Journal of Environmental Economics and Management 61 (2011) 79–94

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

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

Pressure cookers or pressure valves: Do roads lead to deforestation in China?

Xiangzheng Deng [a], Jikun Huang [a], Emi Uchida [b], Scott Rozelle [c,*], John Gibson [d]

[a] Center for Chinese Agricultural Policy, Institute for Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, China
[b] Department of Environmental and Natural Resource Economics, University of Rhode Island, USA
[c] Freeman Spogli Institute, Stanford University, USA
[d] Department of Economics, University of Waikato, New Zealand

ARTICLE INFO

Article history:
Received 10 July 2008
Available online 17 August 2010

Keywords:
Deforestation
Afforestation
Road
Land use
Forest
Forest cover
One kilometer square pixel
OLS
Covariate matching
Jiangxi

ABSTRACT

The effect of roads on forests is ambiguous. Many studies conclude that building and upgrading roads increases pressure on forests but some find that new and better roads may reduce the rate of deforestation. In this paper we use satellite remote sensing images of forest cover in Jiangxi Province, China, to test whether the existence and the size of roads (ranging from expressways to tertiary roads) in 1995 affected the level of forest cover in 2000 or the rate of change between 1995 and 2000. To account for road access for each of our 1 km^2 ("pixel") units of forest cover we measure whether or not and what type of roads penetrate the "watershed" in which the pixel lies. These watersheds allow more plausible measures of accessibility than do traditional "crowfly" distance measures that ignore topography. To account for possible confounding we also use 12 additional covariates: geographic and climatic variables (e.g., elevation, slope, rainfall, temperature, soil properties); demographic and economic variables (e.g., local population and GDP per square kilometer); and distance variables (e.g., distance to the nearest provincial capital). Although simple univariate OLS regressions show that forest levels are lower and deforestation rates higher either when there is a road, or when there is a higher quality road, these results are not robust. Controlling for all of the covariates and also using recently developed covariate matching techniques to estimate treatment effects, we find that roads in China's Jiangxi Province can most safely be described as having no impact on the level of forests and no impact on the rate of deforestation.

1. Introduction

The world lost 3% of its total forested area between 1990 and 2005 [1]. With rising concern over many environmental issues related to deforestation, such as climate change and conservation of biological diversity, increasing efforts are being made by economists, ecologists, geographers and other scientists to understand the causes of deforestation [2–6]. Seeking to identify statistically the determinants of deforestation, the literature finds that various geophysical factors, such as slope and elevation, demographic factors, economic variables and policies and other actions of governments (e.g., setting up of park systems [2]) are all important correlates of forest cover and deforestation [7–9].

* Corresponding author. Fax: +1 650 723 6530.
E-mail address: rozelle@stanford.edu (S. Rozelle).

0095-0696/$ - see front matter
doi:10.1016/j.jeem.2010.04.005

Above all of these factors, however, probably the largest literature—and most rigorous set of empirical papers—has focused on the relationship between roads and deforestation. In many instances roads are found to lead to deforestation [7]. The logic is that when a road enters an area (or when it is widened or improved), pressure will rise on the forest resources in the area and forest cover will fall. An important implication of what we characterize as this ''pressure cooker'' hypothesis is that ''road networks [...] may significantly shape the spatial pattern of remaining forest'' [7]. Hence, road investments in forested areas of tropical developing countries are thought to lead to deforestation.

While evidence for this ''pressure cooker'' hypothesis comes especially from the Amazon [2,5], other studies from Belize [8], the Philippines [10], Honduras [11], Cameroon [12] and Costa Rica [13] also find that roads are associated with deforestation. Yet there are some notable contrary findings in the literature, including a high profile study on deforestation and economic growth in the Brazilian Amazon [14]. These authors find that new roads reduce the rate of deforestation, at least in counties with substantial prior clearing of forest. The suggested causal mechanism is that roads provide a focal point for local development, drawing pressure away from the forested hinterland.

According to this ''pressure valve'' hypothesis, when new or better roads reach a region, access to transportation and new and more convenient linkages to the outside world encourages growth, produces jobs and increases agricultural productivity. If these dynamics are able to refocus the livelihood strategies of households that previously were encroaching on forests into intensive (river-bottom; irrigated) agriculture and off-farm employment, including migration, the pressure on the forest might be reduced. In addition to the Andersen et al. [14] study, other evidence in favor of the pressure valve hypothesis comes from Deininger and Minten [15], who find that roads appear to diminish the negative impact of high poverty levels on forests in Southern Mexico, and Qiao et al. [16], who find that forested area rose when village economies in southern China became more integrated with markets. Clearly, at least in some areas and in some studies, there is evidence that a positive ''development effect'' can lead to less pressure on forests.

In fact, there has been a fairly large literature that discusses the mechanism that may be underlying the pressure-valve hypothesis. Such a phenomenon could arise in part as a result of increased opportunities to purchase inputs that increase or maintain yields [17,18]. In contrast, Zwane [19] demonstrates that if the labor market is imperfect, the poverty–deforestation hypothesis becomes ambiguous. In any case, to the extent that roads—through lower transportation costs—increase agricultural productivity and deepen the labor market by creating new jobs, roads can refocus the livelihood strategies of households that previously were encroaching on forests into intensive (river-bottom, irrigated) agriculture and off-farm migration and therefore reduce pressure on the forest.

The ambiguities about the effect of roads on forests have led to refinements in the data and methods used in more recent empirical studies. For example, Pfaff et al. [6] use a finer spatial scale to re-evaluate the findings of Andersen et al. [14], when they move from using 250 counties to more than 6000 census tracts they find deforestation rises in tracts with new roads and also elsewhere in the same county within 100 km. However, it is not clear how reasonable such ''crowfly'' distance measures are that ignore topography; two tracts in the same county could be separated by considerable physical barriers like mountains. It is also not clear that traditional multiple regression methods of including control variables can provide estimates of treatment effects that are robust to the endogenous placement of roads.

Since existing evidence on the effects of roads on forests is unclear, and has not always benefited from latest refinements in data and methods; new evidence is required, which this paper reports. Specifically, we use satellite remote sensing images of forest cover in Jiangxi Province, China, to test whether the existence of roads and the size of roads (ranging from expressways to tertiary roads) in 1995 affected the level of forest cover in 2000 or the rate of change between 1995 and 2000. To account for road access for each of our 1 km^2 (''pixel'') units of forest cover we measure whether or not and what type of roads penetrate the ''watershed'' in which the pixel lies. These watersheds allow more plausible measures of accessibility than traditional ''crowfly'' distance measures, which ignore topography. To account for confounding from the exclusion of other relevant variables and potentially biased estimates of treatment effects due to the endogenous placement of roads, we also use covariate matching techniques, using 12 additional covariates. Our overall goal is to discover if roads are acting more like pressure cookers—and are associated with lower levels of forests and greater rates of deforestation in Jiangxi—or like pressure valves—and are associated with higher levels of forest and lesser rates of deforestation.

2. Data, case study area and definitions

One of the strengths of our study is the quality of data that we use to estimate the quantity of forested area in a given year and changes in forested area over time. For our purposes, we believe that variables based on satellite remote sensing digital images are the most suitable measures that can be used for detecting and monitoring land use change (LUC) at global and regional scales [3,20–24]. In China land use change—including changes in forested areas—has been tracked by remote sensing data and results of the empirical exercises have been reported in the literature [25–27]. An appendix available at JEEM's online archive of supplementary material provides technical details on the remote sensing data and other additional analyses. This archive can be accessed at 〈http://aere.org/journals〉.

The definition of forested area in our study is a comprehensive one that captures and includes a heterogeneous set of different types of forests. In using the raw land use data, forested area can be sub-divided into four distinct classes: (a) natural or planted forests with canopy cover greater than 30%; (b) land covered by trees less than 2 m high, with a canopy cover greater than 40%; (c) land covered by trees with a canopy cover between 10% and 30%; and (d) the rest of forested area. In this paper, we use a single measure of forested area that includes all four types of forest (a+b+c+d), a definition that should pick up most of the province's area that is covered by trees.

Interpreting the results requires that the setting of the study is clear. Jiangxi Province is located in central, southeastern China. Although completely land-locked and not a member of the group of China's coastal provinces that are driving the nation's modernization, Jiangxi shares a border with the nation's three fastest growing provinces—Zhejiang, Fujian and Guangdong. The proximity of Jiangxi to China's booming coastal provinces has given its rural population access to increasingly lucrative off farm employment opportunities. Its location south of the Yangtze River means it has temperate weather, which provides excellent conditions for forest growth [28]. Therefore, not surprisingly, Jiangxi is a highly forested province. Forests are both economically and environmentally important; the province is part of a traditional timber-producing region that formerly was one of the largest timber producing provinces in China. Jiangxi is also a densely populated province with a large agricultural sector. Appendix B provides a more complete description of the province's forests.

