The Employment Effects of Road Construction in Rural India *

PRELIMINARY - PLEASE DO NOT CITE WITHOUT PERMISSION

Sam Asher†
samuel.asher@economics.ox.ac.uk

Paul Novosad‡

Abstract

Despite a broad consensus that high transportation costs are a large barrier to economic development, many of the world's poor live in rural communities without paved roads, impeding their access to outside markets and public services. Nevertheless, the impact of rural roads on economic activity has proven difficult for economists to assess. We first motivate our empirical approach by demonstrating the presence of large-scale manipulation of administrative records in order to improve the likelihood of road treatment. We then introduce a novel empirical strategy based on alternative government data. We use three estimation strategies that take advantage of the allocation rules of a large-scale rural road construction program in India to estimate the impact of village feeder roads on rural nonfarm economic activity. We first provide OLS estimates that are based on the time when rural roads were constructed. Second, we use a regression discontinuity around population cutoffs for program eligibility. Third, we instrument for road construction using the population rank of villages within districts, which dictates the order of road construction. We find that new paved roads lead to large increases in rural employment, as measured in both firm and demographic censuses. We then provide evidence that these effects differ based on the size of the village, with evidence suggesting that roads lower unemployment in smaller villages while facilitating structural transformation in larger ones.

JEL Codes: O12/P16/D72.

*We are thankful for useful discussions with Alberto Alesina, Josh Angrist, Lorenzo Casaburi, Shawn Cole, Ed Glaeser, Ricardo Hausmann, Richard Hornbeck, Lakshmi Iyer, Asim Khwaja, Michael Kremer, Sendhil Mullainathan, Rohini Pande, Andrei Shleifer and David Yanagizawa. We are grateful to Sandesh Dhungana for excellent research assistance. Mr. PC Mohanan of the Indian Ministry of Statistics has been invaluable in helping us use the Economic Census. This project received financial support from the Center for International Development and the Warburg Fund (Harvard University). All errors are our own.

†Oxford University

‡Dartmouth College

1 Introduction

Universal access to paved roads, much like clean water and consistent electricity, remains an unreached goal in many developing countries, particularly in rural areas. Fifty-four years after independence, 33% of Indian villages did not have a paved approach road in 2001 (Population Census). The absence of such infrastructure raises trade costs and reduces access to both outside markets and government services. The high costs of infrastructure investments mean that both economic and political considerations tend to guide their placement, posing challenges for researchers seeking to understand their effects. In this paper we exploit the allocation rules of a large-scale rural road construction program in India to estimate the impact of feeder roads on rural nonfarm economic activity.

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The government developed specific guidelines to prioritize large, unconnected habitations: those with populations above 1000 were to receive highest priority, followed by those with populations above 500. Eligibility lists were to be prepared at the district level, with the largest habitations not already connected to high quality roads receiving the highest priority in each district. At the outset, about 170,000 habitations were eligible for the program, a number that has grown as the guidelines have been expanded to include smaller habitations. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).

The program rules provide us with three distinct ways to estimate the impact of a new road on rural economic activity. First, we provide OLS estimates of the relationship between PMGSY road construction and employment growth, based on the timing of road construction in villages that eventually received roads under PMGSY. However, endogenous timing of road construction creates unknown bias in these estimates. We address this in two ways. Program rules create discontinuities in the likelihood of receiving a road at populations of 500 and 1000, allowing us to use a fuzzy regression discontinuity approach to estimate the impact. Finally, we take advantage of the fact that planning and implementation was carried out at the district level. The probability of receiving a road early in the program was therefore a function not only of village size, but also of the relative size of a village within its district. We are thus able to instrument for road treatment with the within-district population rank, controlling flexibly for population.

We construct a new dataset that combines data on road construction with village characteristics and economic outcomes. We match Population Census data (1991, 2001, 2011) to Economic Census data (1998, 2005) to measure the economic consequences of road construction during the first eleven years of the PMGSY (2000-11). The Economic Census is a complete enumeration of nonfarm economic establishments in India, covering over 4000 towns and 500,000 villages. It contains, among other variables, data on employment and industrial sector for each establishment, allowing us to estimate the effect of rural roads on employment growth, firm size and formality.

We first provide evidence of significant manipulation of habitation population as reported by PMGSY, particularly around the population cutoffs that determine eligibility. Such manipulation violates any instrument variables or regression discontinuity empirical strategy using official program data. To solve this problem, we incorporate separately collected habitation population data from the National Habitation Survey conducted by the Department of Drinking Water Supply. These population numbers do not appear to be manipulated. Reassuringly, unlike in PMGSY, we find no evidence that the Ministry of Water Resources uses these data to allocate resources.

We interpret the construction of a village feeder road under the PMGSY as a large reduction in the costs of moving goods, capital and people to and from a village. Theoretically, it is unclear what impact such a change in factor mobility should have on village economic activity. For example, out-migration may increase as the costs of travel decrease, or fall as economic opportunity expands in the village. Nonfarm employment may grow as firms serving outside demand now face sufficiently low transportation costs to be competitive in outside markets; alternatively, villages may further specialize in agriculture as trade for outside goods becomes more feasible. Increasing returns to scale, such as those due to fixed costs, would predict that certain firms may only exist when transport costs are low enough to enable them to access markets beyond the confines of a village. Thus road construction may not result in the specialization predicted by simple trade models but rather diversification of the village economy as it becomes integrated into a larger trade network.

We find that the construction of a road results in a significant increase in nonfarm village employment. Instrumental variables estimates predict greater than 50 percent growth in nonfarm employment upon receiving a road, as measured in the Economic Census of 2005. As we assign treatment based on road completion, we interpret these not as temporary increases due to ongoing construction but medium-term changes in the level of village employment. We further investigate these results using the Population Census of 2011, where a fuzzy regression discontinuity design finds further evidence of employment growth as a result of road treatment. As we are able to exploit two population cutoffs (500 and 1000), we are able to test for the effect of habitation size on treatment effect. Our results suggest quite different effects depending on habitation size: smaller habitations seem to reduce unemployment in response to road construction, while larger habitations experience an increase in population and a decrease in agricultural employment. We interpret this evidence as potential signs of structural transformation in the large habitations.

The rest of the paper proceeds as follows: Section 2 summarizes the most relevant literature on transport costs and rural roads. Section 3 provides a description of the PMGSY rural road construction program. Section 4 describes our empirical strategies. Section 5 explains the data used. Section 6 presents results and discussion. Section 7 concludes.

2 Literature

Recent years have seen a renewed interest in the importance of transportation costs in facilitating growth and development. Limao and Venables (2001) use quotes from a shipping company and other sources of data to estimate the impact of inter-country transportation costs on trade flows, concluding that much of the low trade volume in sub-Saharan Africa is due to high transportation costs that result from poor quality infrastructure. Djankov et al. (2010) estimate that every additional day required to ship a container between two countries is associated with a reduction in bilateral trade of more than 1%, in addition to causing a distortion in trade away from time-sensitive exports. Intra-national transport costs exacerbate the challenge of realizing gains from trade; Atkin and Donaldson (2012) estimate that internal trade costs in Ethiopia and Nigeria are 7-15 times larger than in the United States, greatly reducing the benefits of globalization.

