# The Ecological Footprint of Transportation Infrastructure*

Sam Asher†

Teevrat Garg‡

Paul Novosad§

October 2017

**Abstract**

Road construction is thought to result in forest loss, but causal identification has been elusive. Using multiple causal identification strategies, we find that India's rural road construction program, which built new feeder roads to over 100,000 villages and 100 million people, had precisely zero effect on local deforestation. In contrast, when 10,000 kilometers of India's national highway network were upgraded, there was substantial forest cover loss, driven by increased timber demand along the highway corridor. In terms of forests, last mile connectivity had a negligible environmental cost, while expansion of major corridors had important environmental impacts.

*We are grateful for feedback on earlier versions of this paper from Prashant Bharadwaj, Jenn Burney, Jonah Busch, Eric Edmonds, Paul Ferraro, Gordon Hanson, Maulik Jagnani, Erzo Luttmer, Gordon McCord, and seminar participants at UC San Diego, George Washington University, and the 2017 CU Boulder Environmental and Energy Economics Workshop. Ryu Matsuura provided excellent research assistance.

†World Bank, sasher@worldbank.org

‡UC San Diego, teevrat@ucsd.edu

§Dartmouth College, paul.novosad@dartmouth.edu

# I Introduction

Does human economic progress have an unavoidable environmental cost? This is a central question for policymakers pursuing sustainable development and has been a long-standing debate in the economics literature (Arrow et al., 1995; Grossman and Krueger, 1995; Stern, Common and Barbier, 1996; Andreoni and Levinson, 2001; Foster and Rosenzweig, 2003; Dasgupta, 2007; Alix-Garcia et al., 2013). A key pillar of economic development is large-scale investment in transportation infrastructure that reduces the costs of moving goods and people across space. Concern has been expressed about the potential environmental cost of such investments, and of increased trade more generally (Copeland and Taylor, 1994; Antweiler, Copeland and Taylor, 2001; Copeland and Taylor, 2004; Frankel and Rose, 2005), but researchers have struggled to identify causal estimates of the impact of transportation infrastructure on local environmental quality.

The most omnipresent of these investments are road networks. We focus on the impact of the construction of these road networks on forest loss as it is among the primary environmental concerns associated with new road construction. Forest cover loss is globally and locally important, generating global greenhouse emissions (IPCC, 2007; Jayachandran et al., 2017) and local health externalities (Bauch et al., 2015; Garg, 2017). Global forest cover loss accounts for more emissions than the entire U.S. economy and the global transportation sector (IPCC, 2007; U.S. Environmental Protection Agency, 2017).

Because of the high cost and high expected return of investments in roads, their placement typically depends upon various economic and political factors, making causal identification difficult. While a large number of earlier studies have documented changes in forest cover following the construction of new roads, none has addressed the endogeneity of road placement beyond the inclusion of control variables and in a few cases, location fixed effects. Further, the most well-known of these studies have focused on large highways built into the Amazon rainforest; while these highways are important in terms of potential deforestation, their impacts are of uncertain relevance for the set of potential rural roads and highways that policy-makers in developing countries are considering today. These are more likely to be last-mile roads to people not currently connected to the road network (especially in Sub-Saharan Africa) and upgrading of existing transportation corridors into modern highways.

In this paper, we provide causal estimates of the impacts on forest cover of different kinds of road projects, taking advantage of two major transportation investment programs recently undertaken by the Indian government. The first of these was a rural road construction program, under which over 100,000 new paved rural feeder roads were built, providing new connections to over 100 million rural residents. The second was an initiative to upgrade two major transportation corridors: the 6000 km "Golden Quadrilateral" network (GQ) connecting the country's four largest cities, and the comparably-sized "North-South and East-West" network (NS-EW) connecting the country's four cardinal endpoints. Both corridors were already used for cross-city transportation before 2000, but over the following fifteen years they were upgraded into world class divided highways. The rural road project and the national highway project each have had costs in the tens of billions of dollars to date. The tremendous scale of these projects in combination with new administrative and remote sensing data make it possible to evaluate the causal impact of these infrastructure projects.

Theoretically, the effect of road investments on local forest cover can be positive or negative. New roads can increase forest cover loss by: (i) providing external markets for forest resources, especially timber and firewood; (ii) providing external markets for agricultural products, motivating extensification of agriculture into forested land; and (iii) increasing the value of land for settlement and industry, resulting in forest clearing. On the other hand, paved roads could also reduce forest cover loss by (i) improving local household and industry access to substitutes for local forest resources, especially firewood; (ii) providing access to external output and labor markets, lowering the relative returns to clearing forests for agricultural land as well as to harvesting other forest products such as firewood.

We use two identification strategies to evaluate the impact of rural roads, using a validated satellite-based measure of forest cover (Vegetation Continuous Fields or VCF). First, in a regression discontinuity design, we exploit an implementation rule that discontinuously raised the probability of road construction in villages with population above an arbitrary threshold. Second, we use a difference-in-differences specification that exploits the exact timing of road construction. Both approaches show zero effects of new roads on forest cover. The estimates are precise; we can rule out gains larger than 0.6% and losses greater than 0.2% in forest cover up to five years after roads are completed. We find zero effects for sample subgroups defined by baseline forest cover or socioeconomic composition of villages. Nor do we see changes in household firewood use in treated villages. We do identify marginal reductions in forest cover during the road construction period; these reductions are reversed soon after roads are completed, but there is no evidence that forest cover continues to rise. We show that ignoring these construction period effects could lead to biased impact estimates. These roads have no effect on forest cover in spite of significantly altering economic opportunities for people in villages.[1]

In sharp contrast to rural roads, we find that the highway upgrades have had substantial negative effects on forest cover. We use a difference-in-differences design that estimates changes in forest cover in the proximity of a highway around the time that the highway is built, relative to places farther from the highway, as in Ghani, Goswami and Kerr's (2016) study of the impacts of the GQ on manufacturing activity.[2] We exploit the differential timing in the construction of the two highway networks to generate plausible counterfactuals. While both highways were scheduled to be completed by 2005, in practice significant construction did not begin along the NS-EW corridor until 2008, making it a reasonable control group for analysis of the GQ. We find that the expansion of the GQ led to a 20% decline in forest cover in a 100 kilometer band around the highway, an effect that persists for at least 8 years, which is the maximum period that we can evaluate in our sample. We find no change in

[1] Using similar identification strategies, Asher and Novosad (2017) and Adukia, Asher and Novosad (2017) find that these roads cause workers to shift away from agricultural work into wage work outside of villages and increase schooling, respectively.

[2] On the impacts of the Golden Quadrilateral on firms in India, see also Datta (2012) and Khanna (2016).

forest cover along the NS-EW corridor until construction accelerates in 2008, at which point we also observe local net forest cover loss. The timing of relative forest loss around the construction of each corridor supports a causal interpretation of these estimates.

Because forest cover in India is rising on average during the sample period, these are net effects on forest cover, combining increases in deforestation and reductions in afforestation. A conservative back of the envelope calculation assuming that highways simply reduced net forest cover by eight years (rather than causing permanent changes in forests) suggests that the Golden Quadrilateral project generated \$2.4 billion in social carbon costs from forest loss, approximately 25% of the total project cost.

These highways appear to have depleted forest cover by increasing timber demand in their vicinity, which has wide ranging effects into the hinterlands of the corridors. Following the construction of the GQ, we find a substantial upward trend break in employment in proximate firms that use timber and wood as primary inputs, as well as employment in logging firms. Additional tests reject the competing mechanisms; there are no changes in expansion of agricultural land or increases in local firewood consumption along the GQ corridor.

This paper makes three central contributions. First, we generate the first causal estimates of the impact of large scale transportation infrastructure investments on natural resource depletion.[3] In so doing, we contribute to a long literature on the trade-offs and synergies between economic development and environmental conservation.[4]

[3]Studies of cross-sectional correlation between roads and forest cover or forest loss include Chomitz and Gray (1996), Angelsen and Kaimowitz (1999), Pfaff (1999), Cropper et al. (2001), Geist and Lambin (2002), Deng et al. (2011), Barber et al. (2014), Li et al. (2015) and Dasgupta and Wheeler (2016). Pfaff et al. (2007) and Weinhold and Reis (2008) examine forest loss in areas with new roads but do not address the endogeneity of road placement. The closest study to ours is ongoing work by Kaczan (2017), who uses a difference-in-differences design similar to our first strategy, finding that India's new rural roads marginally increased forest cover. The differences may arise because Kaczan (2017) does not distinguish between construction and post-construction periods, and includes villages that never receive roads as part of the control group. We show in Section IV that both of these choices may lead to biased treatment effects.

[4]On the general relationship between economic development and the environment, see Den Butter and Verbruggen (1994), Arrow et al. (1995), Grossman and Krueger (1995), Stern, Common and Barbier (1996), Andreoni and Levinson (2001), Dasgupta et al. (2002), Foster and Rosenzweig (2003) and Stern (2004). On deforestation specifically, see Koop and Tole (1999), Burgess et al. (2012), Alix-Garcia et al. (2013), and Jayachandran et al. (2017). Assunção et al. (2017) provide causal evidence that rural *electrification*

Second, this is the first paper to show that the impact of roads on deforestation is a function of which markets are being connected by those roads. Last-mile rural roads provide connectivity to small local markets, facilitating exits from agriculture but without significantly changing industry's access to forest products (Asher and Novosad, 2017). In contrast, highways dramatically change the geographic distribution of industry (Ghani, Goswami and Kerr, 2016); in India at least, this appears to have important environmental consequences. The difference between these estimates highlights that environmental externalities from transportation infrastructure are not uniform and should be incorporated in cost-benefit analysis accordingly. Our estimates are particularly relevant as the infrastructure agenda in Sub-Saharan Africa and South and Southeast Asia is likely to prioritize exactly the kinds of infrastructure investments that we study here — new feeder roads and expansion of existing corridors — as opposed to the large highways through virgin rainforest that have been the subject of much of the earlier work on roads and deforestation.

Finally, we make a methodological note in the literature on estimating the effects of infrastructure. Large-scale infrastructure often takes many years to build and involves significant land clearing and economic activity during the construction process. In both our examination of highways and of rural roads, we find that deforestation takes place during the construction period; in either case, estimates based strictly on the timing of infrastructure completion would underestimate the deforestation impact of roads.

The rest of the paper is organized as follows. The next section describes India's rural road and highway construction programs. In Section III, we describe the data on forest cover and roads, as well as other secondary datasets used in our analysis. Section IV presents empirical strategy and results describing the impact of rural roads on deforestation. Section V presents the empirical strategy and impacts of highway expansions, and Section VI concludes.

mitigated forest loss in Brazil. For an exhaustive review on drivers of deforestation, see Ferretti-Gallon and Busch (2014). For a literature review on impacts of highways and rural roads on outcomes other than the environment, see Asher and Novosad (2017).

## II Background: Road Construction Programs in India

In 1999 and 2000, the Government of India launched two major road construction programs — one aimed at upgrading several national highway corridors and the other at connecting the remainder of India's population to the road network. Together, these programs marked the largest expansion of road infrastructure in Indian history and came at a joint cost exceeding $50 billion. This section provides background information on both road construction programs.

### II.A Rural Roads

In 2000, the Indian government launched the Pradhan Mantri Gram Sadak Yojana (PMGSY), or the Prime Minister's Village Roads Scheme. The primary objective of the program was to provide new paved roads to previously unconnected villages, although in practice this also involved upgrading low quality roads in already connected villages. By 2015, over 400,000 kilometers of new roads were built, providing access to the national road network to over 100 million rural people in over 100,000 villages.

Rural road construction began toward the end of 2001 and was continuing steadily through the end of the sample period in 2014 (See Appendix Figure A1). Villages were selected for roads based on a set of guidelines issued by a national government body, the National Rural Roads Development Authority. Notably, the program prioritized construction of roads to larger villages; district-level implementation plans were to first target all villages with populations greater than 1000, followed by villages with population greater than 500, and finally those with population greater than 250. The rules were applied on a state-by-state basis, allowing states to move from one threshold to another on their own timelines. In practice, there were several other prioritization guidelines, so that the population thresholds significantly influenced villages' likelihood of getting roads but were were not definitive. For example, thresholds were lower in desert and tribal areas and in places with extremist groups. Additionally, smaller villages could be connected if they were along the least-cost path between larger prioritized villages, and proximate villages could combine their populations to attain the eligibility thresholds. For more details, see Asher and Novosad (2017) and National Rural Roads Development Agency (2005).

### II.B National Highways

In 1999, the Indian government announced a plan for modernizing its major highways, the National Highways Development Project. The first component of the project was the upgrading and widening of the Golden Quadrilateral highway corridor (henceforth, GQ), so named because it connected the four major cities in India: New Delhi, Mumbai, Chennai and Kolkata. The second component was a similar upgrading of the the North-South and East-West corridor (NS-EW), which would connect the furthest corners of the country from Srinagar in the north to Kanyakumari in the south, and from Porbandar in the west to Silchar in the east. Panel A of Figure 1 shows both highway corridors along with the major cities that were connected by them.

While the GQ and NS-EW projects were commissioned around the same time, the government prioritized the implementation of the GQ and construction of the NS-EW was substantially delayed. Construction on the GQ began in 2001; 80% was completed by 2004 and 95% by 2006. In contrast, by 2006 only 10% of the NS-EW corridor was completed, almost half of which was a set of highways which were shared with the GQ (Ghani, Goswami and Kerr, 2016). By 2010, 72% of the NS-EW was completed, and 90% was completed by 2015. The delay in the construction of the NS-EW allows us to use the NS-EW corridor as a reasonable control group for changes in forest cover in the GQ corridor during and immediately following substantial completion of the GQ.

Before these highways were widened and upgraded, the GQ and NS-EW routes were already significant transportation corridors, but their road quality and congestion were highly variable. The upgrading of these networks dramatically improved their quality and reliability; these were the first major long-distance divided highway networks to be developed in India. As documented by Datta (2012), Khanna (2016) and Ghani, Goswami and Kerr (2016), the construction of the GQ changed national supply networks and led to a substantial reallocation of manufacturing firms into the GQ corridor. The economic impact of the NS-EW corridor has so far been little studied due to its later completion date.

## III Data

To estimate effects of new roads on forest cover, we combine five different national data sources. We use a validated high resolution satellite-based measure of forest cover. Data on rural roads come from the administrative implementation data generated by the rural road construction program, and geographic data on new major highway networks come from national highway maps. While these datasets form the basis of our core specifications, we also use data from the 1991, 2001 and 2011 Population Censuses and 3rd through 6th rounds of the Economic Census to control for location characteristics and explore mechanisms of treatment effects. All of these are census datasets that describe the entire population of India and are geocoded to the village, town and subdistrict levels. This section describes the details of how we prepare and combine all of these datasets. Table 1 shows summary statistics for all variables used.

### III.A Forest Cover

Detailed and reliable administrative records on forest cover and deforestation rarely exist, especially in developing countries. However, we are able to obtain high resolution time series estimates of forest cover using a standardized publicly-available satellite-based dataset. Vegetation Continuous Fields (VCF) is available at 250m resolution and provides annual tree cover from 2000-2014 in the form of the percentage of each pixel under forest cover (Townshend et al., 2011). For our primary specification, we define forest cover as the log of the sum of these values over all pixels in a given geographic area, plus one to allow for the inclusion of zeros.[5] Results are robust to using the average percentage of forest cover in each village.

[5]Results are robust to using the inverse hyperbolic sine transformation instead of log plus one.