In such an environment one can see how the entry of roads in recent years could affect the forest—albeit, it is not necessarily clear what the net direction is. With new roads, it would be possible for previously isolated villages to use the new road network to begin to exploit the forest since markets for forest products were more readily available and transportations costs had fallen. On the other hand, with new opportunities for working off the farm and becoming engaged in intensive agricultural activities that do not require farmers to push their land into the mountains, it also is possible to see how farmers might use the roads to turn their backs on the forest, leaving them alone while focusing on non-forest activities. Likewise, tree planting sometimes may be labor-saving. With the opportunity cost of farmers rising (due to new off farm employment opportunities), the roads that make leaving easier, could also make afforestation efforts easier to organize and more incentive compatible with busy farmers.

According to our definition of the forest specified above, Jiangxi today is indeed a highly forested province. Although China's total forested area is reported by the State Forest Administration to be under 20%, our data suggest that Jiangxi's forest cover was 56.28% in 1995. By 2000, according to the data, forested area was 56.23%.[1]

Because our data are from two distinct sets of years, 1995 and 2000, they can be used to track the changes in the forest. Using the data from both years, we can show that the forest cover in Jiangxi decreased only slightly, by 0.05% points, between 1995 and 2000.[2] Although decreasing in the aggregate by only a small margin, Fig. 1, which charts the *differences in the forest cover* across time for each pixel in the data set, demonstrates two things. First, there is both deforestation and afforestation occurring at the same time. Second, based on the data used in Fig. 1, it can be shown that, in fact, the forest in Jiangxi is quite dynamic and changing over a large area. Specifically, the forest area was rising in more than 7.1% of the pixels. In contrast, the forest area was falling in more than 47.1% of the pixels. This means that, although the total change in forest area was low, the total share of the forest that was changing was high (54.2%=7.1%+47.1%). While such numbers suggest that deforestation is occurring, it must be remembered that the nature of deforestation and afforestation sometimes makes it difficult to interpret net changes in a "fair" way. Since the return of the forest often takes longer to occur than the time it takes to cut down the forest, it is possible that what appears like deforestation today, could mean that the forest area will rise in the future. One of the main questions of this paper is to understand if roads are encouraging the observed deforestation, or if they are part of the set of forces that facilitates afforestation.

The basic data for our roads variable come from provincial, county and local maps which are from the CAS data center. The maps are up to date through 1996. The information from the hard copies of the maps was digitized by a working group based in the Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Science, in 1999 and 2000. Although it would be simple to calculate the straight-line ("crowfly") distance from each pixel to the nearest segment of digitized road, such an approach is likely to provide a misleading measure of road access. Almost two-thirds of Jiangxi's land area is classified as mountainous or hilly, so travel from many pixels to the nearest road involves going over mountains, which may be impractical. In such an environment, a more realistic measure of accessibility requires knowledge of the topography, and in particular, of *watersheds*. Our assumption is that travel within a watershed is likely to be less impeded than travel between watersheds, so that if a road enters a watershed all pixels within that watershed are likely to have feasible access to the road.[3]

A second issue also arises concerning what is the appropriate unit into which we should divide our data. Should we use administrative boundaries, such as counties, or should we use natural boundaries, such as watersheds? It is plausible that the timber harvesting decisions are influenced by forest planning (or some other cross watershed policy or economic force) at a

[1] Forest inventory data released by China's State Forest Administration showed that the forest cover rate in Jiangxi had reached 59.7% in 1999. While higher than the rate estimated by the remote sensing data from 2000, the two can be mostly reconciled after accounting for the fact that this level included all the forestry land with a forest canopy of 20% or greater; it also included four types of forested areas: timber forests, economic forests, bamboo forests and scattered forests. Hence, the differences between the estimated rates of forested cover from census and remote sensing data mainly stem from the differences in definitions and the methods by which forested area is classified.

[2] Although the change in forest area in our paper is small, in fact, the rate is comparable to those of a few studies. Robalino et al. [9] used data from Costa Rica, where forest area in the study area declined by 1% (down from 47% to 46%) between 1986 and 1997. Pfaff et al. [6] used data from the Amazon, where 3.7% of forest area in 1976 was deforested by 1987. Mertens and Lambin [12] used data from Cameroon, where the forest cover declined by 7% during 1973–1986.

[3] As a robustness analysis, we also run the analyses by defining the road variable at the pixel level, i.e., whether or not a type(s) of road runs through the pixel itself.

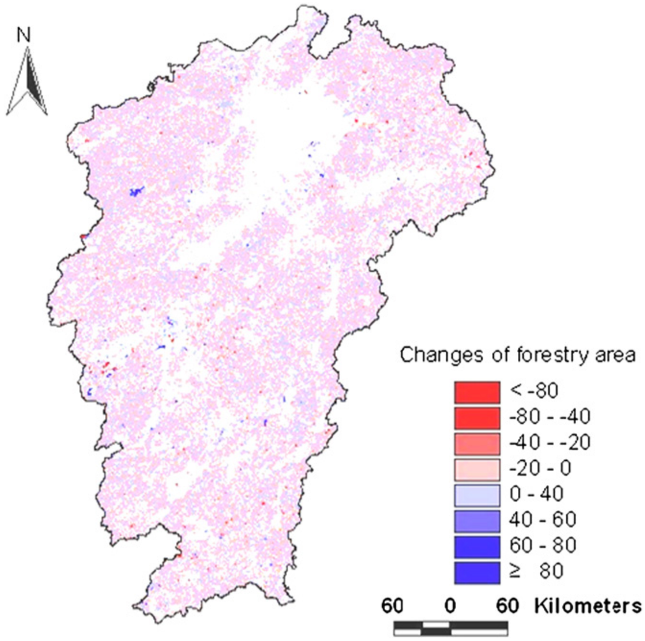

Fig. 1. Measurement the changes of forestry area of Jiangxi in each 1 km GRID pixel, measured in ha.

larger scale than the pixel level. For example, county-level forest policies may affect micro-level harvesting decisions. Planning of large roads also is likely to occur at a much larger scale than pixel. However, there are many other factors that suggest that natural watershed boundaries are more appropriate than administrative boundaries. This is especially true considering that the boundaries of watersheds reflect the inherent winding line of a watershed in which all the natural conditions of land use and the human development activities share the same characteristics. Because of this, local governments almost always use watersheds (instead of administrative boundaries) when they do environmental planning or development planning. Therefore, it seems most natural to use the boundary of watersheds as the analytical units in this study.

As a result of these factors and considerations, to take the digitized road map of the province and turn it into a discrete, pixel-specific measure of *the largest road that penetrates any part of each watershed*, we undertook a four step procedure. The first step began with a detailed GIS map of Jiangxi Province. We used this map—and a watershed delineation function included in ArcGIS—to divide the area of the province into distinct, non-overlapping watersheds. This process was accomplished using information on elevation and the formation of the landscape. In our study (as in most GIS-based watershed analyses) a watershed is defined as a set of spatially contiguous pixels, where if two drops of water were to fall on any two arbitrary pixels and if the drops were allowed to flow out of the watershed, they would both leave the watershed through the same point (or through the same *outlet pixel*).

Previous evidence suggests that the choice of spatial scale can affect results [6,29]. We therefore specified two different ranges for the number of watersheds into which the software attempts to divide the province: 1000–1500 (a finer division of watersheds) and 500–1000 (a coarser division of watersheds). After doing so, we created two maps in ArcGIS, which are used in different sets of our estimates—the main one we use divides Jiangxi into 1474 watersheds (Fig. 2, panel a). In Fig. 2, panel b, we include a map of Jiangxi that was divided into a lesser number of watersheds.[4]

In step two, the digitized road map and the digitized watershed map were merged.[5] Representations of the watershed maps overlaid with the road network are shown in Fig. 3, panels a and b. This merging then allowed each watershed to be assigned to one of four categories, according to *the size of the largest road that runs through the watershed*.

The four types of roads are *expressways* (which are multilane, controlled access highways); *province-level highways* (which are major roads that are typically not controlled access, but which are usually relatively well maintained since the province's highway bureau is charged with their maintenance); *other roads* (which are all major roads—county-level roads—except expressways and province-level highways) and *no road* (or pixels with no major roads—or those with only smaller, town- and village-level roads). After this step every watershed in the province is labeled with one (and only one) of four names: *expressway watershed*; *province-level highway watershed*; *other road watershed*; or a *no road watershed*. If a watershed has both an expressway and a province-level highway (and/or other road) inside its boundary, it is still, by definition, an expressway watershed.

[4] The findings of the paper are not affected by the aggregation level used to define watersheds. In fact, when we divided Jiangxi Province into 654 watersheds (Fig. 2, panel b) and reran all of the estimation exercises, the results of the study are not materially affected.

[5] The maps were drawn so that a given segment of road could not, simultaneously, be in two watersheds at once.

