distance to Bank | 5.0 (4.9) | 6.8 (9.4) | 0.421 | 5.3 (5.8) | 5.2 (6.1) | 0.379 |
| Distance to Weekly Market | 4.8 (4.9) | 5.5 (8.2) | 0.406 | 6.5 (7.4) | 10.2 (22.2) | 0.237 |
| Distance to Block HQ | 12.2 (10.4) | 16.7 (11.7) | 0.965 | 21.9 (23.0) | 20.2 (18.2) | 0.532 |
| Distance to Town | 18.0 (16.4) | 18.5 (17.7) | 0.672 | 16.4 (18.4) | 10.8 (11.0) | 0.112 |
| Gross Area Cultivated (Acres) | 779.1 (593.8) | 1120.3 (787.0) | 0.062 | 1159.4 (1200.8) | 1460.5 (1561.8) | 0.664 |
| Average Harvest Price of Paddy (Rs./quintile) | 535.5 (93.2) | 571.9 (88.1) | 0.926 | 549.5 (88.7) | 542.0 (85.9) | 0.608 |
| Average Harvest Price of Wheat (Rs./quintile) | 443.4 (111.1) | 500.8 (284.2) | 0.493 | 478.2 (167.0) | 503.4 (94.3) | 0.589 |
| Agricultural Wage, Male (Rs.) | 43.6 (8.3) | 51.0 (16.7) | 0.450 | 51.0 (12.9) | 56.1 (15.1) | 0.373 |
| Agricultural Wage, Female (Rs.) | 32.7 (7.7) | 39.5 (12.1) | 0.294 | 39.3 (12.0) | 41.2 (10.1) | 0.191 |
| Non-Agricultural Wage, Male (Rs.) | 77.4 (16.1) | 89.8 (25.6) | 0.640 | 90.5 (30.2) | 94.6 (44.0) | 0.178 |
| Non-Agricultural Wage, Female (Rs.) | 61.5 (78.6) | 51.1 (14.0) | 0.393 | 49.7 (19.5) | 55.9 (24.7) | 0.817 |

*Notes:* Table presents baseline means by treatment status, with standard deviations reported in parentheses. Columns (1)–(2) are restricted to program-eligible villages (N=37, N(treatment)=15), and Columns (4)–(5) are restricted to program-ineligible villages with a neighboring eligible village (N=47, N(treatment)=14). Columns (3) and (6) display *p*-values from a comparison of means across treatment and control villages, obtained from a simple univariate regression with state fixed effects, where standard errors are clustered at the village level. Outcomes are measured in the 1999 REDS village survey.

In my matched sample, there are 6 program-eligible treated villages and 9 program-eligible control villages close to towns, 9 program-eligible treated villages and 13 program-eligible control villages far from towns, 8 incidental treated villages and 15 incidental control villages close to towns, and 6 incidental treated villages and 18 incidental control villages far from towns.

## 6. Results

### *6.1. Validating the parallel trends assumption*

The difference-in-differences specification relies on the assumption that in the absence of PMGSY, the evolution of outcomes in villages that receive roads in earlier years would have parallel trends with that of villages that receive roads in later years. I validate this assumption with three separate exercises — I examine trends in outcomes in the pre-period using data from the Indian Population Census as well as REDS survey waves, and I employ an event study approach using remote sensing data on night time luminosity.

First, I use three rounds of the decennial census (1991, 2001 and 2011) to examine trends in two village-level outcomes — total population and total number of agricultural 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. I classify all villages that received PMGSY roads between 2001–2010 as treated. Panel A of Fig. 3 plots yearly means by treatment status for program-eligible villages (a) and incidental villages (b). There are two features worth highlighting. First, treatment villages are larger than control villages in (a); this is consistent with program rules that dictated larger eligible villages were to receive program roads before smaller eligible villages. Second, there appears to be an uptick in the number of agricultural workers in treated villages in 2011, consistent with an increase in agricultural activity. I conduct a formal test of equality of trends in the pre-period — in particular, I estimate differential changes in outcomes across treatment and control villages prior to the launch of PMGSY using data from 1991 and 2001. As expected, I do not find any differential increase in population and agricultural worker counts in treatment villages relative to control villages prior to the launch of the program (Columns 1-4, Appendix Table B2).

To address concerns of this test potentially being underpowered, I repeat this exercise with all program-eligible villages across India.[21] Panel A (c) of Fig. 3 plots yearly means by treatment status for this sample of 51,424 villages, of which 23,985 are treated between 2001–2010. The patterns I describe above hold in this sample of program-eligible villages across India as well. Treatment villages are larger than control villages, and there appears to be an uptick in the number of agricultural workers in treated villages in 2011, which suggests an increase in agricultural activity. A formal test of equality of trends in the pre-period is presented in Columns 5 and 6 of Appendix Table B2. I find a small, statistically significant difference in total population growth across treatment and control villages in Column 5 (the magnitude of this difference is 0.68%, relative to the control mean). The test yields a precisely estimated null coefficient on the total number of agricultural workers in Column 6, which provides empirical support for the underlying assumption that agricultural activity trends in the

[21] I outline the step-by-step implementation of this exercise in Appendix D.1.

same way across treatment and control villages prior to the launch of the program.

Second, I use two waves of REDS surveys — corresponding to survey years 1982 and 1999 — to test the parallel pre-trends assumption in key agricultural outcomes used in the main analysis.[22] I use specifications that are similar to Eqs. (1) and (3), except that the post dummy here takes the value one for the year 1999. I present results for five agricultural outcomes in Appendix Table B3. I fail to reject the null for four out of the five outcomes, which suggests that there are no differential changes in these outcomes across treatment and control villages prior to the launch of PMGSY.

Third, I estimate an event study version of the difference-in-differences model using night time luminosity data, which has increasingly been used in economics as a proxy of economic output. The event study serves as a way to test the causal interpretation of the difference-in-differences analysis by allowing me to examine trends in pre-PMGSY years. I use an event window that ranges from three years before to three years after the launch of PMGSY, omitting the year 2000. Panel B of Fig. 3 plots the estimated differences and 95% confidence intervals for night time luminosity in treatment villages relative to control villages before and after the launch of PMGSY for program-eligible villages in (d) and incidental villages in (e). I find statistically insignificant differences in night time luminosity between treatment and control villages in the years prior to the launch of PMGSY. I also conduct a formal test of equality of trends in the pre-period in Columns 7 and 8 of Appendix Table B2 — in particular, I estimate the differential changes in night time luminosity across treatment and control villages pre-PMGSY using data from 1997–2000. The coefficients on the interaction term are statistically indistinguishable from zero — I thus conclude that there is no differential change in night time luminosity in treatment villages relative to control villages prior to the launch of PMGSY.

Taken together, these three exercises minimize concerns over confounding nonparallel pre-trends in the subsequent difference-in-differences analysis.

### 6.2. Compliance with program rules

Table 2 presents evidence for compliance of the road-building program with the stated rules. Villages that were classified as unconnected by the program were 56 percent less likely to have any type of road in the village at baseline, as observed in the REDS survey (column 1).[23] This result suggests that rules were adhered to in determining eligibility under the program. Next, for program-eligible villages, there was a 42 percentage point increase in the probability of having a program road in their village by 2006, significant at the 1 percent level (column 2).

### 6.3. Impact on agricultural outcomes

I begin by first accounting for the movement of workers across sectors in response to new connectivity. Table 3 presents estimates of the impact of improvements in rural road infrastructure on households' activity status. Households' activity status is measured using an indicator for whether the household cultivates any agricultural land, and the total number of person-days worked in the casual labor market for non-agriculture by all members of the household.