Infrastructure has long been one of the priorities of economic development policy and research: fully 15% of World Bank spending between 1995 and 2005 was dedicated to infrastructure projects, with 42% of that amount spent in China and India alone (The World Bank, 2007). Recent research has utilized novel identification strategies to investigate the link between the expansion of infrastructure and local economic performance. Banerjee et al. (2012a) examine the impact of railroads in China, finding that while railroads caused a level increase in income, nearby locations grew no faster than farther locations during a nearly 20 year period of rapid economic growth. Storeygard (2012) interacts global oil price shocks with distances to the nearest port to investigate the impact of transport costs on urban economic activity in Africa, finding a significant inverse relationship between transport costs and urban economic output, as proxied by nighttime luminosity. Donaldson (n.d.) develops a multi-region, multi-commodity trade model to assess the impact of railroad construction in colonial India, estimating that the expansion of the railroad network into a region increased real income by approximately 16% and greatly reduced trade costs. Michaels (2008) finds that the construction of the US Interstate Highway System generates sectoral and wage growth consistent with standard trade theory.

Another strand of research estimates the impact of changes in market access rather than average transportation costs generally. Redding and Sturm (2008) find that cities close to the border between West and East Germany experienced population loss relative to cities further from the border whose market access was less impeded by the partition of Germany following World War II. Hornbeck and Donaldson (2012) seek to unite the market access literature with the infrastructure literature by estimating the impact of railroads on agricultural land values using a market access approach, finding that the railroad network had by 1890 more than tripled the total value of agricultural land in the United States.

Of course, there are many reasons why rural roads may have very different economic effects when compared to core infrastructure projects such as interstate highways and long-distance railroads. Casaburi et al. (2013) use a fuzzy regression discontinuity design to examine the effect of village feeder roads on agricultural markets, finding that such roads significantly lower market prices of local agricultural goods. Gollin and Rogerson (2010) build a multi-sector model and calibrate it using data from Uganda to understand the relationship between agricultural productivity, transport costs and economic activity. They predict that high transport costs produce inefficiently high specialization in agriculture, and that investments in road infrastructure would lead to significant reallocation of labor to the nonfarm economy. Khandker et al. (2009) use propensity score matching to evaluate the impact of a rural road program in Bangladesh, estimating that receiving a road lowers village poverty by 5-6% while increasing household consumption by 8-10%, although subsequent work suggests that some of these gains may not persist over time (Khandker and Koolwal, 2011). Most closely related to this paper, Banerjee et al. (2012b) study the impact of the PMGSY on a broad range of outcomes in a sample of 267 villages in Uttar Pradesh. They find that road construction results in greater access to government services, lower consumer prices, higher agricultural prices, increased employment outside of agriculture and less daily migration, with no effect on longer-term migration. While the majority of evidence points towards large economic gains from the construction of rural roads, some studies have suggested that low incomes and population densities in rural areas may not generate sufficient demand for transportation services on rural roads (Raballand et al., 2011).

Methodologically, our use of within-district population rank is similar to Andrabi et al. (2013), who use local population rank to instrument for the placement of a girls' secondary school. Banerjee et al. (2012b) also use within-district population rank to estimate the effects of the PMGSY although, as discussed above, their focus is on the response of households, rather than firms, to the construction of a new road.

3 Context and background

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The focus was on the provision of new feeder roads to localities that did not have access, although in practice many projects under the scheme upgraded pre-existing roads. Originally, the stated goal was to provide all habitations with populations greater than 1000 with connectivity by 2003 and all habitations with population greater than 500 with connectivity by 2007. These thresholds were to be lower in desert and tribal areas, as well as hilly states and districts affected by left-wing extremism.[1]

[1] Habitations are defined as clusters of population whose location does not change over time. They are distinct from, but form parts of, revenue villages used by the Economic and Population Censuses. See National Rural Roads Development Agency (2005) for more details.

Although funded and overseen by the federal Ministry of Rural Development, responsibility for road construction was delegated to state governments. District Rural Road Plans were drafted for every district in India. Funding comes by a combination of taxes on diesel fuel (0.75 INR per liter), central government support and loans from the Asian Development Bank and World Bank. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).[2] The mandate of the program has recently been expanded to include all habitations with populations above 100.

4 Empirical Strategy

Our goal is to estimate the effect of the construction of a new rural road on changes in village-level economic activity. We start by estimating the OLS relationship between road construction and village employment growth using the following estimating equation:

$$Y_{v,s,t} = \beta_0 + \beta_1 * newroad_{v,s} + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,t-1}, \tag{1}$$

where $Y_{v,s,t}$ is log employment growth in village v in state s at time t, and $newroad$ indicates that a new feeder road was constructed in village v at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect.

The problem with this approach is that roads are not allocated at random. Roads are expensive infrastructural investments likely to be demanded by nearly all unconnected villages. Governments may target such investments to locations that are particularly needy, or have high potential for growth, or are politically connected or favored by powerful politicians. Controlling for baseline village characteristics or regional fixed effects will reduce this bias, but it remains a concern at the heart of the literature on infrastructure.

[2] We use an exchange rate of 44.06 INR per USD, the average for 2005

We undertake three strategies to mitigate bias from selection in the placement of roads: (i) using OLS to exploit variation in the timing of PMGSY road construction; (ii) using regression discontinuity to exploit population thresholds that determine road priority; (iii) using an instrumental variable approach to exploit within-district population ranking of villages, which is an additional determinant of road prioritization.

4.1 Road timing (OLS)

The naive OLS is biased because the types of villages that receive new rural roads differ on many unobserved characteristics from the types of villages that do not receive new roads. By limiting our sample to villages that eventually did receive PMGSY roads, and exploiting the timing of road construction, we can eliminate any confounders that differ between villages that were and were not eligibile to receive new roads.

Table 1 shows the number of rural roads built under the PMGSY in each year. Of the 75,399 roads in our sample that were built by 2009, 25,354 were built by 2005, the year in which we measure employment outcomes.[3] We then estimate Equation 1 on this limited sample, where the treatment variable indicates that a road was built before 2005.

Under the assumption that the order in which roads were constructed under the PMGSY is uncorrelated with other factors that affect growth, this would be an unbiased estimate of the effect of rural roads on village outcomes. However, this assumption is tenuous: order of construction is likely to be influenced by both political and economic factors that could bias OLS estimates either upwards or downwards.

4.2 Population priority thresholds (RD)

Timing of road construction is endogenous, which means the OLS estimates of the effects of roads are likely to be biased. In order to overcome this endogeneity, we exploit the population

[3] The 2005 Economic Census was conducted from late 2005 to early 2006.

thresholds intended to guide the allocation of roads under the PMGSY. State implementing officials were instructed to target habitations in the following order: (i) habitations with population greater than 1000; (ii) habitations with populations greater than 500; and lastly, (iii) habitations with populations greater than 250.