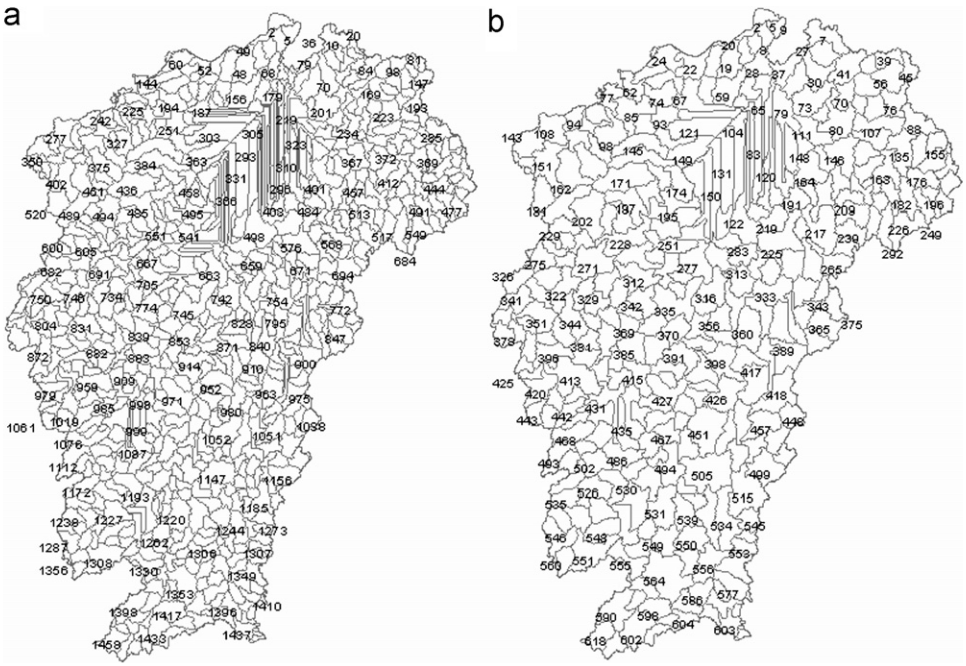

Fig. 2. Boundaries of watersheds in Jiangxi, 1474 watersheds (a) and 654 watersheds (b).

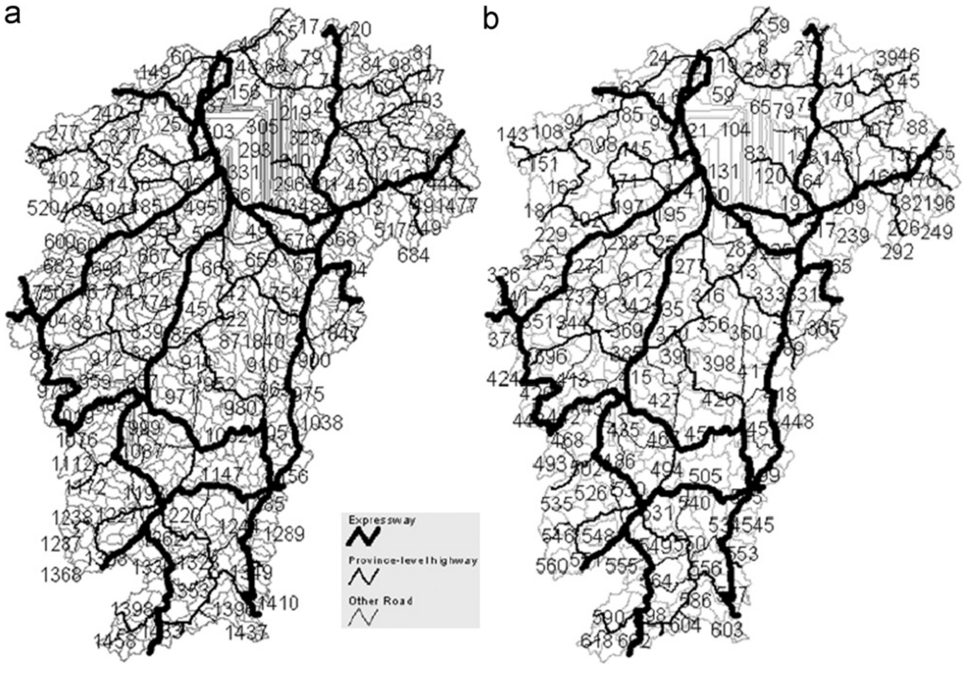

Fig. 3. Watershed maps overlaid with the road network, 1474 watersheds (a) and 654 watersheds (b).

The final step assigns all of the pixels in the watershed the same name as the watershed. This means that all 166,932 pixels in Jiangxi have one of four names—expressway pixel; province-level highway pixel; other road pixel; or no road pixel.[6] Appendix C also provides a more detailed explanation of the pixel labeling process.

[6] It is important to remember that since watersheds are of different sizes and since the distance of any given pixel in a watershed to the road is different, not all pixels within each of the watersheds have equal access to the road. While this feature of the data is unavoidable, it may add some error to

2.1. Other control variables

In addition to information on forested area and roads, other data are used to create control variables for the other factors, all of which have been consistently found to causally affect deforestation in a review paper by Kaimowitz and Angelsen [31]. When looking at empirical literature on the determinants of forest cover, there are four broad categories of variables: geographic and climatic variables [8]; demographic and economic variables [8,30]; measures of distance from the forested plots to different features (such as, distance to the nearest city) [2,9]; and other factors, such as whether or not the pixel is in a protected area. In order to make our analysis as consistent as possible with the existing literature, we have collected information on these four sets of variables.

With our data we are able to create seven measures of geographic and climatic factors. The data for measuring *rainfall* (measured in millimeters per year) and *temperature* (measured in accumulated degrees centigrade per year) are from the CAS data center but were initially collected and organized by the Meteorological Observation Bureau of China from more than 600 national climatic and meteorological data centers. For use in our study, we take the point data from the 15 climate stations in Jiangxi and interpolate them into surface data using an approach called the thin plate smoothing spline method [32]. The *elevation* and *terrain slope* variables, which measure the nature of the terrain of each county, are generated from China's digital elevation model data set that are part of the basic CAS database. Information on the properties of soil also is part of our set of geographic and climatic variables from the CAS data center. Originally collected by a special nationwide research and documentation project (the *Second Round of China's National Soil Survey*) organized by the State Council and run by a consortium of universities, research institutes and soils extension centers, we use the data to specify three variables: the nitrogen content of the soil (*nitrogen*—measured in percent); *available phosphorous* in the top soil (measured in ppm); and *soil pH value*. Using a Kriging algorithm [33], we are able to interpolate the soil information to surface data to get more disaggregated information on the property of the soil over space for each pixel.

We use two demographic and economic variables, *population* and the level of gross domestic product per square kilometer (*GDP*). The demographic data for 1995 and 2000 are from the *Population Statistical Yearbook for China's Counties* [34,35]. Information on GDP for each county for 1995 and 2000 are from the *Socio-economic Statistical Yearbook for China's Counties* [36]. When there are missing data in the yearbook, the information is supplemented by each province's annual statistical yearbook for 1995 and 2000. In order to get pixel-specific measures of the demographic variables we use an approach called the Surface Modeling of Population Distribution framework to interpolate the data across space (measured as persons/square kilometer) [37,38]. The level of GDP (GDP per square kilometer) is also interpolated across space using commonly available GIS algorithms [38–40].

We also created two measures of distance (measured in kilometers), defined separately for each pixel in our sample. *Distance to the provincial capital* is measured as the distance (by the shortest road route) from each pixel to Nanchang, Jiangxi's provincial capital. *Distance to the nearest urban core* is generated by measuring the distance by the shortest road route from each pixel to the nearest county seat or other major urban center.

Finally, we also obtained data for one other factor: a dummy variable that is equal to one if the pixel is in a county that contains a protected area and zero otherwise (*protected_area*). This information is from a survey that we conducted of the Environmental Protection Bureaus of all of the counties in Jiangxi. Descriptive statistics for the control variables are included in Table 1.

Before discussing our formal empirical analysis, we provide a brief preview. Comparing pixels with no roads and pixels with (different types of) roads, changes in the level of forest cover appear negatively associated with the availability of roads. Specifically, between 1995 and 2000, the forested area in pixels with no roads rose by 1.02% points from 32.53% to 33.55%.[7] In contrast, the forested area in pixels with provincial-level highways (other roads) fell by 0.03% points from 52.43% to 52.40% (by 0.19% points from 49.58% to 49.41%).[8]

Interestingly, however, not all roads are alike. Forested area in pixels with expressways rises slightly. Given these descriptive statistics, it appears possible that changes of the forest cover of Jiangxi are being affected by roads.

