CIRANO
Knowledge into action

A TALE OF TWO ROADS: GROUNDWATER DEPLETION IN THE NORTH CHINA PLAIN

UJJAYANT **CHAKRAVORTY**
XIANGZHENG **DENG**
YAZHEN **GONG**
MARTINO **PELLI**
QIAN **ZHANG**

2023s-17
WORKING PAPER
CS

ISSN 2292-0838 (online version)

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

Ujjayant Chakravorty[†], *Xiangzheng Deng*[‡], *Yazhen Gong*[§], *Martino Pelli*[**], *Qian Zhang*[††]

Abstract/Résumé

There is a large literature on the role infrastructure plays in economic development, but few papers document the effect of infrastructure on the sustainability of natural resources. We examine the effect of the arrival of two new national highways on ground water levels in a small agricultural county in the North China Plain - a region that produces most of the nation's food grains. We first develop a conceptual framework to show that farmers located closer to the highways devote more acreage to crops that are water intensive. We then use a unique GIS-referenced dataset of all the 12,160 tube wells in this county to show that highway construction accelerates the drilling of new wells in farms closer to the highway. In addition, there is greater depletion of the groundwater in wells closer to the two highways relative to wells located farther away. Our estimated depletion rates near the two roads are at least 5 times higher relative to mean depletion rates in the North China Plain. We show suggestive evidence that depletion is caused by a switch from subsistence to commercial cropping, and intensification of farming practices closer to the highway. These results suggest that the resource cost of new infrastructure building may be significant and needs to be incorporated in benefit-cost analysis.

Il existe une abondante littérature sur le rôle que jouent les infrastructures dans le développement économique, mais peu d'articles documentent l'effet des infrastructures sur la durabilité des ressources naturelles. Nous examinons l'effet de l'arrivée de deux nouvelles routes nationales sur le niveau des eaux souterraines dans un petit comté agricole de la plaine de Chine du Nord, une région qui produit la plupart des céréales alimentaires du pays. Nous développons d'abord un cadre conceptuel pour montrer que les agriculteurs situés plus près des autoroutes consacrent plus de surface aux cultures à forte intensité d'eau. Nous utilisons

---

* We would like to thank participants at the 2020 ASSA meetings in San Diego, CIRANO, EPIC-Cempa, IIT Delhi, Delhi University, Savoie, Bern, Hawaii, Penn State, New Mexico, Tufts, Manitoba, Carleton, Manchester, the Chinese Academy of Sciences, IISER Bhopal, PACE, SCSE, CAERE, CEA, FAERE, the ISI Delhi annual conference, 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, Disha Gupta, 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. This work is supported by the National Key Research and Development Program (No. 2021YFC3200500) and the fund for building world-class universities at Renmin University (KYGJA2022004).

† Chakravorty: Department of Economics, Tufts University, USA, ujjayant.chakravorty@tufts.edu

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

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

** Pelli: Department of Economics, University of Sherbrooke, CIREQ, CIRANO and GREDI, Canada, martino.pelli@usherbrooke.ca

†† Zhang: College of Land Science and Technology, China Agricultural University, Beijing, China, qian.zhang@cau.edu.cn

ensuite un ensemble unique de données référencées par le SIG sur les 12 160 puits tubulaires de ce comté pour montrer que la construction de l'autoroute accélère le forage de nouveaux puits dans les exploitations agricoles situées à proximité de l'autoroute. En outre, l'épuisement des eaux souterraines est plus important dans les puits situés à proximité des deux autoroutes que dans les puits situés plus loin. Les taux d'épuisement estimés à proximité des deux routes sont au moins cinq fois plus élevés que les taux d'épuisement moyens dans la plaine de Chine du Nord. Nous montrons que l'épuisement est dû au passage d'une culture de subsistance à une culture commerciale et à l'intensification des pratiques agricoles à proximité de l'autoroute. Ces résultats suggèrent que le coût en ressources de la construction de nouvelles infrastructures peut être significatif et doit être intégré dans l'analyse coûts-avantages.

Keywords/Mots-clés: Infrastructure, Roads, North China Plain, Water Resources, sustainability / Infrastructure, Routes, Plaine de Chine du Nord, Ressources en eau, Durabilité

JEL Codes/Codes JEL: O13, O18, Q25

Pour citer ce document / To quote this document

Chakravorty, U., Deng, X., Gong, Y., Pelli, M., & Zhang, Q. (2023). A Tale of Two Roads: Groundwater Depletion in the North China Plain (2023s-17, Cahiers scientifiques, CIRANO.) https://doi.org/10.54932/ZMTE3487

1 Introduction

Infrastructure is key to economic development. Many economic studies have addressed the positive relationship between infrastructure provision and economic development. However, building new infrastructure leads to greater economic activity and faster depletion of the natural resource base - water, forest, and other resources. The benefits of building infrastructure must be weighed against its costs. There are few economic studies of the role infrastructure plays in resource depletion, and almost no causal estimates of its effects. Analysis of these potentially adverse effects can help economists to precisely estimate the benefits and costs of infrastructure provision and assess whether these development projects contribute to broader sustainable goals.

In this paper, we study the effect of two large national highways on groundwater levels in a small (roughly 30 km by 40 km) county in China. This county, Lankao, is among the poorest in the country, known nationally for high rates of poverty and for a historically important battle called The Battle of Lanfeng during the Sino-Japanese War (1937-45). The region is primarily agricultural and part of the North China Plain (NCP), the largest alluvial plain in China and its bread basket. Using data on all the 12,160 wells in the county, we show that the arrival of these highways led to the digging of new wells in areas closer to the roads, and triggered an increase in intensive water use. We provide evidence to suggest that farmers closer to the highways switched from growing cereals to perishable cash crops like melons and vegetables that require more water and can be easily transported to distant markets using the highway. We show that there is intensification of farming practices and labor use in farmer plots closer to the two roads.

Many studies focus on the effect of roads on economic growth and in particular, on firms and productivity. Michaels (2008) finds that the US highway system tends to draw economic activity to counties they cross and raised the relative demand for skilled labor in skill-abundant counties while reducing it elsewhere. Chandra and Thompson (2000) show that interstate highway construction raised earnings in counties which they crossed, while total earnings fell in adjacent counties. Holl (2016) studies the Spanish highway system and, instrumenting its placement using postal routes and Roman roads, finds that highways have a positive effect on firm productivity.[1] The effect of roads on market integration and local

[1]Several papers study the updating of the Golden Quadrilateral Highway (GQH) in India. The GQH connects the four largest cities: New Delhi, Mumbai, Chennai and Kolkata. Datta (2012), Ghani et al. (2013, 2016) and Khanna (2014) find a strong positive impact of the upgrading of the GQH on firms, which grew disproportionally along the network. These effects are found for districts up to 10 kilometers away from the highway and disappear at higher distances.

economic outcomes has been studied by Aggarwal (2018). Faber (2015) studies the Chinese National Trunk Highway System, instrumenting placement with the ideal least cost path and finds a reduction in industrial output from non-targeted peripheral regions. 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. (2020) 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. The causal effect of roads on depletion of forest resources has been studied by Asher et al. (2020), who find that upgrading of highways in India led to forest loss, but last mile rural roads had no such adverse effect.

Lankao county has several old highways but two new ones, G30 and G1511, that were built in the nineties and early this century. These roads connect major cities in China and crisscross a large expanse of the country. Only a tiny segment runs through the county - 11.6 km for G30 and 41.4 km for G1511. First, we develop a simple model of land and water use on plots at various distances from the highway to show that more land and water are allocated to the water-intensive crop closer to the highway. Next, we carry out an extensive margin analysis, which looks at the effect of the arrival of the two highways on the number of new wells dug in areas closer to them. For this analysis, we construct a panel dataset where the unit of observation is a 250m x 250m cell. We know the cumulative number of wells in each cell in each year. We conduct a DiD analysis for each highway, where the treatment is defined as the buffer around the highway which ranges from 500 meters to 10 km. The effect of the roads is positive and statistically significant, i.e., the number of new wells increases adjacent to both roads.