Even if selection into PMGSY treatment is biased by political or economic factors, these factors are not likely to change discontinuously at these population thresholds. If these rules were followed to any degree by state officials, the likelihood of PMGSY treatment will discontinuously increase at these population thresholds, making it possible to estimate the effect of the program using a fuzzy regression discontinuity design.

Under the assumption of continuity at the treatment threshold, the fuzzy RD estimator (Imbens and Lemieux, 2008) estimates the local average treatment effect (LATE) of receiving a new road, for a village with population equal to the threshold:

$$\tau = \frac{\lim_{pop \to T^+} \mathbb{E}[Y_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[Y_v | pop_v = T]}{\lim_{pop \to T^+} \mathbb{E}[newroad_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[newroad_v | pop_v = T]}, \quad (2)$$

where pop_i is habitation population, T is the threshold population, and $newroad_i$ is an indicator variable for whether village v received a new road in the sample period.

A given population threshold increases the probability of receiving a road by a different amount in different states. For example, states with a large number of large, unconnected villages, are more likely to have large first stages at the high threshold of 1,000. Analysis of PMGSY documentation and discussions with public officials have led us to focus on the population thresholds of 500 in Andhra Pradesh, Himachal Pradesh, Karnataka and Rajasthan, and 1000 in Chhattisgarh, Madhya Pradesh, Maharashtra, Punjab, Uttar Pradesh and West Bengal.

We estimate the reduced form fuzzy RD using equation 3:

$$Y_{v,s,t} = \beta_0 + \beta_1(pop_{v,s,t-1} > T) + \beta_2 pop_{v,s,t-1} + \beta_3 pop_{v,s,t-1} * (pop_{v,s,t-1} > T) + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,s,t}, \tag{3}$$

where $Y_{v,s,t}$ is log village employment at time t, T is the population threshold, $pop_{v,s,t}$ is habitation population at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect. Village controls and state fixed effects are not necessary for identification but improve the efficiency of the estimation. The local average treatment effect of a road, identified in a village at the population threshold T, is $\beta_1 + \beta_3 * T$. For ease of exposition, we subtract the threshold value (500 or 1000) from the population variable, such that $T = 0$, and β_1 fully describes the treatment effect.

The fuzzy regression discontinuity approach accurately identifies the treatment effect of rural roads, under the assumption that crossing the population threshold affects the probability of receiving a road, and nothing else of significance. There are two potential threats to this identification strategy. First, if other village characteristics vary discontinuously at the threshold in a way that we are unable to control for (e.g. if participation in other government programs uses the same thresholds), then our estimates will be biased. Second, if the running variable (habitation population) can be manipulated, randomness of assignment at the threshold is violated.

Figure 1 shows the distribution of habitation population as reported to the PMGSY, with implementation cutoffs indicated with vertical lines. There are noticeable discontinuities in density at the implementation cutoffs, suggesting that selection into treatment is not as good as random around these population cutoffs—for example, villages that are politically connected or more strategic may be able to report their population as just above 1000, even if it is not. If this is occurring, the RD approach cannot distinguish the effect of a new road from the effect of being politically connected.[4]

To resolve this issue with the validity of the RD, we examine a second source of habitation population data from the National Habitation Survey conducted in 2003 by the Department of Drinking Water Supply. Figure 2 shows the density function of this dataset of habitation population. In this dataset, there are no noticeable discontinuities at the PMGSY population cutoffs, so it meets the assumptions required for validity of the regression discontinuity approach.[5]

4.3 Population rank (IV)

In addition to the population threshold rules, district-level planning and implementation of the PMGSY meant that priorization was determined not only by population but also by relative population ranking within a district: a village would receive higher prioritization than an equivalent village if it had fewer larger eligible villages in its district. Holding population constant, a village in a district with many larger unconnected villages is less likely to receive a new road under the PMGSY. Under the assumption that, after controlling flexibly for total population, the population rank of a village within a district does not affect a village's growth prospects except through the likelihood of receiving a road through PMGSY, instrumental variable estimation provides an unbiased estimate of the effect of a rural road (Angrist and Lavy, 1999).

[4]An alternate explanation is that the PMGSY data, which covers about 80% of villages in India, covers a greater share of treated than untreated villages. Since villages over the population cutoffs are more likely to be treated, they are more likely to be selected into the PMGSY data. If this mechanism is driving the discontinuity in Figure 1, then the RD remains valid with these data.

[5]However, we do find evidence of rounding in this dataset: there are increases in density at populations that are multiples of 50. Reassuringly, there is not a disproportionate increase at either 500 or 1000. To be conservative, we drop from our dataset all villages containing habitations with populations of multiples of 50 in this dataset.

Our empirical specification is:

$$Y_{v,s} = \beta_0 + \beta_1 * newroad_{v,s} + \beta_2 f(pop_{v,s}) + \zeta X_{v,s} + \eta_s + \epsilon_{v,s} \tag{4}$$

where $Y_{v,s}$ is the outcome of interest in village v in state s, $newroad_{v,s}$ is an indicator for whether the village received a road under the PMGSY, $f(pop_{v,s})$ is a function of village population, $X_{v,s}$ is a vector of village controls and η_s is a state fixed effect. We estimate Equation 4 using $RANK_{v,s}$, the within-district population rank of village v, as an instrument for $newroad_{v,s}$. In alternate specifications, in order to reduce noise, we instrument for $newroad_{v,s}$ with a dummy variable indicating that $RANK_{v,s} < 75$.[6]

The estimation provides an unbiased estimate of the effect of a new rural road on employment growth, so long as the exclusion restriction is not violated: $RANK_{v,d,t-1}$ must affect growth only through the increased likelihood of obtaining a new road under the PMGSY. In Section 6 we discuss robustness checks to ensure satisfaction of the exclusion restriction.

5 Data

5.1 PMGSY

Data on the PMGSY is generated through the Online Management and Monitoring System (OMMS), the software used in program tracking and implementation. These data are not a survey - they are the administrative records of the actual program. Data include but are not limited to road sanctioning and completion dates, cost and time overruns, contractor names, and quality monitoring reports.

PMGSY data are reported at either the habitation or the road level. There is a many-to-many correspondence between habitations and roads: roads serve multiple habitations, and

[6] Our results are robust to different cutoffs for this dummy variable and are available upon request.

habitations may be connected to multiple roads. Habitations are subsets of census villages, which tend to comprise between one and three habitations; approximately 200,000 villages consist of only a single habitation.

5.2 Economic and population census

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 4th and 5th Economic Censuses respectively in 1998 and 2005.[7] The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Economic Census records information on the town or village of each establishment, whether ownership is public or private, the number and demographic characteristics of employees, the sources of electricity and finance, and the caste group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC), which corresponds roughly to a 4-digit ISIC code. More detailed information on income or capital is not included. The main strengths of the data are its comprehensiveness, and rich detail on spatial location and industrial classification of firms.