Among program-eligible households treated by the program (panel A), there is a 15 percentage point decline in cultivation, though this effect is not statistically significant (column 1). When I decompose the effect, I find that the decline is much larger for households in villages close to towns — there is a 40 percentage point reduction in cultivation (column 2), and the F-test $p$-value of joint significance is 0.006. This decline in cultivation is accompanied by an increase of 129 person-days worked in non-agriculture (column 4). For households further away from towns, there is no significant exit from cultivation or change in the number of person-days worked in the casual labor market for non-agriculture. A similar pattern emerges for incidental households treated by the program (panel B) – a decline in cultivation and an increase in person-days worked in non-agriculture for households in villages close to towns – though these effects are not statistically significant. These patterns are similar when I pool direct and incidentally treated households (panel C).[24] The effect size that I find is similar in magnitude to the effect size in Asher and Novosad (2020) — they find a 9 percentage point reduction in the share of workers in agriculture.

Taken together, these results suggest that improvements in road connectivity induce movement out of agriculture,[25] but only in villages within close proximity to the non-agricultural sector. This, in turn, results in a change in composition of households that remain in agriculture in these villages. I focus on remote villages far from towns, where there are no compositional changes, in order to isolate the impacts of road infrastructure on agricultural productivity. To maximize power, I use the pooled Eq. (8) for the remainder of my analysis and I restrict

**Table 2**
Compliance with program rules.

| | Any road in village (1999) (1) | Program road in village (2006) (2) |
|---|---|---|
| Unconnected | −0.421*** | |
| | (0.0768) | |
| Program-Eligible | | 0.423*** |
| | | (0.0837) |
| State FE | No | Yes |
| Dependent Variable Mean | 0.747 | 0.131 |
| Observations | 221 | 221 |

*Notes:* Outcome in Column (1) is an indicator that equals 1 if there is any fair-weather or all-weather road in the village in 1999. Outcome in Column (2) is an indicator that equals 1 if there is a program road in the village in 2006. A village is classified by program administration as unconnected if it is located at least 500 m 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 and has a village population greater than 500. Outcomes are constructed using the 1999 REDS village survey (column 1) and OMMS, the PMGSY administrative database (column 2). Robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[22] Given the 17-year gap, there is significant attrition and household splits between the two waves — as a result, the sample for this exercise is 38% smaller than that used in the main analysis. Further, the survey questionnaire changed significantly across waves — as a result, the set of outcomes I can examine in this exercise are limited. I describe the changes in survey samples and questionnaires across waves in greater detail in Appendix D.2.

[23] Recall that eligibility for a program road was determined by the proximity of a village to a *hard-topped, all-weather* road. Thus, under the program, villages with dirt/fair-weather roads at baseline were considered unconnected.

[24] In Appendix Table B4, I present estimates on the impact of improvements in rural road infrastructure on households' activity status, restricting the sample to program-eligible villages only in panel A, and to incidental villages only in panel B. The pattern of results is very similar to those in Table 3, as described above.

[25] In addition to movement across sectors, improvements in road infrastructure could have also induced individuals and/or households to migrate outside the village. I rule out this mechanism with two pieces of evidence. First, I do not find any trend breaks for total village population by treatment status in Fig. 3, suggesting that village size remained stable. Second, I directly examine the impact of PMGSY on total household size in Column 1 of Appendix Table B5. I find no detectable impact at standard significance levels, suggesting that household size also remained stable.

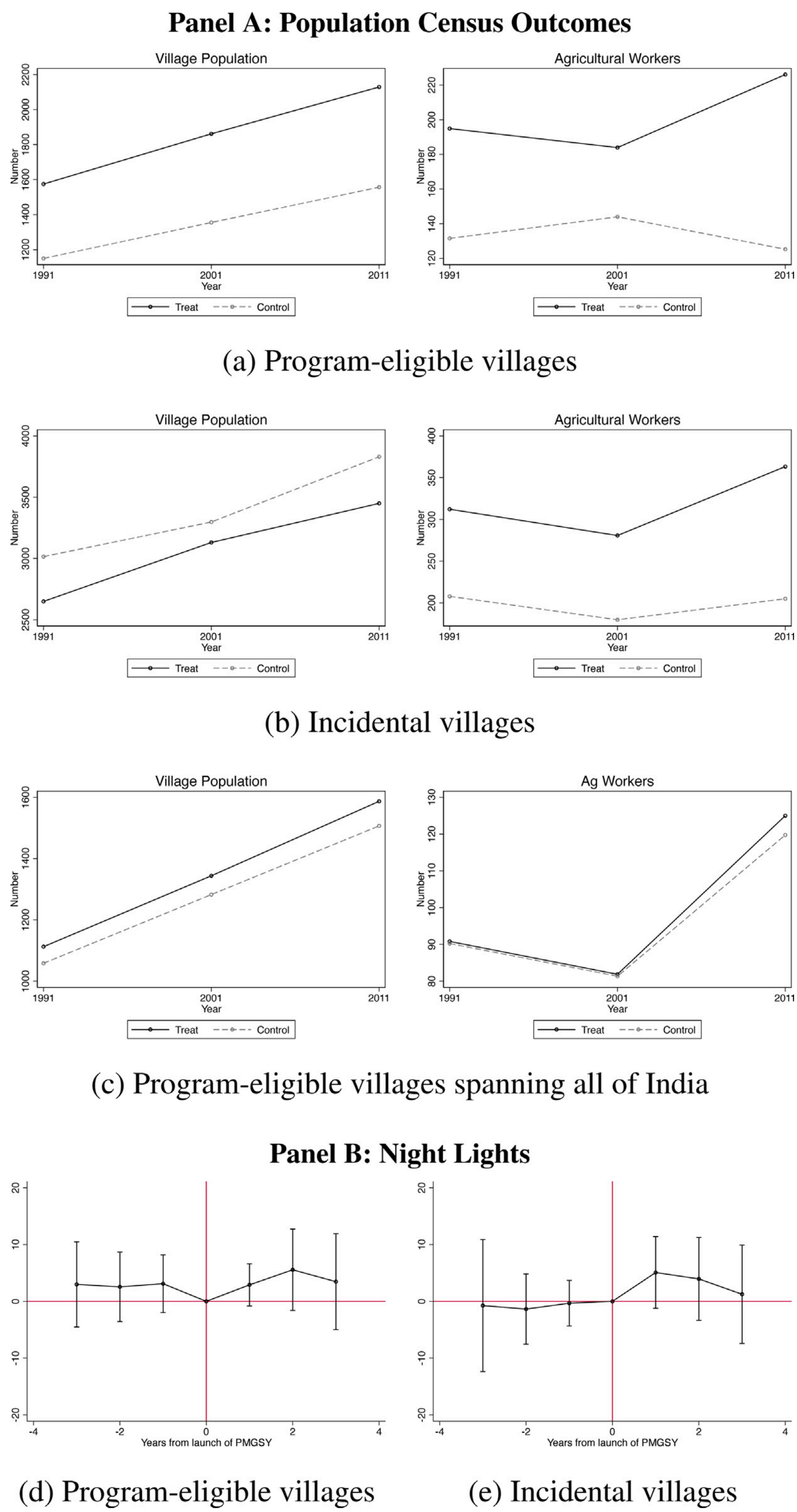


**Fig. 3. Pre-Trends.** Panel A plots trends for two outcomes (village population and agricultural workers) over time for program-eligible villages and incidental villages in the REDS data in (a) and (b) respectively, and for program-eligible villages across India in (c). Means are constructed using village-level data from the 1991, 2001 & 2011 Indian Census Primary Census Abstracts. Panel B plots the estimated differences and 95% confidence intervals for night time luminosity in treatment villages relative to control villages in the years before and after the launch of PMGSY for program-eligible villages in (d) and incidental villages in the REDS data in (e).