It is worth noting that remote sensing applications have difficulty distinguishing between tree cover and other plantations. However, the VCF algorithm that translates imagery into numerical estimates of forest cover employs not only the visible bandwidth but also other bandwidths. For example, VCF uses thermal signatures because forested areas tend to be cooler than non-forested plantation areas, allowing VCF to (partially) distinguish between forest cover and plantations. To the extent that thermal signatures and other correlates can distinguish forests from non-forest plantations, VCF substantially improves upon the Normalized Differenced Vegetation Index (NDVI) that has been widely used in understanding the causes of deforestation (for example, Foster and Rosenzweig (2003)). For all analyses, we restrict the sample of villages to those that had positive forest cover in 2000, a year predating the construction of all roads considered in this research. This is also the earliest year that these forest cover datasets are available.[6]

Some earlier studies have used the Global Forest Cover (GFC) dataset, which describes baseline forest cover in the year 2000, and a binary indicator for the year of deforestation for each pixel, at a 30m resolution, where a pixel is considered deforested if over 90% of forest was lost in a given year (Hansen et al., 2013). GFC is less useful for the study of forest cover in India, because forest change in India is not well described by a binary deforestation indicator. Our VCF measures suggest that forest cover rose 15% over the sample period, an estimate that is confirmed by official and international sources. These gains are not recorded by GFC, nor are partial forest losses. Indeed, GFC data for India reports that over 90% of villages have zero forest change, and shows that nearly all total deforestation in the sample period occurred in the Northeast states, which are both far from the GQ and where few rural roads were built.

We matched forest cover data to village, town and subdistrict boundaries using geographic boundary data purchased from ML InfoMap. In remote parts of India, we received only settlement centroids rather than village boundaries. We generated Thiessen polygons for

[6]Villages with zero forest cover in 2000 very rarely show forest cover in future years. 95% of them have less than 1% forest cover in 2014; the mean of forest cover for pixels with non-zero forest is 12.76% in 2000.

these villages; all results are robust to excluding this set of villages. Panel B of Figure 1 shows a heat map of baseline forest cover in India. While contiguous areas of very dense forest are geographically concentrated, areas with 20-40% of their land covered by forest are found throughout the country.

### III.B Rural Roads

We scraped village-level administrative data describing the construction of rural roads from the program's online management portal.[7] For each road, the data provide the connected villages, the date when the contract for road construction was awarded, and the date of road completion. While data were reported at the sub-village (habitation) level, we aggregated the data to the village level to match our other data sources. We define a village as treated if any habitation in the village was provided with a new road. The data construction and scraping approach is described in detail in Asher and Novosad (2017). The dataset describes over 100,000 new roads built between 2001 and 2014; we limit our sample to areas with non-zero forest cover and no paved road in the baseline year, leaving approximately 65,000 new roads in the analysis sample.[8]

### III.C Highways

Construction dates and geocoordinates for the Golden Quadrilateral and North-South and East-West corridors were generously shared with us by Ghani, Goswami and Kerr (2016). We linked these to the village, town and subdistrict polygons described above by calculating straight line distances from polygon centroids to the nearest point on each highway.

### III.D Population and Economic Censuses

We matched all villages and towns from the 1991, 2001 and 2011 population censuses using a combination of incomplete keys provided by the Registrar General and a set of fuzzy matching

[7] The data is publicly available at http://omms.nic.in.

[8] Results are robust to include upgrades and/or village with no forest cover at baseline. These would be expected to attenuate non-zero treatment effects, this their exclusion would bias us against finding zero effects.

algorithms based on village and town names. The population censuses describe village and town public goods, amenities and household characteristics, including the primary source of cooking fuel. Fuel use is reported as the share of households in a location using firewood (68% of households at baseline), imported fuels (chiefly propane, 8%) or local nonwood fuels (crop residue and dung, 22%) as a primary source of energy. Fuel use is reported at the subdistrict level in 2001 and at the village level in 2011.

The Economic Censuses are complete enumerations of all nonfarm establishments undertaken in 1990, 1998, 2005 and 2013, including informal and non-manufacturing firms. We matched these on village names to the three population censuses using a fuzzy matching algorithm. The Economic Census reports total employment and industry for all firms. We create variables describing total employment in (i) firms engaged in logging and (ii) firms whose primary input is raw lumber, which include sawmilling and planing of wood, manufacture of wooden products such as furniture and wooden containers, manufacture of cork, and manufacture of pulp and paper products. The industry categorization for the 2005 Economic Census places logging firms in the same industry category as firms engaged in the conservation of forest plantations, management of forest tree nurseries and other afforestation categories. We therefore exclude 2005 from analysis of employment in logging firms.

## IV Impacts of Rural Feeder Roads on Forest Cover

Apart from data and statistical power, the main challenge to estimating causal effects of road construction is endogeneity. Because roads are costly to build, their placement is typically correlated with other factors that could also be predictors of deforestation. For example, roads could be targeted to places that are expected to grow or to places that are lagging economically. Road placement may also depend on geographic (e.g. slope, terrain, soil quality) or political factors. Any of these scenarios would bias OLS estimates of the effect of new roads on deforestation.[9] Causal identification of the impact of new roads therefore relies

[9] Appendix Table A1 shows estimates from cross-sectional OLS regressions of village-level log forest cover in 2001 on an indicator variable that takes the value one if a village has a paved road in 2001. While the bivariate relationship is strongly negative and highly statistically significant, the estimate gets progressively

on some kind of variation in road placement or timing that is plausibly exogenous. For rural roads, we rely on (i) an implementation rule that led to a discontinuity in the probability of a village getting a new road based on arbitrary population cutoffs; and (ii) variation in the specific year that a targeted village was treated. Because highways are more expensive and economically important than rural roads, one is unlikely to see arbitrary variation in the placement of highways.

Approximately 100,000 villages became newly connected to the national road network during our sample period. Because these villages are mostly small and isolated, there are likely to be few general equilibrium effects and any impact on forest cover can be expected to be local. We therefore use two empirical strategies that identify causal effects of new roads on deforestation in the direct area of these villages: a regression discontinuity strategy and a difference-in-differences strategy.

### IV.A Rural Roads: Regression Discontinuity Specification

We begin by exploiting the eligibility rule that prioritized villages for new roads based on arbitrary population thresholds. State officials were instructed to target villages in the following order: (1) villages with populations greater than 1000; (2) villages with populations between 500 and 1000; and (3) villages with populations between 250 and 500. While other factors such as political patronage undoubledly played roles in construction decisions, these factors are unlikely to change discontinuously around these population thresholds. Given the imperfect compliance with these eligibility rules (described in Section II), we employ a fuzzy regression discontinuity (RD) design. We limit our RD analysis to states in which administrators adhered sufficiently closely to population threshold rules.[10] Under the assumption of continuity at the treatment threshold, the fuzzy RD estimator calculates the

closer to zero as we add village-level controls and fixed effects, implying substantial selection on observables in the presence of roads. Selection on unobservables is plausibly also important, making the OLS estimates unreliable for causal inference.

[10] We identified these states with the help of officials at NRRDA. They include Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Orissa and Rajasthan. The difference-in-differences analysis below uses all states that built any roads in the sample period.

local average treatment effect (LATE) of a new road for a village at the population threshold:

$$\tau = \frac{\lim_{pop \to T^+} [Forest_v | pop_v = T] - \lim_{pop \to T^-} [Forest_v | pop_v = T]}{\lim_{pop \to T^+} [Treatment_v | pop_v = T] - \lim_{pop \to T^-} [Treatment_v | pop_v = T]} \tag{1}$$

$Forest_v$ is some measure of forest cover in village $v$, $pop_v$ is the baseline village population, $T$ is the population threshold and $Treatment_v$ is the binary indicator for whether village $v$ received a road in the sample period. The treatment effect is the discontinuous change in forest cover at the population threshold scaled by the probability of receiving a road at the population threshold (treatment). Using the approach detailed in Imbens and Lemieux (2008), Imbens and Kalyanaraman (2012) and Gelman and Imbens (2014), we use local linear regression to control for the running variable (village population) on both sides of the threshold and use only data within the optimal bandwidth for the estimation.[11] We use the following two stage least squares specification:

$$\begin{aligned} Treatment_{vj} &= \gamma_0 + \gamma_1 \cdot \mathbb{1}(pop_{vj} \geq T_j) + \gamma_2 (pop_{vj} - T_j) + \gamma_3 (pop_{vj} - T_j) \cdot \mathbb{1}(pop_{vj} \geq T_j) \\ &\quad + \nu_j + \epsilon_{vj} \qquad (2) \\ Forest_{vj} &= \beta_0 + \beta_1 \cdot Treatment_{vj} + \beta_2 (pop_{vj} - T_j) + \beta_3 (pop_{vj} - T_j) \cdot \mathbb{1}(pop_{vj} \geq T_j) \\ &\quad + \mu_j + \eta_{vj} \qquad (3) \end{aligned}$$

$Forest_{vj}$ is forest cover in village $v$ in state $j$, and $Treatment_{vj}$ is an indicator equal to one if a new road was built in village $v$. $pop_{vj}$ is the population of village $v$ and $T_j$ is the treatment threshold used in state $j$.[12] $\mu_j$ and $\nu_j$ are region fixed effects; we use district fixed

[11] Results are robust to different bandwidth choices or a triangle kernel that puts more weight close to the threshold.

[12] The treatment threshold varies with state because some states used a threshold of 500 and others were using a threshold of 1000. States used the lower treatment threshold when they had few villages with population over 1000 that did not already have roads. Officials at the National Rural Roads Development Agency provided us with information on which states were using which cutoffs, which we then verified in the data. Madhya Pradesh used both the 500 and 1000 treatment thresholds for roads built in the same period; we include separate fixed effects for the set of villages in the neighborhood of each threshold. Because the optimal regression discontinuity bandwidth is close to 100, there is no overlapping between these two groups. Few villages around the lowest population threshold of 250 received roads so we do not use this threshold

effects in all specifications, though reassuringly state fixed effects generate virtually identical results. We also add controls for village characteristics in 2001, before any roads were built; like the fixed effects, these are unnecessary for identification but improve precision. Controls include baseline forest cover, indicators for village amenities (primary school, medical center and electrification), the log of total agricultural land area, the share of agricultural land that is irrigated, distance in kilometers from the nearest town, the illiteracy rate and the share of inhabitants that belong to a scheduled caste. This is a cross-sectional regression where $\beta_1$ identifies the effect of new roads on forest cover in a given year. To further improve precision, we stack outcome data from 2010 through 2013 and cluster standard errors at the village level. However, running the test in any of these years separately or for similar sets of years produces virtually identical point estimates with marginally higher standard errors, and thus does not change any of our conclusions.[13]

Appendix Figure A2 shows regression discontinuity balance tests for a set of variables measured in the baseline period; Appendix Table A2 presents the regression estimates on these tests using Equation 2. None of the regression discontinuity estimates are significantly different from zero at baseline. Appendix Figure A3 shows that the density of the running variable is continuous around the treatment threshold (McCrary, 2008).[14]

### IV.B Rural Roads: Regression Discontinuity Results

Figure 2 shows a graphical representation of the regression discontinuity estimates of the impact of rural roads on forest cover. Panel A shows the first stage; the Y axis shows the share of sample villages that received new roads under PMGSY as a function of their population relative to the treatment threshold. Villages above the threshold are about 16% more likely to receive new roads and the discontinuity is evident. Panel B shows the first

for analysis.

[13] We exclude years before 2009 because the first stage for road construction is too weak; the majority of the roads built before 2009 were built by states that did not follow the treatment threshold rules. The first stage is comparable in size from 2010 through 2013 (see Panel B of Figure 2, which shows the first stage in each year).

[14] Asher and Novosad (2017) and Adukia, Asher and Novosad (2017) present further evidence that villages just above and just below the population eligibility thresholds have no significant differences.

stage estimate separately for each outcome year; each point in the figure represents the $\beta_1$ coefficient from Equation 2, where the dependent variable takes the value one if a village received a new road by the year indicated on the X axis. We can see that roads built before 2007 were not prioritized according to the population threshold rule; the first stage of the RD becomes noticeable after 2008 and continues to rise until 2014.

Panel C of Figure 2 plots the average of village-level log forest cover from 2010 to 2013 against the population relative to the treatment threshold. If roads significantly affected local forest cover, we would expect to see a discontinuity at the treatment threshold analogous to that in Panel A; no such treatment effect is evident. Panel D shows the reduced form treatment effect for forest cover in each year separately; as in Panel B, each point is an estimate from a separate regression, where the dependent variable is the log of forest cover for the year on the X axis. If the new rural roads significantly affected forest cover, we would expect to see a change in the coefficient following 2008 when administrators began to adhere to the population implementation rule. Instead, the effect is very close to zero both before and after 2008, indicating that new rural roads had negligible effects on forest cover.

Table 2 shows analogous regression estimates where the dependent variable is average forest cover from 2010 to 2013; standard errors are clustered at the village level. Column 1 shows the first stage estimate of a 16% increase in the probability of road treatment for villages just above the eligibility threshold. Columns 2 and 3 confirm there is no reduced form effect on either log or average forest cover. Columns 4 through 6 test for treatment effects in villages that might be expected to respond more to new roads: (4) villages with above-median baseline forest cover; (5) villages with above-median population shares of constitutionally described "backward" communities (Scheduled Tribes) who often derive livelihoods from forests; and (6) villages with below median assets, who might depend more on forests for fuelwood. There is no evidence of impacts of roads in any of these groups. Columns 7 and 8 show IV estimates on log and average forest cover. The IV estimates respectively rule out a 0.14 gain and a 0.11 loss in log forest cover, or approximately a one percentage point change in average forest cover. Results are robust to different controls or fixed effects and different bandwidth choices.[15] Appendix Table A4 uses the RD specification to show further that there are no changes in household fuel use following completion of a new road.

### IV.C Rural Roads: Difference-in-Differences Specification

The regression discontinuity design generates unbiased results under minimal assumptions, but is limited to estimating a LATE in the neighborhood of the treatment threshold in states that closely followed implementation rules on population thresholds. We can make greater use of our data and obtain tighter treatment estimates using a difference-in-differences specification that exploits the differential timing of road treatment in each village. For this empirical test, we limit the sample of villages to those that received a road at *some* point during the road construction program, and use outcomes in later-treated villages as a control group for villages that were treated earlier. We specifically estimate the following equation:

$$Forest_{vdt} = \beta_1 \cdot Award_{vdt} + \beta_2 \cdot Complete_{vdt} + \boldsymbol{\alpha}_v + \boldsymbol{\gamma}_{dt} + \boldsymbol{X}_v \cdot \boldsymbol{\nu}_t + \eta_{vdt} \tag{4}$$

$Forest_{vdt}$ is a measure of forest cover in village $v$ and district $d$ in year $t$. $Award_{vdt}$ is an indicator that takes the value one for the years where a contract has been awarded for the construction of a road to village $v$ but the road construction is not yet complete. $Complete_{vdt}$ is an indicator that takes the value one for all years following the completion of a new road to village $v$. We separate these two periods because the road construction process may have effects on forest cover (such as clearing of forested area to make room for the physical placement of roads) that are theoretically distinct from the economic effects of a village having a new road. Village fixed effects ($\boldsymbol{\alpha}_v$) control for all village-level time-

[15] Results at many different bandwidths are shown in Appendix Table A3. Appendix Table A6 shows additional specifications. Column 1 adds villages that did not receive roads in the sample period, the specification used in Kaczan (2017). Like Kaczan (2017), we find a positive treatment estimate; however, Column 2 shows that this is not robust to the inclusion of village-specific time trends, indicating that never-treated villages are on different forest cover trends from treated village. Columns 3 and 4 show that our main estimate is robust to village-specific time trends. Column 5 and 6 define the treated area as a circle around the village with a respective radius of 5km and 50km; as in the main specification, we find no treatment effects at these radii.

invariant unobservables, while district-year fixed effects control for any pattern of regional shocks.[16] We also interact a vector of baseline village controls $\boldsymbol{X}_v$ (baseline forest cover, village population and distance from the village to the nearest towns) with year fixed effects. These control for any differential time path of forest cover that is correlated with baseline village characteristics. These controls are particularly important because larger villages are more likely to be treated earlier due to program implementation rules. Standard errors are clustered at the village level to account for serial correlation.

We can interpret $\beta_1$ and $\beta_2$ as the effects of road construction activities and the effects of new roads, respectively; both coefficients describe outcomes relative to the period before any construction began. We restrict our sample from the universe of villages in India to those that had no road in 2000 and had a road completed during the study period. We do this so as not to compare villages that received new roads with those that did not; the endogeneity problem in such a comparison is severe.[17] Identification rests on the assumption that, among the set of villages that received roads in the sample period, there are no other systematic changes specific to villages in the years that roads were awarded and completed that are not caused by the roads themselves.