We obtained location directories for the Economic Censuses, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[8] We were able to match approximately 93% of villages between 1998 and 2005. We also use data from the Population Census of India in 1991 and 2001, which includes village population and other demographic data, as well as information on local public infrastructure (roads, electricity, schools and hospitals). Finally, we incorporate recently available village-level data from the Primary Census Abstract of the Population Census of

[7] The 6th Economic Census is ongoing at the beginning of 2013.

[8] The Economic Census of 1998 was conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

2011, which allows us to look at longer-run impacts of the program on basic demographic and employment outcomes.

We matched PMGSY data to economic and population census data at the village level, using population census codes where they were reported in the PMGSY, and a Hindi-language fuzzy matching algorithm to match village names across the two datasets. We successfully matched over 85% of habitations listed in the PMGSY to their corresponding population census villages.

Table 2 shows village-level summary statistics for the entire sample of villages used in our analysis.

6 Results

6.1 OLS

Table 3 presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

The table shows that villages that received a new PMGSY road by 2005 on average had non-farm employment growth that was 3-12 log points higher than villages that did not receive a new road. The point estimates fall as more controls are included, suggesting that selection of villages for roads is non-random. In particular, the falling coefficient as state fixed effects are included suggests that higher growth states were more likely to implement the program early - this is consistent with reports that state administrative capacity played an important role in early implementation of PMGSY.

6.2 Regression discontinuity

Table 4 presents regression discontinuity estimates of the effect of PMGSY prioritization on a village's likelihood of receiving a road. These regressions can be thought of as the first stage in a two-stage least squares IV estimation. Habitations with a population greater than the population threshold (500, 1000) should have received higher priority than habitations with population under the threshold. The dependent variable in these regressions is either an indicator variable that is set to 1 if a village received a road by 2011 (columns 1 and 3) or a count of the number of years between road completion and the year 2011. The running variable is the population of the largest habitation in the village. The treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to the population threshold used by the states in the sample, as determined by our analysis and conversations with the National Rural Road Development Agency. The threshold value has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect. The results suggest that there is a significant increase in the probability of treatment at the thresholds of 500 and 1000. Figure 3 depicts graphically the increase in probability of receiving a new road when the largest habitation is just above the threshold of 500.

The bandwidth used in all specifications is ten percent, so the sample for the estimation are villages with a largest habitation in the range of 400-600 for the 500 threshold and 800-1200 for 1000, although results are robust to alternate bandwidth choices. Controls and fixed effects are not necessary for identification, but their inclusion increases the efficiency of the estimator.

As discussed earlier, we are interested in the migration and employment effects of the

PMGSY. The Population Census of 2011 Primary Census Abstract contains village-level data on basic demographic characteristics. We choose three outcomes as a starting point for our investigation. The first, log population, captures the net effects of roads on migration, fertility and mortality; however, we consider it to mainly proxy for migration given that we find no evidence that effects of roads on the under 6 population differ from effects on the adult population. The second, the unemployment rate, is defined as the share of adult population considered non-workers in the Census. Finally, the agricultural rate is meant to capture structural transformation, and is defined as the share of main workers employed in agriculture. Table 5 presents results of the reduced form effect of the threshold dummy on total population. Columns 1-3 present results using the 500 threshold and columns 4-6 use the 1000 threshold. Dependent variables are as follows: log population (columns 1 and 4), unemployment rate (columns 2 and 5) and agricultural rate (columns 3 and 6). We find that unemployment drops significantly in the 500 threshold villages, with little evidence for movement of the workforce out of agriculture. In the 1000 villages, on the other hand, we find the threshold dummy associated with a weak increase in the population and a decrease in the share of workers in agriculture. Preliminary results suggest that these results may vary significantly by gender: the reduction in unemployment is strongest among women, while the movement out of agriculture appears concentrated among men.

6.3 IV

In this section we present results from the IV specification described in Section 4.3. We first verify that within-district village population rank appears to be a valid instrument and then discuss results of this estimation strategy.

We define our sample to be villages listed as not having paved roads at the time of the 2001 Population Census. This is for two reasons. First, villages unconnected by paved all-weather roads to the road network are listed as the highest priority under the PMGSY.

Second, we can be confident that roads to such villages are in fact new roads and not upgrades of existing roads, in order to be able to more cleanly interpret the results. The sample is further restricted to villages in districts in which the PMGSY built roads to more than 5% of villages by 2005. Finally, because few villages with very high within-district population ranks (very small villages) receive roads under the PMGSY, we drop villages with population ranks above 300. These sample definitions yield a sample of 3647 eligible villages in our rank IV sample, out of which 663 received PMGSY roads by 2005.

6.3.1 First stage and reduced form

We define our rank instrument two ways. The first is a simple field rank that assigns a village a rank of the number of villages in the sample with populations greater than the village, plus one. The second creates a binary variable out of this rank that takes on the value 1 when a village rank is less than 75. Table 6 presents the results of the first stage, regressing an indicator for the completion of a PMGSY road by 2005 on these two rank instruments. For simplicity of presentation, the rank variable has been divided by 100. Columns 1 and 2 present the effect of rank on the probability that a village receives a road by 2005 with quadratic and quartic polynomial population controls, respectively. A reduction of rank of 100 is associated with a 4 to 5 percentage point increase in the probability of receiving a road, controlling for population. Columns 3 and 4 present the effect of our binary instrument on the likelihood of receiving a PMGSY road by 2005: being in the top 75 villages within one's district is, after controlling for population and other village characteristics, associated with an approximately 5.5 percentage point increase in the probability of receiving a PMGSY road. The results are little changed by adding additional population controls, which is reassuring for the validity of the instruments.

Table 7 presents the reduced form results. The dependent variable in all four columns is log nonfarm employment growth. Columns 1 and 2 show that an increase in rank of 100 (decrease in prioritization for PMGSY road) is associated with a reduction of approximately 7 log points of employment growth, with negligible change when going from quadratic to quartic population controls. Likewise, columns 3 and 4 show that having a rank of less than 75 is associated with an increase of approximately 6 log points of employment growth.

One concern is that the rank instrument may be correlated with prioritization not only in the PMGSY but in other government programs that are also carried out at the district level. We thus define a placebo sample of villages with equivalent rankings as our rank IV sample but in districts that did not construct PMGSY roads in more than 1% of villages. Table 8 presents the results of the same reduced form specification discussed above but in this placebo sample. We find insignificant effects of both the continuous and binary population rank instruments on village nonfarm employment growth. We take this as suggestive, if not conclusive, evidence that the village population rank affects employment growth only through its impact on the likelihood of receiving a PMGSY road.

6.3.2 Rank IV estimates

Table 9 presents the results of the IV estimation based on Equation 4. As in the preceding tables, the first two columns use $RANK$ as the excluded instrument for the construction of a road, while columns 3 and 4 use the binary low rank variable. We find that a new road leads to highly significant positive increases in nonfarm employment growth. Regardless of specification, a new road is associated with an increase of employment growth of over 100 log points.