3. Approach to estimate the effect of access to roads on forests

The basic relationship that we are interested in is:

$$\textit{Forest Area}_{it} = a_0 + a_1 \times \textit{Access to Roads}_{it-j} + e_{it}, \tag{1}$$

where $\textit{Forest Area}_{it}$ is the area of the forest in pixel i in year t; $\textit{Access to Roads}_{it-j}$ is a measure of the nature of the largest road that ran through the watershed which contains pixel i in year $t-j$; and a_1 is our coefficient of interest. We use a lagged

(*footnote continued*)
our explanatory variable. It is for this reason, in part, that we also redo the analysis with road pixels redefined as a "road pixel" only if the road actually passes through the pixel. When doing this analysis, there is no measured effect of roads on the forest.

[7] In addition, it is important to note that this is equivalent to 1.02 ha/1 km^2 pixel.

[8] It is important to note that although it seems like there is not much change in forested area between 1995 and 2000, we should emphasize that these are means and that there is actually high variability within Jiangxi Province as illustrated in Fig. 1 and in the discussion earlier in this section.

Table 1
Descriptive statistics of the variables at pixel level.

Variable	Units	Obs.	Mean	Std. dev.
Dependent variables				
Forest area in 2000	Hectare	6666	56.2	35.4
Change in forest area between 1995 and 2000	Hectare	6666	−0.1	5.5
Geographic and climatic factors				
Elevation	m	6666	251.7	230.0
Terrain slope	deg	6666	2.7	3.0
Nitrogen	%	6666	0.2	0.0
Available phosphorous	ppm	6666	0.8	2.7
Soil pH value	–	6666	4.7	0.8
Temperature	°C	6666	16.9	1.3
Rainfall	mm	6666	1625.2	110.0
Demographic and economic factors				
Population	Persons/km^2	6666	241.1	259.7
GDP	10,000 yuan/km^2	6666	46.1	187.0
Measures of distance				
Distance to the nearest road	km	6666	6.4	5.5
Distance to the provincial capital	km	6666	134.0	70.7
Distance to the near urban core	km	6666	110.2	74.9
Other factors				
Bufferforest10 (forest area in a 10 km^2 buffer)	km^2	6666	59.3	24.0
Protected_area (=1 if protected area, 0 otherwise)	–	6666	0.2	0.4
Road density	m/km^2	6666	53.9	48.6

measure of access to roads to help reduce some of the potential endogeneity bias, since changes in forest cover between 1995 and 2000 should have no direct effect on roads in 1995.

Since we are interested in the impact of whether there is a road in the watershed (or not) as well as the type of road (expressway vs. province-level highway vs. other road), we define $Access\ to\ Roads_{it-j}$ in four different ways. In model 1.1 we will include in our sample only the expressway and province-level highway pixels and $Access\ to\ Roads_{1.1it-j}$ will equal 1 if the pixel is an expressway pixel and will equal 0 if the pixel is a province-level highway. This is called Treatment 1 in the rest of the manuscript. Note, the other road pixels and no road pixels are excluded from the analysis when we use $Access\ to\ Roads_{1.1it-j}$. In the estimation of model 1.1 $a_{1.1}$ will measure the effect on forest area of changing a highway system from a province-level highway to an expressway.

In model 1.2, we will include in our sample only the expressway, province-level and other road pixels and $Access\ to\ Roads_{1.2it-j}$ will equal 1 if the pixel is either an expressway pixel or a province-level highway pixel and will equal 0 if the pixel is an "other road pixel." In the estimation of model 1.2 $a_{1.2}$ will measure the effect on forest area of changing a highway system from some other road to either an expressway or province-level highway. This is called Treatment 2. The roadless pixels are dropped from the analysis when we work with model 1.2.

In models 1.3 and 1.4 we use the full sample. The empirical exercise in model 1.3 will be like that of model 1.2, except we set $Access\ to\ Roads_{1.3it-j}=0$ when the pixels are either other road pixels or no road pixels. In this way, the interpretation of $a_{1.3}$ becomes the effect on forest area of changing a highway system from some other road to either a province-level highway to an expressway or of building a province-level highway or expressway to a previously roadless watershed (Treatment 3). In model 1.4, we set $Access\ to\ Road_{1.4it-j}=1$ if there is any type of road in the watershed and set it to 0 if there is no road in the watershed. The interpretation of $a_{1.4}$ becomes the effect on forest area of building any type of road to a previously roadless watershed (Treatment 4). Table 2 summarizes the different treatments that we will be conducting by estimating models 1.1–1.4.

Eq. (1) is problematic for several reasons. Pixels in watersheds with expressways are likely to differ from those in watersheds without any roads (or with only minor roads) in many ways. They may have easier topography and more productive soils along with unobserved locational advantages, since richer areas (or areas with more development potential) are more likely to attract investment in roads. Hence, applying OLS to Eq. (1) is likely to give biased and inconsistent estimates. Indeed, as discussed above, previous work suggests many other factors that might affect forest area and since some are likely to be correlated with both forest area and access to roads, we can reduce omitted variable bias by controlling for as many variables are possible [2,8,9,30]. This gives the model:

$$Forest\ Area_{it} = a_0 + a_1 \times Access\ to\ Roads_{it-j} + a_2 \times Z_i + e_{it}, \tag{2}$$

where in addition to the variables and parameters in Eq. (1), the specification of Z in Eq. (2), includes seven measures of geographic and climatic variables (*rainfall*, *temperature*, *elevation*, *terrain slope*, *nitrogen*, *available phosphorous* and *soil pH value*); two measures of demographic and economic variables (*population*, *GDP*); two measures of distance variables (*distance to the provincial capital*, *distance to the nearest urban core*) and one other variable (*protected_area,*). Since 9 of the

Table 2
Definition of treatment and control variables for four alternative treatments.

Alternative treatment	Treated—the largest type of road that goes through the watershed is:				Control—the largest type of road that goes through the watershed is:			
	Expressways	Province-level highways	Other roads	No roads	Expressways	Province-level highways	Other roads	No roads
Expressways vs. province-level highways (Treatment 1)	X					X		
Expressway and/or province-level highways vs. other roads (Treatment 2)	X	X					X	
Expressway and/or province-level highways vs. other roads or no roads (Treatment 3)	X	X					X	X
Expressway and/or province-level highways and/or other roads vs. no roads (Treatment 4)	X	X	X					X

12 variables in Z—all except *population, GDP* and *distance to the nearest urban core*—vary only across space, we include only an i subscript on Z.[9]

While adding covariates to an OLS regression (as in Eq. (2)) allows differences in the average values of observed characteristics to be controlled for, many studies show that this is a relatively inflexible and unsuccessful way to deal with the sample selection problem that occurs when observations in non-experimental studies cannot be randomly assigned to "treatment" and "control" groups. On the other hand, matching is an increasingly popular non-experimental evaluation method, with proponents claiming that it can replicate experimental benchmarks when appropriately used [41]. In particular, matching offers a way of structuring non-experimental data to look like experimental data, where for every subject in the "treated" group, the researcher finds comparable subjects in the "control" group.

In other words, while adding Z may help control for some of the factors that might create an omitted variables bias problem when estimating a_1, using ordinary least squares (OLS) assumes that simply conditioning linearly on Z variables suffices to eliminate selection bias. While the linear model can approximate a given non-linear function of the Z arbitrarily well when sufficient higher order terms are included, most of the linear regression models in the deforestation literature do not include higher order terms. Hence, for such models, using the standard OLS estimation approach would also be biased.

4. Matching methodology

The matching method is an alternative approach that can be used to examine the impact of a treatment (in our context, existence of particular types of roads) on an outcome (in our case, forested area) when selection takes place on observable characteristics [42]. As with the OLS model, measuring the effect of roads on forest cover without bias using the matching method assumes that the outcome in the base state (forested area if the pixel was not in a watershed with a particular type of road) is independent of the treatment (in watershed without such roads), conditional on observed covariates Z. In other words, for pixels within subgroups defined by Z, being located in a watershed with roads is unrelated to what the forest cover would be if the pixel were not in a watershed with roads. This is the so-called conditional independence assumption. If this assumption holds, we can say that given the observable covariates, the forest cover of the control pixels are what the forest cover of the treated pixels would have been had they not had the particular types of roads.

Unlike OLS, however, matching works by finding a control pixel that is very similar to the treatment pixel by conditioning on Z variables nonparametrically rather than linearly [43]. Moreover, with matching methods, but not OLS, we can impose "common support," which excludes treated pixels for which we cannot find reliably similar control pixels.

To take advantage of these factors, we follow the recent literature and match every treated pixel with a control pixel using covariate matching and its variants. With covariate matching [44], we estimate the average treatment effect by comparing outcomes between treated observations—pixels in a watershed with a specific type of road—and control observations—pixels in a watershed without the specific type of road.

