A Tale of Two Roads: Groundwater Depletion in the North China Plain*

Ujjayant Chakravorty† Xiangzheng Deng ‡ Yazhen Gong§
Martino Pelli¶ Qian Zhang‖

December 2019

Abstract

Many studies focus on the relationship between infrastructure and economic development, but few study the effect of infrastructure on the sustainability of natural resources. This paper examines the effect of highway building on ground water tables in the North China Plain - which produces most of the country's food grains. We use a unique GIS-referenced dataset of all the 12,000 odd tube wells in one county to show that highway construction accelerates the rate of well drilling in farms near the highway. The highways lead to a faster depletion of groundwater in nearby wells relative to those located farther away. We show that these effects are caused by a switch from subsistence to commercial cropping, and intensification of farming practices closer to the highway.

JEL classification: O13, O18, Q25

Keywords: Groundwater Depletion, Infrastructure, Roads, North China Plains, Water and Development

---

*We would like to thank seminar participants at the University of Manitoba, Carleton University, University of Manchester, the Chinese Academy of Sciences, IISER Bhopal, PACE, SCSE, CAERE, CEA, FAERE, the ISI annual conference on economic growth and development, the Montreal Workshop in Environmental and Resource Economics, the AERE Summer Conference and the Canadian Study Group in Environmental and Resource Economics. We are grateful to Anita Chaudhry, Kyle Emerick and Yang Xie for valuable comments that greatly improved the paper. Noah Castonguay-Khounsombath and Enjun Ma provided outstanding GIS related research assistance. Pelli is grateful for the financial support received from the President's International Fellowship Initiative at the Chinese Academy of Science.

†Department of Economics, Tufts University, Medford, MA, USA. Email: ujjayant@gmail.com.

‡Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. Email: dengxz.ccap@igsnrr.ac.cn.

§School of Environment and Natural Resources, Renmin University of China, Beijing, China. Email: ygong.2010@ruc.edu.cn.

¶Department of Economics, Université de Sherbrooke, Sherbrooke, Quebec, Canada and CIREQ. Email: martino.pelli@usherbrooke.ca.

‖Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. Email: zhangq.ccap@igsnrr.ac.cn.

1 Introduction

There is a growing literature on the relationship between infrastructure provision and economic development (see for instance Faber, 2015; Gertler et al., 2015; Ghani et al., 2016, 2013; Khanna, 2014; Banerjee et al., 2012; Datta, 2012, to mention a few). However, there have not been many studies on the effect of roads on the depletion of natural resources. A few studies looked at the impact of road construction on forest cover, with mixed results (see Pfaff et al., 2018; Deng et al., 2011) In this paper, we use a unique dataset on all the 12,000 plus tube wells in a small county (Lankao) in Henan Province (see Figure 1) to show that there is clear evidence of an increase in the number of tube wells and of an increase in groundwater depletion close to new national highways.

The impact of the construction of a new highway on groundwater extraction works through a change in incentives. Taken to the extreme we equate this to an economy moving from autarky to free trade. All of a sudden you do not need to focus solely on production for your own consumption, but you can send your production to the market. Farmers located close to the new highway have now an easier access to nearby markets, either to sell their output or to buy cheaper input. People have also access to new jobs via a newly available commute to a nearby city. These effects are not new, for instance Donaldson (2015) studies the impact of India's vast railroad network on inter-regional trade. He finds that railroads decreased trade costs and inter-regional price differentials leading to higher real income levels. Proximity to the market could for instance push people to switch to the cultivation of *cash crops*, such as fresh vegetables, which are more profitable but consume more water.

The main goal of the paper is to investigate the existence of a causal relationship between highway construction and groundwater depletion. In order to achieve it, we first analyse the extensive margin, i.e. the impact of highways' construction on the number of new wells in their proximity. We then move on to the intensive margin and analyze the impact of the construction of highways on the depth of the water table in their proximity at one precise point in time, 2011.[1] Finally, we discuss the possible channels through which this relationship may work.

In order to be able to establish a causal relationship between water depletion and highway construction, we need the construction of the highway to be exogenous to economic activity in the county. In order to be able to make this claim, we only focus on newly constructed highways (G30 and G1511). Older highways (i.e. G220, G106 and G310) were born as

[1] The analysis is performed for 2011 because in that year the Chinese government surveyed the water table level in all the wells of Lankao county.

simple roads and have progressively been updated to highways. Historically, humans have settled in places where water was easily accessible and, therefore, the fact that these roads used to go through every single village along their path, increases the likelihood that these highways may be endogenous to the water table and to economic activity. G30 and G1511 were both built between the end of 90s and the beginning of the 00s. The segments of these two highways crossing Lankao are just short bits of these big projects planned in order to connect the est of the country to the west. As we detail later, these highways are plausibly exogenous to the economic activity in the county. This paper will focus mainly on this latter group of highways. In order to convince the reader of the plausible exogeneity of these highways we use also an instrumental variable specification, which confirms our results.

The extensive margin analysis uses a panel, where the units of observation are the cell composing a 250m x 250m grid covering the county. For each cell, we know the cumulative number of wells it contains each year. These data allow us to perform a classical difference-in-difference analysis on each of the two new highways. The treatment is defined as a buffer around each highway going from 500 meters up to 5 kilometres. The evolution of the difference-in-difference coefficient, in both cases, is positive and statistically significant, the number of new wells increases around both highways. When we focus on G30 we observe a stable positive change (up to 10%), while when we look at G1511 we observe a declining (from roughly 15%) coefficient. This sharp difference may be due to the fact that G1511 has been constructed very close to two of the pre-existing highways (G106 and G220).

For the intensive margin analysis, we regress the water table depth in each tube well on the distance from G30, G1511, and the interaction of the two distances. As shown in Figure 2, Lankao county is crossed by the Yellow River and, therefore, when analyzing the depth of the water table we have to take into account the possibility of seepage. In order to control for the effect of the Yellow River, we control for its distance from each tube well and the square of the distance, in order to capture possible non-linearities. Tube wells located closer to the river may also see reduced extraction of water if farmers supplement their groundwater use with surface water. Our results show that an increase of 1 km in the distance of a tube well from G30 leads to a decrease in the depth of the water table by 46.3 centimetres. While an increase in the distance of 1 km from G1511 leads to an increase of 62.4 centimetres. Because government programs focus on drilling tube wells in selected areas within the county at the same time, we control for the decade when the tube well was drilled. These decade fixed effects also control for big policy changes in China.[2]

[2]More details about these policy changes are provided in Section 5.

The argument of the Chinese government is that these highways have been built with specific goals in mind, which do not concern rural areas. Yet, in order to eradicate all possible doubts on the causal interpretation of our specifications, we also run an instrumental variable specification. We instrument the distance of each well to the actual placement of the highway with the distance to the straight line where the highway should have been had the government respected its plan. This instrumental variable specification confirms our results.

The highways could impact groundwater levels through three different channels. Farmers could decide to *i)* switch from traditional culture to cash crops like melons or fresh vegetables, which are more profitable, less durable and require more water; *ii)* intensify the traditional cultivations or *iii)* change input mix in response to a change in the labor supply. We present several pieces of evidence that suggest that all these possible channels are at play.

Section 2 provides some information on groundwater depletion in China. Section 3 contains a brief literature review on the study of roads in economics and some background information on highways in China and more specifically on those passing through Lankao County. Section 4 discusses the data used. Section 5 provides the empirical specification, results and robustness checks. Finally, section 6 discusses possible mechanisms that could explain the results and section 7 concludes the paper.

2 Groundwater use in thr North China Plain

The depletion of groundwater resources is a major problem in China. China is home to more than 20% of the world population but only five to seven percent of its freshwater resources (Qiu, 2010). Grain production is mainly concentrated in the North China Plains (NCP). This region accounts for one-fifth of China's total geographic area, covering the Tianjin municipality, the southern part of the Beijing municipality, a major part of Hebei, Shandong and Henan provinces and the northern part of Jiangsu and Anhui provinces. It includes roughly 340 counties. About 72% of the area is under farming but only 6% has access to surface water (Lu and Fan, 2013). This flat plain is China's bread basket and produces most of its grains, as well as cotton and other crops. It accounts for two-thirds of Chinese wheat production (Lu and Fan, 2013). Most farms practice double-cropping, rotating between summer maize and winter wheat and make extensive use of groundwater irrigation. This region has only a quarter of the nation's water resources yet produces half of its grain.

The NCP is an arid region where about 70% of the rain falls during June and September. Annual rainfall ranges between 400 and 600 mm/year while average evaporation is about

1000 mm/year (Feng et al., 2013). Thus food production in the NCP is largely dependent on groundwater resources, especially in the winter months. Overall about 70% of irrigated water comes from groundwater (Wang et al., 2006). There is evidence that groundwater aquifers in the NCP are depleting and may be under serious threat of overexploitation. Measurements from GRACE satellites suggest that between 2003-2010, the depletion rate was of the order of 8.3 km^3/year (Feng et al., 2013).

Lankao county has a surface area of 1,116 square kilometers and a population density of 744 per square kilometer, much higher than the average density in the country (145 per square km). The endowment of arable land per capita is only 0.09 ha. About 80% of the county's land area is under farming. Lankao is listed as one of the poorest of the 592 counties in China (WantChinaTimes.com, 2014; State Council, 2012). Five national highways and one provincial expressway go through the county, as shown in Figure 2. The Yellow River passes through the northwest corner of the county. Because of the river and of the irrigation canals that channel some of the river water to farms, Lankao has historically enjoyed a relatively good access to surface water for irrigation. Before the 1980s, almost all irrigation relied upon surface water. Since then, the share of groundwater in irrigation has increased at a rapid pace (Wang et al., 2006). During the last decade alone, the government has constructed more than 7,200 tube wells on the county's 74,635 ha of arable land in order to increase agricultural productivity.

3 Highways

The impact of transport infrastructure has been widely studied in economics. The majority of these studies focuses on the impact of roads on economic growth in developing countries and in particular, on their impact on firms and productivity. A few exceptions only look at high-income countries. Michaels (2008) and Chandra and Thompson (2000), for instance, focus on the US highway system. The latter finds that highways tend to draw economic activity to counties which they cross and have a differential impact across industries, while the former observes an increase in trade and in the real demand of skilled workers in skill abundant counties. Holl (2016), focuses on the Spanish highway system and, instrumenting its placement using postal routes and roman roads, finds that highways have a positive effect on firms' productivity.

Faber (2015) studies the China National Trunk Highway System, instrumenting placement with the ideal least cost path and finds a reduction in industrial output from nontargeted peripheral regions. Finally, Gertler et al. (2015) investigate the impact of road quality in Indonesia and find that higher quality leads to job creation in the manufacturing sector and an occupational shift from agriculture to manufacturing. Banerjee et al. (2012) study the impact of county location, especially its proximity to major communication arteries on economic growth in China. They find a small impact of the transportation network on per capita GDP levels across sectors but not on per capita GDP growth.

Several of the most recent studies focus on the updating of the Golden Quadrilateral Highway (GQH) in India. The GQH connects New Delhi, Mumbai, Chennai and Kolkata. Among these studies, we find Datta (2012), Ghani et al. (2013, 2016) and Khanna (2014). These papers find an unequivocally positive impact of the updating of the GQH on firms, which grew disproportionally along the network. These effects are found for districts up to 10 kilometers away from the highways and disappear for district located more than 10 kilometers away from the highway.

Highways in China and the National Trunk Highway Development Program

Over the last few decades, China has been investing heavily in the development of its highway network. Between 1990 and 2006, China invested roughly $40 billion per year in highways development and completed nearly 45,000 km of new highways. This new development of the Chinese highway system has taken place in two phases: *i)* the *kick off* phase, between 1988 and 1997 and, *ii)* the *rapid development* phase, starting in 1998. A big part of this push toward an improved highway network is the *National Trunk Highway Development Program* (NTHDP), consisting of 5 vertical (N-S) and 7 horizontal (E-W) routes with a total of 35,000 km. G30 is part of the NTHDP.

According to government documentation, the NTHDP has been planned "strategically" in order to inter-connect the country through high-speed road corridors connecting all the important centres of activity, rather than based on a detailed economic analysis, considering projected centres of economic growth, traffic growth and distribution. The planning approach for this part of the network has been as follows. First, the tree rooted at Beijing connecting all the provincial capitals. Second, all segments must connect: *a)* all provincial capitals, *b)* all cities with population above 500,000, *c)* all rail hubs, *d)* all ports, *e)* all major airports, and *f)* the old trading routes. The entire land acquisition, resettlement and rehabilitation process for these projects is usually completed within 5 to 6 months.

Highways in Lankao county

Lankao is a small and poor county, but it is crossed by five national highways, as shown in Figure 2. The two main new arteries going through Lankao, and focus of this paper, are the G30 and the G1511, see Figure 1. Both were constructed between 1998 and 2005. To distinguish them from older highways, we will call these two "new highways". As seen in Figure 1, G30 travels east-west for about 4,395 km, connecting Lianyungang in Jiangsu Province to Huoerguosi in Xinjiang Province. It splits slightly to the left of Lankao. The offshoot which goes towards Shandong is G1511. Only a stretch of 11.6 km of G30 goes through Lankao county (see Figure 2).

G1511 follows 41.4 km straight path entering Lankao county from the west and exiting it from the northeast corner. Construction of G30 took place during the period 1998-2001. G1511 is a connector and links G30 to the Rizhao-Nanyang national highway. Both are part of the 28 main national highways according to the *Tenth Five-Year Plan for National Highways* issued by the State Transportation Department. G1511 was constructed between 2003 and 2005.