### IV.D Rural Roads: Difference-in-Differences Results

The difference-in-difference estimates of the impact of rural roads on village-level forest cover are summarized by Figure 3. This graphs show the residual of log forest cover — after taking out fixed effects and controls described above — as a function of the number of years elapsed since a road was completed in a given village. Panel A shows all previously-unconnected villages that received new roads between 2001 and 2014. Panel B restricts the set of villages to those with above median forest cover in 2000. We show only four years before and after road construction because wider windows have more variable sample composition across estimates;

[16] Results are unchanged by replacing these with state-year or subdistrict-year fixed effects.

[17] As we show above, a minority of roads were allocated strictly due to the village population thresholds. There are enough of these to estimate a regression discontinuity test on local compliers, but not enough to assume that all treated villages are selected as good as randomly.

this occurs because we observe different length of pre- and post-periods for different villages depending on their data of treatment.[18] Two patterns are evident in the figure. First, there is a statistically significant reduction in forest cover approximately two years before road construction is complete. Second, forest cover marginally increases in the four years after road completion recovering some or all of the pre-treatment drop.

Given that these rural roads took one to two years to build, this pattern is consistent with a small degree of forest loss (approximately 0.5%) during the road construction period, with partial or complete recovery afterward. We test this directly in Table 3, which shows estimates from Equation 4. Our main estimate in Column 1 shows that villages lose 0.5% of their forest cover during the period between the awarding of a road construction contract and the completion of a road. However, that forest loss is fully restored in the period after the road has been completed; the estimate of 0.002 log points on the completion indicator can be interpreted as the difference in forest cover between the post-road and the pre-award period. Relative to the pre-award period, we can rule out gains larger than 0.6% and declines larger than 0.2% in forest cover. In Column 2, we show that failing to account for the award period would lead to the estimation of a marginal forest cover gain of 0.5% because it would incorrectly attribute the construction period loss to the pretrend. This result highlights the importance of accounting for the construction period when studying the environmental impacts of new infrastructure. Columns 3 and 4 present estimates where forest cover is measured as the average share of each pixel that is covered by forest; results are similar. These estimates are based on different lengths of post-construction periods in different villages, but on average they show effects for four years after treatment.[19]

Table 4 shows these estimates along the same dimensions of heterogeneity described above. Effects are broadly similar whether we cut the sample on baseline forest cover, population share of Scheduled Tribes, or asset poverty. There is thus no evidence that our

[18] Appendix Figure A4 shows a wider time window around treatment; the pattern is the same.

[19] Appendix Table A5 shows that these estimates are robust to a range of specifications including the use of village time trends, subdistrict-year fixed effects (instead of district-year fixed effects) and using a limited sample of roads for which we have at least 4 (or 5) years of both pre-treatment and post-treatment data.

zero results are hiding differential positive and negative effects in different places. The panel estimates confirm the finding in the regression discontinuity analysis, using a different set of villages with a different local average treatment effect; the evidence is clear that new rural roads have had a negligible effect on local forest cover.

## V Impacts of Major Highways on Forest Cover

The national Golden Quadrilateral (GQ) and North-South and East-West (NS-EW) highway corridors were built with the objective of connecting India's major cities and regions; the connection of secondary cities and intermediate places on the route was a secondary priority. We therefore follow the approach used by Ghani, Goswami and Kerr (2016) in focusing on these intermediate locations, comparing changes in forest cover in locations close to expanded highways to locations further away.

The GQ and the NS-EW highway projects were expansions of existing transportation corridors; it is thus essential to distinguish between forest loss along existing highway corridors from deforestation caused by the expansion of those corridors. To do this, we exploit the differential timing of the two corridors; the GQ was 90% finished by 2005, while construction of the NS-EW did not accelerate until 2008. Our main estimates examine changes along the GQ corridor during and immediately after the construction years, using a region far from the GQ as a control group. For this period, we can test for placebo effects along the NS-EW route, or use it as a control group. A secondary set of tests shows that forest cover losses begin along the NS-EW around 2008, as would be expected given the timing of construction.

As a starting point, Figure 4 plots kernel-smoothed local regression estimates of mean forest cover and forest cover change as a function of distance from each highway. Initial forest cover (Panel A) is broadly similar across the two highways. Panel B shows that forest cover on average is rising along both corridors. But relative to the NS-EW (dashed line), forest cover within 100 km of the GQ (solid line) falls substantially between 2000 and 2008. At further distances the effects are similar across the two highways, though there may be smaller relative gains for the GQ. The rest of this section generates formal tests for change, controlling for fixed effects and other factors that may have simultaneously influenced forest change.

### V.A Highways: Empirical Specification

The simplest form of our difference-in-differences specification is described by the following equation:

$$Forest_{ist} = \beta_0 + \beta_1 CLOSE_{is} + \beta_2 POST_t + \beta_3 CLOSE_{is} * POST_t + \epsilon_{ist} \tag{5}$$

In this specification, $i$ indexes a subdistrict in state $s$ and time $t$, $CLOSE_{is}$ is an indicator for subdistricts close to the highway, and $POST$ indicates years following the completion of the highway. $\beta_3$ describes the differential change in forest between locations that are near and far from the highway network after the highway is built, controlling for the same geographic difference before the highway was built. If new highways cause deforestation, we expect $\beta_3$ to be less than zero. We conduct our analysis at the subdistrict level because these represent contiguous regions that cover the whole of India for which we can calculate a range of demographic and socioeconomic controls; we weight results by subdistrict area.[20] There are approximately 4000 subdistricts in India.

We extend this simple specification in three ways. First, because we do not have strong priors on which distances are near and which are far, we use a flexible set of distance indicators that identify effects in a broad range of distance bins, each of which is compared to an outer distance band. This ensures that our result is not dependent upon a particular definition of closeness. Second, because the construction of India's national highways were multiyear projects, we separate the $POST_t$ indicator into multiple periods to capture construction and post-construction effects. Third, we add a wide set of fixed effects and controls

[20] Results from a town- and village-level analysis with subdistrict clusters deliver nearly identical results. We could in principle conduct analysis at the grid cell level, but this would require imputation for control variables not available at the grid cell level.

to improve precision and reduce bias from omitted variables. The most flexible estimating equation is:

$$Forest_{ist} = \sum_{d=1}^{D} \sum_{t=2001}^{2014} \beta_{d,t} \mathbb{1}(DIST \in (d^-, d^+), YEAR = t) + \gamma_{st} + \boldsymbol{X}_i \cdot \boldsymbol{\nu}_t + \eta_{ist} \tag{6}$$

The distance to the highway is divided into $D$ bands, the boundaries of which are indexed by $d$. The sample includes data from an omitted distance category that is furthest from the highway which serves as a comparison group. As above, $Forest_{ist}$ is forest cover in subdistrict $i$ and state $s$ at time $t$, $\gamma_{st}$ is a state-year fixed effect and $\boldsymbol{X}_i$ is a vector of subdistrict controls interacted with year fixed effects. Controls are the same as in Equation 4. We include locations at distances $D + E$ from the Golden Quadrilateral; the outer boundary $E$ is the omitted distance category against which the other estimates can be compared. For regressions, we define the set $D$ to cover ranges 0 to 200 kilometers from the highway network, and use the 200 to 300 kilometer range as the omitted group. Alternate choices of the range of the omitted group, including using the remainder of the country does not appreciably affect our estimates. $\beta_{d,t}$ identifies the change in forest cover from the omitted year 2000 to year $t$, at distance range $d$ from the highway relative to the omitted distance range $E$. The $\beta_{d,t}$ coefficients can thus be directly interpreted as the effect of highway construction on forest cover after $t$ years. If new highways cause proximate deforestation, we would expect $\beta_{d,t}$ to take on negative values for low values of $d$ in the periods $t$ after highway construction has begun. For graphs, we include a set of indicator variables $\beta_{d,2000}$ which describe baseline forest cover as a function of distance from the highway.

We do not include subdistrict fixed effects because we want to generate coefficients on the distance band indicators for the omitted year 2000 — these coefficients describe the baseline differences in forest cover between places that were near and far from the highway. Given the large number of controls, inclusion of subdistrict fixed effects does not appreciably change the treatment estimates. We use state-year fixed effects rather than district-year fixed effects because we wish to test for meaningful effects of distance from highways that may extend beyond the radius of districts. District-year fixed effects would absorb true effects of the GQ that span distances larger than districts.[21] Standard errors are clustered at the subdistrict level to control for serial correlation.

While the regression above fully describes the time and distance structure of deforestation around the construction of the Golden Quadrilateral, coefficients are difficult to interpret if all years are included and distances are narrowly subdivided. In each specification, we therefore either group together years or distances to make results interpretable. Given the highway construction timeline described in Section II, we divide our sample into three periods. 2000 is the only year where we can observe forest cover before construction of the GQ was underway. We define years 2001 to 2004 as the GQ construction period, and 2005-2008 as the post-construction period. The NS-EW was less than 20% completed by the beginning of 2008, with much of that work on the segments very close to the GQ; it is therefore a plausible control group during this period.

To keep tables legible, we report estimates for distances in 50km or 100km bins; we use smaller bin distances for graphs, which allow us to map out the entire distance structure of highway effects.[22] Following Ghani, Goswami and Kerr (2016), we exclude areas within 200km of the nodal towns on the highway routes, as we wish to identify effects of highways rather than effects of growing metropolises. Estimates of NS-EW treatment effects omit areas that are within 200km of the GQ as they are plausibly being treated by the other highway network. We do not omit NS-EW regions from the GQ regressions because NS-EW construction has barely begun during the periods of interest for the GQ analysis; however, regression results are not changed by omitting places within 200km of NS-EW.

Because the Golden Quadrilateral project was upgrading an already existing transporta-

[21] The analysis of Ghani, Goswami and Kerr (2016) is entirely district level, giving us reason to expect meaningful cross-district effects. As expected, the inclusion of district-year fixed effects attenuates our results slightly but does not change the direction of effects nor eliminate statistical significance.

[22] In all cases, results are substantively unchanged by examining 50km or 100km bins. We present results in 50km bins except where doing so would make tables excessively illegible.

tion corridor, we do not expect places close to GQ to be similar to places far from GQ at baseline. Forest cover data is only available from 2000, so we cannot rule out pre-existing trends in deforestation close to the GQ. However, if transportation corridors were already experiencing secular forest losses, we would expect to observe these trends along the NS-EW corridor as well. Estimates on the impact of GQ can therefore be compared to estimates in the same time period on the impact of the yet-to-be-started NS-EW in a triple-diff specification. Our estimates can be interpreted as causal if there is no other phenomenon that is causing changes in deforestation close to GQ during or after GQ construction, but is not causing similar changes in deforestation either in regions far from GQ, or in regions close to NS-EW before NS-EW construction has begun in earnest.

All of these estimates are focused on examining the differences in forest cover changes between locations near and far from the highways. Given that measured forest cover in India has increased during the sample period, it is difficult to determine whether we should interpret negative point estimates close to highways as forest loss along the highway corridor, or as decreased afforestation relative to the counterfactual. We follow the literature in describing differential forest loss as deforestation, and discuss the possibility of displacement effects below.

### V.B Highways: Estimates on Forest Cover

In this section, we present estimates describing the impact of highway expansion on forest cover. All estimates describe changes in forest cover in regions close to expanded highways relative to an omitted distance band at a further distance from the highways.

Panel A of Figure 5 plots coefficient estimates from a single estimation of Equation 6, with distances from the highway network divided into 10km bands, and years divided into a single pre-construction year (2000), the construction period (2001-2004), and two post construction periods (2005-2008 and 2009-2012). All estimates describe the difference between a given 10km distance band from the GQ and the omitted category of 290-300km.[23] The solid black

[23]We include coefficients for the 200-290 km bands in order to plot treatment effects at these ranges.

line describes baseline forest cover as a function of distance from the GQ corridor; baseline forest cover is lower within 50 km of the highway corridor and then takes a relatively constant value from 50-300km from the highway. The remaining lines show that forest cover within 100 km of the GQ declines rapidly during the GQ construction period and then continues to fall in the years following construction. Effects are slightly smaller in the 100-150km bandwidth, and indistinguishable from zero at a distances greater than 150km from the highway. Panel B of the same figure shows treatment estimates along the NS-EW corridor. As predicted, differences from the baseline period do not emerge until the 2009-2012 period, and they decay at similar distances to the GQ distance effects. This distance pattern of forest cover loss is similar to that reported in Pfaff et al. (2007).

Figure 6 presents the time path of treatment effects; here, we plot the indicators on the 0-50 km distance band for each year interaction. Each point on the graph can be interpreted as the difference in forest cover between the 0-50km distance band and the omitted 200-300km distance band in a given year. The solid black line describes the time path of forest cover in places close to the GQ; the dashed red line describes places close to the NS-EW. The graph shows that forest cover falls rapidly during GQ construction and remains low until the end of the sample in 2014. The graph for the NS-EW shows little change in forest cover until 2008 when construction on this corridor begins in earnest. At this point, forest cover begins to fall, though it falls less than the GQ and may recover slightly in the final two years of the sample. Treatment effects along the NS-EW corridor may be smaller than along the GQ both because construction took place slowly and was still incomplete at the end of the sample period in 2014; the network structure of highways mean that the value of any particular segment depends on the completion status of other segments. The NS-EW is also a less important transportation corridor than the GQ, with less traffic, population, and nearby economic activity even at baseline.

Table 5 presents regression estimates from Equation 6, with distances in 50km bands for

Effects in closer bands are very similar if we restrict the distance indicators to 200km and use 200-300km as the omitted group, because there are few differences across years in the 200-290km range.

legibility. Each estimate describes the difference in forest cover between a given distance band and the omitted category of 200-300km. Columns (1) and (2) present estimates of the impact of the GQ respectively on the log of total forest area and on average forest cover in each pixel. The top four rows of the table show estimates of construction period impacts on forest cover. Places within 50km of the new highway network lose 27 log points of forest cover (Column 1) or 1.3 percentage points of forest cover (Column 2, on a base of 7.5%), and the effects shrink at greater distances. The next four rows show similar effects (relative to year 2000) in the post-construction period of 2005-2008. The final four rows show estimates of baseline differences between the GQ and the regions further away; the differences are small relative to the treatment effects, though it is clear from Figure 5 that places within 25km of the highway had less forest to begin with.

The baseline forest cover differences near the highway and the lack of data on forest cover before 2000 raise the concern that unobserved pretrends could explain changes in forest cover during and after the construction of the GQ. To mitigate this concern, we estimate an identical set of specifications examining changes in forest cover around over the same period around the NS-EW corridors, for which little construction took place before 2008. It is reasonable to think that any pretrends in forest cover along existing highway corridors would also exist along the NS-EW corridor.

Columns 3 and 4 of Table 5 show coefficients on indicators of distance to NS-EW in the same time periods used in Columns 1 and 2. There are no detectable changes in forest cover close to the NS-EW in the time period when significant forest loss took place near the GQ. The lack of deforestation along the NS-EW corridor during the GQ construction years alleviates the concern that the GQ treatment effects are driven by generalized deforestation along highway corridors from 2001-2008. These results are robust to instrumenting for highway location using straight line instruments connecting the nodal cities of the highway network, as employed by Ghani, Goswami and Kerr (2016); we present analogous reduced form estimates to the above in Appendix Table A7. These estimates alleviate the concern that the particular routing of the GQ was specifically targeted to places that may have already been losing forest cover.