**Table 3**
Activity status.

| | Engaged in cultivation | | Casual labor days worked in non-agriculture | |
|---|---|---|---|---|
| ***Panel A: Direct Effect*** | (1) | (2) | (3) | (4) |
| Direct Treat x Post | −0.154 | 0.031 | 54.086 | −4.354 |
| | (0.114) | (0.084) | (40.215) | (40.228) |
| Direct Treat x Post x Close | | −0.434*** | | 134.154** |
| | | (0.164) | | (58.430) |
| F-test $p$-value: Sum of coefficients | | 0.006 | | 0.006 |
| ***Panel B: Incidental Effect*** | | | | |
| Incidental Treat x Post | −0.073 | 0.006 | −9.071 | −11.913 |
| | (0.122) | (0.126) | (35.426) | (40.153) |
| Incidental Treat x Post x Close | | −0.172 | | 16.861 |
| | | (0.228) | | (67.828) |
| F-test $p$-value: Sum of coefficients | | 0.422 | | 0.931 |
| ***Panel C: Pooled Effect*** | | | | |
| Treat x Post | −0.113 | 0.018 | 18.062 | −0.643 |
| | (0.092) | (0.093) | (30.203) | (31.712) |
| Treat x Post x Close | | −0.289* | | 63.121 |
| | | (0.163) | | (51.768) |
| F-test $p$-value: Sum of coefficients | | 0.063 | | 0.159 |
| 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. Outcome in Columns (1)–(2) is an indicator that equals 1 if the household engages in any cultivation. Outcome in Columns (3)–(4) is the total number of person-days worked in the casual labor market for non-agriculture. For each outcome, I run the direct regressions in Panel A (Eq. (2) in first column, Eq. (6) in second column), the incidental regressions in Panel B (Eq. (4) in first column, Eq. (7) in second column), and the pooled regressions (Eq. (5) in first column, Eq. (8) in second column) in Panel C. *Close* is an indicator that equals 1 for villages within 10 kilometers of the nearest town at baseline. The $p$-value from a F-test of the joint significance of the two coefficients is reported at the bottom of each panel. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

attention to the coefficient $\delta_1$, the treatment effect on households further away from towns.

Table 4 presents results on households' crop choice.[26] Households often cultivate staple cereal grains such as paddy, wheat and maize for subsistence use, and cultivate commercial non-cereal crops such as pulses, oilseeds, fruits and vegetables for sale in nearby wholesale markets. This is evident among households in my REDS sample at baseline — only 0.3 percent of cultivators with no observed market activity grow non-cereal crops, while 64.2 percent of cultivators engaged in some market activity grow non-cereal crops. Examining crop choice is thus indicative of households' intentions to retain output for home consumption or to sell output at local markets. I examine the following measures of crop choice: 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 HYVs, broken down into two categories — non-cereal and cereal.

I find a significant extensive margin response — there is a 26 percentage point increase in households cultivating any non-cereals (column 1), and an insignificant decline in households cultivating any cereals (column 2). In Appendix Table B7, I break down non-cereal cultivation by crop categories: fruits and vegetables, pulses, oilseeds and cash crops. I find that the uptick in non-cereal cultivation is particularly driven by an increase in the cultivation of pulses. There also appears to be a significant intensive margin response — conditioning on households that cultivate land in both periods, I find a 17 percentage point increase in the share of land cultivated under non-cereals (column 3).[27] Further, there is substantial take up of HYVs for non-cereal crops, significant at the 1 percent level (column 4). While take up of HYVs is also positive for cereal crops, this coefficient is not precisely estimated (column 5). Taken together, these results demonstrate that remote households that gain improved road connectivity begin to cultivate hybrid non-cereal crops on some portion of their land, while continuing to cultivate staple cereal grains. This diversification towards non-cereals is suggestive of households' intentions to market farm output.

Table 5 presents results on households' farm input use, namely productivity-enhancing material inputs and farm labor. Farm inputs are measured using indicators for the purchase of HYV seeds, irrigation (pumps, storage tanks, water), organic manure, chemical fertilizers and the number of labor-days utilized per cultivated acre, broken down into 3 categories — total, hired and family. As farm inputs are likely jointly determined, I allow for correlation between the regressions by estimating the system of equations jointly using seemingly unrelated regressions (SUR).

[26] Appendix Table B6 summarizes the treatment effects of improved rural road infrastructure on households' crop choice (columns 1 to 3) for the direct (panel A) and incidental (panel B) connections respectively.

[27] I explore impacts of improvements in road infrastructure on landholdings in Appendix Table B5. I document no significant change in the total amount of land owned (column 2), a decline in the total amount of land leased out (column 3), and no change in the gross amount of land cultivated (column 4) by households that gain improvements in road connectivity in villages further away from towns. Further, I find no significant changes in landholdings for households that gain improvements in road connectivity in villages close to towns.

**Table 4**
Crop diversification.

| | Indicator: | | Non-Cereal | High-yielding varieties: | |
|---|---|---|---|---|---|
| | Non-Cereal (1) | Cereal (2) | Area Share (3) | Non-Cereal (4) | Cereal (5) |
| Treat x Post | 0.261*** | −0.114 | 0.170* | 0.305*** | 0.177 |
| | (0.099) | (0.110) | (0.090) | (0.096) | (0.125) |
| Household FE | Yes | Yes | Yes | Yes | Yes |
| State-Year FE | Yes | Yes | Yes | Yes | Yes |
| Dep. Var Mean | 0.334 | 0.472 | 0.341 | 0.217 | 0.239 |
| Observations | 8492 | 8492 | 4400 | 8492 | 8492 |

*Notes:* Difference-in-differences regressions. I run the pooled regression in Eq. (8), restricting attention to $\delta_1$, 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 Column (3) is the share of gross land cultivated under non-cereal crops. Outcome in Columns (4)–(5) 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 (3) is conditional on households that are cultivating some land in both periods. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table 5**
Farm inputs.

| | Purchase of material inputs | | | | Labor days per acre | | |
|---|---|---|---|---|---|---|---|
| | HYV seeds (1) | Irrigation (2) | Manure (3) | Fertilizer (4) | Total (5) | Hired (6) | Family (7) |
| Treat x Post | 0.240** | 0.265** | 0.326*** | 0.200* | 28.770*** | 18.585** | 10.185 |
| | (0.117) | (0.102) | (0.076) | (0.111) | (9.550) | (7.449) | (9.803) |
| 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 Eq. (8), restricting attention to $\delta_1$, 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, organic 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. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

I find a significant increase in households' take up of material inputs (columns 1–4).[28] Further, I find substantial expansion of farm labor use, with a 49 percent increase in the total number of labor-days invested per acre of cultivated land, significant at the 1 percent level (column 5). This is driven by a 54 percent increase in the number of labor-days hired per acre of cultivated land (column 6). While I observe an increase in the number of family labor-days employed per acre of cultivated land, this effect is imprecisely estimated (column 7).[29] This suggests that households are not substituting between hired labor and family labor, but instead, increasing investments in both categories. Given strong complementarities between high-yielding variety seeds and other inputs of production[30] (Duflo and Pande, 2007; Foster and Rosenzweig, 2010; Suri, 2011), these results are consistent with households increasingly cultivating high-yielding variety crops and subsequently investing in the complementary material and labor inputs required to grow these crops successfully.

Results in Table 5 are robust to alternative definitions of the farm input variables — in Appendix Table B9, I use the log of total expenditures for each of the inputs. Results are also robust to sample restrictions — in Appendix Table B10, I restrict the sample to villages further away from towns (panel A), to eligible and incidental villages (panel B), and to eligible and incidental villages far from towns (panel C).

Table 6 presents results on four outcomes that measure households' product market activity — an indicator for whether the household engages in any crop sales, an indicator for whether the household engages in the sales of HYV crops, log crop value and log empirical profits. I find a 15 percentage point, or 29 percent increase, in entry into sales of farm output among households that received a program road (column 1). Further, there is a 23 percentage point increase in households selling HYV crops, significant at the 10 percent level (column 2). The positive coefficients in columns 3 to 4 suggest that households possibly experienced gains in crop value and farm profits, though these effects are imprecisely estimated due to large standard errors. Remote households thus transitioned from subsistence

[28] I also examine the impact of improvements in rural road infrastructure on an indicator for the purchase of mechanized farm equipment (tractors/harvesters/threshers). I find no detectable impact along this margin. This result is consistent with Asher and Novosad (2020) who find no change in ownership of mechanized farm equipment in villages that received PMGSY roads.