The other three national highways, the *historic highways*, crossing Lankao county – G106, G220 and G310 – also shown in Figure 2, are not part of the National Trunk System but are upgrades of preexisting roads.[3]

4 Data

We use data on the geographical placement of roads and tube wells. Our database contains detailed information on the location, depth, date of construction and height of the water table for all the 12,160 tubewells in Lankao county constructed between 1955 and 2011. These tube wells are spread out over 389 villages in the 17 townships composing Lankao county.[4] They were dug either by private individuals or by the government. Figure 3 shows the evolution

[3]Highway G106 connects Beijing to Guangzhou City, running from north to south for 2,466 kilometers. G106 was constructed during two phases, a first one in 1956 and a second one in 1988. Highway G220, with a total length of 585 kilometers connects Binzhou City in Shandong Province to Zhenzhou in Henan Province. This highway was constructed in the late sixties and establishes the south-east access from Henan Province to Shandong Province. Finally, G310 runs for 1,613 kilometers from Lianyungang in Jiangsu Province to Tianshui in the Gansu Province. This segment was also constructed in two separate phases, a first one in 1949 and a second phase in 1978-79.

[4]Townships are essentially municipalities and the basic political unit in China. Each township has several villages in its jurisdiction.

in the number of tube wells by decade from 1955 to 2011.[5] The most significant increase in the number of tube wells took place early this century, when the government stepped in and started to dig tube wells. The increase in their number seems relatively stable before 2000. Figure 3 also shows how in the earlier periods wells were mainly dug around canals, and only moving forward they spread to the other parts of the county. Our analysis focuses on the period going from 1980 to 2011, i.e. since the introduction of the Households Responsibility Policy, which allowed individual households to decide what to cultivate, whether to dig a well and how much water to take from it.

In order to run an extensive and an intensive margin analysis, we need two different datasets. We first describe the one used for the extensive margin analysis and then move on to the one dedicated to the intensive margin.

The first step in order to run an extensive margin analysis consists in transforming our dataset in a panel. In order to construct the panel we have to first evenly divide the surface of the county in cells, say of 250 meters by 250 meters.[6] This exercise leaves us with a total of 18,362 cells. Our database contains information on the date at which each of the tube wells was dug. Thanks to these data, we are able to construct a panel, where the unit of observation are the cells of the grid. The variable of interest in this panel is a count variable representing the cumulative number of wells contained in each cell each year. Table 1 reports descriptive statistics for the panel. The number of cells containing zero wells decreases constantly and is equal to 10,691 in 2011, leaving 7,671 cells containing an average of 1.62 wells. Figure 4 and 5 show this evolution. Figure 4 looks at the average number of wells found in the 18,362 cells composing Lankao county. This number evolves from 0.1 in 1980 to roughly 0.7 in 2011. Instead Figure 5 focuses only on cells containing wells, in order to understand whether the increase in the previous picture comes from an increase in the number of cells containing wells or from an increase in the number of wells in the cells already containing wells. It seems that both statistics are increasing. The blue line shows

[5] Over the years, some of the wells go out of use for a variety of reasons. As long as wells are abandoned in a random fashion around the county (i.e. not following any particular pattern) this will not be an issue for our estimations.

[6] Several considerations went into the selection of the cell size. On the one hand, unfortunately, land utilization data for Lankao county is not available and, therefore, we are not able to distinguish between areas where it is feasible to dig a well and areas where it is not (because of pre-existing constructions or else). Taking this fact into account, we know that if the cell dimension is set too small we will plausibly experience a zero inflation in our data (i.e. we will have many cells reporting zero wells, but these would not be real zeros given that it is impossible to dig in these cells). On the other hand, we want to account for potential spatial spillover effects. The cell size was selected at 250m. Yet, since this number was selected ad hoc, we run two robustness tests on it. The results are robust to a change in the size of the cells to 300 by 300 meters and 500 by 500 meters.

that over the years more and more cells contained at least one well, while the red line shows the average number of wells found in these cells. This number is also increasing.

In the extensive margin analysis – for which the only information used about each well is the date of digging – we use all the available observations. The water table depth measurements are more delicate and, therefore, more prone to measurement error. For this reason we proceed to some additional data cleaning and end up running the analysis only on a sub-sample of the wells, the wells for which we have the highest confidence in the water table measurements. The approach used is very simple. First, we eliminate wells if, within a village we do not observe any variation in the water table.[7] Second, we eliminate all outliers. For instance, if in a village containing 30 wells the depth of the water table oscillated between 12 and 16 meters in 29 wells and it is of 1 meter in one well, we eliminate the latter observation. This process leaves us with 7,526 wells which are used for the intensive margin analysis and for which we report descriptive statistics.

Table 2 reports descriptive statistics for the main variables used in in the intensive margin analysis. The first striking fact is that the mean well depth (41.95 meters) is much larger than the mean depth of the water table (13.92 meters). This difference led us to use the latter in our estimation because the well-depth may be a function of other factors such as the cost of drilling, technology and expectations of future depletion. All the data on water table depth was collected in the same year (2011) over a short time span, while well-depth was measured at the time of digging which varies significantly across wells. Figure 6 shows the depth of the water table in the county as measured at the tubewell in 2011. Darker shades correspond to a deeper water table, while lighter shades indicate a shallower one. Note the general pattern imposed by the seepage from the Yellow River. The water table becomes deeper as we move towards the south-east corner of the county. In spite of this, one can still observe a remarkable degree of heterogeneity in the depth of the water table within the county.

The average tube well in the county is situated 22.04 kilometers away from the Yellow River (which flows in the north west corner). Wells on average are located only 9.72 kilometers away from G1511, which crosses almost the full length of the county and 18.77 kilometers away from G30, which is located in the south-west corner of the county. Well density is very high in the county and, therefore, wells are located close to each other, the mean distance between two nearest wells is 0.13 kilometres and the average number of wells in a circle of radius 500 m is about 14.

[7]Since we employ a fixed effect specification, these wells would be dropped anyway during the estimation.

5 Empirical Specification and Results

In order to fully understand the impact of road construction on groundwater depletion, we adopt two separate approaches. First – borrowing the language from trade – we implement an extensive margin analysis, which focuses on whether we observe a change in the number of new wells dug in areas closer to the new highways in the period following the beginning of their construction. Second we perform an intensive margin analysis, which focuses on the level of the water table in each of the wells existing in 2011. This second step compares what happens in wells closer to new highways with respect to wells located further away from the same roads.

5.1 Extensive margin analysis

Specification

This newly constructed panel allows us to run a standard difference-in-difference specification for each of the two highways of interest. In order to cover the first dimension of the difference-in-difference estimation, we construct a treatment dummy (called *Treat*). A cell is defined as treated if it lies within a certain distance from the new road (G30 in the first case and, G1511 in the second), and as non-treated if it lies outside that radius and at the same time also outside the same radius applied to the other highway. The second dimension instead, is covered by a second dummy (called *Post*) which takes value 0 for years previous to the first year of construction of G30 or G1511 and 1 for the following years.[8] The specification takes the following form

$$Wells_{ct} = \alpha + \delta_c + \delta_t + \beta_1 Treat_{ct} + \beta_2 Post_{ct} + \beta_3 Treat_{ct} * Post_{ct} + \varepsilon_{ct} \quad (1)$$

δ_c and δ_t are cell and year fixed effects. $Wells_{ct}$ represents the total number of wells in cell c at time t. Cell fixed effects account for all time invariant characteristics of a cell, such as distance to the river, to other roads, topology, hydrology and other possible confounding factors. As usual, the difference-in-difference effect of interest is given by β_3, the coefficient on the interaction term, which tells us by how much does the number of new wells in the treatment area increases (or decreases) after the construction of the new highway. In order to

[8] We also run a series of robustness by changing the definition of the *Post* variable, meaning that instead of giving it value 1 starting from the first year of construction of the highway, we started from the second year and so on until the year that concluded the works. Results are robust to these changes and available upon request.

satisfy the identification hypothesis of the difference-in-difference approach the specification is run using OLS.[9] Since the dependent variable is a count variable, we also run the main specification using a Poisson model, as a robustness test.

Extensive margin results

G30 Table 3 reports results for the extensive margin analysis of G30. In all specifications standard errors are two-way clustered at the year and cell level. In the case of G30, we defined as treated the area within 2 km of the highway, and as non treated the area between 2 km and 15 km from the two highways (G30 and G1511). Column (1) to (3) report OLS coefficients while column (4) shows coefficients for the Poisson specification. In columns (2) and (3) we add cell and year fixed effects. In this table the *Post* variable takes value 0 for years before 1998 and 1 after. The coefficient of interest is positive and statistically significant at the 5% level across all specifications, including the Poisson one. The coefficient is stable across the three OLS specifications. The number of new wells dug in one of the treated cells is 5.7% higher after the beginning of the construction of G30 with respect to a non-treated cell. We run a falsification test for the common trends hypothesis. Results can be found in column (1) of Table 4. In Table 4, we regress the dependent variable on a trend and the interaction of the trend with the treatment and the usual fixed effects over the period preceding the construction of the highway.[10] The coefficient on the interaction term is statistically insignificant, confirming the common trend hypothesis. Before the construction of the highway, there was no statistically significant difference in the number of new wells dug in the treatment area versus outside of it.

The size of the treated area may play a role in the results. In order to check the robustness of the coefficients obtained we let the size of the treatment area vary between 500 meters and 5 km at intervals of 500 meters.[11] Figure 7 shows the evolution of the difference-in-difference coefficient when expanding the treatment area from 500 meters to 5 kilometers. Figure 7 shows that the effect on the probability to dig new wells is positive in a relatively stable

[9] The use of a non linear linking function (such as poisson or negative binomial) would, for instance, violate the common trend hypothesis, since it does not allow to properly compute the differences-in-differences.

[10] The falsification test takes the following specification

$$Wells_{ct} = \alpha + \delta_c + t + \gamma_1 T_{ct} + \gamma_2 t * T_{ct} + \varepsilon_{ct}$$

where t is a linear time trend.

[11] We limited ourselves to a 5 kilometers band around the highway, since its area of influence is probably not much larger. Yet, we run a robustness up to 10 kilometers, and the effect tend to stabilize after 5 kilometers. These results are available upon request.

way.[12]

Another possible issue that could arise with this identification comes from the contiguity of the treated and non-treated areas. A higher number of wells in the treated cells could reduce the amount of water available in the aquifer and, therefore, decrease the incentive to dig new wells in non-treated cells. As a consequence of this type of externalities our results could be biased. In order to take these possible effects into account, we introduce a buffer area between treated and non-treated cells. Figure 9 shows the coefficient of interest (on the interaction term) when we introduce a buffer varying between 500 meters and 2 kilometres (at 500 meters intervals). As it is clearly shown in the graph, the introduction of the buffer does not affect our results.

G1511 Table 5 reports the difference-in-difference analysis for G1511. Even thought G1511 is a new highway, its location is very similar to the one of two of the old highways, G106 and G220. For this reason, the area around G1511 already contains a higher than average number of wells. In light of this fact we reduced the *ad hoc* treatment area for our baseline estimation to 500 m around G1511. In this case, the non-treated area is everything laying between 500 m of G1511 and 15 km from both highways. The results obtained for G1511 are stable across specifications and larger in magnitude than the ones for G30. After the construction of G1511 the number of new wells dug in the treated area increases up to 14.2%, statistically significant at the 5% level.[13] Falsification test results for G1511 can be found in column (2) of Table 4. The coefficient on the interaction term is statistically insignificant, thus confirming the common trend hypothesis.

Figure 8 shows the variation of the difference-in-difference coefficient for G1511 as we expand the treatment area from 500 m to 5 km.[14] Also in the case of G1511, the coefficient is consistently positive and statistically significant. Yet, we observe a downward trend which was not observed in the case of G30, meaning that the increase in the number of new wells decreases as we increase the size of the treatment area. This decrease may be due to the proximity of G106 and G220, which already increased the number of new wells over the previous years.

[12]The first point – when the treatment area is of only 500 meters – is not statistically significant, yet this may be due to the scarcity of observations so close to the highway. We have to remember that the stretch of G30 going through Lankao is very short.

[13]For both highways, results are robust to a change in the size of the cells to 300 by 300 and 500 by 500 meters.

[14]As for G30, we also run a robustness up to 10 kilometres, and the effect tend to stabilize after 5 kilometres. These results are available upon request.

As for G30, also in the case of G1511 we want to rule out possible externalities between treated and non-treated cells. As before we introduce a buffer varying from 500 meters to 2 kilometres at 500 m intervals. Also in the case of G1511 the coefficients are robust to this modification. Results are shown in Figure 10.

5.2 Intensive margin analysis

Specification

The intensive margin analysis is based on a cross section and, therefore, we need to be extra careful about possible confounder that cannot be accounted for using a fixed effect. Several factors may influence the depth of the water table. Some of these factors are natural, like the geology of the region, the shape and form of the underlying aquifer or proximity to a river, while others may be due to economic activity such as water use for agricultural or industrial use. Our goal is to check whether proximity to a road leads to higher water use and therefore, increased depletion of the groundwater table. As mentioned in the introduction, the underlying mechanism is that a road facilitates access to markets and may lead to more intensive farming practices (Donaldson, 2015). For example, in villages far away from roads, the cost of transporting inputs and outputs may be high, leading farmers to grow subsistence crops for local consumption or for the household. However, in areas closer to major roads, farmers may grow commercially viable crops that require a more timely use of inputs which, thanks to the road, are easier to access (such as fertilizers and pesticides). Farmers may also benefit from quicker and cheaper access to markets for their more perishable products, or have better information on market forces that affect their operations. These cheaper costs of inputs may lead to more intensive cultivation and hence increased use of complementary inputs such as water.