To study the intensive margin effect, we regress the water table depth in tube wells on the distance from the wells to the new roads, controlling for the distance to the Yellow River which passes through the county.[2] Wells close to the river are likely to have a higher water table because of seepage and are possibly pumped less if farmers supplement groundwater with surface water irrigation. We find 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 cm. That is, the water table gets deeper close to the road. An increase of 1 km in the distance from G1511 leads

[2]In our baseline specification we compute distances to the highway using the straight line ("as the crow flies") distance between each well and the closest point on the highway, we also run robustness tests where we use the actual travel distance between the wells and entry/exit ramps on the two highways downloaded from Gaode, a popular Chinese mapping app.

to a decrease of 62.4 cm in the depth of the water table.

The highways we study are part of a national transportation network connecting major cities and only a tiny segment (e.g., 11.6 km out of a total of 4,395 km for G30) passes through our county. We address the issue of the endogenous siting of the highways by running an instrumental variable specification. We instrument the distance of each well to the actual placement of the highway with the distance from the well to the virtual straight line joining the two cities outside the county that the highway actually connects. The results are robust to this specification.

We examine several mechanisms that could cause the drilling of new wells and water depletion along the highways. Using household survey data we show that in farms close to the highway, increased acreage is devoted to perishable crops like melons which are also water intensive. The highway can be used to transport perishable crops to outside markets in a timely fashion. We find evidence of a change in the vegetation index after the highway was constructed suggesting a perceptible change in cropping patterns. Both the price and yield of the traditional crop, wheat, declines as we move away from the road, suggesting a decrease in the intensity of farming, with distance from the roads. Household expenditure on farm machinery also falls with distance from the highway.[3] Calculations based on our estimates suggest that water depletion due to highway G30 may be 5-6 times larger than typical depletion rates in the North China Plain. The effects are higher for G1511 although less precisely estimated. These effects on the resource base are large and need to be included in any benefit cost calculation of infrastructure investments.

Section 2 provides background information on the region and the highways we study. In section 3, we develop a simple model of farming as a function of distance to the highway and show that yields and input use for the water-intensive crop decline with distance, but increase for the traditional crop. Section 4 outlines the data used in the paper. Section 5 provides the empirical specification, results and robustness checks. Section 6 discusses possible mechanisms and section 7 concludes the paper.

2 Background

Groundwater use in the North China Plain The depletion of groundwater resources is a major problem in China. It is home to more than 20% of the world's population but

[3]A similar result is obtained by Asher and Novosad (2020) who find that rural roads in India lead to a decrease in labor allocated to agriculture and a corresponding increase in wage labor.

has only 5-7 percent of the earth's freshwater resources (Qiu, 2010). Grain production is mainly concentrated in the North China Plain (NCP) where our study area is situated. NCP accounts for one-fifth of China's total geographic area, and contains 340 counties. About 72% of the area is under farming but only 6% has access to surface water (Lu and Fan, 2013). The region accounts for two-thirds of Chinese wheat production. Most farms practice double-cropping, rotating between summer corn and winter wheat.

The NCP is an arid region where about 70% of the rain falls during the period June to September. Winter cropping is mainly dependent on groundwater irrigation. 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 are under serious threat of over exploitation. 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).

Our study area, Lankao county has an area of 1,116 square km (430 sq miles) with a population density of about 780 people per square kilometer in 2021, much higher than the average density in the country (about 147 people per square km), see Figure 1. The endowment of arable land per capita is small, around 0.08 ha. The county is located in Henan province, the province with the highest agricultural production nationally. About 70% of the county's land area is under farming and its local GDP depends heavily on agricultural production. Until 2018, Lankao was in the list of most impoverished counties in all of 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 the irrigation canals that channel some of the river water to farms, Lankao has historically enjoyed 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 frenzied pace (Wang et al., 2006). There has been a rapid expansion of private and public tube wells in the county. During 2000-2010, the government constructed more than 7,200 tube wells to help irrigate over 68,000 ha of arable land in the county and increase agricultural productivity.

Highways in the county China has built a large number of new highways in recent years. Between 1990 and 2006, the country invested roughly $40 billion per year in highway development and completed nearly 45,000 km of new highways. 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 length of 35,000 km (World Bank, 2007). This major development of the Chinese highway system has taken place in two phases: the kick off phase, between 1988 and 1997 and the rapid development phase, starting in 1998.

According to the World Bank (2007), the NTHDP has been planned "strategically" in order to connect the country through high-speed road corridors linking the important centers of economic activity, rather than be based on a detailed economic analysis, considering projected centers of economic growth, traffic growth and distribution. The highway network is rooted in Beijing and connects all the provincial capitals. The NTHDP stipulated that all highway 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) old trading routes. The entire land acquisition, resettlement and rehabilitation process for each of these projects was usually completed within 5 to 6 months.

Lankao county is crossed by five national highways, as shown in Figure 2. The two new arteries going through the county we study are G30 and G1511, which are part of the NTHDP (see Figure 1). Both were constructed between 1998 and 2005. To distinguish them from older highways, we will call these two the "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. Only a stretch of 11.6 km of G30 goes through our county (see Figure 2). That is, roughly 0.2% of the length of the highway traverses our county. Construction of G30 took place during the period 1998-2001.

The other highway G1511 is the offshoot of G30 which goes towards Shandong, as seen in the figure. G30 splits into G1511 slightly to the west of Lankao. G1511 follows a 41.4 km near straight line path entering Lankao county from the west and exiting it from the northeast corner. 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. About 9% of G1511 falls within our county. G1511 was constructed between 2003 and 2005.

The other three national highways, the *old 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.[4] The placement of these highways may not be independent

[4]Highway G106 connects Beijing to Guangzhou City, running from north to south for 2,466 kilometers; G106 was constructed during two phases, in 1956 and in 1988; and Highway G220, which is 585 km long connects Binzhou City in Shandong Province to Zhengzhou in Henan Province. This highway was constructed

of the availability of water and economic activity in the county. They were likely built long time ago and evolved as people settled in the area and economic activity expanded. It is possible that these settlement patterns were guided by the availability of water for farming, with areas with shallow groundwater depths being the most attractive.

3 A Model of Crop Choice

In this section we develop a simple framework to show how the distance to the highway may affect farming decisions, especially water and land use. We consider a farmer located at distance x from the highway who has a plot of unit area and allocates this land to two crops, crop 1 which is a water-intensive, commercial crop (e.g., melon) and crop 2, a subsistence crop (e.g., corn).[5] The farmer is a price taker and the main difference between the two crops is in terms of the price the farmer receives. We assume that melons are shipped to markets farther from the county.[6] If p_1 is the equilibrium price of the crop in the market, the farmer receives a price $p_1 - tx$, where x is the distance of the plot from the highway and t is unit transport cost. In our model, plots are only differentiated by their straight line distance to the highway, and travel on the highway is assumed to be the same for all farms.[7] That is, the farther the farm is from the road, the lower the price received by the farmer. Another way to justify this assumption is that melons are perishable and access to the market is more important for melons than for corn which is not perishable. Longer travel times will decrease the price received by the farmer if a portion of the crop is spoilt.[8] For corn which is either consumed by the household or sold locally, the distance to the highway is relatively less important, so the price received p_2 is assumed to be independent of the distance to the

in the late sixties and establishes the south-east access from Henan Province to Shandong Province. Finally, G310 runs for 1,613 km from Lianyungang in Jiangsu Province to Tianshui in the Gansu Province. This segment was also constructed in two separate phases, in 1949 and in 1978-79.

[5]Melon and corn both require irrigation water, but melon typically uses about 2 inches of water per week while corn uses about half of that.

[6]There is evidence that melons are shipped to markets as far as Beijing and the coastal city of Shenzhen, the latter located 1,600 km away from Lankao, see e.g., https://news.cri.cn/20170328/445f2a1d-e2fe-2046-c7fc-d521cc256327.html

[7]Figures 3 and 4 show photographic evidence that these highways are used frequently for local transport, even though the "official" toll exits are located farther away.