Our primary estimates describe the difference between forest change in the proximity of the upgraded highways to forest changes far from the highways. One concern with these estimates is that they could be describing displacement of forest loss from the hinterlands to the highway corridors, or even net afforestation in the hinterlands. This concern arises frequently in studies of transportation projects with national scale, and is typically only resolved by assumption through a structural modeling approach, which is beyond the scope of this paper. This said, large displacement effects are made less plausible by the low quality of the broader road network, and by the high transportation costs during the sample period, weakening market connections with the hinterlands of these highways. While we cannot entirely rule out that there may be some displacement effects, Panel B of Figure 4 suggests that effects are driven by the highway corridor regions rather than the hinterlands; the main feature of this figure is the decline in forest close to the GQ. Further from the GQ, forest cover changes appear to track those in the untreated NS-EW network; the marginal net gains for the GQ in the 150-200km range could also be attributed to declining forest cover on the NS-EW.

#### V.B.1 Mechanisms for Highway Effects

We consider four possible mechanisms for the forest cover loss caused by India's major highway networks: (i) increased demand for timber products by firms due to local growth; (ii) increased demand for firewood due to shifts in household fuel consumption; (iii) expansion of agriculture into previously forested lands; and (iv) clearing of trees for settlements and industry. In this section, we present suggestive evidence that the deforestation along India's major highways is predominantly caused by increased logging driven by local timber demand.

To identify potential mechanisms, we use the regression specification used to identify effects of highways on forest cover (Equation 6), with data from the economic and population censuses which were undertaken in various periods between 1990 and 2013. We use distance bands of 0-100km and 100-200km from the GQ, and the omitted distance band is 200-300km from the GQ. We structure distances this way both for table legibility and because the results above suggest that effects on forest cover are largest within 100km of the highway network, but we find virtually identical results when we use 50km distance bins. The years in the sample are determined by census availability.

We first look at changes in employment in the list of industries that are directly downstream from timber harvesting (described in Section III). Panel A of Figure 7 shows results from a regression of log employment in major wood-consuming sectors on the usual set of year-distance-band fixed effects. We graph the point estimates on the 0-100km coefficient in each year that the Economic Census is available (i.e. the coefficients $\beta_{0-100km,1990}$, $\beta_{0-100km,1998}$, $\beta_{0-100km,2005}$, and $\beta_{0-100km,2013}$ from Equation 6. These estimates can be interpreted as the difference in residual log employment between the 0-100km distance band and the 200-300km distance band, after controlling for state-year fixed effects and the controls described above. In 1990 and 1998 (before the GQ was begun), there is no significant difference between areas close to the highway corridor and areas that are far from it, nor is there a significant trend. By 2005, we see a 5% increase in wood-consuming firms in the GQ corridor relative to the hinterland, which continues to rise through 2013. Panel B shows a similar estimation for employment in logging firms; we omit 2005 because logging firms were not distinguished from firms engaged in afforestation in the 2005 Economic Census. We see a very similar significant increase in logging after the national highway is completed. Appendix Table A8 shows point estimates for all years and distance bands in these regressions; as expected, employment in logging firms is more geographically diffuse, as those firms reach further into the GQ's hinterland. We note that logging firms were more common along the transport corridor even before the GQ was built, but there is zero suggestion of a pretrend that could explain what we see after highway construction. These two graphs suggest that demand for wood from downstream firms is a plausible explanation for local deforestation after construction of the GQ.

To explore whether changes in household fuel consumption could explain these results, we examine the primary source of fuel for cooking. These data are available only in 2001 and 2011. Panels D through F of Figure 7 show the effects in the 0 to 100km distance band separately for the share of households that use respectively firewood, imported fuels, and local non-wood sources like dung and crop residue. We observe marginal increases in firewood and imported wood use, and comparable reductions in use of local non-wood fuels. These effects are not statistically significantly different from zero, nor are they large enough to explain a 20% reduction in forest cover in the neighborhood of the GQ corridor. This said, without data from before 2001 it is impossible to establish pretrends. Table A9 shows the regression estimates.

We next examine the hypothesis that deforestation in the GQ corridor came from extensification of agriculture. For each village, the 1991, 2001 and 2011 Population Censuses report total village area that is cropped. Panel C reports point estimates from Equation 6 for agricultural land share in the 0-100km distance band. If anything, land use shifts away from agriculture in the transportation corridor following construction of the major highway route. Less land was dedicated to agriculture in this corridor even before the GQ was built, but the trend break following highway construction suggests a further shift away from agriculture. This makes it difficult for agricultural expansion to explain the reduction in forest cover in the GQ corridor after 2001. Column 1 of Appendix Table A9 shows the full set of regression estimates.

The last hypothesis is that deforestation has come from the expansion of land dedicated to settlement and industry. It is difficult to test this hypothesis directly because data on land dedicated to settlement and industry only becomes available in the 2011 Population Census. However, it is implausible that settlement and industrial expansion could explain a 20% reduction in forest cover in a distance band as wide as 100km around a 6000km long highway corridor. In 2011, only 6.7% of of rural land was used for settlement and industry.

In conclusion, we find suggestive evidence that expansion of industry demand for timber can explain forest loss in the GQ corridor, and we can rule out agricultural expansion, changes in household fuel consumption and settlement expansion as mechanisms.

### V.B.2 Social Carbon Costs of Highway Upgrading

To contextualize the magnitude of forest cover loss from highway upgrading, we estimate the social cost of the net increase in carbon dioxide emissions, following a process similar to Jayachandran et al. (2017). We estimate the carbon loss per hectare and then multiply that cost by commonly accepted estimates of the social cost of carbon.

We use the Global Forest Watch satellite-based estimate of average biomass in Indian forests of 260 metric tons of biomass per hectare, or 130 metric tons of carbon per hectare (World Resources Institute, 2017). Since a carbon dioxide molecule is 3.67 times as heavy as a carbon atom, the total carbon dioxide emissions from each hectare of forest loss is 477.1 tons of $CO_2$. According to the EPA, the median value of the social cost of carbon (SCC), assuming the EPA median scenario of a 3% discount rate, is \$39 (in 2012 US dollars) per ton of carbon dioxide emissions permanently averted (Jayachandran et al., 2017).[24] Because we can only confirm persistence of highway effects for eight years, we conservatively assume that in the absence of these highways, nearby places would fully reach counterfactual forest cover levels after the eight year period. That is, we assume that the effect of the highways was not permanent forest loss, but merely to reduce forest cover for a period of eight years. Since the SCC value mentioned above is for $CO_2$ emissions permanently avoided, we adjust this figure to reflect only eight years of avoided emissions, providing a substantially lower SCC of \$8.5 per metric ton of $CO_2$ emissions. Multiplying this by the total $CO_2$ emissions from a hectare of forest cover loss, the social cost per hectare of forest cover loss (avoided for eight years) is \$4,055. We multiply this number by the total estimated forest cover loss arising from the highways of approximately one million hectares.

Under these assumptions, we estimate a social carbon cost of net forest loss arising from the GQ highway corridor of \$2.4 billion. The social carbon cost of forest loss from the GQ

[24] More precisely, the social cost of carbon is the social cost of carbon dioxide emissions.

is thus 25% of its \$9.65 billion construction cost (also in 2012 U.S. dollars). The short run deforestation effects of the NS-EW corridor upgrades are smaller than these to date, but given that the highway was not yet completed at the time of writing, the long run costs are still unknown.

## VI Conclusion

The development, maintenance and expansion of transportation infrastructure is an important driver and correlate of economic development around the world. In this paper, we provide causal estimates of the ecological impact of two transportation investments with global significance: India's massive expansions of rural roads and its upgrading of national highways. Using causal identification strategies established in the literature, we find that (1) the new rural roads had negligible effects on forest cover and, (2) the highway expansions had a large negative effect on forest cover, likely driven by the expansion of wood-using industries. Methodologically, we demonstrate the critical importance of accounting for endogeneity and separately estimating the effects of the construction period from the post-completion period.

Globally, road expansion is expected to dramatically increase through the course of the 21st century. Some additional 25 million kilometers (16 million miles) of road infrastructure is projected to be built by 2050, a 60% increase over 2010 levels. Nine out of ten of these roads will be built in developing countries (Laurance et al., 2014). At the same time, tropical forests in developing countries are increasingly under threat. These forests not only provide global carbon benefits but also provide important local ecosystems which support biodiversity as well as the generally poor populations that rely on them (Barrett, Garg and McBride, 2016). Against the background of this tension between economic development and environmental conservation, understanding the relationship between roads and forests is fundamental to a successful strategy for sustainable development.

Crucially, we show that the impact of road construction depends on what those roads connect. In the case of India, while rural roads generated negligible forest cover loss, our calculations place the social carbon cost of highway-related forest cover loss at 29% of project costs. Policymakers considering future road construction should adequately consider the different environmental costs of different types of roads and compare these heterogenous costs against their benefits.

## References


**Adukia, Anjali, Sam Asher, and Paul Novosad.** 2017. "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction." Mimeo.

**Alix-Garcia, Jennifer, Craig McIntosh, Katharine R.E. Sims, and Jarrod R. Welch.** 2013. "The Ecological Footprint of Poverty Alleviation: Evidence from Mexico's Oportunidades Program." Review of Economics and Statistics, 95(2): 417–435.

**Andreoni, James, and Arik Levinson.** 2001. "The Simple Analytics of the Environmental Kuznets Curve." Journal of Public Economics, 80(2): 269–286.

**Angelsen, Arild, and David Kaimowitz.** 1999. "Rethinking the Causes of Deforestation: Lessons from Economic Models." The World Bank Research Observer, 14(1): 73–98.

**Antweiler, Werner, Brian R. Copeland, and Scott M. Taylor.** 2001. "Is Free Trade Good for the Environment?" American Economic Review, 91(4).

**Arrow, Kenneth, Bert Bolin, Robert Costanza, Partha Dasgupta, Carl Folke, Crawford S. Holling, Bengt-Owe Jansson, Simon Levin, Karl-Göran Mäler, Charles Perrings, et al.** 1995. "Economic Growth, Carrying Capacity, and the Environment." Ecological Economics, 15(2): 91–95.

**Asher, Sam, and Paul Novosad.** 2017. "Rural Roads and Structural Transformation." Mimeo.

**Assunção, Juliano, Molly Lipscomb, Ahmed Mushfiq Mobarak, and Dimitri Szerman.** 2017. "Agricultural Productivity and Deforestation in Brazil." Mimeo.

**Barber, Christopher P., Mark A. Cochrane, Carlos M. Souza, and William F. Laurance.** 2014. "Roads, Deforestation, and the Mitigating Effect of Protected Areas in the Amazon." Biological Conservation, 177: 203–209.

**Barrett, Christopher B, Teevrat Garg, and Linden McBride.** 2016. "Well-being dynamics and poverty traps." Annual Review of Resource Economics, 8: 303–327.

**Bauch, Simone C., Anna M. Birkenbach, Subhrendu K. Pattanayak, and Erin O. Sills.** 2015. "Public Health Impacts of Ecosystem Change in the Brazilian Amazon." Proceedings of the National Academy of Sciences, 112(24): 7414–7419.

**Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber.** 2012. "The Political Economy of Deforestation in the Tropics." The Quarterly Journal of Economics, 127(4): 1707–1754.

**Chomitz, Kenneth M., and David A. Gray.** 1996. "Roads, Land Use, and Deforestation: a Spatial Model Applied to Belize." The World Bank Economic Review, 10(3): 487–512.

**Copeland, Brian R., and M. Scott Taylor.** 1994. "North-South Trade and the Environment." The Quarterly Journal of Economics, 109(3).

**Copeland, Brian R., and Scott M. Taylor.** 2004. "Trade, Growth, and the Environment." Journal of Economic Literature, 42(1).

**Cropper, Maureen, Jyotsna Puri, Charles Griffiths, Edward B. Barbier, and Joanne C. Burgess.** 2001. "Predicting the Location of Deforestation: The Role of Roads and Protected Areas in North Thailand." Land Economics, 77(2): 172–186.

**Dasgupta, Partha.** 2007. "The Idea of Sustainable Development." Sustainability Science, 2(1): 5–11.

**Dasgupta, Susmita, and David Wheeler.** 2016. "Minimizing Ecological Damage from Road Improvement in Tropical Forests." World Bank Policy Research Working Paper No. 7826.

**Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler.** 2002. "Confronting the Environmental Kuznets Curve." Journal of Economic Perspectives, 16(1): 147–168.

**Datta, Saugato.** 2012. "The Impact of Improved Highways on Indian Firms." Journal of Development Economics, 99(1): 46–57.

**Den Butter, F.A.G., and Harmen Verbruggen.** 1994. "Measuring the Trade-Off Between Economic Growth and a Clean Environment." Environmental and Resource Economics, 4(2): 187–208.

**Deng, Xiangzheng, Jikun Huang, Emi Uchida, Scott Rozelle, and John Gibson.** 2011. "Pressure Cookers or Pressure Valves: Do Roads Lead to Deforestation in China?" Journal of Environmental Economics and Management, 61(1): 79–94.

**Ferretti-Gallon, Kalifi, and Jonah Busch.** 2014. "What Drives Deforestation and What Stops it? A Meta-Analysis of Spatially Explicit Econometric Studies." Mimeo.

**Foster, Andrew D., and Mark R. Rosenzweig.** 2003. "Economic Growth and the Rise of Forests." The Quarterly Journal of Economics, 118(2): 601–637.

**Frankel, Jeffrey A., and Andrew K. Rose.** 2005. "Is Trade Good or Bad for the Environment? Sorting Out the Causality." Review of Economics and Statistics, 87(1).

**Garg, Teevrat.** 2017. "Ecosystems and Human Health: The Local Benefits of Forest Cover in Indonesia." Mimeo.

**Geist, Helmut J., and Eric F. Lambin.** 2002. "Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical Forests are Disappearing as the Result of Many Pressures, Both Local and Regional, Acting in Various Combinations in Different Geographical Locations." BioScience, 52(2): 143–150.

**Gelman, Andrew, and Guido Imbens.** 2014. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." NBER Working Paper No. 20405.

**Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr.** 2016. "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." Economic Journal, 126(591): 317–357.

**Grossman, Gene M., and Alan B. Krueger.** 1995. "Economic Growth and the Environment." The Quarterly Journal of Economics, 110(2): 353–377.

**Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend.** 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." Science, 342(6160): 850–853.

**Imbens, Guido, and Karthik Kalyanaraman.** 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." Review of Economic Studies, 79(3): 933–959.

**Imbens, Guido W., and Thomas Lemieux.** 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics, 142(2): 615–635.

**IPCC.** 2007. The Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland.

**Jayachandran, Seema, Joost de Laat, Eric F. Lambin, Charlotte Y. Stanton, Robin Audy, and Nancy E. Thomas.** 2017. "Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation." Science, 357(6348): 267–273.

**Kaczan, David.** 2017. "Can roads contribute to forest transitions?" Mimeo.

**Khanna, Gaurav.** 2016. "Road Oft Taken: The Route to Spatial Development." Mimeo.

**Koop, Gary, and Lise Tole.** 1999. "Is There an Environmental Kuznets Curve for Deforestation?" Journal of Development Economics, 58(1): 231–244.

**Laurance, William F., Gopalasamy Reuben Clements, Sean Sloan, Christine S. O'connell, Nathan D. Mueller, Miriam Goosem, Oscar Venter, David P. Edwards, Ben Phalan, Andrew Balmford, et al.** 2014. "A Global Strategy for Road Building." Nature, 513(7517): 229.

**Li, Man, Alessandro De Pinto, John M. Ulimwengu, Liangzhi You, and Richard D. Robertson.** 2015. "Impacts of Road Expansion on Deforestation and Biological Carbon Loss in the Democratic Republic of Congo." Environmental and Resource Economics, 60(3): 433–469.

**McCrary, Justin.** 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." Journal of Econometrics, 142(2): 698–714.

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

**Pfaff, Alexander, Juan Robalino, Robert Walker, Steven Aldrich, Marcellus Caldas, Eustaquio Reis, Stephen Perz, Claudio Bohrer, Eugenio Arima, William Laurance, et al.** 2007. "Road Investments, Spatial Spillovers, and Deforestation in the Brazilian Amazon." Journal of Regional Science, 47(1): 109–123.

**Pfaff, Alexander S.P.** 1999. "What Drives Deforestation in the Brazilian Amazon?: Evidence From Satellite and Socioeconomic Data." Journal of Environmental Economics and Management, 37(1): 26–43.