First, we need to control for seepage of water from the Yellow River which is likely to lead to a higher water table in wells located closer to the river. Figure 11 shows a local polynomial regression of the water table depth as a function of the minimum distance of the well to the river. Note that the water table gets deeper farther from the river, which is to be expected. The effect of seepage may be non-linear with respect to distance (Ghosh et al., 2014), for this reason we introduce it as a quadratic polynomial expression in our specification.

Our empirical specification takes the following form

$$WT_{iv} = \alpha + \delta_v + \delta_d + \beta_1 G1511_{iv} + \beta_2 G30_{iv} + \beta_3 (G30 * G1511)_{iv} + \beta_4 R_{iv} + \beta_5 R_{iv}^2 + \varepsilon_{iv} \quad (2)$$

where i and v denote tube well and village, respectively. WT is the depth of the water table in meters, $G30$, $G1511$ and R represent the distance of each well from G30, G1511 and the Yellow River in kilometres, respectively. Finally, δ_v represents village fixed effects and δ_d decade fixed effects. Finally, ε_{ivt} is the error term.

Since we are dealing with two separate highways at the same time, we have to take into account their interaction. The effect of roads on the water table may change if a well is located further from both roads then if it is located close to one road but very far from the other one. For this reason we introduced an interaction term, which accounts for the position of each well in relation to both roads. We expect to observe a higher depletion of the water table in a well located in close proximity to both roads, with respect to a well located close to one road but very far from the other one. In other words, we expect to obtain a negative sign on β_1 and β_2 but a positive one on β_3. Controlling for village fixed effects allows us to eliminate differences between villages in the form of topography, population density and other village-level characteristics like the quality of the village administration. Decade fixed effects are going to capture big policy changes.[15]

Intensive margin results

Table 6 reports the results for our baseline specification. Standard errors are robust and clustered at the village level. In column (1) we simply introduce the main variables of interest and village fixed effects. In column (2) and (3) we successively add the quadratic polynomial taking into account the distance from the Yellow river and the decade fixed effects, respectively.

The effect of G30 and G1511 on the water table is negative and statistically significant at least at the 10% level as soon as we control for the distance to the Yellow river. Focusing on the full specification, in column (3), we observe that a 1 km increase in the minimum distance from G1511 leads, *ceteris paribus*, to a decrease in the depth of the water table of 62.4 centimetres; while an increase of 1 km in the distance from G30 leads to a decrease of 46.3 centimetres. The coefficient on G1511 is statistically significant at the 10% level, while the one on G30 at the 5% level. The coefficient on the interaction term is positive and statistically significant at the 1% level. This positive coefficient implies that, if the distance

[15]The main policies that we want to capture are the following. In the 80s, the introduction of the Household Responsibility System, with a rental contract period set at not less than 15 years. This lead to a higher land fragmentation, which could make people drill more. In the 90s farmers were allowed to rent in/out land, and this policy change may have affected profit incentives. Finally, in the 00s, the contract period was extended to at least 30 years.

to one highway is kept constant while the distance to the other one is increased, this increase will decrease the total negative impact of highways. In order to fully capture the implications of our main specification, we plot the marginal effect of a change in the distance to either of the two roads. Figure 12 shows the marginal effect of the distance from the two highways on the level of the water table, panel (a) shows the impact of a change in the distance from G1511 on the effect of G30 and panel (b) the impact of a change in the distance from G30 on the effect of G1511. The figure also shows, in each panel, the share of wells at each distance from the road, represented by the shaded gray histogram. The most striking feature of the two pictures, which supports our claim that road construction has an impact on the water table, is the monotonically positive shape of both marginal effects. The impact of highways on the water table is negative and decreases as we move farther from them.

The coefficient on the distance from the Yellow River has the expected positive sign and it is statistically significant at the 5% level. The water table decreases away from the river. From column (3), we see that if the distance from the river increases by 1 km, the water table depth increases by 85 centimetres. The coefficient on the quadratic term is negative, implying that the effect of the river recedes with distance, also this coefficient is statistically significant at the 5% level.

Before moving on to the robustness tests, we run two additional variation of the baseline intensive margin estimations. First, we take one extra step in the generalization of the relationship between the water table depth and the distance of the well from the two highways. We do this by taking a second order Taylor approximation. Second, we dig deeper into the possible endogeneity of road placement and use an instrumental variable approach.

Second order Taylor approximation

The intensive margin specification presented above can be pushed one step further. Instead of simply capturing the intensive margin through the two distances and their interactions, we assume a second order approximation of a non-specified non-linear relationship between the distance from G30 and G1511. This implies a specification of the following form

$$\begin{aligned} WT_{iv} = \alpha + \delta_v + \delta_d \quad &+ \quad \beta_1 G1511_{iv} + \beta_2 G30_{iv} + \beta_3 G1511^2_{iv} + \beta_4 G30^2_{iv} \\ &+ \quad \beta_5 (G30 * G1511)_{iv} + \beta_6 R_{iv} + \beta_7 R^2_{iv} + \varepsilon_{iv} \end{aligned} \tag{3}$$

The only difference from equation (2) are the two squared terms of the distance. Results from this specification are shown in column (4) of Table 6. The marginal effets are now more difficult to compute, for instance, the marginal effect of a change in the distance to G30 for the water table is going to be given by the following expression

$$\frac{\partial WT_{iv}}{\partial G30_{iv}} = \beta_2 + 2\beta_4 G30_{iv} + \beta_5 G1511_{iv} \tag{4}$$

a linear function of both distances. Instead of presenting a 3D graph of the evolution of the marginal effect, using the mean distance from G30, 18.77 kilometres, and the mean distance from G1511, 9.72 kilometres, we compute the average marginal effect, which is equal to -0.112. This confirms our previous findings, as we move marginally further from G30 the depth of the water table decreases. The same exercise using the coefficients on G1511 and the same mean values for G30 and G1511, obtains an average marginal effect equal to -0.068. Therefore, also for G1511, when one moves marginally further the water table becomes shallower.

Instrumental variable specification

Chandra and Thompson (2000) and Michaels (2008) argue that highway placement is exogenous to the countryside in the US and, we could argue that this argument translates also to China. Moreover, Michaels (2008) claims that the shorter is the segment of highway to which we are interested the stronger is the plausibility of its exogeneity. Yet, in order to eradicate all possible doubts on the causal interpretation of our specifications, we run an instrumental variable specification. The argument of the Chinese government is that these highways have been built specifically to fulfil the goals specified in section 3, therefore, if we instrument the distance of each well to the actual placement of the highway with the distance to the straight line where the highway should have been had the government respected its plan, we are going to clear any possibility of endogeneity from our specification.

First, we identified, for each of the two highways, the cities which they were supposed to connect. The segment of G30 crossing Lankao is supposed to connect Kaifeng to Xuzhou, while the segment of G1511 is supposed to connect Kaifeng to Rizhao. Once these city-pairs have been identified, we connect them with straight lines.[16] Figure 13 shows the imaginary straight lines and the actual highways. We then proceed to measure the distance of each well to these imaginary straight lines. These distances are then used in order to instrument the

[16] Alder (2019) computes the optimal placement for G30 and G1511 in this region taking also into account land gradient. According to his computations they both are very close to straight lines.

actual distances and the interaction between the distances to the straight lines to instrument the interaction between the two actual distances.

Table 7 reports the results of this specification. The table takes the same structure as the table reporting the baseline estimates for the intensive margin. Also here, standard errors are robust and clustered at the village level. The instrumental variable results are similar to the OLS results in magnitude and statistical significance and share the same sign. This similarity corroborates our claim of exogeneity in the placement of the two highways of interest.

5.3 Robustness

We run five different additional robustness test on the intensive margin results. First, we test for within village differences in the utilisation of the aquifer. A higher density of wells in one area could lead to a higher depletion rate, unrelated to the proximity to the highway. Second, we check whether the construction of the road affects the hydrology of the region. In order to perform this test, we drop from the analysis the wells that are very close to the new road. Third, we drop from the analysis villages with less than 30 wells, in order to insure a sufficient level of variation. Fourth, we introduce the old highways and, finally, we replace the minimum distance to the road with distance to the highway entry.

Robustness tests are aimed at the intensive margin specification and are presented in Tables 8 and 9. In order to facilitate comparison, each table first shows our baseline results. Standard errors are robust and clustered at the village level in all specifications.

Table 8 controls for well density in the proximity of each well. The presence of a higher number of wells may lower the water table because of the additional extraction of ground water by other wells. This exercise allows us to control for within village confounders. We use various measures of well density, starting with the number of wells within radii of 100 m, 200 m and 500 m, in columns (2), (3) and (4), followed by distance to the nearest well in column (5). In column (6), we jointly control for the number of wells within a radius of 500 m and the distance to the closest well. Our estimates are robust to these controls in terms of magnitude, sign and statistical significance. The coefficients in columns (2), (3) and (4) are positive suggesting that a higher density of wells in the buffers lowers the water table. Yet, only the result on the 500 m radius is statistically significant at the 10% level. That is, a lower density of wells leads to less depletion. As expected, a larger distance to the nearest well leads to a higher water level. Yet, also this coefficient is not statistically significant. When controlling simultaneously for the last two measures, only the coefficient on the number of wells within a 500 meters radius is statistically significant, because it captures also the information provided by the other measure of well density.

In Table 9 we check whether the highway itself may have altered the hydrology of the region and, therefore, the water table. The construction of the highway, by inserting a new structure underground, may have disturbed the water levels in the underlying aquifer. In or der to control for this possible effect, in Table 9, we run the baseline estimation excluding the tubewells which are closer to the road, which would likely be impacted by highway construction. After presenting the baseline estimation in column (1), in column (2) to (4) we eliminate all wells within 50 meters, 100 meters and 500 meters from the highway, respectively. It does not seem that the construction of the highways impacts the water table in their proximity. The coefficients are stable in terms of magnitude, sign and statistical significance. Finally, in columns (5) and (6), we eliminate from our sample all villages containing less than 30 tubewells and less than 40 tubewells. These last two tests significantly decrease the sample size, yet they do not affect the magnitude of the coefficients or their size, but only the precision of the estimates.

Finally, in Table 10 we introduce the distance to the old highways. As one may see from Figure (2) G220 follows basically the same path as G1511, so the two effects may be confounded. When we add both highways, as in column (2), we observe that the sum of the two coefficients has the same magnitude as the coefficient on G1511 obtained in the baseline (column 1). The coefficient on G30 is not affected in terms of sign and magnitude, yet the standard error becomes slightly larger and as a consequence we loose precision in the estimation. The coefficients on G106 and G310, as expected, are highly statistically insignificant. In column (3) we replace the distance from the highways with the distance to the highway exit. Official exits are not the only way to get on these highways, for this reason this is only a robustness test. The sign of the coefficients is maintained, yet we lose in terms of precision of the estimates.

5.4 Discussion

Our results on the extensive as well as on the intensive margin clearly point to an increase in groundwater depletion in proximity of highways in the period following their construction. This increased depletion could operate through three different channels. The first channel is related to crop switches, i.e. diversification. The new highway, providing improved and faster access to the market, creates incentives for farmers to switch their production from the more classical wheat and corn towards more profitable, yet more water intensive, cultures such as melons, fresh vegetables and peanuts. The second channel is culture intensification. This channel is related to the more traditional cultures (corn and wheat), the decrease in transport costs deriving from the new highway may push farmers to intensify their cultivation, for instance by providing more water. Intensification can also happen as a consequence of a decrease in the cost of production inputs. One last possible channel, could be related to migration. The new highway may create new work opportunities. For instance, It may now be easier for people to work farther away, for instance in a city close by, and get jobs outside the agricultural sector. As a consequence, since labor is declining in agriculture, the remaining farmers may increase the other inputs used in production, such as water, trying to adjust their production technology.

In order to provide some evidence on these possible channels through which highways may affect the water table we use three different datasets. The first one, contains data from a household survey. We conducted the survey in Lankao county in the summer of 2014. The survey focuses on 282 households located in 30 villages.[17] The questionnaire focused on household characteristics and on agricultural practices. For this reason, we also collected information on the 1,304 plots of land owned by the surveyed households. The average plot in our survey measures roughly 0.15 ha. The households interviewed operate a total of 602 wells throughout the county. The second dataset comes from the Chinese National Bureau of Statistics and contains information about the total surface sown in Lankao county and total output per year for the main grains and for fresh vegetables. Finally, we use satellite images from the United States Geological Survey (USGS), in order to compute the Normalized Difference Vegetation Index (NDVI) for the county and analyse changes in production patterns.

Crop switch

Let us start with the first channel. Do we observe a switch to *cash crops* in plots closer to the new highways? We analyze this question from three different angles. First using survey data, then with aggregate county-level data, and, finally, using satellite images. Using survey data, we focus on the locations where different crops are cultivated. The households surveyed grow 13 different types of crops: wheat; corn; cotton; potatoes; beans; apple, pear, peach and poplar trees; vegetables; melons; peanuts and garlic. As shown in table 11, the main cultivations in our survey are wheat and corn, accounting for 86% of the plots, only 14% of the plots are cultivated with cash crops, which are more profitable but more water intensive. Let us focus on three specific cash crops: peanuts, vegetables and melons, which alone make

[17] More information on the survey (such as randomization) can be found in the appendix A.1.

up 77.2% of the cash crops cultivated in our sample.