Given the very high costs of infrastructure projects such as roads, it is of great interest to policymakers to understand the economic impact of such projects in terms of money spent rather than per project. The OMMS described in Section 5 contains data on total spending per road, which we use to construct a village road expenditure variable that takes on the value 0 if a road is not built in the village by 2005 and otherwise equals the sum of all PMGSY road spending in habitations contained in that village.[9] As our objective is to estimate the cost of job creation, we define our outcome variable to be the level change in employment between 1998 and 2005, rather than log growth as in the preceding tables. Table 10 presents the results of this estimation. Our estimates range, depending on the instrument used, between approximately 500 and 700 nonfarm jobs per million dollars spent on road construction, which results in a per job cost of between approximately 1430 and 2000 USD (although these results are not statistically significant). Expressed differently, the per job cost of the PMGSY is roughly two to three times India's GDP per capita in 2005 ($731.70). How should we interpret this result? As the Economic Census was conducted in late 2005 and early 2006, it is safe to assume that nearly all of the roads listed as completed by 2005 were finished by the time of data collection; thus, these results should be understood not as temporary jobs related to road construction, but rather short- to medium-term effects. These estimates also assume that the net employment effect of PMGSY spending in locations not receiving roads is zero. This is a strong assumption – one we intend to investigate in future work – as there may be either crowd-in or crowd-out of economic activity in villages and towns that are now better connected to PMGSY villages. Finally, all estimates assume that control villages do not receive any PMGSY road spending, an unlikely result given the multiple years that road construction generally requires, the multiple habitations per PMGSY road project and the many roads that were completed in 2006 and 2007. For this last reason, we consider our estimates to be likely lower bounds on the true effect of road construction in this context.

[9] We generate total road spending village in million USD, using an exchange rate of 44.06 INR per USD, the average for 2005. In the case of roads that connect multiple habitations, spending is allocated equally between the habitations.

7 Conclusion

Access to the outside world via paved roads, easily taken for granted in many rich countries, is far from a reality for many of the world's poor. High transportation costs inhibit gains from the division of labor, economies of scale and comparative advantage. Nevertheless, little is known of the economic effects of road provision on rural economic activity. Theoretically, roads could facilitate migration to urban areas and the specialization of village economies in agriculture; alternatively, lower transportation costs could cause the emergence and expansion of rural economic activities, with potentially large consequences for economic development, urbanization and the spatial distribution of economic activity.

Despite the emphasis of both theorists and development policymakers on the importance of transportation costs, the impact of rural roads on development has proven challenging for economists to estimate. Due to the high cost of infrastructure, roads are likely to be distributed according to political and economic considerations, posing endogeneity problems. In this paper we demonstrate a further complication in the estimation of the effects of public spending: administrative data that determine allocation are susceptible to manipulation, particularly when the stakes are high. We find evidence of such manipulation, which we correct for by incorporating unrelated data not used to determine the use of public funds.

In this paper we estimate the economic impacts of the Pradhan Mantri Gram Sadak Yojana, a large-scale program in India that seeks to provide near universal access to paved "all-weather" roads in rural India. The program design provides two sources of exogenous variation to allow us to overcome the usual challenge of endogeneity of large infrastructure projects. First, the program calls for highest priority to be given to habitations above population thresholds, which may be 250, 500 or 1000 depending on the area. This creates a discontinuity in the probability of receiving a road at these cutoffs, allowing us to use a fuzzy regression discontinuity design to estimate the impact of these roads. A second identification strategy takes advantage of the fact that habitations are prioritized not only by population but also by the relative population rank within a district: a village of a certain size is more likely to receive a road if there are fewer larger villages within its district than an equivalent village that has a lower population rank. We instrument for road construction using this rank, conditioning on population, to provide a second set of estimates of the impact of village feeder roads.

We find that the provision of a new, paved village approach road produces significantly faster nonfarm employment growth. Our multiple identification strategies allow us to test for how local average treatment effects vary with habitation size. Large habitations appear to experience greater population growth and movement out of agriculture as the result of road construction, while smaller habitations show no such effects but do experience significant decreases in unemployment, particularly among women. We provide some of the first well-identified estimates of the cost effectiveness of rural road construction: one job is created for every \$1400 to \$2000 in road construction costs, suggesting very high returns to such investments.

Future work will allow us to further disentangle the channels by which rural roads promote village nonfarm employment. We intend to use the 2012 Economic Census, the collection of which is still in progress, to differentiate between the short- to medium-run effects presented in this paper and sustained, longer-run changes to village economic activity. Finally, we are assembling a unique dataset of village-level poverty that will allow us to test for the distributional effects of roads on job composition, asset accumulation and agricultural investments.

Table 1
Tabulation of Villages Receiving PMGSY Roads by Year

	Year Completed										
Year Sanctioned	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**Total**
2000	6	167	613	1085	995	606	315	422	112	61	4382
2001	0	0	21	721	1391	1382	707	316	162	128	4828
2003	0	0	0	2	404	1252	1156	685	367	268	4134
2004	0	0	0	0	0	300	777	791	866	469	3203
2005	0	0	0	0	0	0	179	1107	1546	1851	4683
2006	0	0	0	0	0	0	1	571	1049	2272	3893
2007	0	0	0	0	0	0	0	0	139	1150	1289
2008	0	0	0	0	0	0	0	0	0	149	149
Total	6	167	634	1808	2790	3540	3135	3892	4241	6348	26561

Table 2
Summary statistics

Variable	Mean	(Std. Dev.)	N
New road	0.049	(0.216)	182095
Employment (1998)	68.08	(100.604)	182095
Employment (2005)	84.235	(120.117)	182095
Ln employment growth	0.213	(0.876)	182095
Firm count (1998)	33.708	(45.795)	182095
Firm count (2005)	45.341	(60.145)	182095
Ln firm count growth	0.289	(0.842)	182095
2001 Population	1420.1	(1025.239)	182095
Pop growth 1991-2001	1.206	(0.29)	182095
Irrigation share	0.43	(0.365)	175468
Ln land area	5.358	(1.089)	175468
Distance from town	20.847	(19.475)	181645
Diversity (1998)	8.717	(6.512)	182095

Table 3

OLS: Employment growth on roads

	(1)	(2)	(3)	(4)
New road before 2005	0.113	0.079	0.058	0.036
	(0.019)***	(0.016)***	(0.015)***	(0.017)**
Baseline log employment	-0.275	-0.328	-0.477	-0.496
	(0.009)***	(0.008)***	(0.013)***	(0.014)***
Population	0.000	0.000	0.000	0.000
	(0.000)***	(0.000)**	(0.000)**	(0.000)***
Share of land irrigated			0.099	0.078
			(0.024)***	(0.026)***
Log(land area)			0.141	0.126
			(0.008)***	(0.008)***
Distance from town			-0.002	-0.002
			(0.000)***	(0.000)***
Baseline number of industries			0.024	0.026
			(0.002)***	(0.002)***
Constant	1.115	1.734	1.305	1.377
	(0.030)***	(0.063)***	(0.078)***	(0.078)***
N	48216	48216	46720	34888
r2	0.13	0.17	0.21	0.22

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

This table presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, as defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

Table 4

First stage: RD estimates of effect of population threshold on probability of new road

	Dummy	Years	Dummy	Years
500 threshold dummy	0.155	0.433		
	(0.038)***	(0.161)***		
1000 threshold dummy			0.074	0.269
			(0.026)***	(0.101)***
N	1604	1604	3808	3808
r2	0.23	0.19	0.11	0.11

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

The table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY prioritization on a village's likelihood of receiving a road. The dependent variable is an indicator variable that is set to 1 if a village received a road by 2011 (columns 1 and 3) and a count of the number of years between road completion and 2011 (columns 2 and 4). The running variable is the population of the largest habitation in the village, and the treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to (i) 500 for columns 1-2 and (ii) 1000 for columns 3-4. The value of the threshold (500, 1000) has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect.