[29] Appendix Table B6 summarize the treatment effects of improved rural road infrastructure on households' farm input decisions (columns 4 to 8) for the direct (panel A) and incidental (panel B) 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. Transplanting, fertilizer and manure application, weeding, and irrigation system management are all labor-intensive processes. Further, as yields are typically higher, the adoption of high-yielding varieties is also associated with an increased demand for labor during harvests. Appendix Table B8 breaks down farm labor use by agricultural operation. I find a significant increase in labor-days per cultivated acre utilized in chemical fertilizer application and weeding. The coefficients on other agricultural operations are positive though imprecisely estimated.

**Table 6**
Market activity.

| | Any crop sales (1) | HYV crop sales (2) | Crop value (3) | Empirical profits (4) |
|---|---|---|---|---|
| Treat x Post | 0.149** | 0.226* | 0.474 | 0.350 |
| | (0.071) | (0.129) | (0.825) | (0.945) |
| Household FE | Yes | Yes | Yes | Yes |
| State-Year FE | Yes | Yes | Yes | Yes |
| Dependent Variable Mean | 0.515 | 0.305 | 5.686 | 5.407 |
| Observations | 8492 | 8492 | 8492 | 8135 |

*Notes:* Difference-in-differences regressions. I run the pooled regression in Eq. (8), restricting attention to $\delta_1$, 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(crop value + 1), where crop value is the sum of values of crops sold and crops retained for self-consumption. Outcome in Column (4) is log(empirical profits + 1). Outcomes in Columns (3) and (4) are winsorized at the top 1 percentile. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to market-oriented, commercial farming following improvements in rural road connectivity.[31]

The effect sizes that I find are generally larger than effect sizes found in other recent micro studies on technology adoption in agriculture, which have largely focused on short-run impacts. For example, Duflo et al. (2011) find a 47–70 percent increase in fertilizer adoption among Kenyan farmers offered free fertilizer delivery early in the season; Cole et al. (2017) find a 12 percent increase in planting of higher-return, higher-risk cash crops among Indian farmers several months after they are provided with rainfall insurance; and Emerick et al. (2016) find that farmers in rural Odisha who received flood-tolerant rice varieties modernize their agricultural production practices a year later — their plots are 14 percent less likely to be cultivated with traditional varieties, and farmers are 22 percent less likely to use the broadcasting method of planting. Relative to these papers, I estimate medium-run impacts. I observe households several years after program roads are built in their villages — the median length of time between the construction of program roads in treated villages and the 2006 REDS round is 4 years. The larger effect sizes I find could potentially be explained by differences in the length of time that has elapsed between the intervention and time of survey.

Further, while the effect sizes I find are large, it is important to highlight that the cost of the road-building intervention is very high.[32] The large effect sizes could thus also be explained by the sheer scale of the road-building intervention, relative to some of these other micro studies. A comparable study in terms of scale evaluates the impact of construction of footbridges in rural villages in Nicaragua; they find that footbridges led to a 60 percent increase in expenditure on farm inputs (Brooks and Donovan, 2020). My findings suggest that there is scope for infrastructure investments to meaningfully transform remote, agrarian economies and influence technology adoption on a very large scale.

### 6.4. Heterogeneous impacts

Given that the adoption of new technologies is associated with upfront fixed costs and scale effects (Foster and Rosenzweig, 2010), I test for heterogeneity in treatment effects by farm size. Using the within-district gross cultivated land distribution at baseline, I split the sample into two subgroups – below and above median cultivators – and fully interact Eq. (8) with a binary variable that takes the value 1 if a household is an above median cultivator at baseline. As shown in Appendix Table B12, effects are largely concentrated among small-scale (below median) cultivators — I find a significant extensive margin response in non-cereal cultivation (row 1, column 1), substantial take up of non-cereal high-yielding varieties (row 1, column 2) and a significant increase in investments in material inputs and farm labor (row 1, columns 3 to 9). As before, this increase in farm labor is driven by a significant increase in the number of labor-days hired per acre of cultivated land (row 1, column 8). I find no detectable impact for above median cultivators — none of the p-values from the F-test of joint significance are less than 0.1. Baseline non-cereal cultivation and usage of agricultural inputs are much higher for this subgroup; this suggests that small-scale cultivators are more likely to be constrained in making productive investments in agriculture in the presence of poor infrastructure.

### 6.5. Robustness checks

*Placebo test.* To further validate the parallel trends assumption underlying my difference-in-differences framework, I conduct a placebo test where I define $Treat$ to be an indicator that equals 1 if the program-eligible 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 program-eligible villages that received 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 the placebo treatment group. Table 7 summarizes the results from this placebo test. Other than the indicators for non-cereal cultivation and HYV use in villages far away from towns, the remaining coefficients are not statistically significant at standard levels, further supporting the pre-trends assumption underlying the difference-in-differences analysis.

*Accounting for multiple hypothesis testing.* Given that a large number of outcome variables are considered in the analysis, I address concerns about multiple hypothesis testing following the approach in Kling et al. (2007). I construct standardized indices for four main families of outcome variables — staple crop cultivation, crop diversification, farm inputs and market activity — using components from Tables 4 through Table 6.[33] Appendix Table B13 presents estimates of the impact of improvements in rural road infrastructure on these four indices.

[31] In Appendix Table B11, I report $\delta_1$ and $\delta_2$ from the pooled regression in Eq. (8) for the series of agricultural outcomes described above. Recall that $\delta_1 + \delta_2$ captures treatment effects of access to improved road infrastructure for households that are in close proximity to towns — in the table, I report p-values from a test of joint significance of the two coefficients. I find a reduction in non-cereal cultivation, an increase in family labor use and a reduction in crop sales for households in villages that are in close proximity to the non-agricultural sector. Given that I find substantial exit from agriculture in these villages, these reduced-form results capture the impacts of improvements in road infrastructure as well as of potential changes in the composition of households who remain in agriculture.

[32] The average per kilometer cost of road construction under the program is estimated to be US$43,300. (Source: PMGSY Online Management & Monitoring System)

[33] Each index is defined to be the equally weighted average of z-scores of its components, where the z-scores are calculated by subtracting the baseline mean and dividing by the baseline standard deviation. The components for the staple crop cultivation include an indicator for whether the household cultivates any cereal crops and an indicator for the use of cereal HYVs. The components for the crop diversification index include an indicator for whether the household cultivates any non-cereal crops and an indicator for the use of non-cereal HYVs. The components for the farm inputs index include indicators for whether the household purchases HYV seeds, irrigation, manure and fertilizer. The components for the market activity index include an indicator for any crop sales, an indicator for HYV crop sales, crop value and empirical profits.

**Table 7**
Placebo test using roads constructed between 2007–2010.

| | Engaged in cultivation | Non-cereal crops | | | Purchase of material inputs | | | | Total labor days per acre |
|---|---|---|---|---|---|---|---|---|---|
| | | Indicator | Area share | HYV | HYV Seeds | Irrigation | Manure | Fertilizer | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Direct Treat x Post | −0.084 | −0.091* | −0.045 | −0.071* | −0.002 | −0.034 | −0.046 | 0.050 | 22.446 |
| | (0.107) | (0.047) | (0.053) | (0.042) | (0.089) | (0.076) | (0.079) | (0.212) | (18.573) |
| Direct Treat x Post x Close | −0.110 | 0.023 | 0.163 | 0.128 | −0.288 | −0.043 | 0.044 | −0.220 | 11.613 |
| | (0.203) | (0.174) | (0.164) | (0.101) | (0.245) | (0.333) | (0.091) | (0.277) | (30.409) |
| $p$-value: sum of coefficients | 0.262 | 0.686 | 0.467 | 0.538 | 0.208 | 0.812 | 0.959 | 0.351 | 0.120 |
| Household FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State-Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 7914 | 7914 | 4038 | 7914 | 7914 | 7914 | 7914 | 7914 | 4038 |

*Notes:* Difference-in-differences regressions. I run the regression in Eq. (6) with $DirectTreat_v$ equals 1 if village $v$ receives a program road between 2007 and 2010. I drop all program-eligible households that received roads before 2007; the relevant control group consists of program-eligible villages that received program roads after 2010 (N(treatment villages)=9, N(control villages)=13). Outcome in Column (1) is an indicator that equals 1 if the household engages in any agricultural cultivation. Outcome in Column (2) is an indicator that equals 1 if the household engages in cultivation of non-cereal crops. Outcome in Column (3) is the share of gross land cultivated under non-cereal crops. Outcome in Column (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. Outcome in Columns (5)–(8) is an indicator that equals 1 if the household purchases high-yielding variety (HYV) seeds, irrigation, manure, and chemical fertilizer respectively. Outcome in Column (9) is the total number of labor-days per cultivated acre. The regressions in Columns (3) and (9) are conditional on households that are cultivating some land in both periods. The $p$-value from a F-test of the joint significance of the two coefficients is reported at the bottom of the panel. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Consistent with the findings in Section 6.3, I find no detectable impact on the index for staple crop cultivation (column 1) and large positive impacts across the other three indices (columns 2 through 4).