The first thing that we observe is that the average plot size for these crops is larger than the average plot size of wheat or corn fields, 0.187 ha on average versus 0.145 ha.[18] For the sake of our analysis we need to remember that peanuts are significantly different from vegetables and melons. Peanuts are small and light, but more importantly are less perishable than vegetables or melons and, therefore, probably do not respond to the same incentives. We would expect the cultivation of vegetables and melons to take more advantage of the proximity of a highway, with respect to the cultivation of peanuts. This is exactly what we observe in our data. The average distance to one of the two new highways for peanuts fields is 8.13 km, while the average distance for vegetables and melons fields is only 0.82 km and 1.13 km, respectively. The average distance for corn or wheat fields is 4,94 km and 5.33 km. These averages are all statistically different from one another. It seems that the location of the different cultivations throughout the county supports the first channel evoked earlier. We may be observing some diversification taking place, yet, given the cross-sectional nature of our survey data, we cannot determine whether these cultivations where already there prior to the construction of the highway or not.

The aggregate data from the Chinese National Bureau of Statistics tell us a similar story. Unfortunately, data on total output from of fresh vegetables is not available, yet we have data on total area dedicated to the cultivation of fresh vegetables. Figure (14) shows the evolution of this cultivation. We can observe a big expansion starting in the late 1990, right after the beginning of the works for the construction of G30, stabilizing at the end of its construction and beginning of the construction of G1511. By running a likelihood ratio test (Wald test) we quickly verify that this time series does experience a structural break in 2001.[19]

Finally, we use satellite data from the United States Geological Survey in order to reconstruct the NDVI for the summer period. We only have one observation per year for the summer period between 1998 and 2010.[20] The frequency of publicly available image acquisitions for the Lankao region is relatively low, and therefore, the data points do not represent the exact same date every year. The images for each year have been selected as a function of availability of high quality satellite images for the summer period. We compute NDVI values for each 150 x 150 meters cell in the county (a total of 75,828 cells). In Figure 15 we

[18]This difference is statistically significant at the 1% level.

[19]The χ^2 value of the test is 35.24.

[20]Sufficient quality satellite images for Lankao county are not available for the summers of 1999, 2006 and 2009.

can see NDVI for 1998 and 2010. We can notice two things, an increase in the number of settlements and in their size and an improved management of the Yellow River. Figure 16 shows a scatter plot of the NDVI for each cell in the county divided by year. The average summer NDVI seems to be relatively stable, and, as expected given the agricultural nature of the majority of activities in Lankao, the majority of the NDVI values are above 0.

The traditional summer cultivation in Lankao is corn. The NDVI for a mature field of corn is higher than the one for a field of vegetables (REFERENCE). This is due to the high density of leaves that characterize corn fields and to the fact that in a vegetable field, be it melons, or other fresh vegetables, a large proportion of the soil remains visible from above. In light of this, we would expect that if crop switching occurs around the new highways the NDVI values should decrease. In order to test for this hypothesis, we run a difference-in-difference specification similar to the one used for the extensive margin. Since the satellite data we have begin in 1998, we are only able to perform this exercise for G1511. We run the following specification:

$$NDVI_{ct} = \varphi + \eta_c + \eta_t + \varphi_1 Treat_{ct} + \varphi_2 Post_{ct} + \varphi_3 Treat_{ct} * Post_{ct} + u_{ct} \tag{5}$$

where η_c and η_t are cell and year fixed effects. $NDVI_{ct}$ represents NDVI in cell c at time t. As before, the difference-in-difference effect of interest is given by φ_3, the coefficient on the interaction term, which tells us by how much does the NDVI in the treatment area change after the construction of the new highway. As in the extensive margin analysis for G1511 we set the boundary for the treatment area at 500 meters around the highway.

Table 12 shows the results of this analysis. As before, standard errors are two-way clustered at the cell and year level. Columns (1) through (3) shows the results from a specification without fixed effects, with cell fixed effects and finally with also year fixed effects. φ_3 is very robust across specifications, after the construction of G1511, in the treated area, NDVI was 3.6% lower. This negative coefficient constitutes extra evidence of the switch towards cash crops happening after the construction of the new highway.[21] In column (4) we run a falsification test. We focus only on the year preceding the construction of G1511 and test whether we observe a difference in the NDVI trend between treated and non-treated cells. The interaction between the trend and the treatment is statistically insignificant, confirming the common trend hypothesis necessary for a difference-in-difference specification.

[21] To ensure that the decrease in NDVI in proximity to the new highway is not due to the highway itself, we also run the specification dropping all the observations within 150 meters of the highway (size of a cell) and by increasing the treated area to 650 meters. The coefficient becomes slightly smaller but stays negative and statistically significant at the 1% level (-0.22).

We cannot link any of the three pieces of evidence presented above to the highways construction with certainty. Yet, if we consider them all together: *i)* cash crops are found on average closer to the new highways, ii) we observe a structural break in the trend of are cultivated in fresh vegetables in 1998, and *iii)* we observe a decrease in NDVI close to the new highway, it seems that they all point in the same direction.

Intensification

Let us now move on to the second channel, do we observe an intensification in the cultivation of wheat and corn? We tackle this question from two sides. First, we look at whether output changes, do we observe an intensification of production – higher yields – in fields located closer to the two highways of interest? This first effect may derive from the fact that the privileged access to the market allows farmers to get a higher price for their harvest. Second, we look at inputs and more specifically their cost. The presence of a highway changes people's habits. Since people may now travel more, they may gain access to cheaper inputs, such as seeds or fertilizer.

In order to answer these questions we are going to *(i)* analyze how the yield and the price charged evolve as we move away from G30 and G1511, and *(ii)* look at how the cost of seeds, fertilizer, pesticide and herbicide evolve. We use the following specification

$$Y_{pi} = \alpha + \beta_1 G30_i + \beta_2 G1511_i + \beta_3 (G30 * G1511)_i + \beta_4 X_{pi} + \varepsilon_{pi} \tag{6}$$

where the dependent variable, Y, is yield, price, seeds cost, fertilizer cost, herbicide cost and pesticide cost. p denotes a plot served by a given well i. Since several plots may be served by the same well, we cluster standard errors at the well level. X contains plot level controls: plot quality (compared to the average plot) and plot slope. Here again, $G30$ and $G1511$ represent the distances from G30 and G1511, respectively. This time we measure the distances in kilometres in order to facilitate the reaading of the coefficients. ε_{pi} is the error term. The results of the estimation of equation (6) are reported in table 13 for yield and prices and in tables 14 and 15 for cost of inputs for wheat and maize, respectively.

Let us start with table 13. Columns (1) and (2) report results for wheat, while columns (3) and (4) are dedicated to corn. We first present the yield regression for each crop and then the price regression. Wheat yield and price seems to decline as we move away from the two highways, in a strongly statistically significant way for G1511, and in a non-statistically significant way for G30. Even thought the interaction term is positive and statistically significant, it is very small in magnitude, simply telling us is that the positive effect of the highway becomes less and less important the more the distance grows. The results for corn are mixed and non-statistically significant. Wheat seems to support our hypothesis of an intensification of cultures in proximity of new highways, that could lead to a heightened water consumption.

The higher wheat yields closer to the new highways could depend on several reasons. First, and in relation to our previous results, we could speculate that farmer located closer to highways now have increased incentives to obtain higher yields, because of lower transport costs. Having a better access to the market makes it worthwhile to spend a few extra yuan for pumping the water in order to obtain a higher yield. Another reason could also be that after the construction of the highways, farmer located close to it had an improved access to better seeds (or other inputs), yielding higher amounts of wheat.

While we cannot investigate seeds quality, we proceed to study the evolution of the price of inputs paid as we we move away from the highways. The results presented in table 14 and 15 are not very precise, with the exception of fertilizer costs (reported in column 2), yet they all seem to move in the same direction, showing an increase in the cost of inputs as we move away from the two new highways. The two tables are organized in the same fashion presenting from column (1) to (4) results for seeds, fertilizer, pesticide and herbicide costs. While the increase in the cost of fertilizer as we move away from either of the two highways is statistically significant at least at the 10% level, all the other coefficients, in spite of being positive, are not statistically significant.

Migration

The last of the three possible channels is related to migration or more precisely commuting. The presence of a major road simplifies the task of looking and getting a job outside of the countryside. After the construction of a new highway it becomes easier to get on a bus and commute to a nearby city in order to get a higher paying job. A decrease in the rural labor force may force farmers to select a different combination of inputs (labor, capital, water) in order to maintain production to the same level. In order to investigate this channel, in table 16, we look at water use and capital investment (in agricultural machineries) per mu as we move away from the two new highways. The table seems to show that also this channel is at play: as we move away from the two highways water use and capital investment are decreasing , mostly in a statistical significant way. This result may indicate a need to change the input mix employed in production in proximity to the highways, this need subsides as we move away.

The various pieces of evidence collected in this section seem to tell us that all the channels mentioned above are at play and their combination leads to an increased depletion of groundwater. Given the data limitations, we cannot identify the channels with certainty, yet we have shown, using different datasets, that in proximity of G30 and G1511 we observe diversification and intensification of cultures and a change in the input mix.

6 Concluding remarks

There is a large literature on infrastructure and economic development but no studies on the relationship between infrastructure building and resource depletion. It is likely that while infrastructure such as roads brings economic activity into a region, it also leads to the depletion of the natural resource base. This paper shows that there is clear evidence of depletion of the water table in Lankao County, China, close to the two national highways that pass through the county. The relationship is found through two different exercises. First, we perform an extensive margin analysis to investigate whether the construction of new highways increases the probability of digging new wells. Second, we analyze the intensive margin, in order to see how the water table is affected by the presence of new highways.

These results may be driven by the fact that a new road facilitates access to the market, in the same way railways do, see Donaldson (2015). An easier access to the market pushes individuals or communities to engage in agricultural activities that are more commercial in nature, such as the cultivation of cash rather than subsistence crops, or the use of modern varieties of seeds and multi-cropping, requiring more use of groundwater irrigation. Our findings suggest that the true benefits of road-building may need to account for these depletion effects, which are likely to impact the long-term economic productivity of the region.

References

Alder, S. (2019). Chinese Roads in India: The Effect of TRansport Infrastructure on Economic Development. Technical report.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. NBER Working Papers 17897, National Bureau of Economic Research, Inc.

Chandra, A. and Thompson, E. (2000). Does Public Infrastructure Affect Economic Activity?: Evidence from the Rural Interstate Highway System. *Regional Science and Urban Economics*, 30(4):457–490.

Datta, S. (2012). The impact of improved highways on Indian firms. *Journal of Development Economics*, 99(1):46–57.

Deng, X., Huang, J., Uchida, E., Rozelle, S., and Gibson, J. (2011). Pressure cookers or pressure valves: Do roads lead to deforestation in China? *Journal of Environmental Economics and Management*, 61(1):79 – 94.

Donaldson, D. (2015). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*.

Faber, B. (2015). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*.

Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H., and Xia, J. (2013). Evaluation of Groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) Data and Ground-Based Measurements. *Water Resources Research*, 49:2110–2118.

Gertler, P., Gonzalez-Navarro, M., Gracner, T., and Rothenberg, A. (2015). Road Quality, Local Economic Activity, and Welfare: Evidence from Indonesia's Highways. Technical report, mimeo.

Ghani, E., Goswami, A., and Kerr, W. (2013). The Golden Quadrilateral Highway Project and Urban/Rural Manufacturing in India. Policy Research Working Paper Series 6620, The World Bank.

Ghani, E., Goswami, A., and Kerr, W. (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.

Ghosh, N., Mishra, G., Sandhu, C., Grischek, T., and Singh, V. (2014). Interaction of Aquifer and River-Canal Network near Well Field. *Groundwater*.

Holl, A. (2016). Highways and Productivity in Manufacturing Firms. *Journal of Urban Economics*, 93(C):131–151.

Khanna, G. (2014). The Road Oft Taken: The Route to Spatial Development. Technical report, mimeo.

Lu, C. and Fan, L. (2013). Winter wheat yield potentials and yield gaps in the North China Plain. *Field Crops Research*, 143:98–105.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, 90(4):683–701.

Pfaff, A., Robalino, J., Reis, E., Walker, R., Perz, S., Laurance, W., Bohrer, C., Aldrich, S., Arima, E., Caldas, M., and Kirby, K. (2018). Roads & SDGs, tradeoffs and synergies: Learning from Brazil's Amazon in distinguishing frontiers. *Economics - The Open-Access, Open-Assessment E-Journal*, 12:1–26.

Qiu, J. (2010). China Faces Up to Groundwater Crisis. *Nature*, 466(308).

State Council (2012). List of national key counties for poverty alleviation. http://www.cpad.gov.cn/publicfiles/business/htmlfiles/FPB/fpyw/201203/175445.html.

Wang, J., Huang, J., Blanke, A., Huang, Q., and Rozelle, S. (2006). *The Development, Challenges and Management of Groundwater in Rural China. Groundwater in Developing World Agriculture: Past, Present and Options for a Sustainable Future*, chapter The Development, Challenges and Management of Groundwater in Rural China, pages 38–61. International Water Management Institute.

WantChinaTimes.com (2014). Has anything changed for Lankao's poor since Jiao Yulu's death? wantchinatimes.com.