Table 5

Reduced form: RD estimates of effect of population threshold on population and employment

	(1)	(2)	(3)	(4)	(5)	(6)
500 threshold dummy	-0.007	-0.020	0.012			
	(0.014)	(0.010)**	(0.020)			
1000 threshold dummy				0.016	-0.001	-0.011
				(0.010)	(0.005)	(0.010)
N	1604	1604	1600	3808	3808	3808
r2	0.91	0.22	0.41	0.78	0.57	0.50

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

The table presents regression discontinuity estimates from Equation 3 of the effect of PMGSY prioritization on outcomes in 2011. The treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to (i) 500 for columns 1-3 and (ii) 1000 for columns 4-6. The value of the threshold (500, 1000) has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the reduced form treatment effect. Outcomes are log employment growth (columns 1, 4), unemployment rate (2, 5) and agricultural employment rate (3, 6), as discussed in the paper.

Table 6

First stage effect of rank on probability of receiving road

	1	2	3	4
Pop rank	-0.127	-0.126		
	(0.014)***	(0.014)***		
Top pop			0.107	0.108
			(0.014)***	(0.014)***
Ln baseline employment	-0.002	-0.002	-0.001	-0.002
	(0.011)	(0.011)	(0.011)	(0.011)
2001 Population	0.000	0.001	0.000	0.002
	(0.000)	(0.001)	(0.000)**	(0.001)*
2001 Population ^{2}	-0.000	-0.009	-0.000	-0.019
	(0.000)	(0.012)	(0.000)**	(0.012)*
2001 Population ^{3}		0.036		0.070
		(0.046)		(0.046)
2001 Population ^{4}		-0.005		-0.009
		(0.006)		(0.007)
Pop growth 1991-2001	-0.024	-0.024	-0.023	-0.024
	(0.010)**	(0.010)**	(0.010)**	(0.010)**
Irrigation share	0.060	0.061	0.061	0.061
	(0.026)**	(0.026)**	(0.026)**	(0.026)**
Ln land area	0.009	0.009	0.009	0.009
	(0.009)	(0.009)	(0.009)	(0.009)
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Diversity (1998)	0.003	0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
N	3647	3647	3647	3647
r2	0.13	0.13	0.12	0.12

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

This table presents first stage regressions using within-district population rank as a predictor of receipt of a PMGSY road by 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 7

Reduced form effect of rank on employment growth

	1	2	3	4
Pop rank	-0.106	-0.100		
	(0.031)***	(0.032)***		
Top pop			0.059	0.060
			(0.031)*	(0.031)*
Ln baseline employment	-0.675	-0.675	-0.672	-0.674
	(0.023)***	(0.023)***	(0.023)***	(0.023)***
2001 Population	0.001	0.003	0.001	0.004
	(0.000)***	(0.002)	(0.000)***	(0.002)*
2001 Population ^{2}	-0.001	-0.021	-0.001	-0.028
	(0.000)***	(0.022)	(0.000)***	(0.022)
2001 Population ^{3}		0.071		0.095
		(0.086)		(0.086)
2001 Population ^{4}		-0.009		-0.012
		(0.012)		(0.012)
Pop growth 1991-2001	0.005	0.005	0.006	0.005
	(0.027)	(0.027)	(0.027)	(0.027)
Irrigation share	0.056	0.057	0.060	0.060
	(0.051)	(0.051)	(0.051)	(0.051)
Ln land area	-0.000	-0.001	0.000	-0.001
	(0.020)	(0.020)	(0.020)	(0.020)
Distance from town	-0.002	-0.002	-0.002	-0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Diversity (1998)	0.011	0.011	0.011	0.011
	(0.004)***	(0.004)***	(0.004)***	(0.004)***
N	3647	3647	3647	3647
r2	0.37	0.37	0.37	0.37

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 8

Reduced form effect of rank on employment growth (placebo sample)

	1	2	3	4
Pop rank	0.003	0.003		
	(0.033)	(0.033)		
Top pop			0.052	0.052
			(0.045)	(0.045)
Ln baseline employment	-0.582	-0.581	-0.584	-0.584
	(0.024)***	(0.024)***	(0.024)***	(0.024)***
2001 Population	0.001	0.003	0.001	0.003
	(0.000)***	(0.003)	(0.000)***	(0.003)
2001 Population ^{2}	-0.001	-0.019	-0.001	-0.019
	(0.000)***	(0.026)	(0.000)***	(0.026)
2001 Population ^{3}		0.067		0.066
		(0.101)		(0.101)
2001 Population ^{4}		-0.009		-0.009
		(0.014)		(0.014)
Pop growth 1991-2001	-0.079	-0.078	-0.077	-0.076
	(0.070)	(0.070)	(0.070)	(0.070)
Irrigation share	-0.117	-0.118	-0.122	-0.123
	(0.056)**	(0.056)**	(0.056)**	(0.057)**
Ln land area	-0.026	-0.026	-0.026	-0.026
	(0.021)	(0.021)	(0.021)	(0.021)
Distance from town	-0.001	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Diversity (1998)	-0.002	-0.002	-0.002	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)
N	2733	2733	2733	2733
r2	0.35	0.35	0.35	0.35

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005 for a placebo sample of villages in districts not receiving many PMGSY roads. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in less than 1% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 9
IV effect of road on employment growth

	1	2	3	4
New road	0.834	0.792	0.552	0.553
	(0.259)***	(0.263)***	(0.296)*	(0.294)*
Ln baseline employment	-0.673	-0.673	-0.672	-0.673
	(0.024)***	(0.024)***	(0.023)***	(0.023)***
2001 Population	0.001	0.002	0.001	0.003
	(0.000)***	(0.003)	(0.000)***	(0.002)
2001 Population ^{2}	-0.001	-0.013	-0.001	-0.017
	(0.000)*	(0.024)	(0.000)**	(0.023)
2001 Population ^{3}		0.043		0.056
		(0.095)		(0.091)
2001 Population ^{4}		-0.005		-0.007
		(0.013)		(0.013)
Pop growth 1991-2001	0.025	0.024	0.019	0.019
	(0.021)	(0.021)	(0.023)	(0.023)
Irrigation share	0.006	0.009	0.026	0.026
	(0.057)	(0.057)	(0.056)	(0.056)
Ln land area	-0.007	-0.008	-0.005	-0.006
	(0.021)	(0.021)	(0.020)	(0.020)
Distance from town	-0.002	-0.002	-0.002	-0.002
	(0.001)**	(0.001)**	(0.001)**	(0.001)**
Diversity (1998)	0.009	0.009	0.009	0.010
	(0.005)*	(0.005)**	(0.004)**	(0.004)**
N	3647	3647	3647	3647
r2	0.25	0.27	0.32	0.32