[8]Since the county is small in size and the farthest farm may only be 10-15 km from the highway, it is likely that transport on "last mile" rural roads is more time consuming. It is also possible that the produce is first moved from farms in the interior to locations near the highway from where they are trucked to distant markets, raising the cost of shipping with distance from the highway.

highway.[9]

The production functions for the two crops are given by $y_1(w_1, L_1)$ and $y_2(w_2, L_2)$, where w and L denote water use and the share of the plot allocated to the crop, respectively, while the subscript denotes the crop, 1 (melon) and 2 (corn). Each of these production functions have the usual signs: rising with decreasing returns to scale and globally concave with respect to water and land, i.e., $\frac{\partial y_i}{\partial w_i} > 0$, $\frac{\partial y_i}{\partial L_i} > 0$, $\frac{\partial^2 y_i}{\partial w_i^2} < 0$, $\frac{\partial^2 y_i}{\partial w_i^2} < 0, i = 1, 2$. Concavity implies that $\frac{\partial^2 y_i}{\partial w_i^2}\frac{\partial^2 y_i}{\partial L_i^2} - [\frac{\partial^2 y_i}{\partial w_i \partial L_i}]^2 > 0$, $i = 1, 2$. As done by other studies, we assume that land and water are complements, i.e., $\frac{\partial^2 y_i}{\partial w_i \partial L_i} > 0$, $i = 1, 2$. We further assume that for the second crop corn, land and water are weakly complementary, i.e., that is, $\frac{\partial^2 y_2}{\partial w_2 \partial L_2}$ is small.[10]

Let c be the unit cost of water. The depth of the aquifer is assumed to be the same at all locations, as in a bathtub. The farmer at location x takes prices as given and allocates land and water to each crop to maximize profits as follows:

$$\max_{w_1, w_2, L_1, L_2} \Pi = (p_1 - tx)y_1(w_1, L_1) - cw_1 + p_2 y_2(w_2, L_2) - cw_2 \tag{1}$$

subject to the constraint $L_1 + L_2 \leq \bar{L}$, which can be substituted into the objective function to get

$$\max_{w_1, w_2, L_1} \Pi = (p_1 - tx)y_1(w_1, L_1) - cw_1 + p_2 y_2(w_2, \bar{L} - L_1) - cw_2. \tag{2}$$

Differentiating with respect to w_1, w_2 and L_1 we get the following set of first order conditions:

$$(p_1 - tx)\frac{\partial y_1}{\partial w_1} - c = 0 \tag{3}$$

$$p_2 \frac{\partial y_2}{\partial w_2} - c = 0 \tag{4}$$

$$\textit{and } (p_1 - tx)\frac{\partial y_1}{\partial L_1} - p_2 \frac{\partial y_2}{\partial L_1} = 0. \tag{5}$$

These three conditions can be totally differentiated (see Appendix) to yield comparative statics results with respect to changes in the distance of the plot from the road. Below we

[9] In reality, even the price of corn may be lower with distance from the highway because some corn may also be sold in markets farther away. We neglect the effect of distance on the corn price, to keep the model tractable. Our empirical results will show that the price of wheat, another major staple crop grown mainly in the winter, does decline with distance, but the magnitude is relatively small. The effect on corn prices is indeterminate.

[10] Similar assumptions have been made by other studies, such as Hornbeck and Keskin (2014)

summarize the main results:

Proposition *Both water use and land allocated to the water intensive crop decline away from the highway. Water and land use increases for the less water intensive crop, with distance to the highway. Aggregate water use in plots declines with distance. Higher water use implies that along the extensive margin, more wells will be dug closer to the highway.*

The intuition behind these results is straightforward. The price of the water intensive crop melon, is highest close to the highway. Hence both water and land allocated to it, decline on plots away from the road. Since total land area is the same everywhere, the share of land allocated to corn increases away from the road. Aggregate water use increases closer to the road, because the price of melons has a positive gradient while the price of corn does not change. On each plot, more water is used near the highway. hence *ceteris paribus*, more wells will be drilled closer to the highway than far. In the rest of the paper we will test these predictions with administrative and survey data.[11]

4 Data

We use data on the geographical placement of roads and tube wells. Our database contains information on the location, depth, year of construction, and depth of the water table for all the 12,160 tube wells in Lankao county constructed between 1955 and 2011. The Census measuring the depth of the water table in the wells was only done during 2011. These tube wells are spread over 389 villages in the 17 townships in the county.[12] They were dug either by private individuals or by the government. Figure 5 shows the growth in the number of tube wells by decade from 1955 to 2011.[13] The largest increase in tube wells took place early this century, when the government started to dig new tube wells. From the figure, the increase in their number seems relatively stable before 2000. It shows how in the earlier

[11]Note that the model is simple and for tractability reasons, we have abstained from adding several important features. We have no output market effects. In reality a large shift out of corn to melons may raise the price of corn and make it attractive to produce. This is one reason why we may observe corn being grown in the county. Many farmers may produce corn because it is a subsistence crop and provides food security to the household. We also did not consider other inputs such as fertilizers, pesticides and labor. To the extent that these inputs are complements to water, the predictions for their use may be similar. This issue is discussed again when we consider mechanisms in Section 6.

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

[13]Over time, some wells have stopped working, possibly due to a lack of maintenance or changes in the underlying aquifer hydrology. As long as wells are abandoned in a random fashion in the county (i.e., not following any particular pattern) this does not pose a problem for our estimation strategy.

periods wells were dug in the northwest of the county, an area that is closer to the river and so likely had a higher water table, but later spread to other parts of the county. Our analysis focuses on the period 1980 - 2011, immediately after the introduction of China's Household Responsibility System in the late 1970s, which replaced communal farming and allowed individual households to decide what to cultivate, choose inputs such as water and chemicals, and sell farm output in markets (Lin, 1988).

In order to conduct the extensive margin analysis, we transform our dataset into a panel. We divide the county into cells of 250 m by 250 m, resulting in 18,362 cells.[14] Our database contains information on the date on which each of the wells was dug. The variable of interest is a count variable representing the cumulative number of wells by cell-year. Table 1 reports descriptive statistics for the panel. The number of cells containing zero wells decreases steadily over time and equals 10,691 in 2011 (i.e. 58.2% of the cells in the county), with 7,671 cells containing an average of 1.62 wells. The maximum number of wells observed in a cell in 2011 is 12, yet only 1% of cells contain 5 wells or more. Figures 6 and 7 show the number of wells by year. Figure 6 shows the average number of wells in the 18,362 cells in the county. This number grows from 0.1 in 1980 to roughly 0.7 in 2011. Figure 7 shows the conditional mean – the number of wells in cells which have a well. The blue line indicates that over time cells with non-zero wells increased, while the red line shows the mean number of wells in cells with a positive number of wells, which went up as well. The two figures together show that the number of wells is increasing not only because more wells have been dug in cells already containing wells, but also because wells were dug in cells that previously did not contain any well.

The extensive margin analysis uses the date of digging of the wells in our sample. For the intensive margin, we use data on the depth of the water table, which are prone to measurement error. We use this data only for a subsample of wells. We eliminate all wells in a village if within the village we do not observe any variation in the water table.[15] Secondly, we eliminate all outliers. For instance, if in a village containing 30 wells the depth of the water table ranged between 12 and 16 meters in 29 wells and 1 meter in one well, we eliminate

[14] Several factors played a role in the choice of cell size. Land utilization data for Lankao 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 (e.g., because of pre-existing construction). A smaller grid size will lead to zero inflation in the data, i.e. we will have many cells reporting zero wells, but these will not be real zeros if it is impossible to dig in these cells (e.g., village ponds). We would like to account for potential spatial spillover effects, which suggests a larger grid. The cell size was thus selected to be 250 m square. Our results are robust to a change in the size of the cells to 300 m and 500 m squares, as shown in the Appendix

[15] Since we employ a fixed effects specification, these wells will be dropped anyway.

the latter observation. This process leaves us with 7,526 wells used in the intensive margin analysis. We report descriptive statistics in Table 2.

Note that the mean well depth (41.95 meters) is much larger than the mean depth of the water table in the well (13.92 meters). We 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 and costs. All the data on water table depth was collected in the same year (2011), while well-depth was measured at the time of digging and varies significantly across wells.[16] Figure 8 shows the depth of the water table in the county as measured at the well in 2011. Darker shades correspond to a deeper water table, while lighter shades indicate a shallower depth. Note the general pattern imposed by the seepage from the Yellow River which lies in the north-west. The water table is deeper as we move towards the south-east corner of the county.