Tables

Table 1: Descriptive statistics extensive margin

		Cells with a positive number of wells				
Year	**Number of cells with zero wells**	**Number of cells with non-zero wells**	**Mean**	**St. deviation**	**Min**	**Max**
1981	16848	1514	1.09	0.32	1	4
1982	16804	1558	1.09	0.32	1	4
1983	16756	1606	1.10	0.35	1	4
1984	16682	1680	1.12	0.37	1	4
1985	16498	1864	1.14	0.39	1	4
1986	16407	1955	1.15	0.41	1	4
1987	16341	2021	1.15	0.42	1	4
1988	16264	2098	1.16	0.43	1	4
1989	16172	2190	1.16	0.43	1	4
1990	15983	2379	1.18	0.46	1	4
1991	15895	2467	1.19	0.46	1	4
1992	15755	2607	1.19	0.47	1	4
1993	15616	2746	1.20	0.48	1	4
1994	15495	2867	1.21	0.50	1	5
1995	15176	3186	1.24	0.52	1	5
1996	14993	3369	1.25	0.54	1	5
1997	14856	3506	1.25	0.55	1	6
1998	14582	3780	1.27	0.57	1	6
1999	14489	3873	1.28	0.57	1	6
2000	14118	4244	1.31	0.61	1	6
2001	13956	4406	1.31	0.61	1	6
2002	13816	4546	1.32	0.62	1	6
2003	13591	4771	1.35	0.65	1	6
2004	13415	4947	1.35	0.65	1	6
2005	13099	5263	1.37	0.69	1	8
2006	12882	5480	1.39	0.70	1	8
2007	12748	5614	1.41	0.72	1	8
2008	12462	5900	1.43	0.75	1	8
2009	11698	6664	1.50	0.82	1	12
2010	10946	7416	1.58	0.91	1	12
2011	10691	7671	1.62	0.96	1	12

Notes: The sample contains 18,362 cells.

Table 2: Descriptive statistics

Variable	Mean	St. Dev.	Min	Max
Well Depth (meters)	41.95	8.33	18.00	250.00
Water Table (meters)	13.92	6.07	1.00	40.00
Distance to highway 1511 (km)	9.72	6.42	0.00	25.00
Distance to highway 30 (km)	18.77	8.39	0.00	37.00
Distance to Yellow River (km)	22.04	11.49	0.00	46.00
Distance to primary canals(km)	2.05	1.95	0.00	11.00
Number of wells within 100m radius	1.52	0.90	1.00	9.00
Number of wells within 200m radius	3.30	2.03	1.00	17.00
Number of wells within 500m radius	14.24	7.21	1.00	54.00
Distance to the nearest well(km)	0.13	0.08	0.00	1.00

Notes: The sample contains 7,526 wells.

Table 3: Difference-in-difference estimation for the effect of the construction of G30

	Probability of digging a new well			
	OLS			Poisson
	(1)	(2)	(3)	(4)
Treat	-0.026^{**}			
	(0.012)			
Post	0.238^{***}	0.238^{***}		1.963^{***}
	(0.031)	(0.031)		(0.030)
Treat*Post	0.057^{***}	0.057^{**}	0.057^{**}	0.267^{***}
	(0.020)	(0.021)	(0.021)	(0.071)
Grid cell FE	no	yes	yes	yes
Year FE	no	no	yes	yes
Observations	398,195	398,195	398,195	155,279

Notes: The variable *Treat* is a dummy taking value 1 if the cell is in the treatment area, which has been fixed at 2 km around G30. The non-treated area includes all cells situated at least 2 km away from G30 and from G1511 and no more than 15 km from either of the two highways. The variable *Post* is a dummy taking value 1 for years following the beginning of the construction of G30 (1998). The interaction of the two variables captures the difference-in-difference effect. Columns (1) to (3) perform a standard OLS estimation, while in column (4) we present the results of a Poisson specification. Standard errors in parentheses are two-way clustered at the cell and year level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Common trend for G30 and G1511

	Number of wells in a cell	
	G30	G1511
	(1)	(2)
Treat*Trend	0.0003	0.002
	(0.001)	(0.001)
Grid cell FE	yes	yes
Year FE	yes	yes
Observations	218,365	337,436

Notes: The variable *Treat* is a dummy taking value 1 if the cell is in the treatment area, which has been fixed at 2 km around G30 and 500 m around G1511. The non-treated area includes all cells situated at no more than 15 km from either of the two highways. The variable *Trend* is a simple linear time trend. These estimations are restricted to the sample before the beginning of the construction of the two highways, before 1998 for G30 and before 2003 for G1511. Standard errors in parentheses are two-way clustered at the cell and year level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Difference-in-difference estimation for the effect of the construction of G1511

	Number of wells in a cell			
	OLS			Poisson
	(1)	(2)	(3)	(4)
Treat	0.052***			
	(0.017)			
Post	0.272***	0.272***		1.929***
	(0.037)	(0.037)		(0.025)
Treat*Post	0.142***	0.142***	0.142***	0.108**
	(0.038)	(0.040)	(0.038)	(0.045)
Grid cell FE	no	yes	yes	yes
Year FE	no	no	yes	yes
Observations	475,478	475,478	475,478	193,781

Notes: The variable *Treat* is a dummy taking value 1 if the cell is in the treatment area, which has been fixed at 500 m around G1511. The non-treated area includes all cells situated at least 500 m away from G1511 and from G30 and no more than 15 km from either of the two highways. The variable *Post* is a dummy taking value 1 for years following the beginning of the construction of G1511 (2003). The interaction of the two variables captures the difference-in-difference effect. Columns (1) to (3) perform a standard OLS estimation, while in column (4) we present the results of a Poisson specification. Standard errors in parentheses are two-way clustered at the cell and year level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Effect of Distance of Well from G30 and G1511

	Dep. variable: Water table			
	(1)	(2)	(3)	(4)
Distance to G1511 (km)	-0.548	-0.628^{*}	-0.624^{*}	-0.979^{**}
	(0.348)	(0.366)	(0.363)	(0.453)
Distance to G30 (km)	-0.442^{*}	-0.466^{**}	-0.463^{**}	0.844^{**}
	(0.230)	(0.215)	(0.215)	(0.417)
Distance to G1511 squared				0.015
				(0.013)
Distance to G30 squared				-0.034^{***}
				(0.009)
Dist G30*Dist G1511	0.027^{*}	0.052^{***}	0.052^{***}	0.033^{*}
	(0.015)	(0.017)	(0.017)	(0.018)
Distance to river		0.848^{**}	0.850^{**}	0.582^{**}
		(0.350)	(0.349)	(0.277)
Distance to river squared		-0.024^{**}	-0.024^{**}	-0.011
		(0.011)	(0.011)	(0.008)
Village F.E.	yes	yes	yes	yes
Decade F.E.	no	no	yes	yes
Observations	7,525	7,525	7,525	7,525

Notes: All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. Column (1) contains only distances from the two new highways (G30 and G1511), their interaction and village fixed effects. In columns (2) we add distance to the Yellow River and its squared value, while in column to (3) we also add decade fixed effects. These fixed effects are supposed to capture the following policy changes. In the 80s, the introduction of the Household Responsibility System, with a rental contract period set at not less than 15 years. This lead to a higher land fragmentation, which could make people drill more. In the 90s farmers were allowed to rent in/out land, and this policy change may have affected profit incentives. Finally, in the 00s, the contract period was extended to at least 30 years. Finally, in column (4), we also add the squares of the distances to the two new highways. All distances are measured in kilometres. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Effect of Distance of Well from G30 and G1511

	Dep. variable: Water table					
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to G1511 (km)	-0.548	-1.181^{**}	-0.628^{*}	-0.579^{*}	-0.624^{*}	-0.592^{*}
	(0.342)	(0.586)	(0.360)	(0.339)	(0.358)	(0.332)
Distance to G30 (km)	-0.442^{*}	-0.694^{***}	-0.466^{**}	-0.446^{***}	-0.463^{**}	-0.451^{***}
	(0.226)	(0.239)	(0.212)	(0.144)	(0.212)	(0.141)
Dist G30*Dist G1511	0.027^{*}	0.103	0.052^{***}	0.061^{*}	0.052^{***}	0.059^{*}
	(0.015)	(0.119)	(0.017)	(0.032)	(0.017)	(0.032)
Distance to river			0.848^{**}	1.016^{*}	0.850^{**}	0.985^{*}
			(0.345)	(0.542)	(0.344)	(0.526)
Distance to river squared			-0.024^{**}	-0.031	-0.024^{**}	-0.029
			(0.011)	(0.020)	(0.011)	(0.020)
Village F.E.	yes	yes	yes	yes	yes	yes
Decade F.E.	no	no	no	no	yes	yes
Observations	7,525	7,525	7,525	7,525	7,525	7,525
F-stat G1511		38.73		74.72		75.31
F-stat G30		115.47		2864.12		2952.90
F-stat G30*G1511		0.82		10.09		10.13

Notes: All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. Column (1), (3) and (5) contain the baseline OLS results from table 6. Columns (2), (4) and (6) contain the equivalent results using an instrumental variable strategy. The instrument used is the distance of the well from the straight line connecting the two cities that each of the two highways has been built to connect. All distances are measured in kilometres. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Controlling for Well Density

	Dependent variable Water table					
	Baseline	No. of wells within 100m	No. of wells within 200m	No. of wells within 500m	Distance to closest well	Both measures
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to G1511 (km)	−0.624*	−0.623*	−0.623*	−0.609*	−0.622*	−0.609*
	(0.363)	(0.363)	(0.364)	(0.362)	(0.364)	(0.362)
Distance to G30 (km)	−0.463**	−0.462**	−0.463**	−0.459**	−0.462**	−0.459**
	(0.215)	(0.215)	(0.215)	(0.214)	(0.215)	(0.214)
Dist G30*Dist G1511	0.052***	0.052***	0.052***	0.052***	0.052***	0.052***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Distance to river (km)	0.850**	0.850**	0.850**	0.852**	0.850**	0.852**
	(0.349)	(0.349)	(0.349)	(0.347)	(0.348)	(0.347)
Distance to river^2 (km)	−0.024**	−0.024**	−0.024**	−0.024**	−0.024**	−0.024**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Number of wells in 100m buffer		0.038				
		(0.033)				
Number of wells in 200m buffer			0.010			
			(0.021)			
Number of wells in 500m buffer				0.018*		0.018*
				(0.010)		(0.009)
Distance to closest well (km)					−0.365	0.085
					(0.496)	(0.427)
Village F.E.	yes	yes	yes	yes	yes	yes
Decade F.E.	yes	yes	yes	yes	yes	yes
Observations	7,525	7,525	7,525	7,525	7,525	7,525

Notes: All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. Column (1) contains the baseline intensive margin results. In columns (2) to (4) we control for within village variation in groundwater utilisation by controlling for the number of wells within a radius of 100m, 200m and 500m around each individual well. In column (5) we control for the same effect using the distance to the closest well. Finally, in column (6) we control for the number of wells within a 500m radius and the distance to the nearest well at the same time. All distances are measured in kilometres. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Estimation without Wells close to the Highway

	Dependent variable Log Water table$_{iv}$					
	Baseline	Elimin. wells within 50m of the highway	Elimin. wells within 100m of the highway	Elimin. wells within 500m of the highway	Elimin. villages with less than 30 wells	Elimin. villages with less than 40 wells
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to G1511 (km)	−0.624*	−0.621*	−0.625*	−0.640*	−0.694	−0.708*
	(0.363)	(0.365)	(0.366)u	(0.380)	(0.423)	(0.358)
Distance to G30 (km)	−0.463**	−0.466**	−0.468**	−0.485**	−0.505**	−0.320
	(0.215)	(0.216)	(0.217)	(0.225)	(0.245)	(0.220)
Dist G30*Dist G1511	0.052***	0.053***	0.053***	0.056***	0.061***	0.070***
	(0.017)	(0.017)	(0.017)	(0.019)	(0.020)	(0.021)
Distance to river (km)	0.850**	0.871**	0.872**	0.947**	1.074**	0.570
	(0.349)	(0.351)	(0.352)	(0.366)	(0.423)	(0.550)
Distance to river^2 (km)	−0.024**	−0.025**	−0.025**	−0.027**	−0.030**	−0.026*
	(0.011)	(0.011)	(0.011)	(0.012)	(0.013)	(0.016)
Village F.E.	yes	yes	yes	yes	yes	yes
Decade F.E.	yes	yes	yes	yes	yes	yes
Observations	7,525	7,499	7,482	7,269	4,991	3,873

Notes: All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. Column (1) contains the baseline intensive margin results. In columns (2) to (4) we eliminate wells that are located within 50m, 100m and 500m from the highway in order to account for the possible changes in the aquifer due to the construction of the highway. In columns (5) and (6) we limit the analysis to villages containing at least 30 and at least 40 wells. All distances are measured in kilometres. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Robustness: other highways and highway exit

	Dep. variable: Water table		
	Baseline	Others	Exit
	(1)	(2)	(3)
Distance to G1511	-0.624^{*}	-1.416^{***}	
	(0.363)	(0.317)	
Distance to G30	-0.463^{**}	-0.456	
	(0.215)	(0.304)	
Distance to G220		0.882^{***}	
		(0.236)	
Distance to G106		-0.044	
		(0.281)	
Distance to G310		-0.055	
		(0.300)	
Distance to G30 exit			-0.194
			(0.189)
Dist G30*Dist G1511	0.052^{***}	0.048^{***}	
	(0.017)	(0.013)	
Distance to river	0.850^{**}	0.923^{***}	0.212
	(0.349)	(0.242)	(0.214)
Distance to river squared	-0.024^{**}	-0.022^{***}	-0.004
	(0.011)	(0.007)	(0.006)
Village F.E.	yes	yes	yes
Decade F.E.	yes	yes	yes
Observations	7,525	7,525	7,525

Notes: All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. Column (1) contains the baseline intensive margin results. In column (2) we add distance to the three old highways, G220, G106 and G310. In column (3) we measure the distance to the highway exit in Lankao county. All distances are measured in kilometres. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Crop distance to the highway

Variable	Obs	Mean	St. Dev.	Min	Max
Proportion of crops					
Fresh vegetables		0.01	0.09	0	1
Melons		0.01	0.12	0	1
Peanuts		0.08	0.28	0	1
Wheat		0.46	0.50	0	1
Corn		0.40	0.50	0	1
Other cash crops		0.03	0.18	0	1
Minimum distance from G30/G1511 (km)					
Fresh vegetables	16	0.816	0.453	0.250	1.858
Melons	23	1.130	1.308	0.250	6.876
Peanuts	133	8.126	5.553	0.047	17.84
Wheat	736	5.327	4.425	0.028	17.840
Corn	633	4.942	3.883	0.028	17.533

Notes: The sample contains 602 wells, managed by 276 households and servicing 1,304 plots of land. The variable *Other cash crops* contains: cotton, potatos, beans, oil, apples, peaches, poplar, pears and garlic.