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

This table presents instrumental variable regression results for which the dependent variable is log employment growth (1998-2005) and the endogenous regressor is a dummy variable indicating the construction of a PMGSY road by 2005, instrumented by a population rank probit model. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 10

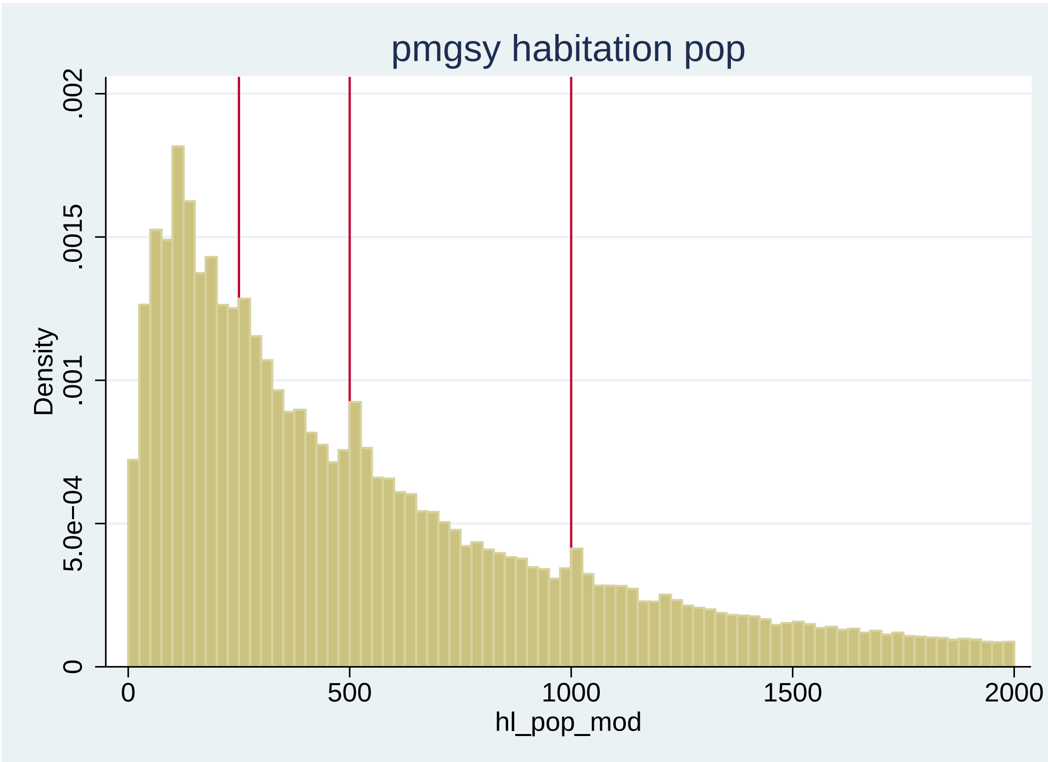
IV effect of road spending on level change in employment

	1	2	3	4
Road cost (million USD)	215.161	229.722	212.388	236.573
	(177.771)	(177.872)	(225.578)	(223.131)
Employment (1998)	-0.348	-0.348	-0.348	-0.348
	(0.075)***	(0.075)***	(0.075)***	(0.075)***
2001 Population	0.035	0.264	0.035	0.262
	(0.014)**	(0.217)	(0.014)**	(0.217)
2001 Population ^{2}	-0.041	-2.369	-0.041	-2.358
	(0.039)	(2.091)	(0.039)	(2.091)
2001 Population ^{3}		9.818		9.775
		(8.481)		(8.487)
2001 Population ^{4}		-1.455		-1.450
		(1.224)		(1.225)
Pop growth 1991-2001	0.879	0.898	0.874	0.911
	(1.214)	(1.221)	(1.249)	(1.252)
Irrigation share	3.249	3.374	3.253	3.365
	(4.457)	(4.466)	(4.451)	(4.461)
Ln land area	-3.539	-3.546	-3.533	-3.561
	(1.830)*	(1.835)*	(1.835)*	(1.835)*
Distance from town	-0.096	-0.093	-0.096	-0.094
	(0.078)	(0.079)	(0.079)	(0.079)
Diversity (1998)	-0.385	-0.395	-0.385	-0.395
	(0.488)	(0.489)	(0.489)	(0.490)
N	3589	3589	3589	3589
r2	0.14	0.14	0.15	0.14

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

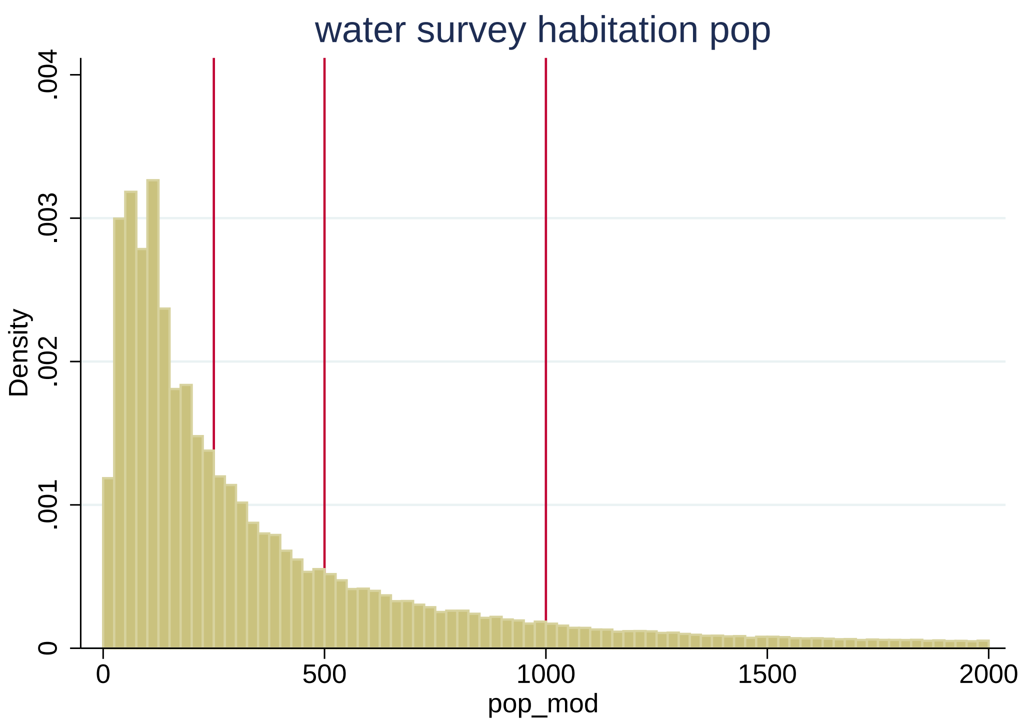
This table presents instrumental variable regression results for which the dependent variable is level employment growth (1998-2005) and the endogenous regressor is the amount of spending on a PMGSY road by 2005. For roads are not completed by 2005, this number is 0. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Figure 1



The figure shows the histogram of habitation population as reported in the PMGSY Online Monitoring and Management System. The vertical lines show the program eligibility cutoffs at 250, 500 and 1000.