*Alternate specifications of proxy for access to non-agriculture.* Throughout my analysis, I use a binary indicator for villages within 10 kilometers of the nearest town at baseline to proxy for access to non-agriculture. I test the sensitivity of my results to two alternate specifications of this proxy in Appendix Table B14 – 5 (columns 1 and 2) and 15 kilometers (columns 3 and 4) – and I find that the results are largely consistent. Among directly treated households in villages within 5 kilometers of the closest town, I find a significant decline in households' engagement in cultivation (F-test $p$-value of 0.013), accompanied by an increase in person-days worked in the casual labor market for non-agriculture (F-test $p$-value of 0.003). These effects are larger in magnitude than the effects I find with the 10 kilometer cutoff in Table 3, lending further support to the argument that non-agricultural opportunities are spatially concentrated close to towns. When I move to the 15 kilometer cutoff, the impacts are less precisely estimated but the sign of the coefficients point in the right direction. In Appendix Table B15, I estimate the pooled treatment effects of improvements in rural road infrastructure on households' crop choices and farm input decisions using a 5 (panel A) and 15 kilometer (panel B) cutoff; I find that the results are largely consistent with my findings in Tables 4 and 5.

*Sample selection.* In constructing my balanced panel, I restrict my sample to households that appeared in both waves of the REDS and aggregate partitioned households back to the 1999 household unit level. In Appendix Table B16, I test for differential attrition and splitting of households across treatment and control villages. None of the coefficients are significant at standard levels, which suggests that there was no differential attrition or splitting of households as a result of the road-building program. Finally, in Appendix Table B17, I show that my results are robust to excluding all households that split over time.

## 7. Potential channels

In this section, I explore numerous potential channels that might explain my results: (1) increased flows of agricultural laborers, (2) increased flows of traders and extension workers, (3) increased access to agricultural markets where farm inputs and output are sold, (4) increased access to credit markets, and (5) greater information diffusion and social learning.

*Access to agricultural labor.* Improvements in road infrastructure can facilitate the movement of agricultural workers across space. Previous work has discussed the role of labor availability in influencing adoption of new agricultural practices; in particular, they attribute low take up of labor-intensive production technologies such as hybrid seeds to low rural labor supply as well as to seasonality in the demand for labor (Hicks and Johnson, 1979; Feder et al., 1985).[34] By connecting adjacent village labor markets across space, road infrastructure effectively increases the rural labor supply an agricultural household can draw from. This in turn enables households to switch to labor-intensive technologies and intensify agricultural hiring as needed to support farm operations. I present three pieces of evidence consistent with improved road infrastructure increasing the mobility of agricultural workers across newly connected labor markets.

First, when I restrict my analysis to the incidental sample only, I find that the effects on crop diversification, input use and labor hiring persist (Appendix Table B6). Given that a majority of villages in the incidental sample were already connected to a market center at baseline, a new road linkage to a previously unconnected village is unlikely to have a first order effect on the movement of goods. However, given that labor markets are local due to limited labor mobility, the reduced form effects could be driven by a change in access to agricultural workers from a previously isolated village labor market.

Second, if agricultural workers are indeed moving across space, I expect to see variation in labor hiring by tightness of the labor market — workers should flow from regions with labor surpluses to regions with labor shortages. While I am unable to directly observe worker flows, I can test for heterogeneity in treatment effects along this margin by exploiting relative labor densities across villages. In this case, I expect workers to flow from villages with relatively higher labor density (proxy for "labor surplus" regions) to villages with relatively lower labor density (proxy for "labor shortage" regions). I thus hypothesize that the treatment effects on hired labor will be larger in villages with relatively lower labor density at baseline. I compute baseline labor density (defined as total number of main workers/village area using the 2001 Indian Census) for each village in my sample and for all surrounding villages within a close radius. I then construct two

[34] For example, Feder et al. (1985) write, "New technologies may increase the seasonal demand for labor, so that adoption is less attractive for those with limited family labor or those operating in areas with less access to labor markets."

**Table 8**
Heterogeneity in labor use, by relative labor density at baseline.

| | Total labor days per acre | | | | Hired labor days per acre | | | |
|---|---|---|---|---|---|---|---|---|
| | Below Mean=1 (1) | Below Mean=0 (2) | Lowest=1 (3) | Lowest=0 (4) | Below Mean=1 (5) | Below Mean=0 (6) | Lowest=1 (7) | Lowest=0 (8) |
| Treat x Post | 34.518** | 2.875 | 62.665* | 21.490 | 20.037 | 12.518 | 36.736*** | 9.348 |
| | (13.915) | (8.999) | (36.966) | (15.437) | (12.912) | (9.506) | (12.170) | (7.022) |
| Household FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State-Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Dependent Variable Mean | 64.145 | 50.653 | 63.377 | 57.974 | 39.454 | 26.915 | 40.901 | 32.710 |
| Observations | 2304 | 1906 | 732 | 3478 | 2304 | 1906 | 732 | 3478 |

*Notes:* Difference-in-differences regressions. I run the pooled regression in Eq. (8), restricting attention to $\delta_1$, the treatment effect on households further away from towns. I construct two measures of relative labor density: (1) an indicator that takes the value 1 if the labor density of a village is below the mean labor density of its surrounding villages (columns 1 and 5), and 0 otherwise (columns 2 and 6); (2) an indicator that takes the value 1 if the labor density of a village is lower than the labor density of its surrounding villages (columns 3 and 7), and 0 otherwise (columns 4 and 8). Outcome in Columns (1)–(4) is the total number of labor-days per cultivated acre. Outcome in Columns (5)–(8) is the total number of hired labor-days per cultivated acre. The regressions are conditional on households that are cultivating some land in both periods. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

measures of relative labor density: an indicator that takes the value 1 if the labor density of a village is below the mean labor density of its surrounding villages, and an indicator that takes the value 1 if the labor density of a village is lower than the labor density of all its surrounding villages. Results from this heterogeneity analysis are presented in Table 8. Consistent with my hypothesis, I find that the positive treatment effects on labor use (total labor-days and hired labor-days per acre of cultivated land) are driven by "labor shortage" villages that have relatively lower labor densities at baseline (columns 1, 3, 5 and 7) using either constructed measure.

Third, I examine the reduced form effect of improved road connectivity on households' ownership of bicycles and scooters, two important modes of transport in this setting, in Table 9.[35] I construct indicators that equal 1 if the household reports owning a bicycle and scooter respectively. I find a 35 percent increase in bicycle ownership among households that received a program road, significant at the 5 percent level (column 1), and a non-significant increase in scooter ownership (column 2). This suggests that households increasingly invest in assets that allow for greater mobility upon gaining access to improved rural road infrastructure. Undoubtedly, bicycle ownership extends beyond enabling households to seek outside employment — for example, it might also increase access to banking services, or enable employers to search for workers beyond village boundaries. I now turn to supplementary survey evidence from the field to shed light on the role of infrastructure in influencing agricultural production in previously connected villages. I conducted a short survey with 114 male cultivators in 18 rural villages – all of which have all-weather road access and are greater than 10 km from the closest town — across 3 districts in Odisha, India.[36] The cultivators in the survey sample are similar on key observables to cultivators in the REDS sample at baseline; for example, the average amount of cultivated land is 2.5 acres (2.4 acres in REDS sample), and 53 percent cultivate non-cereal crops (56 percent in REDS sample).