Wells on average are located only 9.72 km away from G1511, which crosses almost the full length of the county and 18.77 km away from G30, which is located in the south-west corner of the county. The mean distance of a well from the Yellow River is 22.04 km. Well density is high and, therefore, wells are located close to each other. The mean distance between two neighboring wells is 0.13 km and on average, there are roughly 14 wells in a circle of radius 500 m. Figure 9 summarizes the timeline for construction of the highways and the date of the well census.

5 Empirical Specification and Results

The conceptual framework we developed showed that aggregate water use is higher and hence more wells are likely to be dug on plots closer to the highway. At the extensive margin, we check if the arrival of the roads leads to an increase in the number of new wells drilled in cells near the two highways. Later in the intensive margin analysis, we focus on the level of the water table in the wells in 2011 as a function of the distance from the highways.

[16] We are unable to capture seasonal variation in the water table since we do not know precisely when the depth was measured. However, this problem may be addressed by the village fixed effects specification since it is likely that given the set-up costs of visiting a given village, water levels in wells within a village were measured at the same time.

5.1 Extensive margin analysis

The extensive margin analysis is split in two parts. First, we present a differences-in-differences (DiD) specification in which the construction of each road is considered separately. Second we perform a staggered DiD in which both highways are considered together.

DiD specification

The panel set-up allows us to run a standard DiD specification for each of the two highways of interest. We construct a treatment dummy, called *Treat*. A cell is defined as treated if it lies within a certain radius from the new road (G30 or G1511), and as non-treated if it lies outside the radii of both roads. A second dummy (called *Post*) takes the value zero for years before the first year of construction of G30 or G1511 and one for the years following.[17]

The specification takes the following form

$$Wells_{ct} = \alpha + \delta_c + \delta_t + \beta_1 Treat_{ct} + \beta_2 Post_{ct} + \beta_3 Treat_{ct} * Post_{ct} + \varepsilon_{ct} \quad (6)$$

where $Wells_{ct}$ represents the total number of wells in cell c at time t, and δ_c and δ_t are cell and year fixed effects. 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. These fixed effects also capture the influence of the old highways. The parameter of interest is given by β_3, the coefficient on the interaction term, we expect to be positive (as predicted by the theoretical model), and gives the change in the number of new wells in a treatment cell after the construction of the new highway. We run an OLS specification.[18] We choose to be agnostic with respect to the area of influence of each highway, i.e., the treatment area and, therefore, take a semi-parametric approach in which we allow the treatment area to increase from 500 m around each highway to 10,000 m at 50 m intervals. We prefer this approach rather than using a continuous function of the distance and therefore imposing a linear relationship, since it has the advantage of capturing any structural breaks in the relationship between the number of wells and the distance to the highway.

[17] We run a series of robustness tests by changing the definition of the *Post* variable – starting from two years prior to construction until the road was completed. Results are robust to these changes and are presented in Tables 18 and 19 of the Appendix.

[18] The use of a non linear linking function (such as Poisson or negative binomial) would, for instance, violate the common trends hypothesis, since it does not allow us to correctly compute the DiD estimate.

Extensive margin results

G30: Figure 10 shows the change in the DiD coefficient, β_3, when expanding the treatment area from 500 meters to 10 km at 50 m intervals. After a small increase over the first 4 km, the coefficient becomes stable.[19] Since the impact stabilizes at the 4 km threshold, we show the results for the estimation of equation (6) for a treatment area of 4 km.

The first three columns of Table 3 report 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, *treated* is defined as the area within 4 km of the highway, and *non-treated* is the area outside a 4 km band around each highway (G30 and G1511), see Figure 11 for a graphic representation of the treated area around G30. The area that falls within the radius of G1511 is dropped from the estimation since it cannot be cleanly attributed to either the treated or non-treated area for G30. In columns (2) and (3) we add cell and year fixed effects. In this table the *Post* variable takes value zero for years before 1998 and one after (G30 was built between 1998-2001). The coefficient of interest is positive and statistically significant at the 1% level across all specifications. It is stable across the three OLS specifications. The number of new wells dug in one of the treated cells is 0.055 higher after the beginning of construction of G30, with respect to a non-treated cell. Considering that the average cell contains 0.257 wells, an increase by 0.055 corresponds to an increase in the average number of wells in the treated area by 21.4%.

We run a falsification test for the common trends hypothesis for a treatment area varying between 500 and 4000 m. Results can be found in Figure 12. The coefficients shown in this graph are obtained by regressing 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, i.e., for the period 1980-98.[20] The coefficient on the interaction term is statistically insignificant up to almost a 3 km treatment area, confirming the common trend hypothesis. Beyond the 3 km mark we observe a slight difference (one order of magnitude smaller than the DiD coefficient, roughly 0.002 vs 0.05) between the treated and non-treated areas. Before construction of the highway, there was no statistically significant difference in

[19] The coefficients on the left of the graph are estimated less precisely because only a small number of cells is within a treated area with a radius of 500 m around the highway. At 500 m the number of treated cells is 238, representing only the 1.36% of the total, when we increase the treated area to 4000 meters the number of cells in the treated area increases to 1,540, the 11.38% of the cells in the estimation.

[20] The falsification test takes the following form

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

where t is a linear time trend.

the number of new wells dug in the treatment area relative to control.

A possible threat to our identification arises from the contiguity of the treated and non-treated areas. A higher number of wells in the treated cells could reduce the water available in the aquifer and decrease the incentive to dig new wells in non-treated cells. This would violate SUTVA and bias our estimates. We introduce a buffer area between treated and non-treated cells. Figure 13 shows the coefficient of interest (on the interaction term) when we introduce a buffer the width of which varies between 50 m and 2 km at 50 m intervals, we fix the treatment area at 2 km, so that treatment plus maximum buffer equals 4 km as in the baseline. The introduction of the buffer zone does not affect our results.

G1511: We undertake a similar analysis for highway G1511. Here the *Post* variable takes the value of one starting in 2003. Figure 14 shows the variation of the DiD coefficient for G1511 as we expand the treatment area from 500 m to 10 km at 50 m intervals. The coefficient is consistently positive and statistically significant. Yet, we observe a downward trend which was not observed in the case of G30, suggesting that the growth in the number of new wells declines in treatment relative to control. This decrease may be due to the closeness of G1511 to the pre-existing highways G106 and G220, which may have led to new wells being dug in years before the arrival of G1511. Just like for G30, the coefficient stabilizes around the 4 km mark.

Columns (4) to (6) of Table 3 report results for the DiD analysis for G1511. Even though G1511 is a new highway, its location is very close to two of the pre-existing highways, G106 and G220. Possibly for this reason, the area around G1511 already contains a higher than average number of wells. Again the treatment band for our baseline estimation is set to 4 km around G1511. The non-treated area is the land mass lying outside of the 4 km band around G1511 and G30. The results obtained for G1511 are stable across specifications and similar in magnitude to the ones for G30. After the construction of G1511 the number of new wells dug in the treated area increases by 0.047 wells, statistically significant at the 1% level.[21] Given that the mean number of wells in a cell is 0.276, a coefficient of 0.047 corresponds to an increase of about 17%. The smaller increase found for G1511 is not surprising, given the other highways already present in the vicinity.

Falsification test results for G1511 are based on data until 2003 and shown in Figure 15. The coefficient on the interaction term is statistically insignificant, only for treatment areas narrowly around G1511. Upon expanding the treatment area we notice a small difference

[21] For both highways, results are robust to an increase in cell size from 250 m to 300 m and 500 m. The results are shown in Tables 16 and 17 of the Appendix.

between treated and non-treated, probably due to the presence of pre-existing highways. Again we rule out possible externalities between treated and non-treated cells by introducing a buffer varying from 50 m to 2 km at 50 m intervals. The coefficients are robust to this modification, as shown in Figure 16.