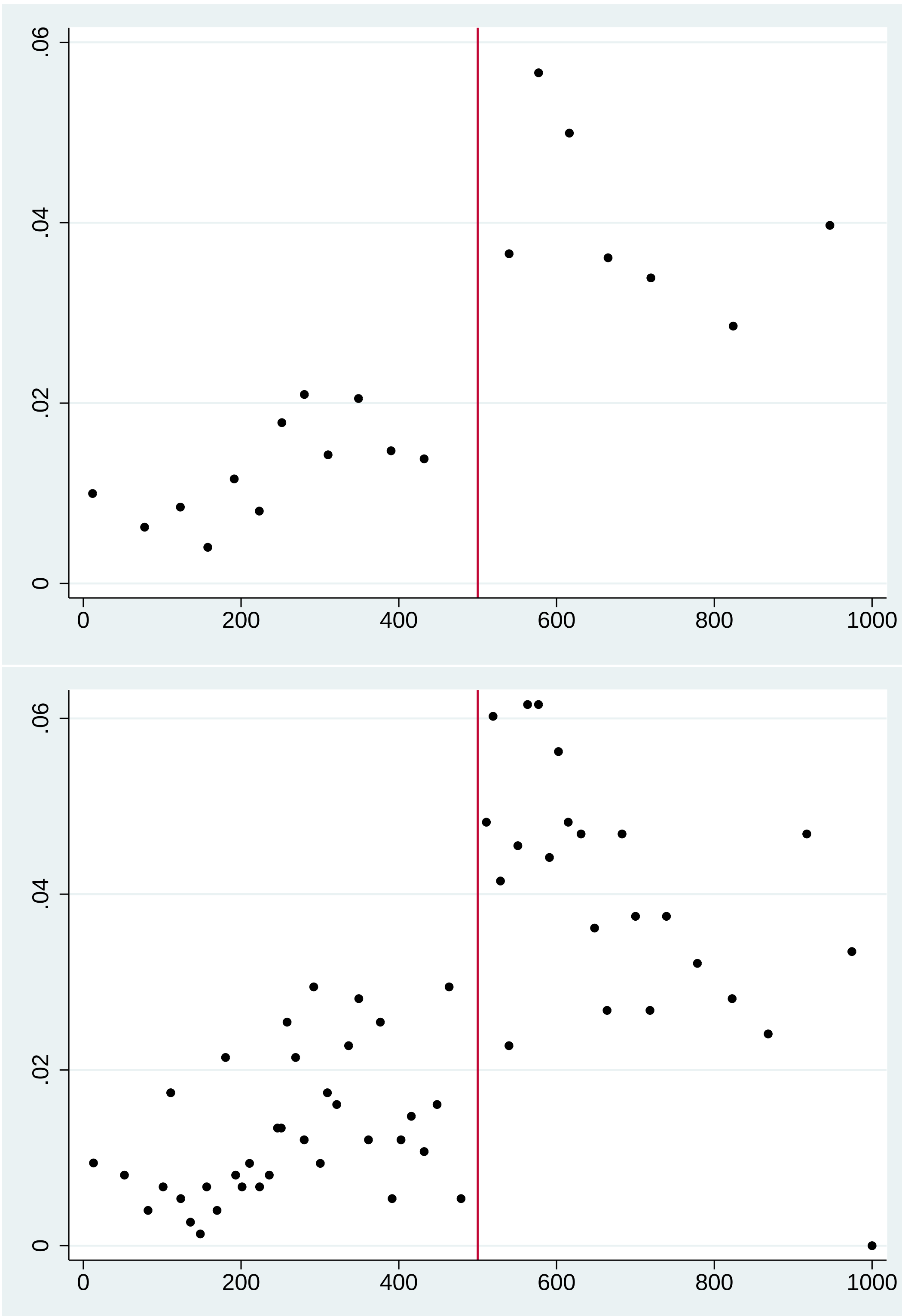
Figure 2



The figure shows the histogram of habitation population as reported in the National Habitation Survey conducted in 2003 by the Department of Drinking Water Supply. The vertical lines show PMGSY program eligibility cutoffs at 250, 500 and 1000.

Figure 3

First stage: population threshold and new rural roads



The figure shows the share of habitations that received a road, by population. Each point represents approximately 1000 habitations in the top panel, and 300 habitations in the bottom panel. The PMGSY instructed states to target roads to habitations with population greater than 500, the value indicated by the solid vertical line.

References

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Students today, teachers tomorrow: Identifying constraints on the provision of education," *Journal of Public Economics*, April 2013, *100*, 1–14.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 1999, *114* (2).

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2012.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012.

___ , Santosh Kumar, and Rohini Pande, "Connectivity and Rural Development: Examining India's Rural Road Building Scheme," 2012.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," 2013.

Djankov, Simeon, Caroline Freund, and Cong S. Pham, "Trading on time," *The Review of Economics and Statistics*, 2010, *92* (1), 166–173.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review.*

Gollin, Douglas and Richard Rogerson, "Agriculture, Roads and Economic Development in Uganda," 2010.

Hornbeck, Richard and Dave Donaldson, "Railroads and American Economic Growth: A "Market Access" Approach," 2012.

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

Khandker, Shaidur R. and Gayatri B Koolwal, "Estimating the Long-term Impacts of Rural Roads A Dynamic Panel Approach," 2011.

___ , Zaid Bakht, and Gayatri B. Koolwal, "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, 2009, *57* (4), 685–722.

Limao, Nuno and Anthony J Venables, "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

Michaels, Guy, "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 2008, *90* (November), 683–701.

Ministry of Rural Development, "Working Group on Rural Roads in the 12th Five Year Plan," Technical Report 2012.

National Rural Roads Development Agency, "Pradhan Mantri Gram Sadak Yojana - Operations Manual," Technical Report, Ministry of Rural Development, Government of India 2005.

Raballand, Gael, Rebecca Thornton, Dean Yang, Jessica Goldberg, Niall Keleher, and Annika Muller, "Are Rural Road Investments Alone Sufficient to Generate Transport Flows? Lessons from a Randomized Experiment in Rural Malawi and Policy Implications," 2011.

Redding, By Stephen J and Daniel M Sturm, "The Costs of Remoteness : Evidence from German Division and Reunification," *American Economic Review*, 2008, (1999), 1766–1797.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," 2012.

The World Bank, "A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005," Technical Report, The World Bank, Washington, DC 2007.