Despite having road connectivity, 42 percent of cultivators express facing difficulties in finding enough hired laborers to work on their land in the past agricultural cycle (Fig. 4, Panel A), with a majority attributing this to a high demand for workers at the same time. In contrast, only 5 percent and 7 percent of cultivators express having difficulty procuring agricultural inputs and selling farm output respectively. This is consistent with seasonality in the demand for labor, and suggests that establishing a reliable supply of agricultural labor in the peak season remains to be a concern among cultivators with access to a road network.

I present cultivators with a series of questions to better understand their labor hiring practices. 92 percent 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 percent report knowing laborers inside the village better, while 37 percent 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 percent report having to pay a higher wage in cash, while 82 percent report having to pay more in kind. This emphasis on in-kind transfers could also potentially explain why the casual daily agricultural cash wage does not appear to respond in Table 9.

Next, I present cultivators with a series of hypothetical scenarios pertaining to agriculture (Fig. 4, Panel B). In the first hypothetical scenario, I describe the option of hiring agricultural labor from two equidistant villages – one with a hard-topped road and one with a dirt road – and elicit responses on hiring preferences. While 99 percent of respondents report that workers from both villages would be equally hardworking, 76 percent express a strict preference for hiring from the village with a hard-topped road. In the second hypothetical scenario, 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 percent of cultivators report that it would be easier for them to find workers to hire for work on their land and 29 percent report that they would be able to hire workers more frequently, while only 8 percent report that it would be less expensive to hire labor. This suggests the existence of underlying frictions in the rural labor market – for example, transaction costs associated with hiring agricultural labor (Binswanger and Rosenzweig, 1984) – that are alleviated by improvements in connectivity. Taken together, these results suggest that even in villages that have existing road connectivity at baseline, establishing a reliable supply of agricultural labor remains to be an important concern during the agricultural peak season.

*Access to extension workers and traders.* Improvements in road infrastructure can increase access to key agricultural agents such as extension workers and traders (Owens et al., 2003; Anderson and Feder, 2007). The most recent round of the REDS survey included modules on households' sources of information on issues pertaining to agriculture – types of agricultural credit, input and output prices – as well as households' crop production practices. Since this module was only added to the 2006 REDS survey, I am unable to run the difference-in-differences analysis on this set of outcomes. Instead, I plot the frequency at which households report extension workers (panel A), traders (panel B) and village elders/neighbors/friends (panel C)

[35] Car ownership is very low in this setting - only 0.1 percent 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 5.3 on commuting patterns among rural Indian workers.

[36] This survey was conducted in November 2016 in 3 neighboring districts: Khordha, Nayagarh and Cuttack.

**Table 9**
Farm prices & asset ownership.

| | Assets | | | Input prices | | | Output prices | |
|---|---|---|---|---|---|---|---|---|
| | Bicycle (1) | Scooter (2) | Mobile phone (3) | HYV Seeds (4) | Fertilizer (5) | Casual agri labor (6) | Paddy (7) | Wheat (8) |
| Treat x Post | 0.160** | 0.003 | −0.029 | 5.881 | 1.294 | −2.888 | 0.860 | −1.179 |
| | (0.076) | (0.005) | (0.092) | (4.108) | (1.163) | (6.457) | (1.245) | (1.488) |
| Household FE | Yes | Yes | Yes | No | No | No | No | No |
| Village FE | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Dependent Variable Mean | 0.449 | 0.018 | 0.055 | 23.697 | 5.880 | 42.836 | 5.133 | 6.171 |
| Observations | 8492 | 8492 | 8492 | 2443 | 5298 | 4530 | 3103 | 2704 |

*Notes:* Difference-in-differences regressions. I run the pooled regression in Eq. (8), restricting attention to $\delta_1$, the treatment effect on households further away from towns. Outcomes in Columns (1)–(3) are indicators that equal 1 if the household owns a bicycle, a scooter or a mobile phone respectively. Outcomes in Columns (4)–(5) are imputed prices of HYV seeds and fertilizer (Rs./unit) respectively. Outcome in Column (6) is the casual agricultural daily cash wage (Rs./day), measured at the household-member level. Outcomes in Columns (7)–(8) are imputed unit prices of paddy and wheat (Rs./kg) respectively. All regressions include state-year fixed effects. Observations are weighted using 1999 sample weights. Standard errors, clustered at the village level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

as their most important source of information in Appendix Figure A3, and the frequency at which households report changing practices in the past 7 years in Appendix Figure A4. While these results are merely suggestive given the lack of baseline data, the absence of stark differences across treatment and control groups suggests that a change in access to extension workers and traders is not a primary driver of the reduced-form effects.

*Access to agricultural markets.* Improvements in road infrastructure can also lower transport costs of goods and subsequently increase access to input and product markets. To test this hypothesis, I estimate the effects of improved rural road connectivity on the imputed price of high-yielding variety seeds, fertilizer and the cash wage for casual daily agricultural laborers in Table 9.[37] I do not find any statistically significant impact on the prices of these farm inputs (columns 4 to 6).[38] Next, I estimate effects on the imputed price of the two most commonly grown crops – paddy and wheat – in Table 9. Similarly, I find no detectable impact on the prices of farm outputs (columns 7 and 8). At first glance, this null result on prices appears to be surprising. However, given that improvements in road infrastructure impact both supply and demand – for example, it lowers transport costs for traders supplying seeds as well as for farmers demanding seeds – the effect on output prices is theoretically ambiguous ex-ante.[39] The null effect on agricultural cash wages is also consistent with the presence of nominal wage rigidity in markets for casual daily agricultural labor, which has been documented in the rural Indian setting (Kaur, 2019).

*Access to credit markets.* Alternatively, improvements in road infrastructure can increase access to credit, possibly relaxing liquidity constraints faced by farmers (Croppenstedt et al., 2003; Karlan et al., 2014). I explore the effects of improved road connectivity on households' demand for credit from five sources – banks, shopkeepers, moneylenders, and friends and family networks inside and outside the village – in Appendix Table B18. While imprecisely estimated, the point estimates are suggestive of an increase in demand for formal credit. It is worth noting that this result does not shed light on possible changes in the total volume of credit available to farmers.[40]

*Access to information.* Another possibility is that improvements in road infrastructure enhance information flows. Prior work has shown that information and communication technologies (ICTs) such as mobile phones have played an important role in the dissemination of price information and in linking buyers and sellers in rural settings (Jensen, 2007; Aker, 2010). I test for differences in mobile phone ownership by treatment status in Table 9; I find no significant effect (column 3). This suggests that while the rollout of program roads coincided with rapid growth of ICTs,[41] this growth did not vary differentially by treatment status under the program. Further, given that the agricultural technologies I consider have existed in this setting for many years – the introduction of high-yielding seeds dates back several decades to the Green Revolution in the 1960s – it is unlikely that low take up is driven by the lack of knowledge of proper use or the expected returns associated with these technologies.

*Social learning.* Finally, improvements in road infrastructure can facilitate social learning. Previous work has documented that learning from others is an important driver of technology adoption in agriculture (Foster and Rosenzweig, 1995; Conley and Udry, 2010). I test for heterogeneity in treatment effects across villages with different levels of non-cereal cultivation at baseline. If social learning is important, I should expect to find larger treatment effects in villages with below median non-cereal cultivation at baseline, relative to above median villages where non-cereals are already prevalent among farmers.[42] I fully interact Eq. (8) with a binary variable that takes the value 1 if the village has above median non-cereal cultivation at baseline in Appendix Table B20. I find statistically significant impacts for households in below median villages at baseline. These impacts are also significant for households in above median villages – the triple interaction terms reflecting differences in the subgroups are mostly small and insignificant – suggesting that social learning is not a key driver of the reduced form effects.