Covariate matching, the method created by Abadie and Imbens [44], matches directly on covariates. In our analysis, we choose to match the two nearest neighbors with the same (similar) covariates (Z), where the variables in Z are the same as in the OLS model. Within these pixels, we can then directly estimate $E(Y_{i1}|T_i=1, Z_i)$ and $E(Y_{i0}|T_i=1, Z_i)$. This approach means that once we have a matched sample, we compare the forested area of the treated pixel with the forested area of the

[9] To avoid over-controlling, we do not include in the Z_i matrix of Eq. (2) three explanatory variables that are independent controls for roads—*distance to the nearest road*$_i$, *road density*$_i$ or *bufferforest10*$_i$. These variables are introduced, however, in the robustness analysis. These variables are measured as follows. The variable *distance to nearest road* measures the distance from each grid cell to the nearest road of any type. We generated the variable *road density* within the watershed by measuring "the length of all roads per square kilometer (m/km^2)". We also created a variable *bufferforest10* to identify if a pixel is surrounded by forested area or not. At the very least, this measure helps us gage how far to the "edge of the forest" is any given pixel—which also helps control for the distance that a pixel is from a road.

control pixel. We also report the estimated coefficients that use the post-matching bias correction factor also developed by Abadie and Imbens [44]. This correction factor is needed to correct for the conditional bias in finite samples when there are three or more continuous variables. Based on recent work that demonstrates that bootstrapping standard errors are invalid with non-smooth nearest neighbor estimators [48], we use Abadie and Imbens's variance formula for nearest-neighbor estimators. With covariate matching, we report the results using two weighting matrices. One approach uses the inverse variance weighting scheme; the other uses the Mahalanobis metric weighting scheme.

5. Two definitions of the dependent variable: forest area and change in forest area

In addition to using matching, we also have a data set that can help us control for part of the unobservables that might be leading to biased estimates of our parameters of interest. Specifically, we have two dependent variables of interest, in contrast with many studies on roads and deforestation that just use cross-sectional data on forest area. Such a cross-section may merely illustrate the correlation between where roads and forests are (i.e., which areas are more remote and which areas are more developed). There may be concerns about giving a causal interpretation to the estimated coefficients from such a model, which will be biased if the observable covariates are correlated with unobservable fixed effects, γ_i:

$$Forest\ Area_{it} = a_0 + a_1 \times Access\ to\ Roads_{it-j} + a_2 \times Z_i + \gamma_i + e_{it}, \tag{3}$$

One way to deal with these time-invariant unobservable variables is to estimate a first differenced model, which is partially possible since we have two years of land use data for each pixel in Jiangxi Province. In other words we can difference out some of this bias by estimating:

$$Forest\ Area_{i2000} - Forest\ Area_{i1995} = a_0 + a_1 \times Access\ to\ Roads_{it-j} + a_2 \times Z_i + e_{it}, \tag{4}$$

which compares the change in forested areas for pixels with and without access to roads, and so is a difference-in-differences estimator.[10] Therefore, in the rest of our analysis, we report estimates for all of the models using two dependent variables—forest area in 2000 and the change in forest area between 1995 and 2000.

6. Spatial scale issues

The basic unit of observation in our study is the 1 km^2 pixel, of which there are 166,932 in Jiangxi. There is a high correlation in forest cover between neighboring pixels and a lesser but still statistically significant correlation in the residuals of the OLS estimates of Eq. (2).[11] At the very least, this spatial autocorrelation can lead to inefficiency and invalid hypothesis testing procedures but it may also cause biased and inconsistent parameter estimates if spatial interactions are present such that a spatially lagged dependent variable belongs in the model [45].

We take three approaches to dealing with this spatial autocorrelation problem. First, rather than using all pixels we take a 1-in-25 sample by choosing only the pixels at the vertices of a 5 km $\times$ 5 km grid. This approach, which is widely used in econometric studies of deforestation, greatly reduces the spatial autocorrelation [4,7].[12] Second, since we use matching methods this eliminates even more of the spatial autocorrelation because every treated pixel is matched with a control pixel from a *different* watershed. Except for the extreme case where the two matched pixels share a common watershed boundary, the pixels are unlikely to be adjacent neighbors.

Our third strategy is to also estimate the model at the watershed level (in addition to estimating the model at the pixel level). This is somewhat akin to sampling in that it involves a loss of information, this time due to aggregation. However, since the effective sample size in the presence of spatial autocorrelation is much lower than the number of pixels (given that neighboring pixels have highly similar information) this may be appropriate. We note that aggregating (or smoothing) pixel data results in the loss of intra-watershed spatial variability. The greater the variability in forest cover within the watershed (i.e., the smaller the spatial scale at which the process operates), the less accurate the aggregate as an estimate for the dependent variable. Moreover, the within watershed variability is not likely to be constant across units, resulting in heteroskedasticity [46]. In our data, in fact, we have high variability in forest cover and forest cover changes within each watershed, so we should be concerned about the bias due to this factor. But when we compare estimation results for the 1474 watershed and 654 watershed aggregations, we find it is unlikely that the findings of the paper are affected by the aggregation level used to define watersheds.

[10] In principle the right-hand side variables could also be first-differenced, so that the effect of change in road access on change in forest cover which was estimated. However this fully differenced model is not feasible in the current case since we have only a single cross-section for the right-hand side variables. This restriction may not matter too much since many of the right-hand side variables are likely to be time-invariant over the short time period studied. Moreover, just differencing the dependent variable reduces the impact of the unobservable fixed effects.

[11] The Moran *I* statistic is 0.73 for the dependent variable and 0.49 for the residuals. Intuitively, this statistic is equivalent to the slope coefficient of a linear regression of the weighted average value of forest cover (residuals) for the pixels surrounding the *i*th pixel on forest cover (residual) in pixel *i*.

[12] The Moran *I* statistic falls to 0.47 for the dependent variable and 0.13 for the residuals.

7. Summary of our estimation approach

Given the proceeding discussion, in order to estimate the effect on forest area of access to roads, we take the following approach. First, we estimate Eqs. (1) and (2) and using OLS. Next, we use a covariate matching approach, using two algorithms—"covariate matching using an inverse variance weighting scheme" and "covariate matching using a Mahalanobis weighting scheme." We use these estimators to analyze the effect of roads on the forest and do so holding constant (in the case of our OLS estimators) and matching on (in the case of our covariate matching estimators) a set of covariates that are described above.

In order to check the robustness of our results, we first report the models excluding four variables that may be correlated with our treatment variable and then add them one by one for robustness checks with one of the *Access to Road* variables (Treatment 4). Furthermore, we estimate all our equations using both level of forest area in 2000 ($Forest\ Area_{i2000}$) and the changes in forest area $Forest\ Area_{i2000}$–$Forest\ Area_{i1995}$=Δ $Forest\ Area_i$) both using pixels and watersheds as the units of observation.

8. Results

The simple linear regression using OLS (with no controls and when we divide Jiangxi into 1474 watersheds) produces results that are similar to those found in the descriptive analysis above (Table 3, row 1). Regardless of the definition of the roads variable (Treatments 1–4), the larger the road in the mid-1990s (or if there is any road vs. no road), the lower the forest area in 2000. The signs on the coefficient of the roads variable are negative and significant in seven of the eight columns. In other words, when we use any treatment variable and when using observations at either the pixel or watershed level, there is a negative and significant relationship between roads in one period (1995) and forests in the next period (2000). Examining the magnitude of the coefficients demonstrates that the presence (or size) of a road, when we do not control for other factors, is associated with land that has 4.951–13.780% less forested area.

Importantly, however, as soon as we add the 12 covariates defined above to the OLS model, the negative association between roads in the mid-1990s and the level of the forest area in 2000 disappears (Table 3, row 2).[13] Regardless of the treatment or the level of observation (watershed or pixel), point estimates of the relationship between roads and forest area are positive in seven out of eight models, and two are statistically significant. In only one case is the coefficient on the road variable negative and in this case the *t*-ratio is small (0.25—Table 3, row 2, column 8). The two positive (and significant) coefficients (2.830—column 1; and 2.141—column 5) also are quite small.[14] Therefore, the most accurate interpretation of the findings when we estimate Eq. (2) is that roads have no impact when it comes to influencing Jiangxi's forest area. In other words, it seems clear from our findings that the results are at least not negative. What is also clear from our findings is that the previous negative relationship between roads and forest area (Table 3, row 1) is an artifact of an overly restrictive specification that excludes relevant covariates.

The absence of an impact of roads in the mid-1990s on Jiangxi's forest area in 2000 is reinforced when we use the two covariate matching approaches to estimate treatment effects (Table 3, rows 3 and 4).[15] Regardless of the treatment variable and whether we use pixels or watersheds as our units of observation, there are no cases where we find a negative and significant impact of roads on forest area. In fact, in 13 of 16 different models the signs on the road variables are positive (although they are significant in only one case). Hence, we can most accurately classify the result as having no effect. In other words, in our analysis there is no evidence to suggest that roads are creating pressures on forest area. In this way, our results are more similar to those of Andersen et al. (2002), who find a positive relationship between roads and forests than findings of others who have estimated a negative relationship.

[13] Although for most of our results we present only the coefficient of interest (that is, the coefficient that measures the impact of Treatment 1 or 2 or 3 or 4 on the forest or forest change), the interested reader is directed to JEEM's online archive (http:/aere.org/journals). This archive includes several tables that present all of the coefficients.