**Stern, David I.** 2004. "The Rise and Fall of the Environmental Kuznets Curve." World Development, 32(8): 1419–1439.

**Stern, David I., Michael S. Common, and Edward B. Barbier.** 1996. "Economic Growth and Environmental Degradation: the Environmental Kuznets Curve and Sustainable Development." World Development, 24(7): 1151–1160.

**Townshend, J., M. Hansen, M. Carroll, C. DiMiceli, R. Sohlberg, and C. Huang.** 2011. "User Guide for the MODIS Vegetation Continuous Fields product Collection 5 version 1."

**U.S. Environmental Protection Agency.** 2017. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2015. Washington, D.C.

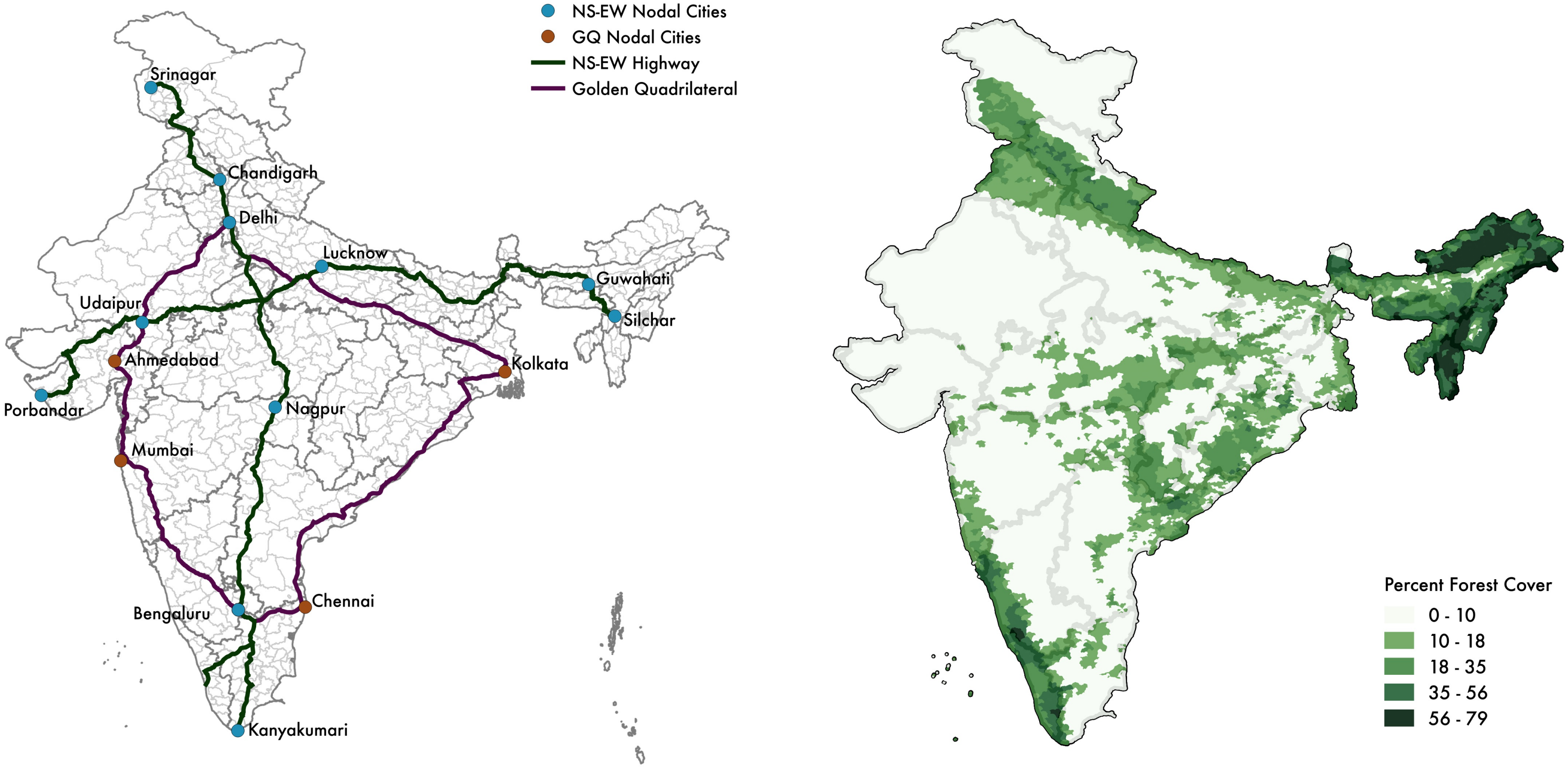
**Weinhold, Diana, and Eustaquio Reis.** 2008. "Transportation Costs and the Spatial Distribution of Land Use in the Brazilian Amazon." Global Environmental Change, 18(1): 54–68.

**World Resources Institute.** 2017. "Global Forest Watch." Accessed: 2017-09-21.

**Figure 1**

Panel A: Highway Construction (2001-2014)

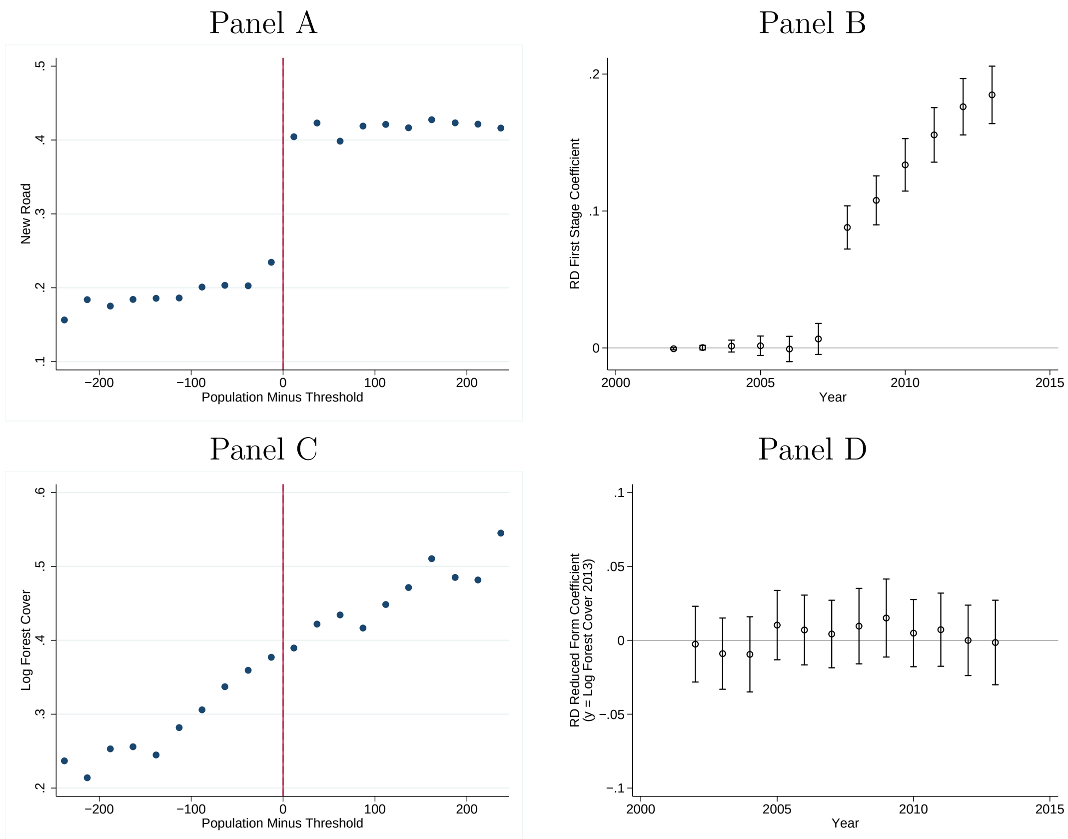
Panel B: Forest Cover (2001)




Panel A shows a map of the Golden Quadrilateral and North-South/East-West corridor highways. Panel B shows a heat map of forest cover in 2001. Areas are shaded according to average share of each pixel that is covered by forest.

# Figure 2

Regression Discontinuity Estimates of Impact of Rural Roads on Forest Cover

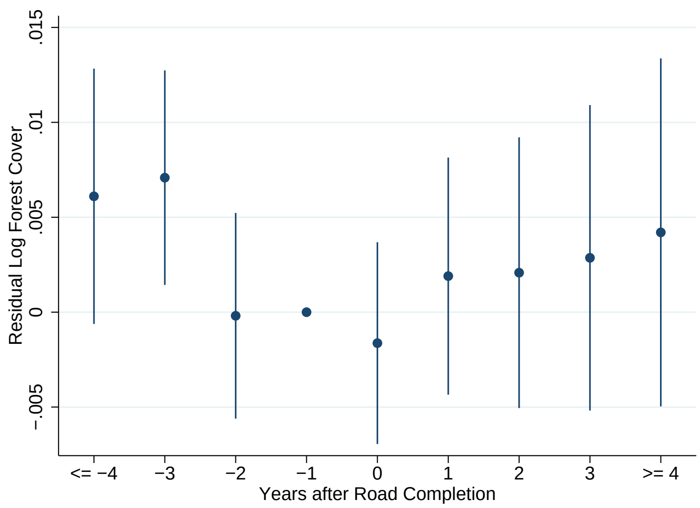



The figure shows regression discontinuity estimates of the impact of new rural roads on local deforestation. Panel A shows the first stage probability of a village receiving a new road before 2014 as a function of its population relative to the population threshold. Each point shows the mean of the Y variable in a given population bin. Panel B shows the first stage RD estimate of a village receiving a new road by the year indicated on the X axis. Each point is an estimate from an RD first stage regression. Panel C is analogous to Panel B; the dependent variable is the log of forest cover in 2014. The points show the mean of this variable in each population bin; population is shown relative to the population treatment threshold. Panel D shows reduced form RD estimates of the impact of being above the population threshold on forest cover in each year on the X axis. All estimates in Panels B and D use the same specification as Table 2, and include district-population threshold fixed effects and a control for baseline forest cover.

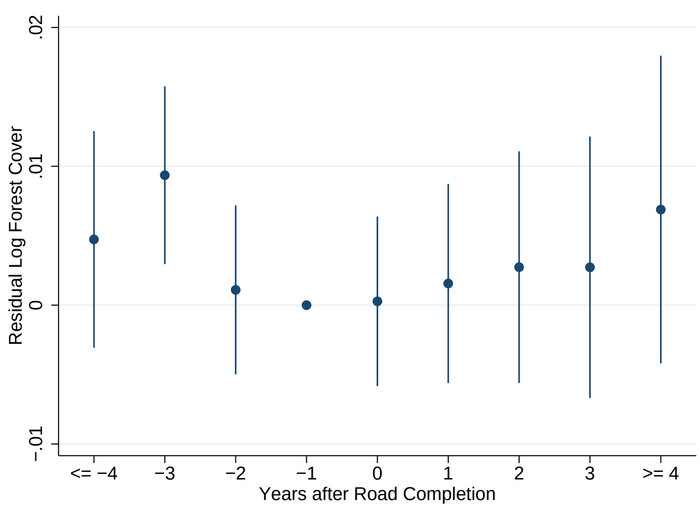
**Figure 3**

Difference-in-Differences Estimates of Impact of Rural Roads on Forest Cover

Panel A: Full Sample




Panel B: High Baseline Forest

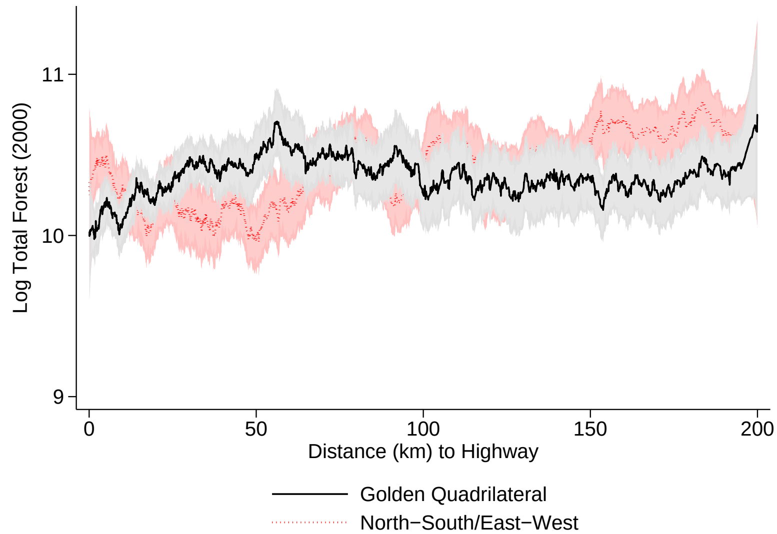



The figure shows year-by-year estimates of log forest cover in villages that received new roads between 2001 and 2013. Villages are grouped on the X axis according to the year relative to road completion. Each point thus shows the average value of log forest cover in villages in a given year relative to the treatment year, controlling for village fixed effects, district*year fixed effects, baseline population * year and baseline log forest cover * year interactions. Standard errors are clustered at the village level. The year before road completion is omitted (t=-1); forest cover is thus shown relative to this period.

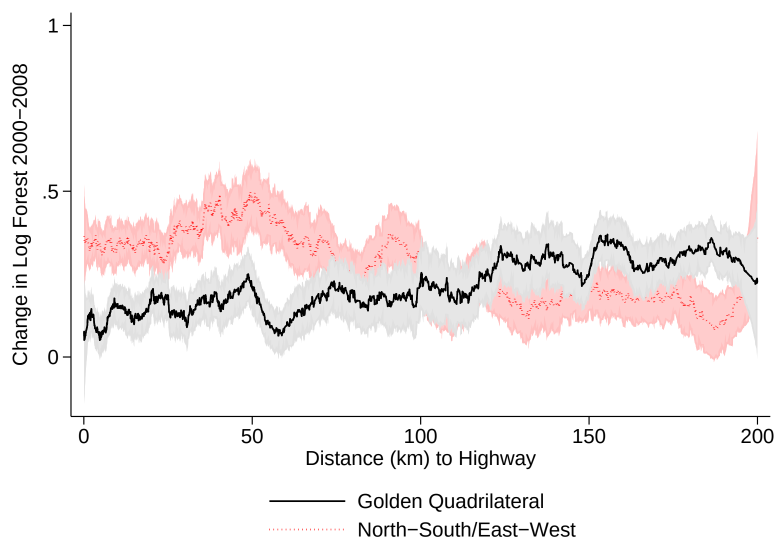
**Figure 4**

Forest Cover and Forest Cover Change
Along Highway Corridors (2000-2008)

Panel A: Pre-Expansion Forest Cover (2000)




Panel B: Forest Cover Change (2000-2008)



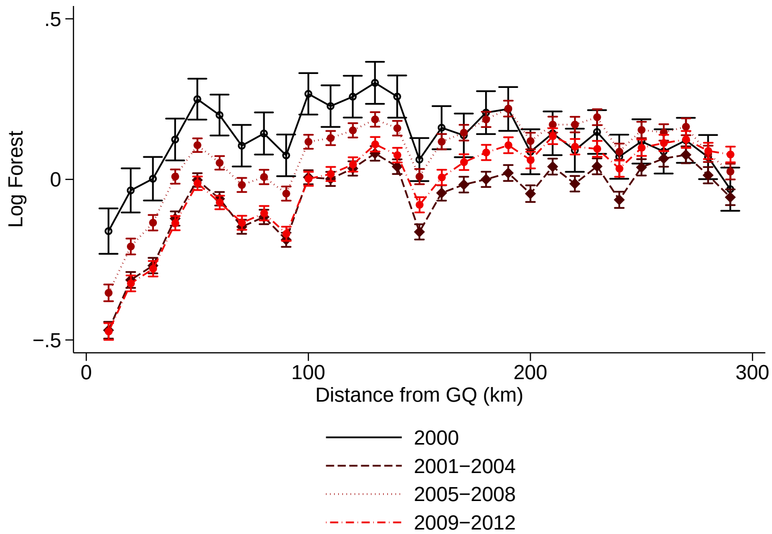

---

Panel A shows a kernel-smoothed regression of log subdistrict forest cover in 2000 on distance to the corridors where the Golden Quadrilateral and North-South/East-West highways will be expanded. Panel B plots kernel-smoothed regression estimates of change in log subdistrict forest cover from 2000 to 2008 against distance to each highway network. By 2008, there was very little construction on the NS-EW corridor, so we treat it here as a control group. The plots display means that are unadjusted for any fixed effects or controls. 95% confidence intervals are displayed in the shaded areas.