[37] Price observations are at the household level, and wage observations are at the household-member level. I only observe prices and wages for households that buy material and labor inputs and sell farm output. Prices are imputed by taking the total value divided by the total quantity of the input purchased/output sold. I do not observe where the inputs are purchased, and where the output is sold. To maximize power, I use all observations in a pooled specification with village fixed effects.

[38] I cannot rule out the possibility that input quality may also be changing. Given data limitations, I am unable to directly test for changes in input quality.

[39] Casaburi et al. (2013) provide a good summary of the various models of market structure that have different predictions for price effects in response to infrastructure investments.

[40] Road construction itself could have relaxed liquidity constraints faced by households if they were hired by contractors to work in construction of program roads. Anecdotally, this does not seem to have been the case as contractors required skilled workers who were able to operate the machinery required to pave the roads. As such, they typically brought full-time migratory workers from other districts to work on these projects. As a robustness check, I run my analysis excluding all households who were treated in the year right before the survey. Given the timing, this group of households would have experienced the largest impact if road construction itself generated income that relaxed households' liquidity constraints. The results hold when I drop these households in Appendix Table B19, suggesting that wage income from road construction is not an important confound.

[41] Between 1999 and 2006, mobile phone ownership among REDS households grew from 7 percent to 42 percent.

[42] Foster and Rosenzweig (2010) define learning as "taking place when new information affects behavior and results in outcomes for an individual that are closer to the (private) optimum. Thus, in an environment where there is no new information, learning is unlikely, while in a setting in which a new technology or input is introduced, learning should be important."

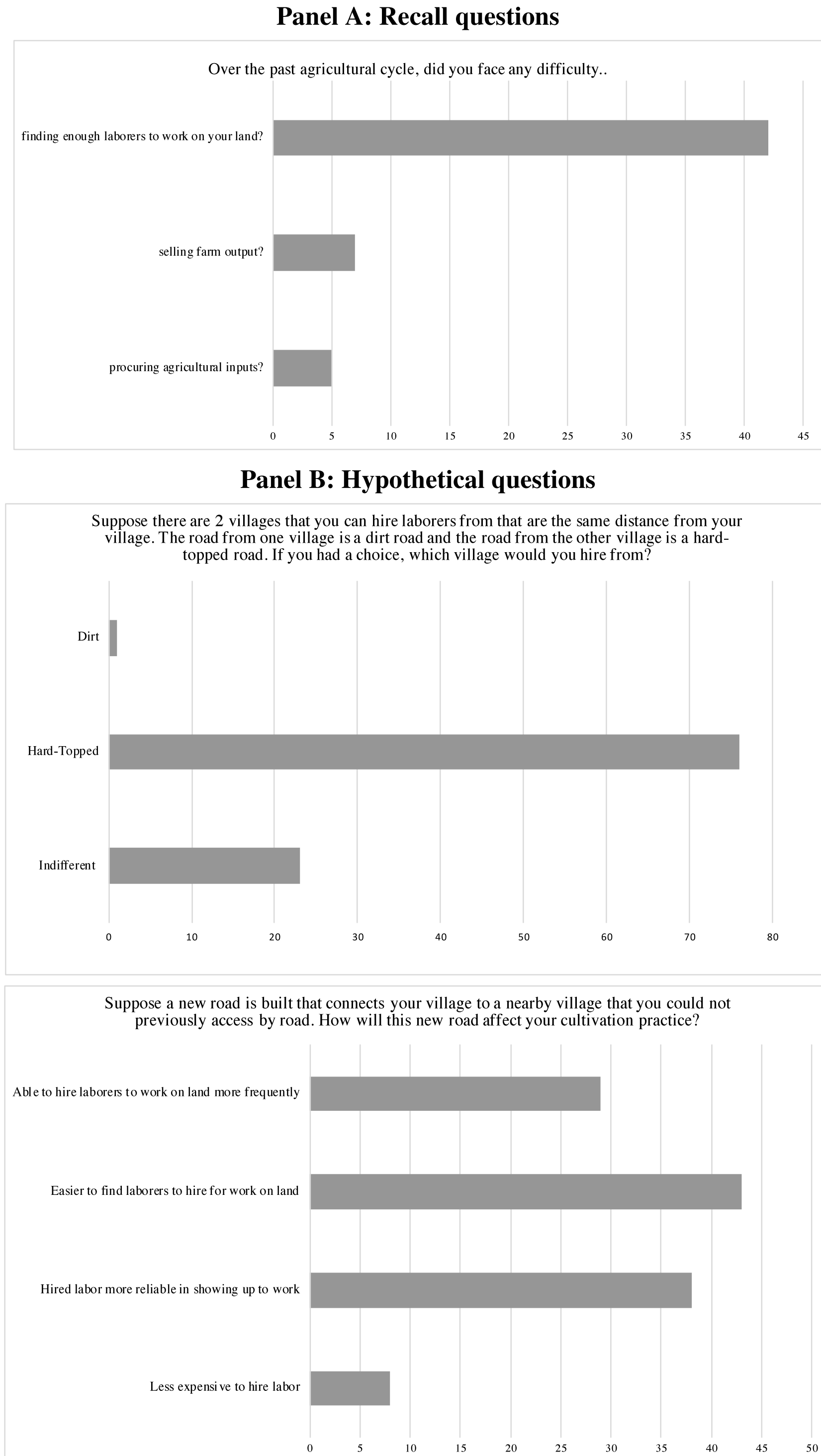


**Fig. 4. Survey evidence.** This figure summarizes surveys conducted with 114 male cultivators in remote villages with road connectivity across 3 districts in rural Odisha, India. Respondents were asked a series of questions pertaining to the past agricultural cycle (panel A) as well as several hypothetical scenarios on agriculture (panel B). The $x$-axis shows the percentage of respondents choosing a given answer.

While my results are consistent with labor mobility playing an important role, I cannot definitively rule out other alternative channels at play. However, I have provided several pieces of evidence that, taken together, suggests that they are unlikely to fully account for the reduced form effects described in Section 6.

## 8. Conclusion

This paper provides causal evidence for the effects of improvements in transport infrastructure on production decisions in agriculture for remote villages in rural India. The provision of hard-topped, all-weather roads led to crop diversification, modernization of cultivation practices through the adoption of improved inputs and technologies, increased labor hiring and commercialization of farm output. Further evidence is consistent with these effects operating through an increase in mobility of agricultural workers across connected village labor markets. Taken together, this evidence suggests that poorly integrated rural labor markets can serve as a substantial barrier to productive investments in agriculture.

Given a strong policy interest in both the provision of infrastructure and the adoption of agricultural technologies as means towards sustainable poverty reduction – the Central Government of India allocated USD 14 billion to the roads sector and USD 11 billion to fertilizer subsidies for rural Indian farmers in the 2016–2017 Budget[43] – this suggests that there are potential gains that can arise from coupling infrastructure projects with other commonly used policy instruments such as fertilizer and irrigation subsidies.[44] Further, given that treatment effects are heterogeneous across space, spatial targeting should be made an important consideration when designing such policy interventions. These findings also demonstrate that the gains that accrue to agricultural households extend beyond the households targeted by the program, highlighting the importance of incorporating network effects in the design of transport policy.

The Rural Access Index developed at World Bank estimates that over one billion people, 98 percent of them in developing countries, live more than 2 kilometers away from the nearest all-weather road. Improving mobility through infrastructure thus remains a key priority in economic development policy not only in India, but across many other developing countries. These findings can be used to inform an ongoing policy debate on whether costly infrastructure investments in remote areas can be justified by the benefits that accrue to agricultural households.

## CRediT authorship contribution statement

**Yogita Shamdasani:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, writing - review & editing, Visualization, Supervision, Project administration.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102686.

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

[44] As an example, Gebresilasse (2019) documents strong complementarities between rural roads and extension services in Ethiopia. He finds that rural road infrastructure and extension services are ineffective in isolation, but lead to large productivity gains when coupled together.