Table 12: Difference-in-difference estimation for the effect of the construction of G1511 on NDVI

	NDVI			
	OLS			Falsif.
	(1)	(2)	(3)	(4)
Treat	0.049^{***}			
	(0.006)			
Post	0.053	0.053		
	(0.094)	(0.094)		
Treat*Post	-0.036^{***}	-0.036^{***}	-0.036^{***}	
	(0.009)	(0.009)	(0.010)	
Trend*Post				0.003
				(0.006)
Grid cell FE	no	yes	yes	yes
Year FE	no	no	yes	yes
Observations	599,820	599,820	599,820	239,928

Notes: The dependent variable is the Normalized Difference Vegetation Index (NDVI), which can vary between -1 and 1, -1 if there is no vegetation and 1 for a maximum vegetation cover. The variable *Treat* is a dummy taking value 1 if the cell is in the treatment area, which has been fixed at 500 m around G1511. The non-treated area includes all cells situated at least 500 m away from G1511 and from G30 and no more than 15 km from either of the two highways. The variable *Post* is a dummy taking value 1 for years following the beginning of the construction of G1511 (2003). The interaction of the two variables captures the difference-in-difference effect. Columns (1) to (3) perform a standard OLS estimation, column (4) presents the results of a falsification test for the common trend hypothesis, and is therefore based only on data before 2003. Standard errors in parentheses are two-way clustered at the cell and year level. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Wheat and Maize: yield and price

	Dep. variable:			
	Wheat (winter)		Corn (summer)	
	yield	price	yield	price
	(1)	(2)	(1)	(2)
Distance to G1511 (km)	−17.862***	−0.014***	−0.299	0.002
	(5.872)	(0.005)	(9.610)	(0.003)
Distance to G30 (km)	−0.688	−0.001	3.039	0.0005
	(1.585)	(0.001)	(2.328)	(0.001)
Dist G30*Dist G1511	0.0004*	$4.03e{-7}$**	−0.0002	$-9.47e{-8}$
	(0.0002)	($1.69e{-7}$)	(0.0003)	($9.6e{-8}$)
Plot controls	yes	yes	yes	yes
Observations	734	593	631	590

Notes: Yield is expressed in jin, where 1 jin corresponds to 0.5 kg. The price is per jin. In this table, distances are expressed in kilometres and not metres in order to facilitate the interpretation of the coefficients. Standard errors in parentheses are clustered by well, since several plots may be served by the same well. Plot controls include plot quality (compared to the average plot), and plot slope. *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Cost per ha: Winter Wheat

	Dep. variable:			
	Seeds cost	Fertilizer cost	Pesticide cost	Herbicide cost
	(1)	(2)	(3)	(4)
Distance to G1511 (km)	4.561	7.987*	0.571	0.448
	(6.977)	(4.088)	(1.286)	(0.662)
Distance to G30 (km)	0.632	2.687***	0.471	0.216
	((1.144)	(0.985)	(0.394)	(0.170)
Dist G30*Dist G1511	−0.0002	−0.0003**	−0.0001	−0.00004*
	(0.00022)	(0.00015)	(0.00005)	(0.00002)
Plot controls	yes	yes	yes	yes
Observations	731	727	735	735

Notes: Quantities are per mu (15 mu constitute a hectare). Plot controls include plot quality (compared to the average plot), and plot slope. In this table, distances are expressed in kilometres and not metres in order to facilitate the interpretation of the coefficients. Standard errors in parentheses are clustered by well, since several plots may be served by the same well. *** p<0.01, ** p<0.05, * p<0.1.

Table 15: Cost per ha: Summer Corn

	Dep. variable:			
	Seeds cost	Fertilizer cost	Pesticide cost	Herbicide cost
	(1)	(2)	(3)	(4)
Distance to G1511 (km)	3.684	7.349*	0.515	0.591
	(8.163)	(4.391)	(1.427)	(0.762)
Distance to G30 (km)	0.202	2.117**	0.258	0.308
	(1.257)	(1.031)	(0.371)	(0.192)
Dist G30*Dist G1511	−0.0001	−0.0003**	−0.0001	−0.00005*
	(0.0002)	(0.0002)	(0.00005)	(0.00003)
Plot controls	yes	yes	yes	yes
Observations	630	625	632	632

Notes: Quantities are per mu (15 mu constitute a hectare). Plot controls include plot quality (compared to the average plot), and plot slope. In this table, distances are expressed in kilometres and not metres in order to facilitate the interpretation of the coefficients. Standard errors in parentheses are clustered by well, since several plots may be served by the same well. *** p<0.01, ** p<0.05, * p<0.1.

Table 16: Wheat and Maize: change in the production technology

	Dep. variable:			
	Wheat (winter)		Corn (summer)	
	Water use	Capital investment	Water use	Capital investment
	(1)	(2)	(1)	(2)
Distance to G1511 (km)	−13.264	−33.320***	−4.615	−34.725***
	(9.253)	(12.393)	(10.192)	(12.348)
Distance to G30 (km)	−10.268***	−15.627**	−7.740***	−10.337
	(2.253)	(7.298)	(2.358)	(6.844)
Dist G30*Dist G1511	0.0005	0.001***	0.00001	0.001***
	(0.0003)	(0.0004)	(0.0004)	(0.0004)
Plot controls	yes	yes	yes	yes
Observations	736	704	633	605

Notes: Quantities are per mu (15 mu constitute a hectare). Plot controls include plot quality (compared to the average plot), and plot slope. In this table, distances are expressed in kilometres and not metres in order to facilitate the interpretation of the coefficients. Standard errors in parentheses are clustered by well, since several plots may be served by the same well. *** p<0.01, ** p<0.05, * p<0.1.

Figures

Figure 1: Location of Lankao county

Figure 2: Map of Lankao county

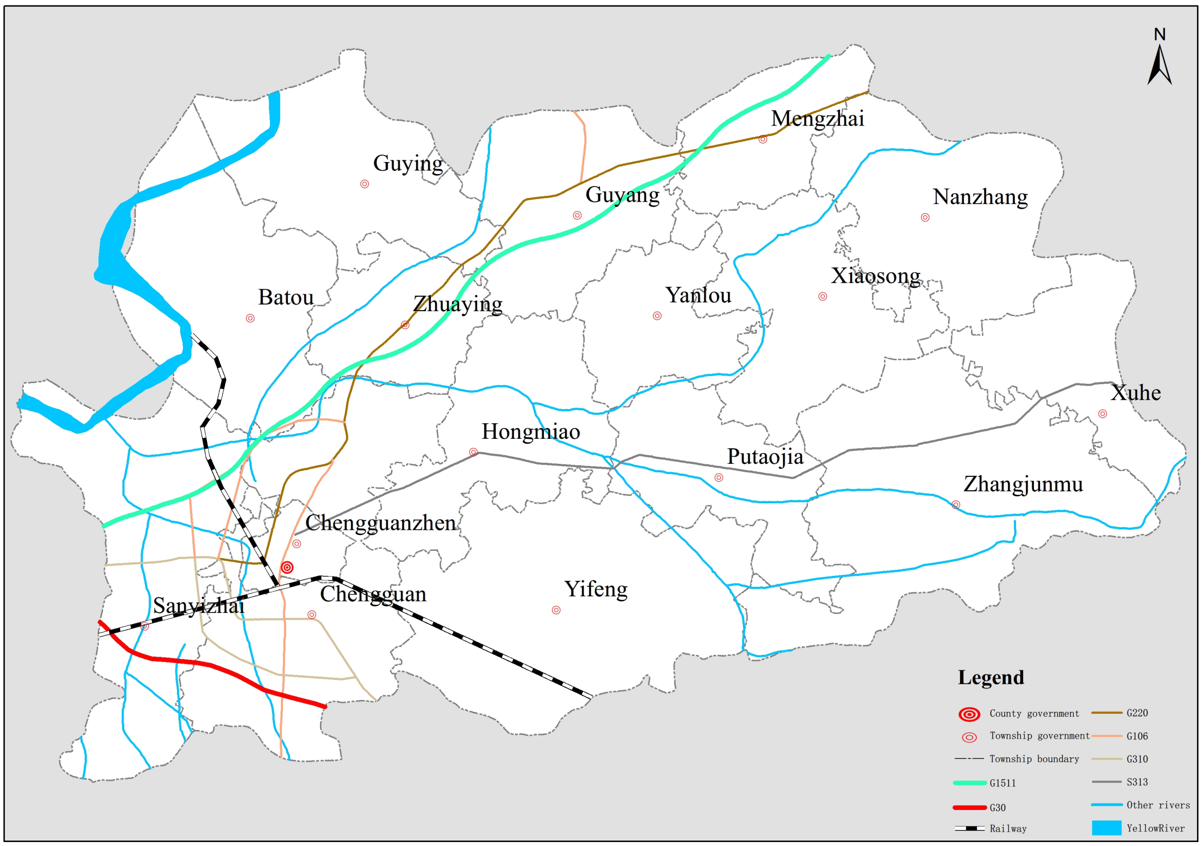
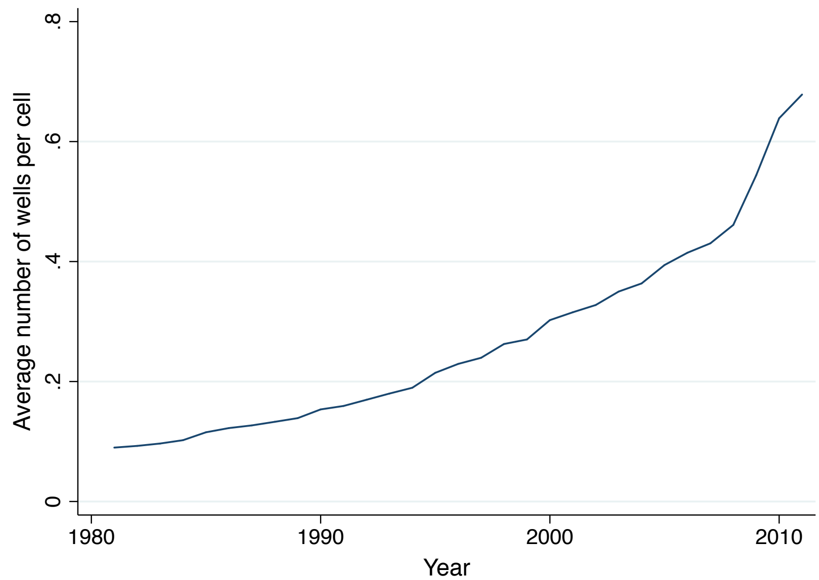


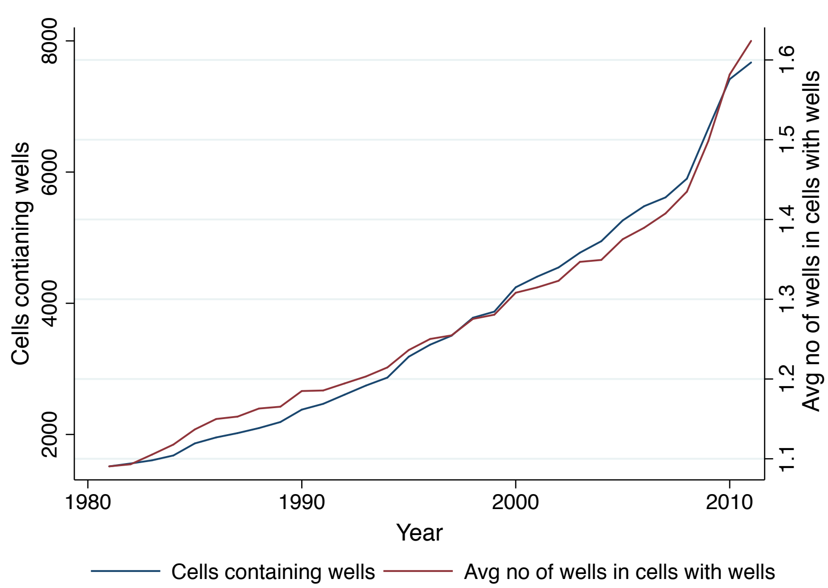
Figure 3: Tubewells evolution (by decade) from 1955 to 2011

Figure 4: Evolution of the average number of wells per cell from 1980 to 2011



Notes: We look at the average number of wells found in the 18,362 cells composing Lankao county. This number evolves from 0.1 in 1980 to roughly 0.7 in 2011.

Figure 5: Evolution of the cells containing wells from 1980 to 2011



Notes: In this graph we focus only on cells containing wells. The blue line shows that over the years more and more cells contained at least one well, while the red line shows the average number of wells found in these cells. This number is also increasing.