[14] Because we might be concerned that there are different effects of roads on different types of forests, we have rerun the model with forest type "a" (closed-canopy forest) as the dependent variable instead of total forest area (a+b+c+d). The results (not shown for brevity) show that roads are actually associated with slightly slower deforestation of closed-canopy forests. Importantly, although the coefficient was negative and significant, it should be noted that the magnitude of the coefficient is small enough almost to be called zero or immaterial). Although for brevity we do not include the results, we also found no evidence of the impact of roads on other types of forests (forest types "b," "c" or "d") in Jiangxi Province.

[15] Matching on the propensity score is another common technique. In contrast with the covariate matching method, which uses metrics such as inverse distance weighting to measure the closeness between the treated and the control, the propensity score matching method requires that the researcher first estimate a discrete choice model using the covariates and then use the predicted scores to match the treated and the control. We choose to focus on the covariate method for several reasons. Although in the literature there is no absolute dominance between covariate matching and propensity score matching, Zhao [47] demonstrates that covariate matching is robust under a number of different settings. Moreover, Zhao also shows that covariate matching is preferable when sample size is small and when the correlation between covariates and the participation indicator is not high. In our case, the sample size is not large—especially when we match treated watersheds (N=559) with control watersheds (N=265). In addition, when we estimate a logit model to describe the treatment, pseudo-R^2 is not high, suggesting that the correlation between covariates and the treatment indicator is not high. Finally, we also favor covariate matching since bootstrapping the standard error in propensity score matching does not provide standard errors with correct coverage [48]. As a robustness check, we estimated the models using propensity score matching and the results were essentially the same as the covariate matching method.

Table 3
Results from ordinary least squares regression approach and covariate matching analyzing the effect of roads on forest cover level in 1474 watersheds in Jiangxi Province.

Dependent variable: forest cover in 2000								
	Expressways vs. province-level highways (Treatment 1)		**Expressway and/or province-level highways vs. other roads (Treatment 2)**		**Expressway and/or province-level highways vs. other roads or no roads (Treatment 3)**		**Expressway and/or province-level highways and/or other roads vs. no roads (Treatment 4)**	
	Pixel	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**
OLS, no control	−7.814 (5.38)***	−1.863 (0.98)	−8.543 (6.99)***	−4.951 (1.89)*	−13.780 (11.84)***	−6.851 (3.70)***	−12.156 (11.63)***	−8.455 (4.26)***
OLS, with covariates	2.830 (2.11)**	2.720 (1.31)	0.780 (0.60)	1.527 (1.00)	2.141 (1.76)*	1.462 (1.02)	1.519 (1.74)	−0.406 (0.25)
Covariate matching (inverse variance)	2.764 (1.79)*	3.324 (1.37)	0.689 (0.51)	1.649 (1.04)	0.384 (0.15)	0.869 (0.96)	0.319 (0.41)	0.188 (0.12)
Covariate matching (Mahalanobis)	1.924 (1.13)	1.835 (0.93)	−1.017 (0.97)	1.296 (0.72)	0.540 (0.38)	−0.157 (0.13)	0.128 (0.11)	−0.979 (0.09)
N treated	1240	153	2908	328	2908	328	5076	559
N available controls	1191	131	995	120	1606	479	1590	265

Notes: In this analysis we were unable to use all pixels due to the fact that the number of observations is too large for conventional statistical software to handle. Therefore, we created a data set that was composed of every 25th pixel chosen by regular grid sampling. Absolute value of *t*-statistics in parentheses for OLS and *z* statistics otherwise. Calipers restrict matches to units within 0.5 standard deviations of each covariate. Covariate matching includes all covariates.

*** Significant at 1% level,

** Significant at 5% level,

* Significant at 10% level.

In consideration of avoiding the problem of ''over-controlling,'' we specified the regression that was used in creating the results in Table 3 in a way so that the analysis did not include any other roads-like variables in the regression. In other words, in Eq. (2) we did not include pixel-specific measures of road access: *distance to the nearest road*, *road density* or the *bufferforest10*. It could be, however, that these measures of access to the roads can help control for some degree of heterogeneity, which will make the rest of our analysis sharper. To examine this, we perform a robustness analysis by adding these controls (one at a time) to our original specification (i.e., instead of having 12 control variables we have 13 control variables, one of which is an independent, pixel- or watershed-specific measure of access to roads). For the sake of brevity, we conduct the robustness analysis using Treatment 4 only. When doing so, both results from the OLS models (with the 13 covariates) and the matching model (matching on 13 variables) demonstrate that the measured effect of roads on the forest is still zero (Table 4, rows 2–5).

The positive relationship between roads and forest area is in a strict statistical sense also supported by the simple linear regression of the change in forest area (Δ *Forest Area*) on roads (Table 5, row 1). In all of the eight models (considering different treatments, levels of observation—pixels or watersheds), the signs on the road variable are positive. In two of the models, the coefficients are positive and significantly different than zero. Considering the magnitude of these two coefficients, the nature of the roads in the mid-1990s leads to an increase in forest area of between 0.34% and 0.57% points. Given the small magnitude of the change, while statistically a larger road might be associated with a positive change, it might be most accurate to say that the magnitude of the change is small—perhaps closer to having no impact than having a positive in a substantive sense.

After controlling for covariates and after implementing our two matching schemes (Table 5, rows 2–4), our interpretation of the findings using the change of forest area as the dependent variable continues to be consistent with the findings when using the level of forest area as the dependent variable. While the point estimates in 20 of 24 models are positive, they are significant only in three (of the 24 models). In the rest of the cases, whether the sign of the coefficient is positive or negative, statistically there is no discernible relationship between roads and change of forest area.

The estimation results in Table 5 are carried out with the base model (which does not have any other roads-like variable in the regression). The robustness analysis (which is carried out by adding three controls (one at a time) using Treatment 4 also can be interpreted as showing that there is no impact of roads on the forests in Jiangxi. These results are in Table 6.

Finally, before concluding that the findings from Jiangxi (that is, there is no impact of roads on forests) differ from that of many previous studies, we want to make sure that the results hold up to more conventional analysis. To show this, instead of using our variable of interest (whether or not and what type of roads penetrate the watershed or pixel), we want to examine the coefficients on more traditional measures of road access (e.g., measures based on straight line distance to a pixel or watershed, etc.) and using more standard regression approaches (i.e., OLS instead of matching methods). In this sensitivity exercise, we use two types of data sets (pixels, watersheds) and four different measures of roads (watershed is roaded; distance to roads; road density; and bufferforest10). The results of the sensitivity analysis (not shown for brevity) using the more traditional measures of roads also show that there is no impact of roads on change in forest in Jiangxi. Therefore, the substance of the results using our new approach is also found using more traditional methodological approaches.[16]

9. Summary and conclusions

In this paper we have sought to estimate the impact of roads on forest area and change in forest area in Jiangxi, one of the most forested provinces in China. Using satellite remote sensing data to track changes over time at the 1 km^2 pixel level, we have found that Jiangxi experienced a fall in forest cover in the 1990s—although only from 56.3% to 56.2%. Although there was not much change on average, pixel-level maps showing changes in forested area across the province reveal that there is a lot of heterogeneity. In some parts of the province, forest area rose between 1995 and 2000; in other parts it fell. In the rest of the paper we analyzed the determinants of these changes, focusing mostly on the role of roads.

To estimate the impact of the presence (or the size) of roads on the forest, we developed an empirical framework in which we assigned pixels—the level of observations on which we can observe forested area—labels indicating whether or not the 1×1 km^2 land area is easily accessible by a major road (and indicating the size of the road that entered the watershed in which the pixel resided). Holding constant a set of carefully defined geographic and climatic factors, demographic and economic variables, distance variables and other factors, we sought to measure the net impact of the nature of the road network in the mid-1990s on level of the forest in 2000 and change of the forest area between 1995 and 2000. Using both standard OLS with covariates and two covariate matching methods, we found no evidence that roads contributed to deforestation. In other words, according to the findings from three different empirical approaches, when roads

[16] We realize the finding of zero impact (as we have in this paper) can be suspect. One could get a negative result if the experiment were poorly designed. In this paper, the high quality nature of the data minimizes the possibility that our results are due to errors in our data. Our robustness checks (of which there are many) also support the fact that we are indeed measuring the impact of roads on forest accurately. However, it is possible that if there were a high degree of spatial variability combined with a small amount of variation in forest cover that the power of the analysis could be low. In this light, caution in interpreting the results should be exercised.

Table 4
Results of effect of roads on forest cover level from various approaches using alternative specification for checking robustness of results, 1474 watersheds (treatment variable: expressways and/or province-level highways and/or other roads vs. no roads—i.e., Treatment 4).