Staggered DiD specification: We analyze the extensive margin using a staggered DiD specification.[22] This methodology can be used because the two highways are relatively far apart and by selecting a 2 km buffer, no cell is going to fall in the treatment area for both. We construct a treatment variable that takes a value of one if the cell falls either within a 2 km band around G1511 or G30. We then construct two separate *Post* variables: $Post^{G30}$ takes a value of one after the construction of G30 (i.e., after 1998) and $Post^{G1511}$ equals one after the construction of G1511 (i.e., after 2003). The full specification takes the form

$$Wells_{ct} = \alpha + \delta_c + \delta_t + \beta_1 Treat_{ct} + \sum_j \gamma_j Post^j_{ct} + \sum_j \varphi_j Treat_{ct} * Post^j_{ct} + \varepsilon_{ct} \tag{7}$$

where $j = G30, G1511$. This specification allows us to contemporaneously measure the impact of the construction of the two highways. The vector of coefficients φ_j identifies the DiD effect for each of the two highways.

Table 4 presents the results for equation (7). The impact of the construction of the two highways is similar in magnitude to the estimates obtained when they are considered separately, 0.038 for G30 and 0.052 for G1511. Both coefficients are statistically significant at the 1% level. Considering the mean value of 0.268 wells per cell, these increases correspond to a 14% and 19.4% increase in wells for G30 and G1511, respectively.

Event study: We need to ensure that we are not capturing a joint effort of the government to promote irrigation by digging more wells and building the highway network. First, in Figure (17) we plot the coefficients of the year dummies from a simple regression of the number of wells on cell and year fixed effects. The figure shows an intensifying drilling effort over time. To check whether this increased drilling effort is related to the new highways, we

[22] We also run the test suggested by de Chaisemartin and D'Haultfœuille (2020) to ensure that our treatment is not heterogeneous over time or across cells. Our estimations do not have a problem of negative weights.

perform an event study analysis of the following form

$$Wells_{ct} = \alpha + \delta_c + \delta_t + \gamma_t \sum_{t=1980}^{2011} (Treat_{ct} \text{ x } D_t) + \varepsilon_{ct}. \quad (8)$$

By interacting the year dummies with the treatment dummies we observe the change in the drilling effort (number of wells) within a year between treated and non-treated cells. These coefficients are reported in Figure 18. Until 1996 there was no significant difference in the number of wells drilled in the treated and non-treated areas, i.e. the increase in drilling was even across the whole county. However a couple of years before G30 was built, there is clearly a bump up in the number of wells dug in the treated area, likely a result of anticipation of the arrival of the highway. The positive trend continues through G30 and G1511 and tapers off after both highways are situated.

Intensive margin analysis

Specification

The intensive margin analysis is based on a cross section and, therefore, we need to be careful about possible confounders that cannot be accounted for by using fixed effects. 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 agriculture or industry. Our goal is to check whether proximity to a road leads to higher water use and therefore, increased depletion of the groundwater table. The underlying mechanism is that a road facilitates access to markets and may lead to more intensive farming practices (see e.g., Donaldson, 2015). In villages far away from the two roads, the cost of transporting inputs and harvested output may be high, leading farmers to grow subsistence crops for local consumption or for the household.[23] However, in areas closer to the roads, farmers may grow commercially viable crops that require a more timely use of inputs which are easier to access, such as fertilizers and pesticides. Farmers may benefit from quicker and cheaper access to markets for their perishable products, and have better information on market forces that affect their operations. These benefits of the highway, in terms of better output prices and lower input costs may lead to more intensive cultivation and hence increased use of

[23] Road quality within the village may be low or paved roads may be non-existent in which case local transport costs would be high.

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 19 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), so we include a quadratic polynomial in our specification. Controlling for village fixed effects allows us to eliminate differences between villages in the form of topography, geographic location, and other village-level, time-invariant characteristics. Decade fixed effects act as an alternative spatial fixed effect capturing variation due to major policy changes.[24]

Our empirical specification then takes the form

$$WT_{iv} = \alpha + \delta_v + \delta_d + \beta_1 G1511_{iv} + \beta_2 G30_{iv} + \beta_3 (G30 * G1511)_{iv} + \beta_4 R_{iv} + \beta_5 R_{iv}^2 + \varepsilon_{iv} \quad (9)$$

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 kilometers, respectively. Finally, δ_v represents village fixed effects, δ_d decade fixed effects and ε_{ivt} is the error term.

Since we are dealing with two separate highways at the same time, we must account for their interaction. The effect of roads on the water table may change if a well is located farther from both roads then if it is located closer to one road but far from the other one. We introduce an interaction term, which accounts for the position of each well in relation to both roads. We expect to observe increased depletion of the water table in wells located in close proximity to either road, hence the marginal effect is expected to be negative, i.e., a negative sign on β_1 and β_2. However, this marginal effect will become less negative as the well moves farther from any given road keeping its distance to the other road constant, hence β_3 is expected to be positive.

Intensive margin results

Table 5 reports the results for our baseline specification. Standard errors are robust and clustered at the village level. In column (1) we introduce the main variables of interest and

[24]The main policies that we would like to take into account are the following: in the 80s, the introduction of the Household Responsibility System, with an initial rental contract period set at not less than 15 years. This led to greater land fragmentation, which could induce more drilling. In the 90s farmers were allowed to rent land, and this policy change may have affected profit incentives. Finally, in the early 2000s, the contract period was extended to at least 30 years.

village fixed effects. In columns (2) and (3) we add the quadratic polynomial taking into account the distance from the river and decadal fixed effects, respectively.

The effect of G30 and G1511 on the water table is negative and statistically significant at least at the 10% level when we control for the distance to the river. 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 cm (roughly a 4.5% decrease with respect to mean water depth); while an increase of 1 km in the distance from G30 leads to a decrease of 46.3 cm (3.3% decline relative to the mean water depth). The coefficient on G1511 is statistically significant at the 10% level, and 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 to one highway is kept constant while the distance to the other is increased, this will decrease the net 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 on the water table in Figure 20. Panel (a) shows the impact of a change in the distance from G1511 while keeping the distance to G30 fixed and panel (b) the effect of a change in the distance from G30 while keeping the distance to G1511 fixed. In each panel, the share of wells at each distance from the road, is shown by the shaded gray histogram. Note 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 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 cm (roughly a 6.1% increase with respect to mean water depth). The coefficient on the quadratic term is negative, implying that the effect of the river recedes with distance, this coefficient being statistically significant at the 5% level.

We run two additional variations of the baseline intensive margin estimations. We generalize the relationship between the water table depth and the distance of the well from the two highways by taking a second order Taylor approximation. Second, we account for the possible endogeneity of road placement by using an instrumental variable.

Second order Taylor approximation

Instead of simply capturing the intensive margin through the two distances and their interactions, we assume a second order approximation of an unspecified non-linear relationship between the distance from G30 and G1511. This implies a specification of the following form

$$WT_{iv} = \alpha + \delta_v + \delta_d + \beta_1 G1511_{iv} + \beta_2 G30_{iv} + \beta_3 G1511^2_{iv} + \beta_4 G30^2_{iv} + \beta_5 (G30 * G1511)_{iv} + \beta_6 R_{iv} + \beta_7 R^2_{iv} + \varepsilon_{iv}. \quad (10)$$

The only difference from equation (9) are the two squared terms for distance. Results from this specification are shown in column (4) of Table 5. The marginal effect of a change in the distance to G30 on the water table depth is given by the expression

$$\frac{\partial WT_{iv}}{\partial G30_{iv}} = \beta_2 + 2\beta_4 G30_{iv} + \beta_5 G1511_{iv} \quad (11)$$

which is 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 km, and the mean distance from G1511, 9.72 km, we compute the average marginal effect, which equals -0.112. This confirms our previous findings, as we move further from G30 the depth of the water table decreases. Using the coefficients on G1511 and the mean values for G30 and G1511, we obtain an average marginal effect equal to -0.068. Again, the water table is shallower as one moves away from G1511.

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 to China as well. Michaels (2008) also suggests that longer highways connecting distant cities are less likely to be influenced by local economic conditions. In our case, the length of the highways passing through Lankao county is quite small relative to their total length (0.3% of G30 traverses our county) hence local factors are likely to be of even less importance. Since the highways were built to connect major cities (World Bank, 2007), we can 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 situated, had the government joined the two cities by a straight line road.