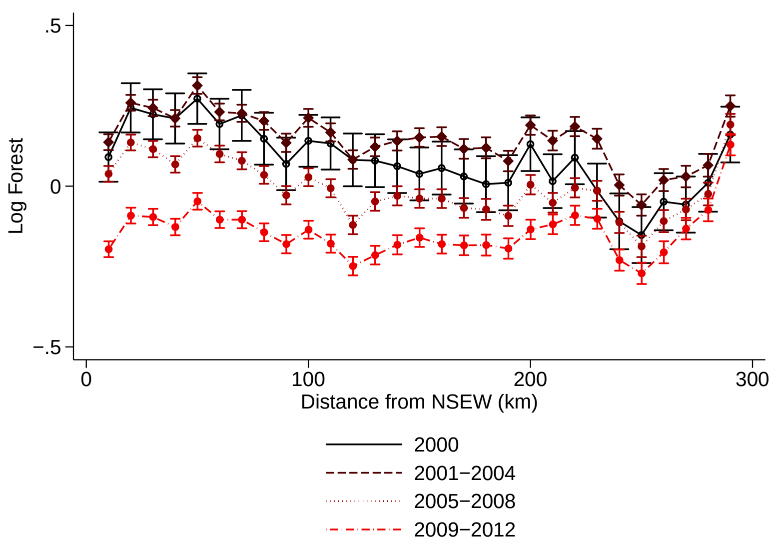
**Figure 5**

Difference-in-Differences Estimates of Impact of Highways on Forest Cover, by Distance Bands

Panel A: Golden Quadrilateral



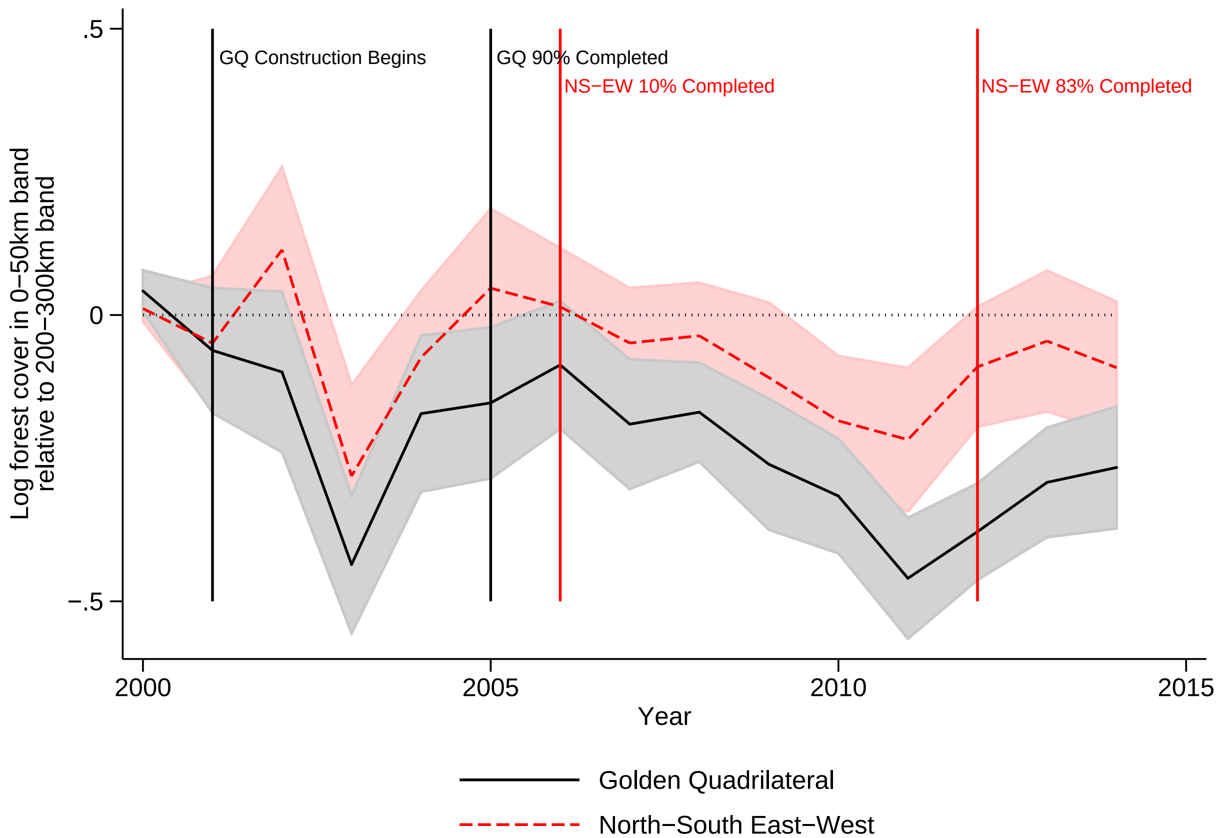

Panel B: North-South/East-West




The figure shows point estimates from Equation 6, with distance from the Golden Quadrilateral highway network (Panel A) and distance from the North-South/East-West highway network (Panel B) divided into 10km bands. Each point on the graph shows, for a given set of years (shown in the legend), the average value of log forest cover at a given distance band from the given highway network, relative to the omitted distance band of 290 to 300 km from the highway. All estimates control for state*year fixed effects, baseline population * year and baseline log forest cover * year interactions. Standard errors are clustered at the subdistrict level.

**Figure 6**

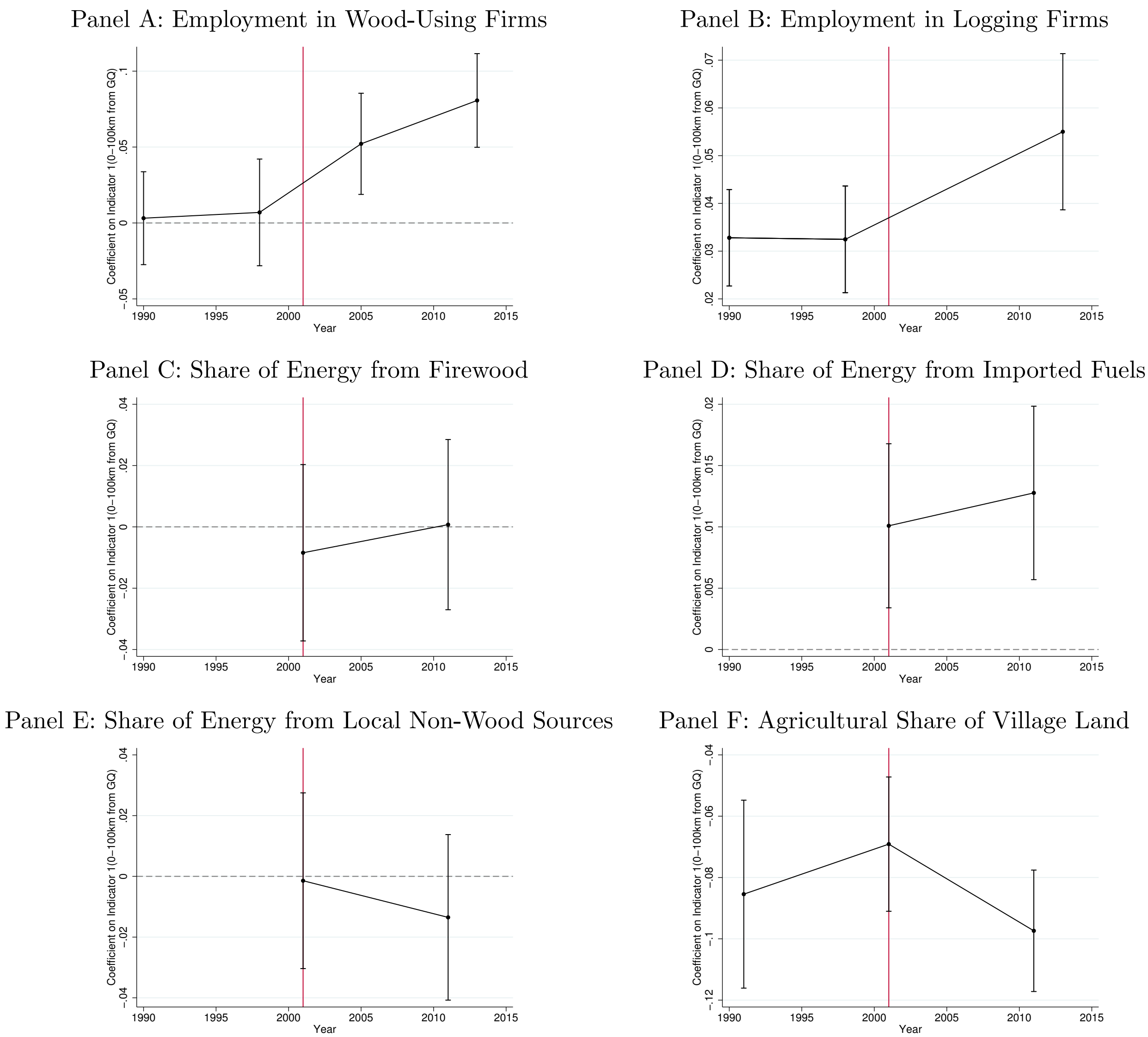
Difference-in-Differences Estimates of Impact of Highways on Proximate Forest Cover, Over Time




The figure shows treatment effects of the two highway networks over time. The solid black line shows residual log forest cover in the area that is 0-50km from the Golden Quadrilateral highway corridor. The dashed red line shows residual log forest cover in the area that is 0-50km from North-South/East-West highway corridor. The point estimates come from the distance-year interactions in Equation 6. The omitted category is the 200-300km distance band. Intermediate distance bands are omitted for clarity, but their effects can be seen in Table 5. All estimates control for state-year fixed effects, and interactions between year fixed effects and baseline population, baseline log forest cover and distance from the nearest town with population over 100,000. Progress dates on the construction of the two highways are displayed in the graph. Standard errors are clustered at the subdistrict level.

## Figure 7

Mechanism Tests for Impact of Highways on Deforestation




The figure shows point estimates from Equation 5, with distances from the Golden Quadrilateral highway network specified in 100km bands. Each figure shows the point estimate on the 0-100km distance indicator, interacted with the year shown in the X axis. The omitted category is the set of places that are 200-300 kilometers from the Golden Quadrilateral network. The dependent variables in Panels A and B are log employment in respectively wood-consuming firms and in logging firms. Logging was not specified in the 2005 Economic Census so this point is omitted. In Panels C through E the dependent variable is the share of households' cooking fuel that takes the form of (C) firewood; (D) imported fuels, primarily propane; and (E) crop residue and animal waste. In Panel F, the dependent variable is the share of village land dedicated to agriculture. All estimates are from regressions with state-year fixed effects and standard errors are clustered at the subdistrict level. Appendix Table A9 shows the full set of estimates from the regressions that produced these graphs.

**Table 1**

Summary Statistics

Village-level Statistics

| | Mean | Standard Deviation | Observations |
|---|---|---|---|
| New road before 2011 | 0.18 | 0.38 | 257256 |
| Road completion year | 2007 | 2 | 45459 |
| Population share with no assets (2002) | 0.69 | 0.31 | 169387 |
| Population share Scheduled Tribes (2001) | 0.22 | 0.39 | 257256 |
| Agricultural share of village land (2001) | 0.64 | 0.28 | 372246 |
| Share energy from firewood (2001) | 0.67 | 0.26 | 409298 |
| Share energy from imports (2001) | 0.07 | 0.09 | 409298 |
| Share energy from local nonwood (2001) | 0.26 | 0.26 | 409298 |

Subdistrict-level Statistics

| | Mean | Standard Deviation | Observations |
|---|---|---|---|
| Average forest cover (2000) | 12.76 | 14.66 | 4019 |
| Average forest cover (2014) | 14.69 | 14.49 | 4019 |
| Distance to Golden Quadrilateral | 218.50 | 212.33 | 4019 |
| Distance to North-South East-West | 191.48 | 155.68 | 4019 |
| Employment in wood-using firms | 141.41 | 299.77 | 4019 |
| Employment in logging firms | 9.78 | 92.66 | 4019 |

The table shows summary statistics for the samples used for village- and subdistrict-level analyses. Road completion year is shown only for villages that received new roads between 2001 and 2011. The sample for the first four village-level variables consists of the set of villages that did not have a road at baseline. The sample for agricultural land and energy shares consists of all villages with non-zero forest cover at baseline.

**Table 2**

Regression Discontinuity Estimates of Impact of
Rural Roads on Forest Cover

| | First Stage | Reduced Form | | | | | IV | |
|---|---|---|---|---|---|---|---|---|
| | Any Road | Log Forest | Avg Forest | High Baseline | High ST | Low Assets | Log Forest | Avg Forest |
| Above Population Threshold | 0.163*** | 0.003 | 0.042 | -0.006 | 0.007 | 0.017 | | |
| | (0.010) | (0.011) | (0.060) | (0.013) | (0.014) | (0.017) | | |
| New Road | | | | | | | 0.016 | -0.062 |
| | | | | | | | (0.065) | (0.539) |
| N | 89476 | 89476 | 89476 | 44880 | 44388 | 35520 | 89476 | 89476 |
| r2 | 0.25 | 0.80 | 0.56 | 0.69 | 0.83 | 0.78 | 0.80 | 0.36 |

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

The table shows regression discontinuity treatment estimates of the effect of new village roads on local forest cover, estimated with Equation 2. In Column 1, the dependent variable is an indicator that takes the value one if a village received a new road in the sample period. Above Population Threshold is an indicator for a village population being above the treatment threshold. Columns 2 through 6 show reduced form estimates of the effect of being above the treatment population threshold. The dependent variables in Columns 2 and 3 respectively are log village forest cover and average covered share of each village pixel; the data source is Vegetation Continuous Fields. Columns 4 through 6 run the log forest cover specification on subgroups defined respectively by (i) above-median forest cover villages; (ii) above median share of Scheduled Tribes in a village; and (iii) below median baseline village assets. Columns 7 and 8 show IV estimates of the treatment effects of new roads, using respectively log and average forest cover as dependent variables. The sample for Columns 2 through 8 includes forest cover estimates for years 2011 through 2013 for increased precision; robust standard errors are clustered at the village level to account for serial correlation. All estimates include district-population threshold fixed effects and a control for baseline forest cover.

**Table 3**

Difference-in-Differences Estimates of Impact of Rural Roads on Forest Cover

| | Log Forest | | Average Forest | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Award Period | -0.005*** | | -0.035*** | |
| | (0.002) | | (0.013) | |
| Completion Period | 0.002 | 0.005*** | 0.009 | 0.014 |
| | (0.002) | (0.002) | (0.015) | (0.012) |
| District-Year F.E. | Yes | Yes | Yes | Yes |
| Village F.E. | Yes | Yes | Yes | Yes |
| N | 689745 | 689745 | 689745 | 689745 |
| r2 | 0.94 | 0.94 | 0.92 | 0.92 |

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

The table shows difference-in-differences estimates of the impact of new village roads on local forest cover. We define forest cover as log village forest cover (Columns 1 and 2) and average covered share of each village pixel (Columns 3 and 4); the data source is Vegetation Continuous Fields. The sample consists strictly of villages that received new roads between 2001 and 2013, and were not accessible by paved road in 2001. Award Period is an indicator variable that takes the value one for years after a road contract was awarded and before the road was completed. Completion period is an indicator variable that marks the years after a village's new road was built. All regressions include district*year fixed effects, village fixed effects, baseline population * year fixed effects, and baseline forest * year fixed effects. Standard errors are clustered at the village level to correct for serial correlation.