Figure 6: Depth of watertable in tubewells in 2011

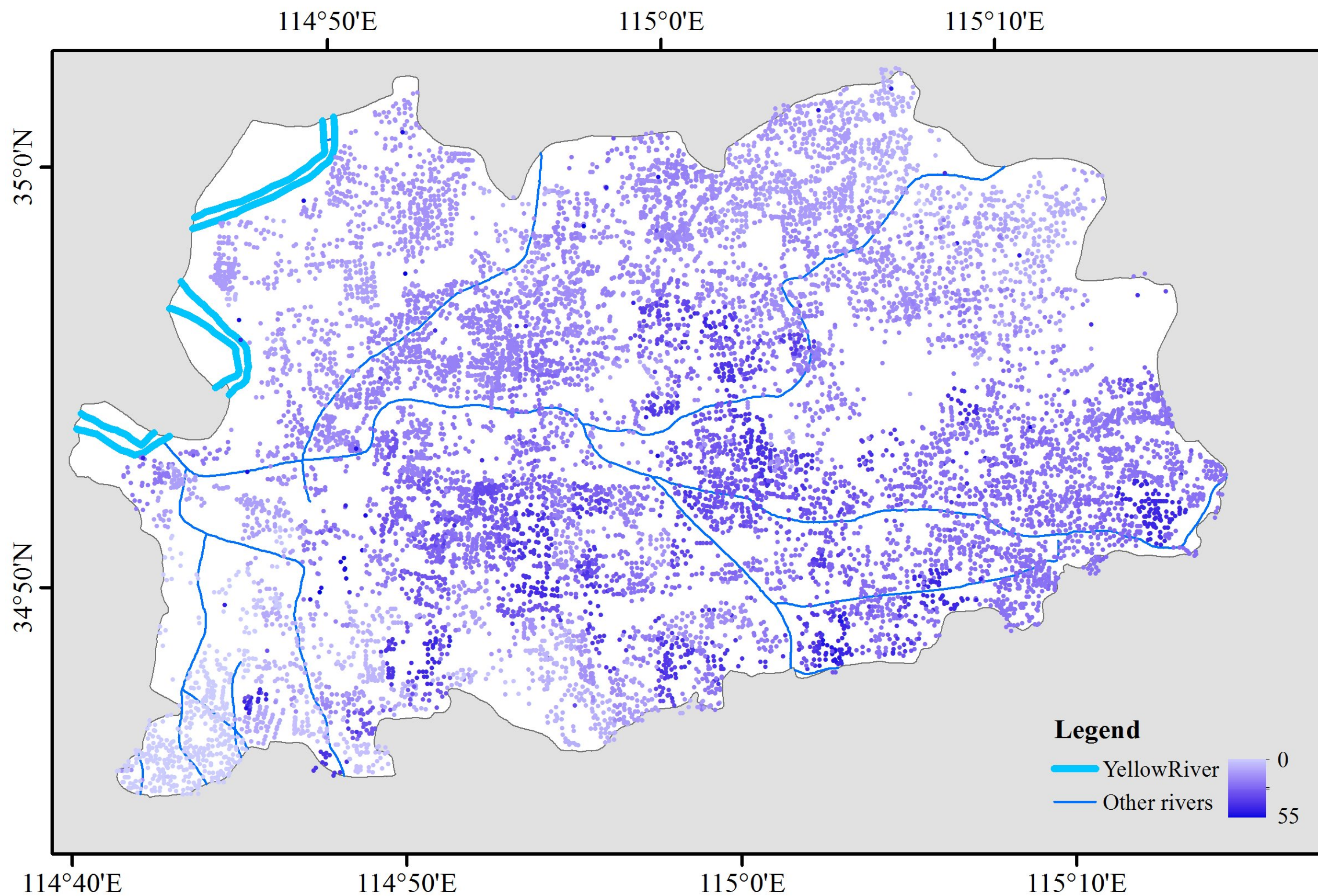


Figure 7: Variation of the DID coefficient as the boundary of the treatment moves away from G30

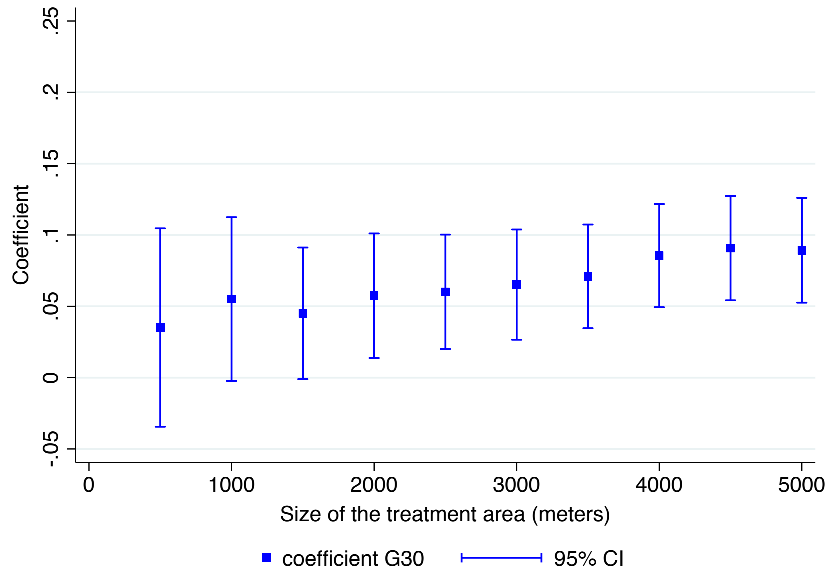


Figure 8: Variation of the DID coefficient as the boundary of the treatment moves away from G1511

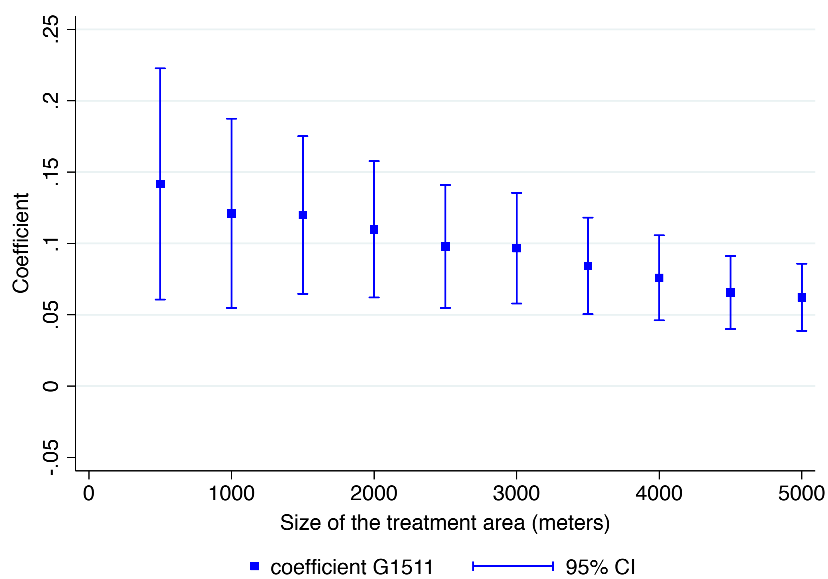


Figure 9: Variation of the DID coefficient as the buffer between treatment and control increases G30

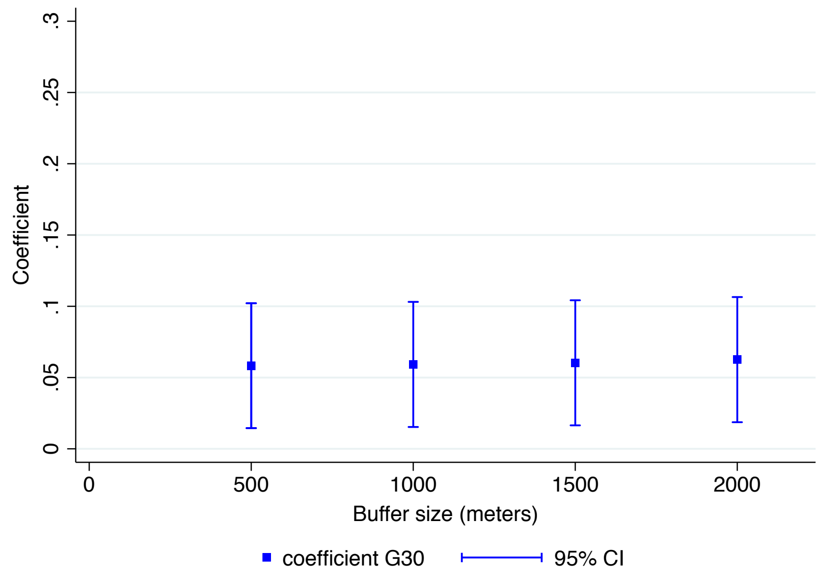


Figure 10: Variation of the DID coefficient as the buffer between treatment and control increases G1511