The segment of G30 crossing Lankao was built to connect Kaifeng to Xuzhou, while the segment of G1511 connects Kaifeng to Rizhao (see Figure 21). We connect these cities with straight lines.[25] The figure shows both the real and virtual highways. The distance of each

[25] Alder (2019) computes the optimal placement for G30 and G1511 in this region taking into account land

well to these virtual roads is used to instrument the actual distance from the well to the road.

Table 6 reports the results of this specification. 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. The F-statistic for each individual first stage is above the critical value of 10. Moreover, given the presence of three endogenous variables, we also implement the methodology developed by Sanderson and Windmeijer (2016) to check for weak instruments. We obtain a chi-square value of 10.27 with a respective p-value of 0.0164, confirming that our instruments are not weak.

Robustness

First, we test for within-village differences in the drawdown of the aquifer. A higher density of wells in one area could lead to a higher depletion rate, unrelated to proximity to the highway. Second, we check whether the construction of the road affects the local hydrology in the region. In order to perform this test, we drop from the analysis the wells that are in close proximity to the new road. Third, we drop villages with less than 30 wells, in order to ensure a sufficient level of within-village variation. Fourth, we introduce the "old" roads into the sample, and fifth, we replace the minimum distance to the road with the distance to the actual location of the entrance/exit ramp on the highway. Finally, we also control for spatial correlation in the error term.

Robustness tests are aimed at the intensive margin specification and are presented in Tables 7 and 8. 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 7 controls for well density near each well. The presence of a larger number of neighboring wells may lower the water table because of competitive extraction of ground water by 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 of each well, in columns (2), (3) and (4), followed by distance to the nearest well in column (5). In column (6), we control both 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 buffer zone lowers the water table. Yet, only the result on the 500 m radius is statistically significant at the 10% level.

gradient. His calculations suggest that they both are very close to being straight lines.

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. This coefficient is not statistically significant. When controlling simultaneously for both measures, only the coefficient on the number of wells within the 500-meter radius is statistically significant, because it also captures the information provided by the other measure of well density used (distance to closest well).

In Table 8 we check whether the highway itself may have altered the hydrology of the region and, therefore, the water table. The construction of a major highway with deep underground foundations may have impacted the water levels in the underlying aquifer, leading to a rising water level away from the road (akin to a large object being dropped into a bathtub). In Table 8, we run the baseline estimation excluding the tubewells which are closer to the road, and likely to be impacted by highway construction. After showing the baseline estimation in column (1), in columns (2) to (4) we eliminate all wells within 50 m, 100 m and 500 m from the highway, respectively. Clearly, highway construction does not seem to affect the water table nearby. 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 and 40 tubewells, respectively. These tests sharply reduce 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 9 we introduce distance to the old highways. Figure 2 shows that G220 follows the same path as G1511, so the effects of the two roads may be confounded. When we add both highways in column (2), 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 larger and we lose precision in the estimation. The coefficients on G106 and G310, as expected, are highly statistically insignificant. In column (3) we replace the straight line distance to the highways with the driving distance to the highway off/on ramp. There is one such ramp in the whole county for G30 and three for G1511 as shown in Figure 22. First, we measure the driving distance from each village centroid to each ramp following the fastest route using ArcGIS and Gaode (a popular Chinese driving App), then we add to it the straight line distance from each well to the village centroid. Since entrance and exit points for the same ramp are adjacent to each other, we take their mean location. We choose the shortest such distance from each well to G30 and to G1511. We use these distances in a specification similar to equation (9). The sign and the magnitude of the coefficients are maintained and the estimates, with the exception of the one for G1511, are statistically significant at conventional levels. In column

(4) we control for spatial correlation in the error term, this makes all the coefficients much more statistically significant.

6 Discussion of Mechanisms

Our results on the extensive as well as on the intensive margin clearly point to an increase in groundwater depletion close to the two roads following their construction. This increased depletion could operate through many different channels. In this section, we provide evidence for a few major channels through which depletion may occur.

Agricultural production in Lankao county follows two growing seasons: winter and summer. The two major staple crops grown are wheat and corn. For reasons of food security, the government used to encourage the production of wheat. Wheat is grown mostly during the winter and harvested in early June in Lankao. Farmers have more freedom to grow alternative crops during the summer season.[26] Historically, farmers produce corn in the summer. The arrival of the new roads could push them to switch to more profitable crops in the summer season.

Thus the first channel we examine is whether agricultural diversification or crop-switching is observed during the summer growing season. The new highways, by providing improved and faster access to the market, create incentives for farmers to switch production from corn towards more profitable, but more water-intensive and perishable, commercial crops such as melons and vegetables. The second channel is intensification of farming in the region during the winter growing season. The decrease in transport costs from better market access may induce farmers located closer to the roads to increase their use of inputs such as water, fertilizer and pesticides. A third channel may be increased work opportunities outside the region. Labor from the region may commute or migrate to other areas for work, in which case farmers closer to the highway may decrease their use of labor and increase use of substitute inputs such as capital and machinery. Of course, the opposite may happen too: with increased access, the county may import labor from neighboring regions, especially close to the highways, and labor use closer to the roads may increase. However, this is less likely since Lankao is a poor county and unlikely to successfully compete in larger labor markets

[26] The document Grain Production Plan in Major Production Areas of Henan Province for the period 2008-2020 lists a series of measures to ensure that local grain targets are met. Measures include promoting high-yield seed varieties and advanced cultivation technology, technical extension services to farmers, land consolidation, and developing infrastructure including government-funded well-drilling. See General Office of Henan Provincial Government, 2010. Advice on Implementing The Plan for Developing Key Food Production Areas in Henan Province (No. 114 of Yu Zheng Ban 2010) Zhengzhou, China.

along the highways.

To provide evidence on these channels, we use three different datasets. We conducted a household survey in Lankao county in the summer of 2014 in which we interviewed 300 randomly selected households located in 30 villages.[27] The questionnaire focused on household characteristics and agricultural practices. We collected information on the 839 plots of land cultivated by the survey households. The average plot in our survey measures roughly 0.14 ha (or 2.03 mu, one hectare is about 15 mu). The households interviewed operate a total of 602 wells throughout the county, of which roughly 40% are government owned. The second dataset comes from the Chinese National Bureau of Statistics and contains information on the total agricultural acreage in Lankao County and total annual output of grains and vegetables. Finally, we use satellite images from the United States Geological Survey (USGS), to compute the Normalized Difference Vegetation Index (NDVI) for the county and analyse changes in production patterns.[28]

Crop switching

Do we observe a switch to cash crops on plots closer to the new highways? We address this question from three different angles. First using survey data, then with aggregate county-level data, and, finally, using satellite imagery. Using survey data, we focus on the locations where crops are cultivated. The surveyed households grow 9 different types of crops: wheat, corn, cotton, potatoes, beans, vegetables, melons, peanuts, and garlic. As shown in Table 11, the main crops grown are wheat and corn, accounting for 86% of the plots. Only 14% of the plots grow cash crops, which are more profitable but more water-intensive. Three crops – peanuts, vegetables, and melons, account for 77.2% of the cash crops cultivated in our sample.

The mean plot size for these three cash crops is larger than the mean plot size of wheat or corn fields, 0.187 ha vs 0.145 ha.[29] Peanuts are fundamentally different from vegetables and melons because they are less perishable and can be stored. We expect that farmers located close to the highway will aim to benefit from their proximity to timely transportation access and choose to grow vegetables and melons rather than peanuts. If crop-switching occurs in the summer, then corn, rather than wheat,will be replaced. Table 11 shows that the mean

[27] The number of households used in the analysis is 276 since for 24 households we are not able to obtain information about the identification number of the wells they are using. More information on the survey (such as randomization) can be found in Appendix B.

[28] The NDVI is a reflectance-based vegetation index computed from remote sensing images.

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

distance to the nearest highway is 8.13 km for plots that grow peanuts, while for vegetable and melon plots, it is only 0.82 km and 1.13 km, respectively. The mean distance for corn and wheat plots is 4.94 km and 5.33 km. respectively. These values are all statistically different from one another. The table shows that while 736 plots are cultivated with wheat, only 633 are under corn cultivation. As expected, farmers switch during the summer growing season and not during the winter despite subsidies for growing improved varieties of wheat, corn, cotton, rice and peanuts.[30] These observations seem to show that crop diversification has indeed occurred in Lankao based on closeness to the highway. However, given the cross-sectional nature of our survey data, we cannot determine whether these trends were already in place before the highways arrived.