Dependent variable: forest cover in 2000								
	Base model: without *distance to road* **variable; without** *road density* **variable; without** *bufferforest10* **variable**		**With** *distance to road* **variable**		**With** *road density* **variable**		**With** *bufferforest10* **variable**	
	Pixel	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**
OLS, no control	−12.156 (11.63)***	−8.455 (4.26)***	−12.156 (11.63)***	−8.455 (4.26)***	−12.156 (11.63)***	−8.455 (4.26)***	−12.156 (11.63)***	−8.455 (4.26)***
OLS, with covariates	1.519 (1.74)	−0.406 (0.25)	0.702 (1.49)	3.385 (2.21)**	−0.119 (0.37)	0.815 (0.41)	0.437 (0.51)	−0.248 (0.23)
Covariate matching (inverse variance)	0.319 (0.41)	0.188 (0.12)	1.151 (1.53)	3.709 (3.14)***	−0.231 (0.93)	1.971 (0.97)	0.121 (0.06)	0.449 (0.45)
Covariate matching (Mahalanobis)	0.128 (0.11)	−0.979 (0.09)	0.110 (0.41)	2.611 (0.99)	0.424 (0.62)	0.307 (0.58)	0.056 (0.21)	0.324 (1.55)
N treated	5076	559	5076	559	5076	559	5076	559
N available controls	1590	265	1590	265	1590	265	1590	265

Notes: Treatment 4 stands for the variable "expressways and/or province-level highways and/or other roads vs. no roads". In this analysis we were unable to use all pixels due to the fact that the number of observations is too large for conventional statistical software to handle. Therefore, we created a data set that was composed of every 25th pixel chosen by regular grid sampling. Absolute value of *t*-statistics in parentheses for OLS and *z* statistics otherwise. Calipers restrict matches to units within 0.5 standard deviations of each covariate.

*** Significant at 1% level.

** Significant at 5% level.

Table 5
Results from ordinary least squares regression approach and covariate matching analyzing the effect of roads on changes in forest cover in 1474 watersheds in Jiangxi Province.

Dependent variable: forest cover in 2000–forest cover in 1995								
	Expressways vs. province-level highways (Treatment 1)		**Expressway and/or province-level highways vs. other roads (Treatment 2)**		**Expressway and/or province-level highways vs. other roads or no roads (Treatment 3)**		**Expressway and/or province-level highways and/or other roads vs. no roads (Treatment 4)**	
	Pixel	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**
OLS, no control	0.602 (2.34)***	0.581 (1.61)	0.137 (0.48)	0.142 (0.61)	0.256 (1.83)*	0.307 (1.01)	0.219 (1.26)	0.171 (0.89)
OLS, with covariates	0.640 (1.99)**	0.457 (1.37)	0.111 (0.53)	0.053 (0.15)	−0.328 (1.39)	0.141 (0.95)	−0.119 (0.69)	−0.392 (1.38)
Covariate matching (inverse variance)	0.174 (0.46)	0.629 (2.01)**	−0.232 (0.76)	0.071 (0.10)	0.272 (1.66)	0.179 (0.83)	0.160 (0.90)	0.324 (1.19)
Covariate matching (Mahalanobis)	0.277 (0.81)	0.615 (3.03)***	0.059 (0.21)	0.059 (0.36)	0.363 (1.48)	0.152 (0.69)	0.185 (0.35)	0.304 (1.05)
N treated	1240	153	2908	328	2908	328	5076	559
N available controls	1191	131	995	120	1606	479	1590	265

Notes: In this analysis we were unable to use all pixels due to the fact that the number of observations is too large for conventional statistical software to handle. Therefore, we created a data set that was composed of every 25th pixel chosen by regular grid sampling. Absolute value of *t*-statistics in parentheses for OLS and *z* statistics otherwise. Calipers restrict matches to units within 0.5 standard deviations of each covariate. Covariate matching includes all covariates.

*** Significant at 1% level.

** Significant at 5% level.

* Significant at 10% level.

Table 6
Results of effect of roads on changes in forest cover from various approaches using alternative specification for checking robustness of results, 1474 watersheds (treatment variable: expressways and/or province-level highways and/or other roads vs. no roads—i.e., Treatment 4).

Dependent variable: forest cover in 2000–forest cover in 1995								
	Base model: without *distance to road* **variable; without** *road density* **variable; without** *bufferforest10* **variable**		**With** *distance to road* **variable**		**With** *road density* **variable**		**With** *bufferforest10* **variable**	
	Pixel	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**	**Pixel**	**Watershed**
OLS, no control	0.219 (1.26)	0.171 (0.89)	0.219 (1.26)	0.171 (0.89)	0.219 (1.26)	0.171 (0.89)	0.219 (1.26)	0.171 (0.89)
OLS, with covariates	−0.119 (0.69)	−0.392 (1.38)	0.082 (0.41)	−0.071 (0.36)	−0.820 (0.54)	−0.214 (1.03)	0.130 (0.89)	−0.102 (0.66)
Covariate matching (inverse variance)	0.160 (0.90)	0.324 (1.19)	−0.168 (0.46)	0.129 (0.35)	−0.163 (0.74)	−0.450 (0.29)	0.475 (0.51)	0.467 (1.57)
Covariate matching (Mahalanobis)	0.185 (0.35)	0.304 (1.05)	0.037 (0.22)	−0.069 (0.55)	0.205 (1.28)	−0.225 (1.52)	0.195 (0.76)	0.253 (1.71)
N treated	5076	559	5076	559	5076	559	5076	559
N available controls	1590	265	1590	265	1590	265	1590	265

Notes: Treatment 4 stands for the variable "expressways and/or province-level highways and/or other roads vs. no roads". In this analysis we were unable to use all pixels due to the fact that the number of observations is too large for conventional statistical software to handle. Therefore, we created a data set that was composed of every 25th pixel chosen by regular grid sampling. Absolute value of *t*-statistics in parentheses for OLS and *z* statistics otherwise. Calipers restrict matches to units within 0.5 standard deviations of each covariate.
***significant at 1% level, **significant at 5% level, *significant at 10% level.

were either widened or improved or when roads penetrated into a watershed, they did not appear to be adding any pressure to the forests. If anything, there is weak evidence that roads relieved a bit of pressure off the forests.

While our results run against the hypothesis that roads are the trigger of deforestation, they are consistent with some areas outside China. One way to reconcile these findings is that a negative relationship is found when roads penetrate into an area that has not been settled for very long or in which population densities are low. In this case, roads are a tool to haul away timber and non-timber forest products and reduce the cost of those wishing to use the land resources (e.g., for farming or livestock). These activities could account for the negative relationship between roads and forest area.

However, when an area is long settled, and when population densities are fairly high, it is possible that when roads enter an area (or are improved), they can act as a way to reduce the cost of moving out of the region or reduce the cost of technologies that will encourage more intensive cultivation/use of land resources in non-forested areas. If this were the case, roads could lead to less pressure on the forest, eventually rising forest area.

Clearly if these two interpretations are true, it is possible that our findings are accurately portraying the situation in Jiangxi. Jiangxi Province has been settled for thousands of years. The population pressures in many regions of the province become quite high. With China's fast growth, there are increasing opportunities for out-migration. Hence, if roads in any way encourage out-migration, and if out-migration reduces pressure on forests [16], then our results may be reasonable. Of course, Jiangxi has valuable forest resources. Roads would still reduce the cost of exploiting the forests. Because we are looking only at forest area, we are not able to discern if the quality of the forest is rising or deteriorating. It could be that roads are both increasing certain pressures and reducing them at the same time. If so, then the absence of a measured effect is showing the net position of several forces. Future research might want to try to understand the impact on value of the forest and other eco-services provided by forests.

Because the scope of this paper is so ambitious, we must necessarily draw some limits. First, we recognize that we have grouped different types of forests into a single measure. As a result, we are unable to measure the transition between different types of forests (e.g., from dense to sparse). If roads affect this transition, but do not lead to deforestation, we will be ignoring these impacts.

In addition, we are also able to measure only the impact of the stock of roads in 1995 on the forest. We do not have data on which of the roads are new and which have been around for a period of time. Consequently, the available data do not allow us to separately identify the impacts of roads separately from the impacts of all of the history of human settlement and activity, which is likely correlated with where the roads have been placed over time. Both this historical human activity and roads may shape the current spatial pattern of the forest, making it hard to identify the impact of roads per se. However there is less reason to believe that this omitted history will affect the recent change in forest cover, so the similarity of our results showing neutral impacts of roads on both the level and change in forests reduces concern about possible biases in our estimated impact of roads. Future research could consider searching for instrumental variables that show why roads developed over time in some parts of Jiangxi but not others, although with the current scale of road building and improvement in China a more forward-looking research strategy may be more productive.

Acknowledgments

This research was supported by the National Scientific Foundation of China (70873118, 70821140353), the Chinese Academy of Sciences (kzcx2-yw-305-2, KZCX2-YW-326-1) and the Ministry of Science and Technology of China (2006BAC08B03, 2006BAC08B06; 2008BAK50B06, 2010CB950904). Emi Uchida acknowledges support from Chinese Academy of Sciences Research Fellowship for International Young Scholars and the Rhode Island Agricultural Extension Service (AES# 5160).