**Table 4**

Rural Roads and Deforestation:
Heterogeneity of Difference-in-Differences Estimates

| | Baseline Forest | | ST Share | | Asset Poverty | |
|---|---|---|---|---|---|---|
| | High | Low | High | Low | Poor | Not Poor |
| Award Period | -0.005* | -0.005*** | -0.003 | -0.006*** | -0.004 | -0.006*** |
| | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) |
| Completion Period | -0.002 | 0.002 | 0.000 | 0.003 | 0.001 | 0.001 |
| | (0.004) | (0.002) | (0.003) | (0.003) | (0.004) | (0.003) |
| N | 342345 | 346860 | 344760 | 344580 | 264810 | 424545 |
| r2 | 0.86 | 0.92 | 0.93 | 0.95 | 0.94 | 0.95 |

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

The table shows difference-in-differences estimates of the impact of new village roads on local forest cover, along three dimensions of heterogeneity. Forest cover is defined as log village forest cover; the data source is Vegetation Continuous Fields. Columns 1 and 2 respectively show estimates for villages with above and below median baseline forest cover. Columns 3 and 4 respectively show estimates for villages and above and below median population share of members of Scheduled Tribes. Columns 5 and 6 respectively show estimates for below- and above-median shares of households who report no assets in the 2002 Below Poverty Line survey. The sample consists strictly of villages that received new roads between 2001 and 2013, and were not accessible by paved road in 2001. Award Period is an indicator variable that takes the value one for years after a road contract was awarded and before the road was completed. Completion period is an indicator variable that marks the years after a village's new road was built. All regressions include district*year fixed effects, village fixed effects, baseline population * year fixed effects, and baseline forest * year fixed effects. Standard errors are clustered at the village level to correct for serial correlation.

# Table 5
## Difference-in-Differences Estimates of Impact of Highways on Forest Cover

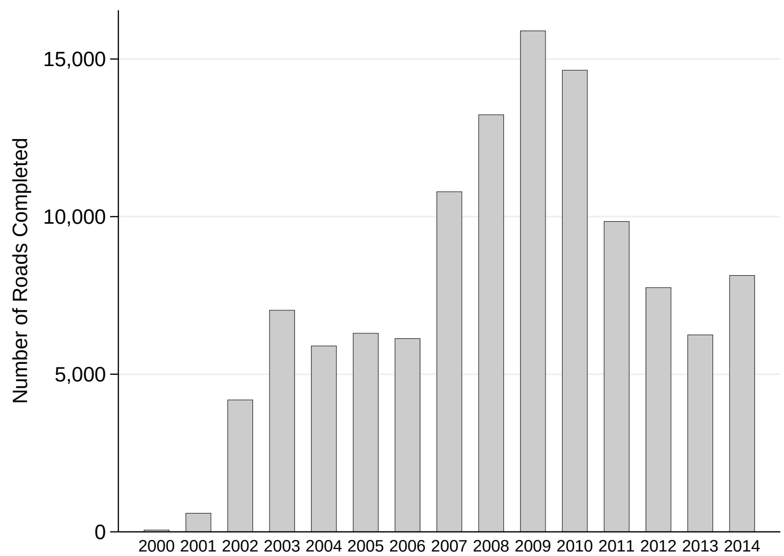
| | GQ (Treatment) | | NSEW (Placebo) | |
|---|---|---|---|---|
| | Log Forest | Avg Forest | Log Forest | Average Forest |
| GQ Construction Period * (0-50km) | -0.265*** | -1.306*** | -0.038 | 0.183 |
| | (0.055) | (0.248) | (0.072) | (0.305) |
| GQ Construction Period * (50-100km) | -0.278*** | -1.215*** | -0.001 | 0.077 |
| | (0.056) | (0.247) | (0.065) | (0.286) |
| GQ Construction Period * (100-150km) | -0.221*** | -1.085*** | -0.001 | -0.176 |
| | (0.051) | (0.235) | (0.060) | (0.267) |
| GQ Construction Period * (150-200km) | -0.102** | -0.444** | 0.023 | 0.004 |
| | (0.044) | (0.198) | (0.057) | (0.264) |
| GQ Post Period * (0-50km) | -0.210*** | -1.161*** | -0.013 | 0.413 |
| | (0.061) | (0.253) | (0.062) | (0.319) |
| GQ Post Period * (50-100km) | -0.185*** | -1.023*** | 0.022 | 0.106 |
| | (0.060) | (0.249) | (0.061) | (0.308) |
| GQ Post Period * (100-150km) | -0.131** | -0.855*** | 0.019 | -0.200 |
| | (0.058) | (0.226) | (0.060) | (0.294) |
| GQ Post Period * (150-200km) | -0.008 | -0.198 | 0.028 | -0.012 |
| | (0.051) | (0.197) | (0.060) | (0.301) |
| Distance 0-50km | 0.022 | 0.145 | -0.014 | 0.265 |
| | (0.021) | (0.100) | (0.020) | (0.722) |
| Distance 50-100km | 0.021 | 0.151 | -0.021 | 0.394 |
| | (0.020) | (0.099) | (0.018) | (0.717) |
| Distance 100-150km | 0.026 | 0.133 | -0.006 | -0.582 |
| | (0.019) | (0.083) | (0.021) | (0.784) |
| Distance 150-200km | 0.015 | 0.075 | -0.000 | -0.560 |
| | (0.014) | (0.060) | (0.016) | (0.697) |
| N | 26766 | 26766 | 19062 | 19062 |
| r2 | 0.89 | 0.91 | 0.92 | 0.85 |

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

The table shows treatment estimates for the impact of the construction of the GQ highway network on forest cover in the proximity of the highway, according to Equation 6. We define forest cover as log village forest cover (Columns 1 and 3) and average covered share of each village pixel (Columns 2 and 4); the data source is Vegetation Continuous Fields. The distance variables are indicators that identify places within a given distance band from the GQ (Columns 1 and 2) or the NS-EW highway network (Columns 3 and 4). The omitted category is the band of places at a distance of 200-300km from the highway network. These distance band indicators are then interacted with time period indicators. The construction period (rows 1 through 4) is 2001 to 2004. The post period (rows 5 through 8) is 2005 to 2008. Columns 3 and 4 estimate a placebo specification with distances to the NS-EW highway network, where construction had barely begun by 2008. The sample includes data from 2000 to 2008; 2000 is the omitted period. We omit years after 2008 as the placebo group is treated in those years. In Columns 3 and 4, we exclude places within 150km of the GQ network to prevent sample contamination. All estimates include state-year fixed effects and standard errors are clustered at the subdistrict level to account for serial correlation.

# A Appendices: Additional Tables and Figures

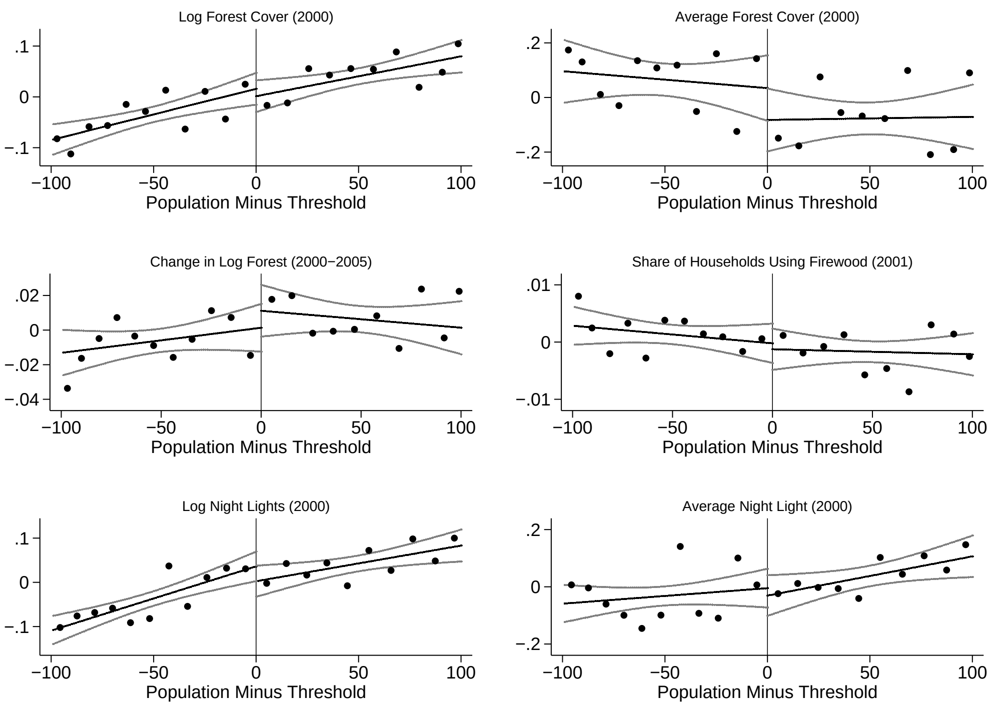
**Figure A1**
Road Program Completion Dates




The figure shows the number of roads completed under the PMGSY road construction program, by year.

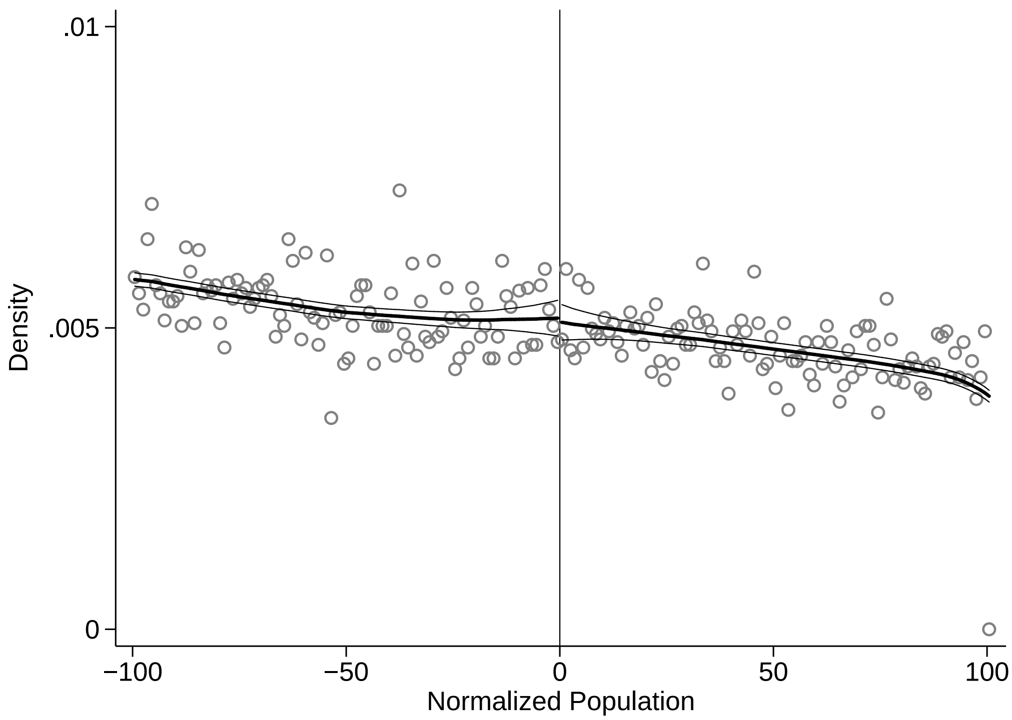
**Figure A2**

Regression Discontinuity Balance Tests




The figure displays a graphical form of the regression discontinuity balance test. Each graph shows the means of a variable measured at baseline in bins defined by population relative to the rural road program treatment threshold. The linear fits and standard errors are estimated from Equation 2. The vertical line shows the treatment threshold; the jump in the fit at this line is the regression discontinuity treatment estimate. The dependent variable in each panel (left-to-right, top-to-bottom) is (A) log forest cover in 2000; (B) average forest cover in 2000; (C) change in log forest cover from 2000 to 2005; (D) the share of households whose primary cooking fuel is firewood; (E) the log of night light luminosity in 2000; and (F) average night light luminosity in 2000. In Panel C, we omit villages with roads built before 2006, to ensure that balance estimates are not contaminated by the small number of treated villages before this date. All estimates include district-population threshold fixed effects.

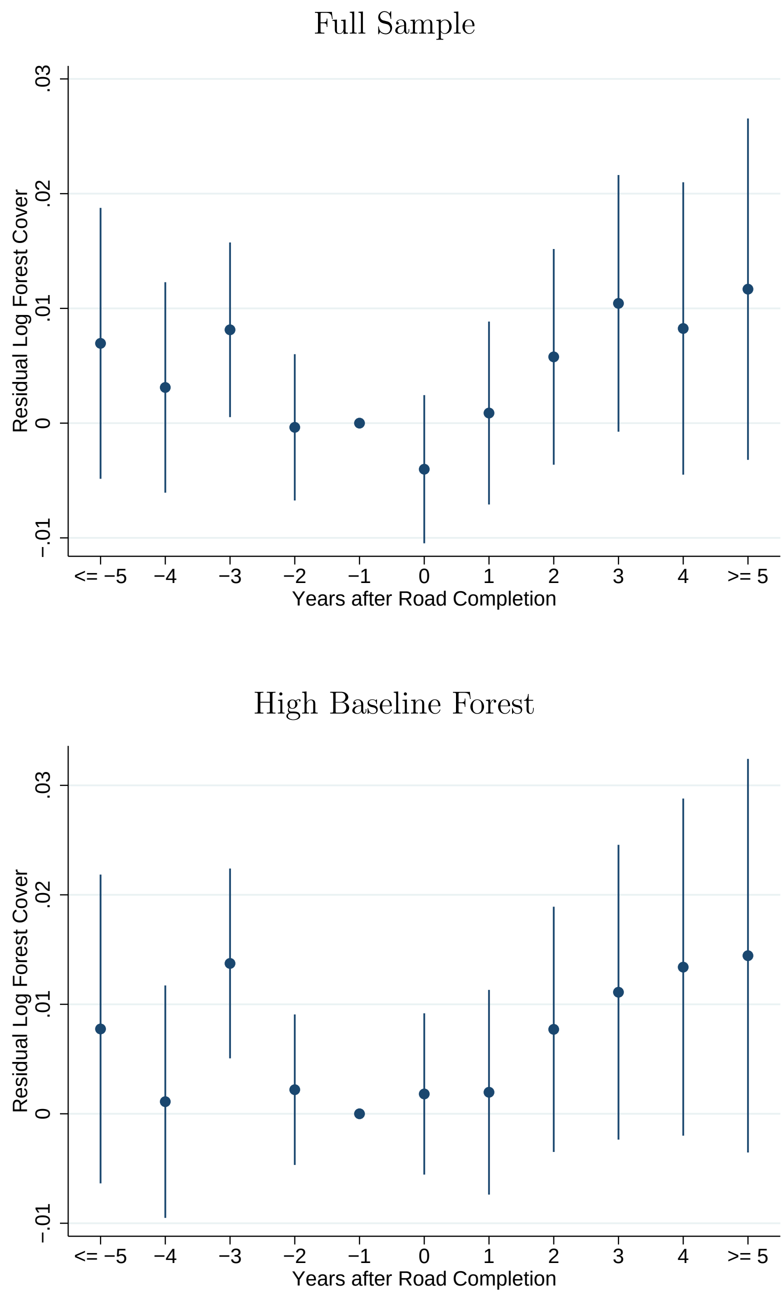
**Figure A3**
Regression Discontinuity Density Test




The figure displays a graph from a regression discontinuity density test (McCrary, 2008). The X axis shows the population relative to the road program treatment eligibility threshold. The Y axis shows a kernel estimate of the density of villages in a given normalized population band. The lines display non-parametric fits to the density function along with 95% confidence intervals.