## References

Aggarwal, S., 2018. Do rural roads create pathways out of poverty? Evidence from India. J. Dev. Econ.
Aker, J.C., 2010. Information from markets near and far: Mobile phones and agricultural markets in Niger. Am. Econ. J. Appl. 2 (3), 46–59.
Ali, R., 2011. Impact of rural road improvement on high yield variety technology adoption: Evidence from Bangladesh, Working Paper.
Allen, T., Atkin, D., 2016. Volatility and the gains from trade, Working Paper.
Anderson, J.R., Feder, G., 2007. Agricultural extension. Handb. Agric. Econ. 3, 2343–2378.
Asher, S., Lunt, T., Matsuura, R., Novosad, P., 2019. The socioeconomic high-resolution rural-urban geographic dataset on India (SHRUG), Working Paper.
Asher, S., Novosad, P., 2020. Rural roads and local economic development. Amer. Econ. Rev.
Bandiera, O., Rasul, I., 2006. Social networks and technology adoption in Northern Mozambique. Econ. J. 116 (514), 869–902.
Banerjee, A., Duflo, E., Qian, N., 2021. On the road: Transportation infrastructure and economic growth in China. J. Dev. Econ.
Benjamin, D., 1992. Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. Econometrica 287–322.
Binswanger, H.P., Rosenzweig, M.R., 1984. Contractual Arrangements, Employment, and Wages in Rural Labor Markets in Asia. Yale University Press.
Breza, E., Kaur, S., Shamdasani, Y., 2021. Labor rationing, NBER Working Paper 28643.
Brooks, W., Donovan, K., 2020. Eliminating uncertainty in market access: The impact of new bridges in rural Nicaragua. Econometrica.
Bryan, G., Chowdhury, S., Mobarak, A.M., 2014. Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. Econometrica 82 (5), 1671–1748.
Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., Padro i Miquel, G., 2015. The value of democracy: Evidence from road building in Kenya. Amer. Econ. Rev. 105 (6), 1817–1851.
Bustos, P., Caprettini, B., Ponticelli, J., 2016. Agricultural productivity and structural transformation. evidence from Brazil. Amer. Econ. Rev. 106 (6), 1320–1365.
Casaburi, L., Glennerster, R., Suri, T., 2013. Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone, Working Paper.
Cole, S., Fernando, N., 2021. Mobile'izing agricultural advice: Technology adoption, diffusion and sustainability. Econ. J.
Cole, S.A., Giné, X., Vickery, J.I., 2017. How does risk management influence production decisions? Evidence from a field experiment. Rev. Financ. Stud. 30 (6), 1935–1970.
Conley, T.G., Udry, C.R., 2010. Learning about a new technology: Pineapple in Ghana. Amer. Econ. Rev. 100 (1), 35–69.
Croppenstedt, A., Demeke, M., Meschi, M.M., 2003. Technology adoption in the presence of constraints: the case of fertilizer demand in Ethiopia. Rev. Dev. Econ. 7 (1), 58–70.
Dercon, S., Gilligan, D.O., Hoddinott, J., Woldehanna, T., 2009. The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. Am. J. Agric. Econ. 91 (4), 1007–1021.
Donaldson, D., 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. Amer. Econ. Rev. 108 (4–5), 899–934.
Donaldson, D., Hornbeck, R., 2016. Railroads and American economic growth: A "market access" approach. Q. J. Econ.
Duflo, E., Kremer, M., Robinson, J., 2011. Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. Amer. Econ. Rev. 101 (6), 2350–2390.
Duflo, E., Pande, R., 2007. Dams. Q. J. Econ. 122 (2), 601–646.
Emerick, K., de Janvry, A., Sadoulet, E., Dar, M.H., 2016. Technological innovations, downside risk, and the modernization of agriculture. Amer. Econ. Rev. 106 (6), 1537–1561.
Fafchamps, M., 1993. Sequential labor decisions under uncertainty: An estimable household model of West-African farmers. Econometrica 1173–1197.
Feder, G., Just, R.E., Zilberman, D., 1985. Adoption of agricultural innovations in developing countries: A survey. Econom. Dev. Cult. Chang. 33 (2), 255–298.
Foster, A.D., Rosenzweig, M.R., 1995. Learning by doing and learning from others: Human capital and technical change in agriculture. J. Political Econ. 1176–1209.
Foster, A.D., Rosenzweig, M.R., 2010. Microeconomics of technology adoption. Annu. Rev. Econ. 2, 395–424.
Gebresilasse, M., 2019. Rural roads, agricultural extension, and productivity, Working Paper.
Gertler, P.J., Gonzalez-Navarro, M., Gracner, T., Rothenberg, A.D., 2016. Road quality, local economic activity, and welfare: Evidence from Indonesia's highways, Working Paper.
Gibson, J., Olivia, S., 2010. The effect of infrastructure access and quality on non-farm enterprises in rural Indonesia. World Dev. 38 (5), 717–726.
Gibson, J., Rozelle, S., 2003. Poverty and access to roads in Papua New Guinea. Econ. Dev. Cult. Chang. 52 (1), 159–185.
Henderson, J.V., Storeygard, A., Weil, D.N., 2011. A bright idea for measuring economic growth. Am. Econ. Rev. Pap. Proc. 101 (3), 194–199.
Hicks, W.W., Johnson, S., 1979. Population growth and the adoption of new technology in colonial Taiwanese agriculture. J. Dev. Stud. 15 (4), 289–303.
Jacoby, H.G., 2000. Access to markets and the benefits of rural roads. Econ. J. 110 (465), 713–737.
Jedwab, R., Moradi, A., 2016. The permanent effects of transportation revolutions in poor countries: Evidence from Africa. Rev. Econ. Stat. 98 (2), 268–284.
Jensen, R., 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian Fisheries sector. Q. J. Econ. 122 (3), 879–924.
Jones, M.R., Kondylis, F., Loeser, J.A., Magruder, J., 2020. Factor market failures and the adoption of irrigation in Rwanda, Working Paper.
Karlan, D., Osei, R.D., Osei-Akoto, I., Udry, C., 2014. Agricultural decisions after relaxing credit and risk constraints. Q. J. Econ. 597–652.
Kaur, S., 2019. Nominal wage rigidity in village labor markets. Amer. Econ. Rev.
Khandker, S.R., Bakht, Z., Koolwal, G.B., 2009. The poverty impact of rural roads: Evidence from Bangladesh. Econom. Dev. Cult. Chang. 57 (4), 685–722.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.
LaFave, D., Thomas, D., 2016. Farms, families, and markets: New evidence on completeness of markets in agricultural settings. Econometrica 84 (5), 1917–1960.
Mu, R., Van de Walle, D., 2011. Rural roads and local market development in Vietnam. J. Dev. Stud. 47 (5), 709–734.
Munshi, K., 2004. Social learning in a heterogeneous population: Technology diffusion in the Indian green revolution. J. Dev. Econ. 73 (1), 185–213.

Munshi, K., Rosenzweig, M., 2016. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. Amer. Econ. Rev. 106 (1), 46–98.

Owens, T., Hoddinott, J., Kinsey, B., 2003. The impact of agricultural extension on farm production in resettlement areas of Zimbabwe. Econom. Dev. Cult. Chang. 51 (2), 337–357.

PMGSY, 2001. Manual for the preparation of district rural road plans. http://pmgsy.nic.in/pmg61.asp.

PMGSY, 2005. Operations manual. https://pmgsy.nic.in/sites/default/files/pdf/opman_feb.pdf.

PMGSY, 2012. Program guidelines. http://omms.nic.in/ReferenceDocs/PMGSY_Guidelines.pdf.

Porteous, O., 2019. High trade costs and their consequences: An estimated dynamic model of african agricultural storage and trade. Am. Econ. J. Appl. 11 (4), 327–366.

Shrestha, S.A., 2020. The effects of rural roads on farmland values: Evidence from the topography-based road network in Nepal. Econom. Dev. Cult. Chang.

Singh, I., Squire, L., Strauss, J., et al., 1986. Agricultural Household Models: Extensions, Applications, and Policy. Johns Hopkins University Press.

Suri, T., 2011. Selection and comparative advantage in technology adoption. Econometrica 79 (1), 159–209.