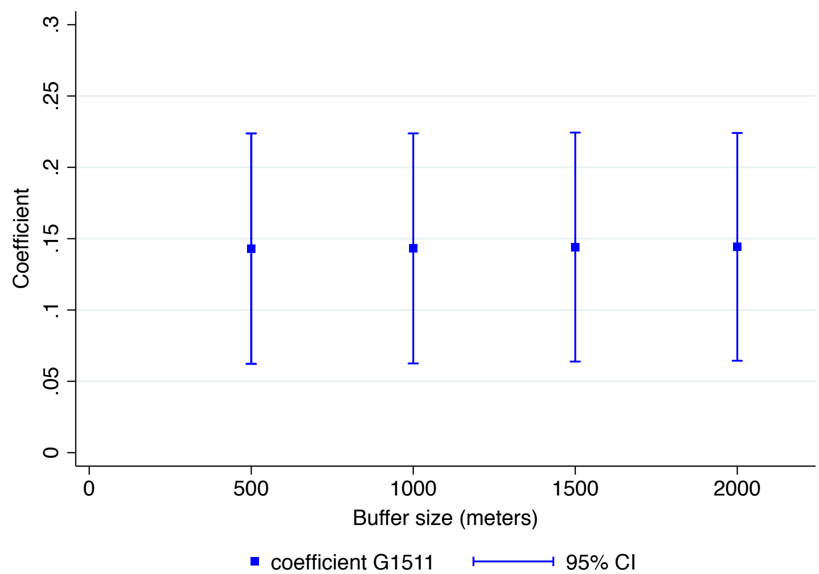


Figure 11: Depth of Water Table in Wells with Distance from Yellow River

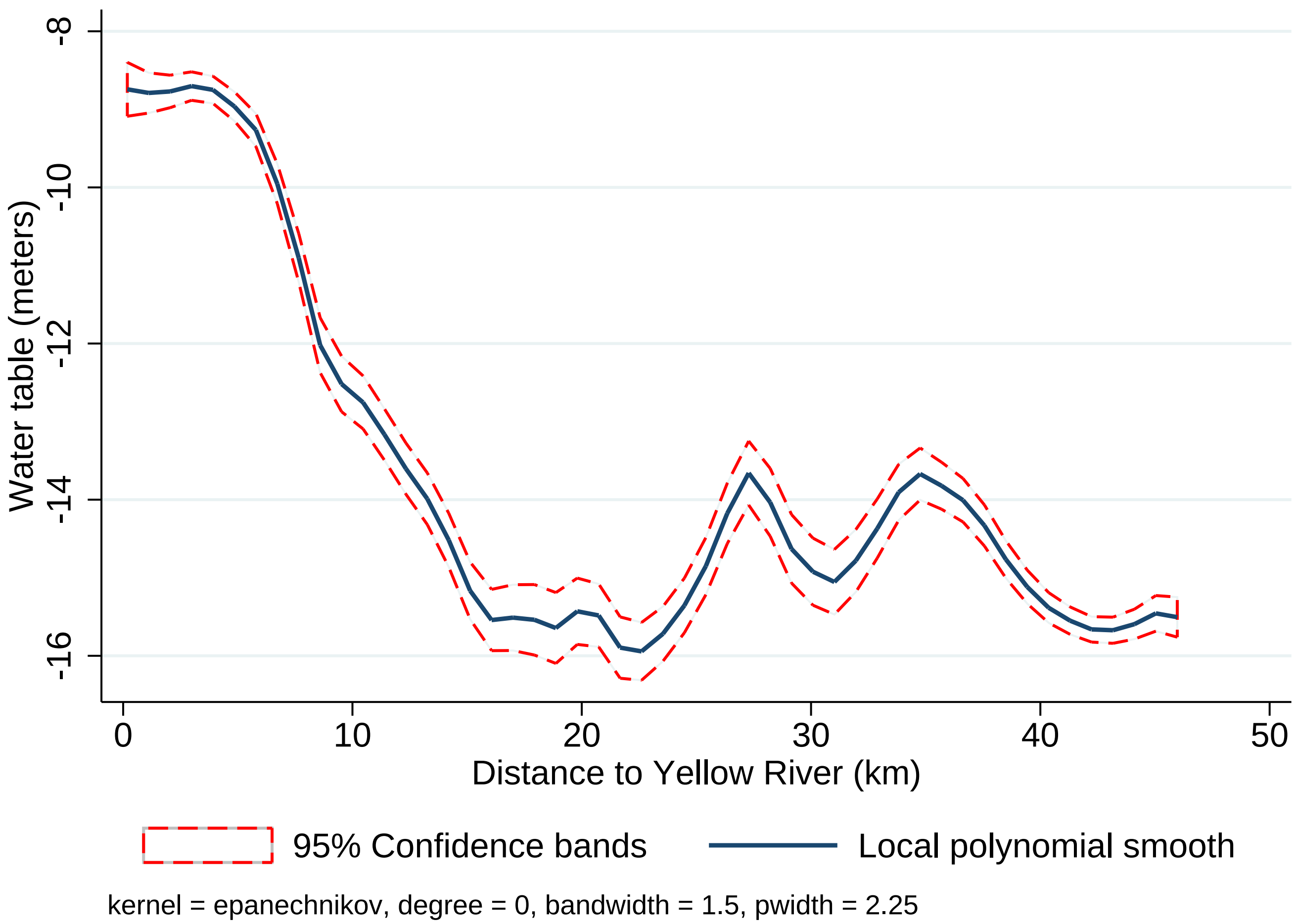


Figure 12: Marginal effect of the distance from G1511 and G30

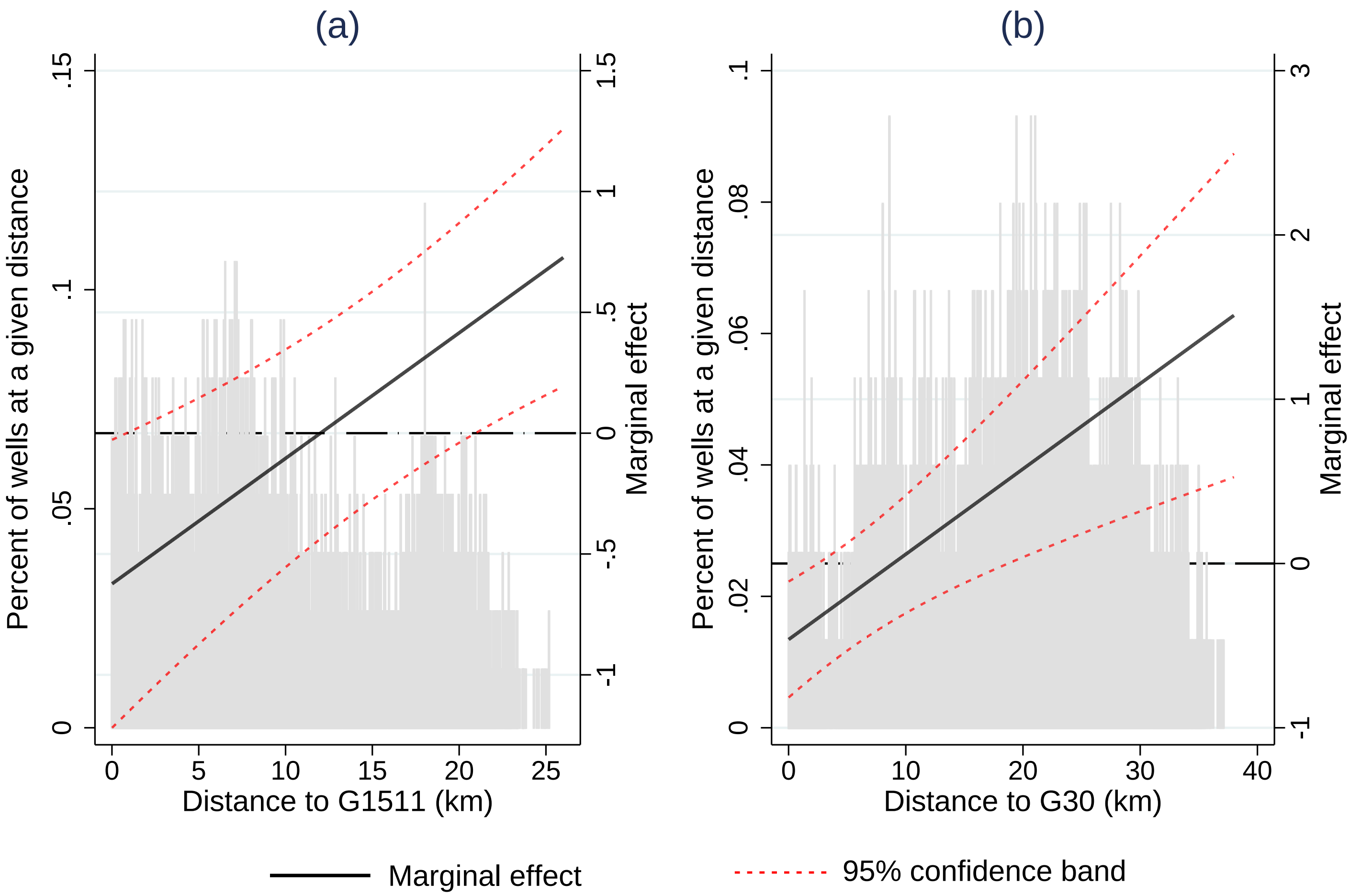
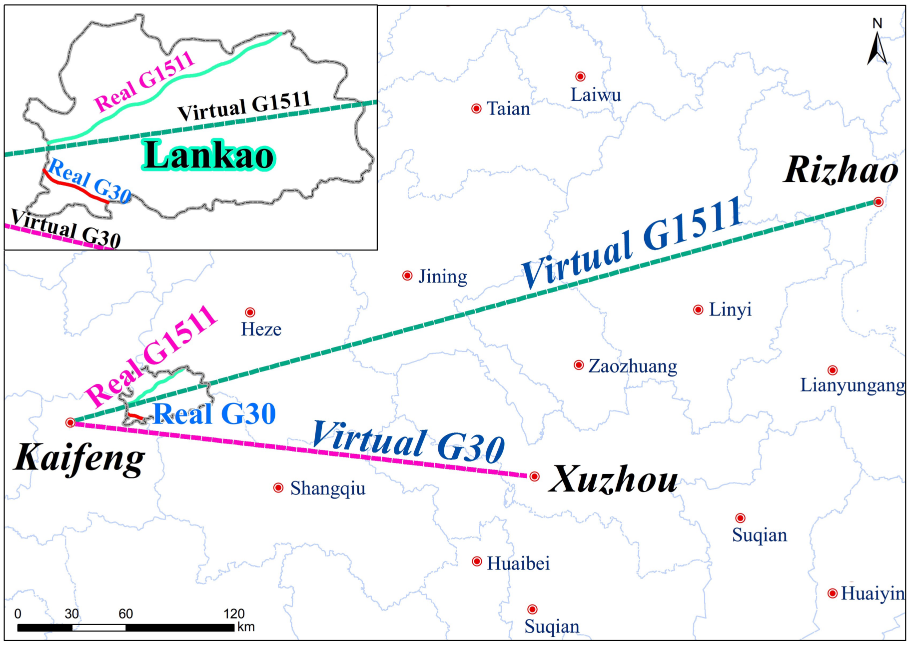


Figure 13: G1511 and G30 and the corresponding straight lines



7 Appendix A.1: Survey

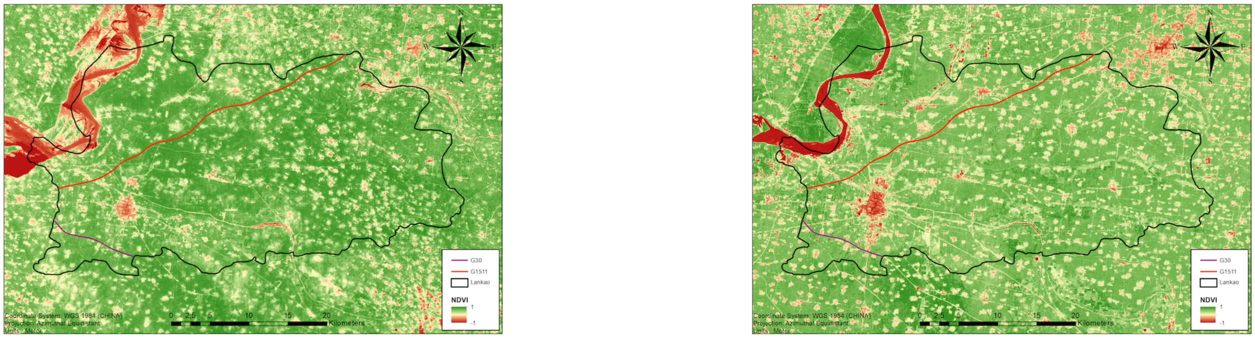
We conducted questionnaire surveys with family heads of 300 households in Lankao County in summer of 2014. To choose survey respondents, we used random sampling methods. We randomly selected 30 villages from a total number of 429 villages in Lankao. These 30 villages are from 13 townships of a total number of 16 townships in the county. To randomly choose these villages, we first put all 429 villages on a list. We then generated a random number from a normal distribution and took that random number as a starting point. That starting point was the first village we chose. We then moved down to a village whose number of 14 away from the first village. With the same iteration, we chose a total number of 30 villages. In each selected village, 10 households were randomly chosen from village rosters. For each selected family, the family head was asked to participate in personal interviews conducted by our survey team.

The questionnaire used in our surveys contains the following key information: [1] family demographic information; [2] detailed information related to each of all wells used by the family in 2013. Such information includes the ownership of each well used by the family, the type of pumps used in each well, plots irrigated by each well and ownership of each well etc. In addition, information of coordinates (longitude and latitude) was collected for each

Figure 14: Total area cultivated with fresh vegetables in Lankao county (1986-2009)

Notes: The two vertical lines denote the years in which construction started for G30 (1998) and G1511 (2001). Source: Chinese National Bureau of Statistics.

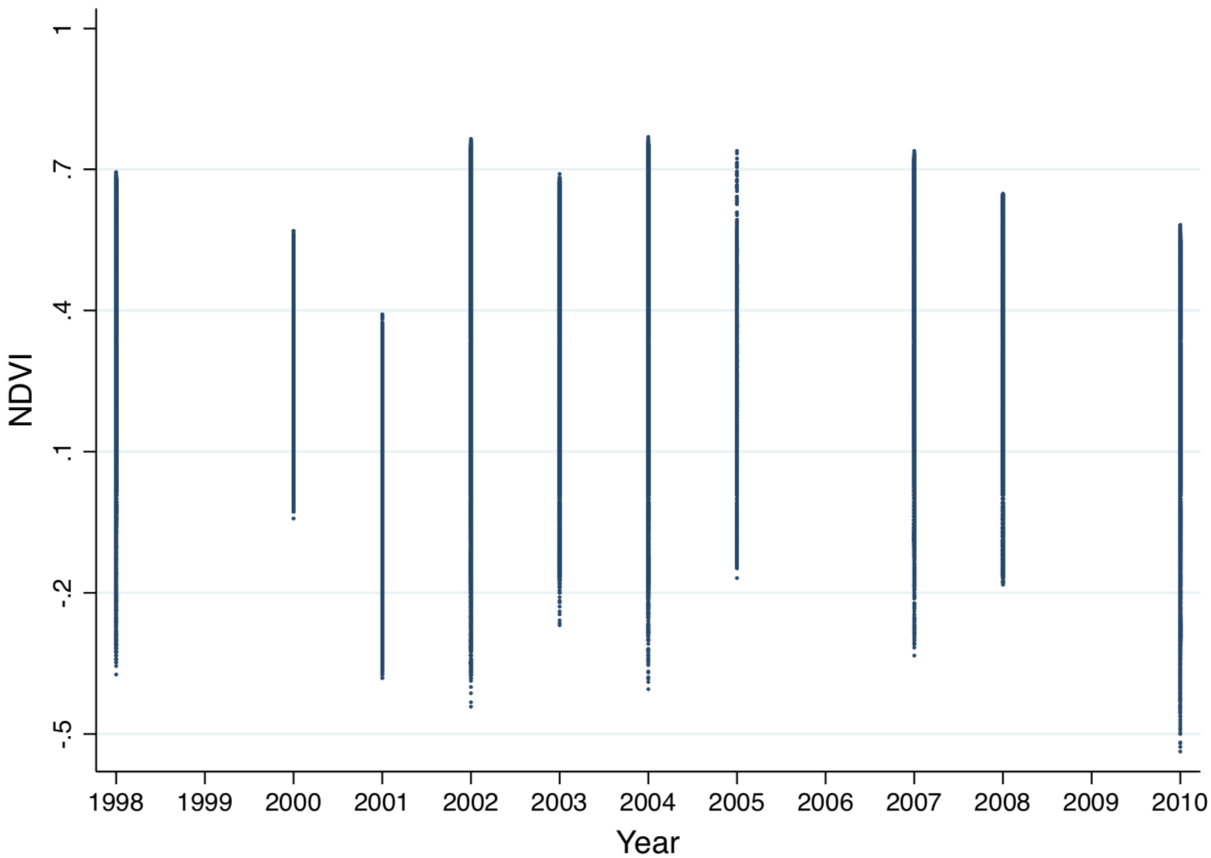
Figure 15: NDVI, 1998 and 2010



Notes: The panel on the left shows NDVI for Lankao county during the summer of 1998, while the panel on the right shows the same picture for the summer of 2010. These images are constructed using satellite data from the United States Geological Survey (USGS). NDVI values are computed for each of the 75,828 150 x 150 meters cells composing Lankao county. NDVI varies between -1, no vegetation – represented by the color red – and +1, maximum vegetation cover – represented by the color green.

well used by the family in 2013. [3] detailed information related to each of all pumps used by the family, such as types of pumps, the well corresponding to each pump, pumping time etc.; [4] detailed information related to each of all plots operated by the family in 2013. For each plot, the following detailed information was collected: slope, soil fertility, sources of irrigation water (including no irrigation), irrigation methods; wells used for irrigation (the corresponding well mentioned in [2]). For each plot, if the family has surface water to irrigate their plot, the following information was collected: frequency and time of irrigation, total cost of irrigation. For each plot, if the family used ground water to irrigate the plot, the following information was collected: frequency and time of irrigation. If the family had more than one irrigation in 2013, the following further information was collected: time of each irrigation. For each plot, information on each of all crops grown on the plot was collected. For each crop grown on each plot, information of frequency and time of irrigation during the crop growing season, input used and its yield was collected.

Figure 16: NDVI by cell and year for Lankao (1998-2010)



Notes: NDVI by year for the 75,828 cells (150x150 meters) composing Lankao county. Data for 1999, 2006 and 2009 are missing because images of sufficient quality for the desired period were not available. The data used to compute the NDVI come from the United States Geological Survey. NDVI varies between -1 – indicating no vegetation in the cell – to 1 – indicating a full cover.