The aggregate data from the Chinese National Bureau of Statistics tell us a similar story. Unfortunately, data on total output of vegetables is not available, but we have data on total area under vegetable production. Figure 23 shows the area under cash crops in the county. We observe a major expansion in cultivated area before the construction of the highways, which flattens out gradually post-construction. By running a likelihood ratio (Wald) test we can verify that this time series experiences a structural break in 2001.[31]

Finally, we use satellite data from the United States Geological Survey to reconstruct the NDVI for the summer growing season, specifically for the month of August. We only have one observation per year for the summer between 1998 and 2010.[32] The frequency of publicly available images for the Lankao region is relatively low, and therefore, the data points do not represent the exact same date every year. To facilitate accurate comparison across years we ensure that each observation falls within the first two weeks of August. We compute NDVI values for each 150 x 150 m cell in the county (a total of 75,828 cells). Figure 24 shows the NDVI measure for 1998 and 2010. We observe an increase in the number of settlements and their size. The red footprint of the river becomes more defined and we see an increase in vegetation in what appears to have been swamps in 1998, suggesting improved water management near the Yellow River. Figure 25 shows a scatter plot of the NDVI for each cell in the county divided by year. The average summer NDVI seems relatively stable. Since Lankao is mainly an agricultural county, most NDVI values are above zero.

Corn is the traditional summer crop in the county. The NDVI for a mature field of corn

[30] General Office of Lankao County Government, 2013. Notice on the Issuance of the Implementation Plan for Subsidies for Improved Varieties of Wheat, Cotton, Corn, Rice and Peanut in 2013 (No 94 of Lan Zheng Ban 2013). Lankao, China.

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

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

is higher than for a field of vegetables, because of the high density of leaves that characterize corn fields. In fields with vegetables, a higher share of the soil remains visible from above. If crop switching occured after the arrival of the new highways, the NDVI values should decrease. We run a DiD specification on this panel similar to the one used for the extensive margin. Since our satellite data begins in 1998, we are only able to perform this exercise for G1511. We run the following specification:

$$NDVI_{ct} = \varphi + \eta_c + \eta_t + \varphi_1 Treat_{ct} + \varphi_2 Post_{ct} + \varphi_3 Treat_{ct} * Post_{ct} + u_{ct} \tag{12}$$

where η_c and η_t are cell and year fixed effects and $NDVI_{ct}$ represents NDVI in cell c at time t. The DiD coefficient of interest is given by φ_3, which indicates the magnitude of change in NDVI in the treatment area after the construction of the new highway. As in the extensive margin analysis for G1511, we set the boundary for the treatment area at 4 km from the highway.

Table 12 shows the results. Standard errors are two-way clustered at the cell and year level. Columns (1) through (3) show results from a specification without fixed effects, with cell fixed effects and finally with cell and year fixed effects. The value of φ_3 is robust across specifications. After the construction of G1511, NDVI was 2.9% lower in the treated area. This negative coefficient provides additional evidence of the switch towards cash crops that occurred after the construction of the new highway.[33] 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.

We cannot directly link any of the individual pieces of evidence presented above to highway construction. But in the aggregate, to summarize: (i) cash crops are cultivated closer to the new roads, (ii) crop-switching occurred during the summer growing season, (iii) there is a structural break in the time trend of acreage under vegetables in 1998, and (iv) we observe a decrease in NDVI close to the new highway.

Intensification of agriculture

If farmers receive higher output prices or pay lower input costs, we should observe inten-

[33] To ensure that the decrease in NDVI close to the new highway is not due to the highway itself, we re-run the model after dropping all the observations within 150 meters of the highway (size of a cell). The coefficient becomes slightly smaller but stays negative and statistically significant at the 1% level (-0.22).

sification in farming practices close to the new roads. This should take place, especially during the winter growing season, when farmers are less likely to switch toward more profitable crops. We check whether there is variation in (i) crop yields and price and (ii) expenditures on seeds, fertilizer, pesticides and herbicides with distance from the two highways. We use the following specification

$$Y_{pi} = \alpha + \beta_1 G30^p_i + \beta_2 G1511^p_i + \beta_3 (G30^p * G1511^p)_i + \beta_4 X_{pi} + \varepsilon_{pi} \tag{13}$$

where the dependent variable, Y, represents crop yield, output price, expenditures on seeds, fertilizer, herbicides and pesticides, and 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. The vector X contains plot level controls: soil type (clay, loam, sand or other), plot quality (compared to the average plot), plot type (paddy or dry field) and plot slope.[34] $G30^p$ and $G1511^p$ represent the distance from G30 and G1511 in kilometers, and ε_{pi} is the error term. The results of the estimation of equation (13) for crop yields and output price are reported in Table 13 and for input expenditures for wheat and corn in Table 14.

Columns (1) and (2) in Table 13 report results for wheat, and columns (3) and (4) for corn. We observe statistical significance only for the winter crop: wheat. Wheat yields and prices decline as we move away from the two highways. They are statistically significant for G1511, and of the correct sign but noisy for G30. The interaction term is positive and statistically significant for wheat yields. The estimate is small in magnitude, suggesting that the positive effect of the highway becomes less important with distance. The results for corn are not statistically significant, likely because in the summer farmers find it easier to switch crops rather than do intensive production of corn.

While intensification closer to the roads may imply higher application of inputs, the price of these inputs may rise for plots remotely located. Thus the results on input expenditures may go either way. The table shows that expenditure on fertilizers increases with distance to roads, and is statistically significant. The interaction terms are of the correct sign, and significant for fertilizers. If output prices decline with distance from the road, there is little incentive to increase input use. Hence we can conclude that the higher expenditure on fertilizers and pesticides is likely due to these farmers paying a higher price for inputs away from the highway. Table 14 also shows that the price and quantity of seeds does not change

[34] The soil types are defined according to the water retention capacity of soil: sand has the lowest capacity and clay has the highest. For slope of plot, the plot is defined as flat if farmers can use machinery such as seeding machines and tractors on the plot.

with distance to the roads.

Water, capital and labor

Table 15 shows that for both winter wheat and summer corn, water consumption on plots declines with distance. The estimate is only statistically significant for G30. Spending on electricity for irrigation also declines, and both coefficients are statistically significant. The stock of capital (e.g., agricultural machinery) per mu (1 ha = 15 mu) declined with distance to the roads, both coefficients being statistically significant. Annual investment in the last year (2013) before our survey was also higher closer to the roads. Capital investments, including chemical inputs such as fertilizer and pesticides, manure, seeds, hired draft animals, agricultural machinery such as machines for seeding and harvesting, and labor input (family/hired/exchange), are expected to be higher on plots which grow commercial rather than subsistence crops, reflecting complementarities in production. Higher profits from cash crops closer to roads may indicate lower liquidity constraints and higher investment as well. Both total labor use and family labor use was higher in plots closer to the road, significant for G1511, but not for G30. Labor use could go either way. Access to the highway may lead to local labor working outside the county, raising local wages and leading to higher capital use. However, access to markets may lead to higher demand for labor (both family labor and migrant workers) closer to roads, which seems to be the case in our study area.

Estimating Depletion

According to Feng et al. (2013), satellite data suggests that groundwater depletion in the North China Plains averages 2 cm per year. We compute the additional depletion due to the construction of the highway using our intensive margin estimation. We make three simplifying assumptions: *i)* the interaction term is zero, *ii)* the area of influence of a highway is limited to 5 km on either side, and *iii)* depletion occurs at the same rate across the whole North China Plain.