References

[1] FAO, State of the World's Forests, Rome, 2007.
[2] A. Pfaff, What drives deforestation in the Brazilian Amazon? Evidence from satellite and socioeconomic data, Journal of Environmental Economics and Management 3 (1998) 26–43.
[3] H.J. Geist, E.F. Lambin, What drives tropical deforestation? LUCC Project Report Series No. 4, International Project Office, University of Louvain, Louvain-la-Neuve, Belgium, 2001.
[4] G.C. Nelson, D. Hellerstein, Do roads cause deforestation? Using satellite images in econometric analysis of land use, American Journal of Agricultural Economics 79 (1997) 80–88.
[5] D. Lawrence, D.R. Foster, Recovery of nutrient cycling and ecosystem properties following shifting cultivation: regional and stand-level constraints, in: B.L. Turner II, J. Geoghegan, D. Foster (Eds.), Integrated Land-Change Science and Tropical Deforestation in the Southern Yucatán: Final Frontiers, Clarendon Press of Oxford University Press, Oxford, 2004, pp. 81–104.
[6] A. Pfaff, J.A. Robalino, R. Walker, E.J. Reis, S. Perz, C. Bohrer, S. Aldrich, E. Arima, M. Caldas, W. Laurance, K. Kirby, Road investments, spatial intensification and deforestation in the Brazilian Amazon, Journal of Regional Science 47 (2007) 109–123.
[7] M. Cropper, J. Puri, C. Griffiths, Predicting the location of deforestation: the role of roads and protected areas in North Thailand, Land Economics 77 (2001) 172–186.
[8] K.M. Chomitz, D.A. Gray, Land Roads, Use, and deforestation: a spatial model applied to Belize, World Bank Economic Review 10 (1996) 487–512.
[9] J. Robalino, A. Pfaff, G.A. Sanchez-Azofeifa, Estimating spatial interactions in deforestation decisions, in: A. Kontoleon, U. Pascual, T. Swanson (Eds.), Frontiers in Biodiversity Economics, Cambridge University Press, 2007, p. 38.
[10] D.S. Liu, L.R. Iverson, S. Brown, Rates and patterns of deforestation in the Philippines: applications of geographic information system analysis, Forest Ecology and Management 57 (1993) 1–16.
[11] A.K. Ludeke, R.C. Maggio, L.M. Reid, An analysis of anthropogenic deforestation using logistic regression and GIS, Journal of Environmental Management 31 (1990) 247–259.
[12] B. Mertens, E.F. Lambin, Spatial modelling of deforestation in Southern Cameroon: spatial disaggregation of diverse deforestation processes, Applied Geography 17 (1997) 143–168.
[13] S.A. Sader, A.T. Joyce, Deforestation rates and trends in Costa Rica, Biotropica 20 (1988) 11–19.
[14] L.E Andersen, C.W.J. Granger, E.J. Reis, D. Weinhold, S. Wunder, The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon, Cambridge University Press, Cambridge, 2002.
[15] K. Deininger, B. Minten, Determinants of forest cover and the economics of protection: an application to Mexico, American Journal of Agricultural Economics 84 (1997) 943–960.
[16] F. Qiao, S. Rozelle, Tenure of forest land and the development of forestry sector, Problems of Agricultural Economy **7** (1998) 23–29.
[17] J.S. Shortle, D.G. Abler, Agriculture and the environment, in: J.C. van den Bergh (Ed.), Handbook of Environmental and Resource Economics, Edward Elgar, Cheltenham, UK, 1999, pp. 159–176.

[18] World Bank, World Development Report, World Bank, Washington, DC, 1992.
[19] A.P. Zwane, Does poverty constrain deforestation? Econometric evidence from Peru, Journal of Development Economics 84 (2007) 330–349.
[20] B.L. Turner, R.H. Moss, D.L. Skole (Eds.), Relating Land Use and Global Land-Cover Change: A Proposal for an IGBP–HDP Core Project, Land-Use/Land-Cover Change, IGBP–HDP Working Group, 1993.
[21] L.O. Fresco, R. Leemans, Land Use and Cover Change (LUCC) Open Science Meeting Proceedings, in: B.L. Turner II, D. Skole, A.G.V. Zeijl-Rozema, V. Haarmann (Eds.), LUCC Report Series No. 1, Barcelona International Project Office, LUCC, Institut Carogràfic de Catalunya, 1996.
[22] E.F. Lambin, X. Baulies, N. Bockstael, G. Fischer, Land-use and land-cover change (LUCC): implementation strategy, IGBP Report No. 48 and IHDP Report No. 10, IGBP and IHDP, Stockholm/Bonn, 1999.
[23] K. Kok, The role of population in understanding Honduran land use patterns, Journal of Environmental Management 72 (2004) 73–89.
[24] X. Deng, J. Huang, S. Rozelle, E. Uchida, Cultivated land conversion and potential agricultural productivity in China, Land Use Policy 23 (2006) 372–384.
[25] X. Li, A.G.O. Yeh, Accuracy improvement of land use change detection using principal components analysis: a case study in the Pearl River Delta, Journal of Remote Sensing 1 (1997) 282–289.
[26] P. Shi, Method and Practice of Study on Land Use/Land Cover Changes, Science and Technology Press, Beijing, China, 2000.
[27] Y. Sato, K. Yamamoto, Population concentration, urbanization, and demographic transition, Journal of Urban Economics 58 (2005) 45–61.
[28] X. Deng, S. Huang, J. Zhan, T. Zhao, A methodology framework for regional land-use changes studies based on Landsat TM images: a case study in Northeast China, Journal of Population, Resources and Environment 2 (2004) 3–11.
[29] D.R. Anderson, K.P. Burnham, W.L. Thompson, Null hypothesis testing: problems, prevalence, and an alternative, Journal of Wildlife Management 64 (2000) 912–923.
[30] M.L. Cropper, C.W. Griffiths, M. Mani, Roads, population pressures, and deforestation in Thailand, 1976–89, Policy Research Working Paper 1726, World Bank, Washington, DC, 1997.
[31] D. Kaimowitz, A. Angelsen, Economic Models of Tropical Deforestation: A Review, Indonesia, Center for International Forestry Research, 1998.
[32] A.D. Hartkamp, K. De Beurs, A. Stein, J.W. White, Interpolation techniques for climate variables, NRG-GIS Series (1999) 99–101.
[33] A. Kravchenko, D.G. Bullock, A comparative study of interpolation methods for mapping soil properties, Journal of Agronomy 91 (1999) 393–400.
[34] Ministry of Public Security of China, China Counties and Cities' Population Yearbook, Chinese Public Security University Press, Beijing, 1996.
[35] Ministry of Public Security of China, China Counties and Cities' Population Yearbook, Chinese Public Security University Press, Beijing, 2001.
[36] NBSC, China Social–Economic Statistical Yearbooks for China's Counties and Cities, China Statistics Press, Beijing, 2001.
[37] T.X. Yue, Z.M. Fan, J.Y. Liu, Changes of major terrestrial ecosystems in China since 1960, Global and Planetary Change 48 (2005) 287–302.
[38] X. Deng, H. Su, J. Zhan, Integration of multiple data sources to simulate the dynamics of land systems, Sensors 8 (2008) 620–634.
[39] C.N.H. Doll, J.P. Muller, C.D. Elvidge, Night-time imagery as a tool for global mapping of socio-economic parameters and greenhouse gas emissions, Ambio 29 (2000) 157–162.
[40] C.N.H. Doll, J.P. Muller, J.G. Morley, Mapping regional economic activity from night-time light satellite imagery, Ecological Economics 57 (2006) 75–92.
[41] R.H. Dehejia, S. Wahba, Propensity score-matching methods for nonexperimental causal studies, Review of Economics and Statistics 84 (2002) 151–161.
[42] P.R. Rosenbaum, D.B. Rubin, The central role of propensity score in observational studies for causal effects, Biometrika 70 (1983) 41–55.
[43] D.A. Black, J.A. Smith, How robust is the evidence on the effects of college quality? Evidence from matching, Journal of Econometrics 121 (2004) 99–124.
[44] A. Abadie, G.W. Imbens, Large sample properties of matching estimators for average treatment effects, Econometrica 74 (2006) 235–267.
[45] L. Anselin, Local indicators of spatial association, Geographical Analysis 27 (1995) 93–115.
[46] L. Anselin, Spatial effects in econometric practice in environmental and resource economics, American Journal of Agricultural Economics 83 (2001) 705–710.
[47] Z. Zhao, Sensitivity of propensity score methods to the specifications, Economics Letters 98 (2008) 309–319.
[48] A. Abadie, G.W. Imbens, On the failure of the bootstrap for matching estimators, Econometrica 76 (2008) 1537–1557.