# Figure A4

Difference-in-Differences Estimates of Impact of Rural Roads on Forest Cover (Long Panel)




The figure shows year-by-year estimates of log forest cover in villages that received new roads between 2001 and 2013. The figure is identical to Figure 3, but with an additional estimate for the 5th year before and after treatment. Villages are grouped on the X axis according to the year relative to road completion. Each point thus shows the average value of log forest cover in villages in a given year relative to the treatment year, controlling for village fixed effects, district*year fixed effects and baseline population * year and baseline log forest cover * year interactions. Standard errors are clustered at the village level.

**Table A1**

OLS Regressions of Forest Cover on Rural Road Indicators

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Paved Road in 2001 | -0.162*** | -0.270*** | -0.041*** | -0.025*** | -0.018** |
| | (0.010) | (0.010) | (0.009) | (0.009) | (0.009) |
| Population | | 0.476*** | 0.862*** | 0.896*** | 0.910*** |
| | | (0.013) | (0.012) | (0.012) | (0.012) |
| Population$^2$ | | -0.037*** | -0.085*** | -0.092*** | -0.095*** |
| | | (0.004) | (0.003) | (0.003) | (0.003) |
| Distance in km to town of 10,000 | | | | | 0.007*** |
| | | | | | (0.000) |
| Distance in km to town of 100,000 | | | | | -0.000 |
| | | | | | (0.000) |
| Constant | 3.405*** | 3.083*** | 2.786*** | 2.764*** | 2.533*** |
| | (0.004) | (0.008) | (0.007) | (0.007) | (0.013) |
| Fixed Effects | None | None | State | State | District |
| N | 270871 | 270871 | 270871 | 270871 | 270871 |
| r2 | 0.00 | 0.02 | 0.19 | 0.28 | 0.28 |

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

The table shows estimates from OLS regressions of village-level log forest cover in 2001 on an indicator variable that takes the value one if a village has a paved road in 2001. Column 1 presents the bivariate estimates, and Columns 2 through 5 present estimates with progressively greater numbers of controls and fixed effects. Forest cover is calculated from Vegetation Continuous Fields. Population is measured in millions of people.

**Table A2**

Regression Discontinuity Balance Tests

| Variable | RD Estimate |
|---|---|
| Log Forest (2000) | -0.012 |
| | (0.027) |
| Average Forest (2000) | -0.092 |
| | (0.101) |
| Share Cooking with Firewood | -0.001 |
| | (0.003) |
| Log Forest Change (2000-2005) | 0.012 |
| | (0.012) |
| Mean Night Light (2000) | -0.020 |
| | (0.060) |
| Log Night Light (2000) | -0.033 |
| | (0.029) |
| Number of Observations | 55222 |

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

The table shows estimates from a regression discontinuity balance test. We run the regression discontinuity specification defined by Equation 2 on variables measured before any rural road construction took place, and report the reduced form treatment estimates. Row 4 (Log Forest Change 2000-2005) tests for pretrends in forest cover; we exclude villages with roads built before 2006 for this sample. All estimates include district-population threshold fixed effects.

## Table A3

Regression Discontinuity Estimates of Impact of
Rural Roads on Forest Cover (Alternate Bandwidths)

| | | Log Forest (2013) | | | | Average Forest (2013) | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Above Population Threshold | 0.015 | 0.003 | 0.002 | -0.002 | 0.124 | 0.042 | 0.035 | 0.017 |
| | (0.015) | (0.011) | (0.009) | (0.008) | (0.085) | (0.060) | (0.050) | (0.043) |
| Bandwidth | 50 | 100 | 150 | 200 | 50 | 100 | 150 | 200 |
| N | 45112 | 89476 | 133908 | 178292 | 45112 | 89476 | 133908 | 178292 |
| r2 | 0.80 | 0.80 | 0.80 | 0.80 | 0.56 | 0.56 | 0.57 | 0.57 |

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

The table shows reduced form regression discontinuity estimates of the impact of rural roads on forest cover. The specifications are comparable to those in Columns 2 and 3 of Table 2 but with alternate bandwidths defined in the bandwidth row of the table. The sample includes forest cover estimates for years 2011 through 2013 for increased precision; robust standard errors are clustered at the village level to account for serial correlation. All estimates include district-population threshold fixed effects and a control for baseline forest cover.

**Table A4**

Regression Discontinuity Estimates of Impact of Rural Roads on Household Fuel Use

| | Imports | Local Non-Wood | Firewood |
|---|---|---|---|
| Above Population Threshold | -0.002 | 0.002 | 0.000 |
| | (0.002) | (0.006) | (0.006) |
| N | 22318 | 22318 | 22318 |
| r2 | 0.28 | 0.42 | 0.42 |

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

The table shows reduced form regression discontinuity treatment estimates of the effect of new village roads on village-level household fuel use, estimated with Equation 2. The dependent variable is the share of households in a village that use imported fuel sources (primarily propane, Column 1); dung and crop residue (Column 2); and firewood (Column 3) as primary fuel sources for cooking. The dependent variables are measured in 2011. In addition to district-population threshold fixed effects, controls include the baseline fuel share reported in 2001 (at the subdistrict level) and forest cover in 2000.

**Table A5**

Difference-in-Differences Estimates of Impact of Rural Roads on Forest Cover (Robustness Tests)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Award Period | -0.004** | -0.006*** | -0.010*** | -0.007*** |
| | (0.002) | (0.002) | (0.003) | (0.002) |
| Completion Period | 0.001 | -0.000 | -0.002 | -0.002 |
| | (0.002) | (0.002) | (0.004) | (0.003) |
| District-Year F.E. | Yes | No | Yes | Yes |
| Subdistrict-Year F.E. | No | Yes | No | No |
| Village F.E. | Yes | Yes | Yes | Yes |
| Village Time Trends | Yes | No | No | No |
| Panel Sample | Full | Full | +/- 5 Years | +/- 4 Years |
| N | 689745 | 683025 | 374385 | 482130 |
| r2 | 0.95 | 0.96 | 0.94 | 0.94 |

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

The table shows difference-in-differences estimates of the impact of new village roads on local forest cover, under alternate sample definitions. We define forest cover as log village forest cover; the data source is Vegetation Continuous Fields. Specifications are identical to those in Table 3 with the following changes. Column 1 includes village-specific time trends. Column 2 uses subdistrict-year fixed effects instead of district-year fixed effects. Column 3 restricts the sample to villages with roads for which we can observe at least 5 years of data before road completion and 5 years after. Column 4 does the same, with 4 years. The sample consists strictly of villages that received new roads between 2001 and 2013, and were not accessible by paved road in 2001. Award Period is an indicator variable that takes the value one for years after a road contract was awarded and before the road was completed. Completion period is an indicator variable that marks the years after a village's new road was built. All regressions include district*year fixed effects, village fixed effects, baseline population * year fixed effects, and baseline forest * year fixed effects. Standard errors are clustered at the village level to correct for serial correlation.

**Table A6**

Difference-in-Differences Estimates of Impact of
Rural Roads on Forest Cover (Alternate Specifications)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Award Period | 0.004*** | -0.002 | -0.005*** | -0.004** | -0.005** | -0.001 |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) | (0.001) |
| Completion Period | 0.017*** | 0.000 | 0.002 | 0.001 | 0.003 | -0.000 |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| District-Year F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Village F.E. | Yes | Yes | Yes | Yes | Yes | Yes |
| Village Time Trends. | No | Yes | No | Yes | No | No |
| Village Definition | Boundary | Boundary | Boundary | Boundary | 5 km radius | 50 km radius |
| N | 3359370 | 3359370 | 689745 | 689745 | 688275 | 688275 |
| r2 | 0.95 | 0.96 | 0.94 | 0.95 | 0.94 | 0.97 |

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

The table shows difference-in-differences estimates of the impact of new village roads on local forest cover, under alternate sample definitions. We define forest cover as log village forest cover; the data source is Vegetation Continuous Fields. Specifications are identical to those in Table 3 with the following changes. Column 1 includes villages that did not receive PMGSY roads at any time as part of the control group. Column 2 adds village-specific time trends to show that the positive treatment estimate in Column 1 is driven by differential trends in never-treated villages. Columns 3 and 4 repeat these two specifications with the standard set of villages to show that village-specific time trends do not affect our main estimates. Column 5 estimates the standard specification from Table 3 with treated villages only, but the dependent variable includes forest cover in a 5km radius from the village centroid. Column 6 uses a 50km centroid. Award Period is an indicator variable that takes the value one for years after a road contract was awarded and before the road was completed. Completion period is an indicator variable that marks the years after a village's new road was built. All regressions include village fixed effects, baseline population * year fixed effects, and baseline forest * year fixed effects. Standard errors are clustered at the village level to correct for serial correlation.

# Table A7

Difference-in-Differences Estimates of Impact of Highways on Forest Cover: Straight Line Instrumental Variables

| | GQ (Straight Line) | | NSEW (Straight Line) | |
|---|---|---|---|---|
| | Log Forest | Avg Forest | Log Forest | Average Forest |
| GQ Construction Period * (0-50km) | -0.144*** | -0.442*** | -0.013 | 0.379 |
| | (0.026) | (0.142) | (0.055) | (0.314) |
| GQ Construction Period * (50-100km) | -0.181*** | -0.749*** | -0.010 | 0.199 |
| | (0.027) | (0.142) | (0.052) | (0.297) |
| GQ Construction Period * (100-150km) | -0.129*** | -0.495*** | 0.014 | -0.336 |
| | (0.029) | (0.164) | (0.048) | (0.272) |
| GQ Construction Period * (150-200km) | -0.086*** | -0.362** | -0.030 | -0.157 |
| | (0.031) | (0.178) | (0.044) | (0.268) |
| GQ Post Period * (0-50km) | -0.113*** | -0.422*** | -0.015 | 0.528 |
| | (0.026) | (0.133) | (0.053) | (0.332) |
| GQ Post Period * (50-100km) | -0.110*** | -0.518*** | -0.061 | -0.051 |
| | (0.027) | (0.137) | (0.054) | (0.329) |
| GQ Post Period * (100-150km) | -0.057** | -0.138 | -0.027 | -0.374 |
| | (0.028) | (0.154) | (0.051) | (0.316) |
| GQ Post Period * (150-200km) | -0.050* | -0.186 | -0.051 | -0.361 |
| | (0.029) | (0.162) | (0.042) | (0.320) |
| Distance 0-50km | 0.019*** | 0.179*** | -0.056*** | -0.166 |
| | (0.007) | (0.051) | (0.017) | (0.801) |
| Distance 50-100km | 0.033*** | 0.282*** | -0.057*** | 1.903** |
| | (0.008) | (0.055) | (0.016) | (0.886) |
| Distance 100-150km | 0.033*** | 0.273*** | -0.045*** | 0.480 |
| | (0.007) | (0.051) | (0.016) | (0.819) |
| Distance 150-200km | 0.024*** | 0.195*** | -0.019* | -0.342 |
| | (0.005) | (0.039) | (0.011) | (0.866) |
| N | 26397 | 26397 | 14958 | 14958 |
| r2 | 0.90 | 0.90 | 0.94 | 0.86 |

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

The table shows reduced form estimates from regressions of distance band * time period interactions on forest cover. Distance bands are calculated to straight line approximations of the Golden Quadrilateral (Columns 1 and 2) and North-South/East-West (Columns 3 and 4) highway corridors. The estimating equation is Equation 6. We define forest cover as log village forest cover (Columns 1 and 3) and average covered share of each village pixel (Columns 2 and 4); the data source is Vegetation Continuous Fields. The omitted distance category is the set of subdistricts at a distance of 200-300km from each set of straight line approximations. The construction period (rows 1 through 4) is 2001 to 2004. The post period (rows 5 through 8) is 2005 to 2008. Columns 3 and 4 estimate a placebo specification with distances to the NS-EW highway network, where construction had barely begun by 2008. The sample includes data from 2000 to 2008; 2000 is the omitted period. We omit years after 2008 as the placebo group is treated in those years. In Columns 3 and 4, we exclude places within 150km of the GQ network to prevent sample contamination. All estimates include state-year fixed effects and standard errors are clustered at the subdistrict level to account for serial correlation.

# Table A8

Mechanism Tests for Impact of Highways on Deforestation: Employment in Wood-Using Firms

| | Wood Use | Logging |
|---|---|---|
| (0-100km from GQ) * 1(Year == 1990) | 0.003 | 0.033*** |
| | (0.016) | (0.005) |
| (100-200km from GQ) * 1(Year == 1990) | -0.028* | 0.018*** |
| | (0.014) | (0.003) |
| (0-100km from GQ) * 1(Year == 1998) | 0.007 | 0.032*** |
| | (0.018) | (0.006) |
| (100-200km from GQ) * 1(Year == 1998) | -0.044*** | 0.023*** |
| | (0.015) | (0.005) |
| (0-100km from GQ) * 1(Year == 2005) | 0.052*** | |
| | (0.017) | |
| (100-200km from GQ) * 1(Year == 2005) | -0.011 | |
| | (0.014) | |
| (0-100km from GQ) * 1(Year == 2013) | 0.081*** | 0.055*** |
| | (0.016) | (0.008) |
| (100-200km from GQ) * 1(Year == 2013) | -0.009 | 0.041*** |
| | (0.012) | (0.008) |
| N | 1037954 | 766458 |
| r2 | 0.17 | 0.06 |

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

The table shows estimates of the impact of the Golden Quadrilateral highway on log employment in timber-related firms. The table shows the full specifications used to generate Figure 7, panels A and B. The dependent variable is log employment in firms for which timber is the primary input (sawmilling, pulp and paper, manufacture of wooden containers, wooden furniture and cork boards, Column 1) and log employment in logging firms (Column 2). Each row shows the interaction of an indicator for a given distance band from the Golden Quadrilateral, interacted with an indicator for a given Economic Census year. The omitted distance category is 200-300km. The estimates thus show the difference between log employment in each sector/year/distance band with log employment in the same sector/year at distsance 200-300km from the highway. Logging is not specifically identified in the 2005 Economic Census, so this estimate is omitted. All regressions include state-year fixed effects and cluster standard errors at the subdistrict level.

**Table A9**

Mechanism Tests for Impact of Highways on Deforestation:
Land and Fuel Use

| | Ag Land Share | Fuel: Firewood | Fuel: Imported | Fuel: Local Non-wood |
|---|---|---|---|---|
| (0-100km from GQ) * 1(Year == 1991) | -0.085*** | | | |
| | (0.016) | | | |
| (100-200km from GQ) * 1(Year == 1991) | -0.080*** | | | |
| | (0.014) | | | |
| (0-100km from GQ) * 1(Year == 2001) | -0.069*** | -0.008 | 0.010*** | -0.001 |
| | (0.011) | (0.015) | (0.003) | (0.015) |
| (100-200km from GQ) * 1(Year == 2001) | -0.066*** | 0.051*** | -0.003 | -0.049*** |
| | (0.009) | (0.012) | (0.003) | (0.012) |
| (0-100km from GQ) * 1(Year == 2011) | -0.097*** | 0.001 | 0.013*** | -0.014 |
| | (0.010) | (0.014) | (0.004) | (0.014) |
| (100-200km from GQ) * 1(Year == 2011) | -0.054*** | 0.060*** | 0.004 | -0.064*** |
| | (0.008) | (0.012) | (0.003) | (0.012) |
| N | 813691 | 638458 | 638458 | 638458 |
| r2 | 0.30 | 0.36 | 0.22 | 0.40 |

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

The table shows estimates of the impact of the Golden Quadrilateral highway on land and fuel use. The table shows the full specifications used to generate Figure 7, panels C through F. The dependent variable is the share of village land dedicated to agriculture (Column 1); the share of households in a village that use firewood (Column 2); imported fuel sources (primarily propane, Column 3); and dung and crop residue (Column 4) as primary fuel sources for cooking. Each row shows the interaction of an indicator for a given distance band from the Golden Quadrilateral, interacted with an indicator for a given Population Census year. The omitted distance category is 200-300km. The estimates thus show the difference between the outcome variable in each sector/year/distance band with the outcome variable in the same sector/year at distsance 200-300km from the highway. All regressions includes state-year fixed effects and cluster standard errors at the subdistrict level.