From the above estimate, the baseline depletion rate on the 5 km stretch on one side of a km length of highway is 100,000 m^3 of water per year.[35] For G30, the coefficient of depletion away from G30 is -0.463 (from column 3 of Table 5), meaning that for every km away from the highway the depth of the water table decreases by 46.3 cm over a 10 year time span, between the end of construction of G30 in 2001 and measurement of the water

[35]=0.02 m x 5000 m x 1000 m. Taking into account both sides of the highway will not change the ratios because of symmetry.

table in 2011. Assuming an area of influence of 5 km, the additional depletion due to the road is 231.5 cm.[36] The volume of extra depletion is the triangle moving away from the highway, as shown in Figure 26. For each linear km of highway the extra depletion equals 0.579 million cubic meters per year. Thus G30 seems to generate depletion that is 5.8 times larger than the baseline depletion of 100,000 cubic meters. The standard deviation of the estimate is 0.215, which using an analogous calculation, leads to a depletion of 0.269 million cubic meters. Thus the range of depletion within one standard deviation of the mean is [0.31,0.848] million cubic meters.

How much of crop switching is needed to generate this estimated depletion? A switch from corn to melon increases water consumption by roughly 2.6 cm (1 inch) per week during the melon growing season - the crops are watered for 8 to 10 weeks, with a mean of 9 weeks. If farmers start growing melon over the whole area of influence of G30 (a 5 km band around the road) this would result in an additional water consumption (over the summer season) of 1.17 million cubic meters of water per km of highway.[37] Roughly, a conversion of 49% of the plots from corn to melons would produce this extra depletion.[38] Therefore, in order to observe the estimated depletion with a probability of at least 68%, about 26% to 72% of the plots around G30 need to switch to growing melons.[39]

The coefficient for G1511, while estimated less precisely, is bigger in size. For an additional km away from G1511 the water table decreased by 62.4 cm over the six years between the end of construction of G1511 in 2005 and measurement of the water table in 2011. The 5 km area of influence assumption yields an additional depletion of 312 cm, which (over the 5 km area) leads to a loss of 7.8 million m^3 of water per linear km of highway, or roughly 1.3 million cubic m per year. Thus G1511 generates depletion roughly 13 times higher than average for the North China Plain as a whole. This estimate is much bigger than for G30, although the coefficient for G1511 is noisier.

Do these numbers correspond to what we observe? While we are not able to provide a definitive answer, our survey data provides some support for the depletion rates estimated here. The 5 km bands around the two highways correspond to an area of 116 square kilometres for G30 and 414 square kilometres for G1511 - the 10 km buffer multiplied by the length of the road segments, 41.4 km and 11.6 km for G1511 and G30 respectively. This 530 sq km area represents 47.5% of the total county area of 1,116 square km. In our survey,

[36]=46.3 cm/km x 5 km.

[37]=5000 m x 1000 m x 0.026 m.

[38]=0.579/1.17.

[39]=0.31 and 0.848 divided by 1.17.

cash crops account for about 10% of the total area under cultivation. Since cash crops are cultivated on average about one km away from the two roads (see Table 11) we assume for simplicity that they are all cultivated within the 5 km band. Thus 10% of all plots in the county represent 111.6 sq km, which is 21% of the 5 km band. This number, while noisy, is close to the 26-72% switch in acreage needed to generate the depletion we observe.[40] In summary, if 10% of the plots within the 5 km radius of the roads shifted to cash crops, the depletion caused would be close to the lower end of the range we estimate.

7 Concluding remarks

There is a large literature on infrastructure and economic development but few 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 is one of the first to document this effect by studying the impact of two major national highways on water levels and wells in a small county in the North China Plain. We show that these two national highways led to increased digging of tubewells in this major farming region and depletion in water levels closer to the roads. We perform several robustness checks to account for spatial externalities and instrument the highways by drawing virtual straight line highways joining major cities outside the county. Our results hold in these cases.

We show evidence of agricultural intensification closer to the highways and higher input use and output prices and yields. Farmers closer to the two roads grow more cash crops such as melons and peanuts and less of subsistence crops such as wheat and corn. Investment in water use and agricultural machinery also increases closer to the highway. Using our estimates of the depletion rate of the groundwater table away from the two highways, we estimate that the depletion rate in the county is 6-13 times greater than the average depletion in the entire North China Plain. At least for one of the highways (G30), we estimate that these higher depletion rates can be achieved with an error of one standard deviation if 26-72% of the plots in a 5 km band around the highway switched from the traditional crop (wheat) to producing the commercial crop (melon). These estimates are consistent with our survey findings on the share of plots growing cash crops in the county.

These findings suggest that the cost of resource depletion from building new infrastructure

[40] We can not obtain more precise estimates because our survey was randomized over the whole county and not within treatment and non-treatment areas.

may not be small. Local depletion rates closer to highways may be several times higher. A key question is, whether these effects are large relative to the benefits that accrue from the highway and whether infrastructure building needs to be accompanied by other policy instruments that reduce their adverse impacts on the resource base. In the specific case we study, policies such as Pigouvian taxes or rationing of groundwater could serve to reduce over-pumping of scarce groundwater. More research needs to be done to answer these types of questions.

A key question is whether these findings have external validity. Since the county we study is predominantly agricultural, the results may be relevant for the rest of the North China Plain. Moreover, China is unique in the sense that the geographical mobility of farmers is severely restricted, so the effects of depletion will be stronger than in other contexts, where mobility is greater.

References

Aggarwal, S. (2018). Do Rural Roads Create Pathways Out of Poverty? Evidence from India. *Journal of Development Economics*, 133(C):375–395.

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

Asher, S., Garg, T., and Novosad, P. (2020). The Ecological Impact of Transportation Infrastructure. *The Economic Journal*, 130(629):1173–1199.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110(3):797–823.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 145:102442.

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.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

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.

Hornbeck, R. and Keskin, P. (2014). The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought. *American Economic Journal: Applied Economics*, 6(1):190–219.

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

Lin, J. (1988). The Household Responsibility System in China's Agricultural Reform: A Theoretical and Empirical Study. *Economic Development and Cultural Change*, 36(3):199–224.

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.

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

Sanderson, E. and Windmeijer, F. (2016). A weak Instrument F-test in linear IV Models with Multiple Endogenous Variables. *Journal of Econometrics*, 190(2):212–221.

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.

World Bank (2007). China Study Tour by New Delhi Office Transport Team. Technical report, New Delhi Office Transport Team.

Tables

Table 1: Descriptive statistics – extensive margin

Year	Cells with zero wells	Cells with non-zero wells				
	# cells	# cells	Mean	Std. dev.	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

Note: This table shows descriptive statistics for the extensive margin analysis sample. The sample contains 18,362 cells of 250 x 250 meters between 1981 and 2011. The table shows the evolution through time of the number of cells with and without wells. For the cells with wells, we show the evolution of the average number of wells within each cell.

Table 2: Descriptive statistics – intensive margin

Variable	Mean	Std. Dev.	Min	Max
Water table (m)	13.92	6.07	2.00	40.00
Well depth (m)	41.95	8.33	18.00	250.00
Distance to highway G1511 (km)	9.72	6.42	0.00	25.00
Distance to highway G30 (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 in 100m of well	1.52	0.90	1.00	9.00
Number of wells in 200m of well	3.30	2.03	1.00	17.00
Number of wells in 500m of well	14.24	7.21	1.00	54.00
Distance to the nearest well (km)	0.13	0.08	0.00	1.00
Number of wells per village	44.76	23.28	2.00	108.00

Note: This table shows descriptive statistics for the intensive margin analysis sample. The intensive margin sample – after taking out outliers – contains 7,526 wells across 234 villages. Wells are eliminated if, within a village we do not observe any variation in the water table - this may be due to measurement error. Since our main specification contains village-level fixed effects, these wells would have been dropped from the analysis in any case. We lose data for 155 villages of the original 389 in the county.

Table 3: DiD estimation – Effect of G30 and G1511

	Number of wells in a cell					
	G30			G1511		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat^{G30}	−0.003					
	(0.010)					
Post^{G30}	0.241***	0.241***				
	(0.035)	(0.035)				
$\text{Treat}^{G30}*\text{Post}^{G30}$	0.055***	0.055***	0.055***			
	(0.019)	(0.019)	(0.019)			
Treat^{G1511}				0.081***		
				(0.009)		
Post^{G1511}				0.274***	0.274***	
				(0.041)	(0.041)	
$\text{Treat}^{G1511}*\text{Post}^{G1511}$				0.047***	0.047***	0.047***
				(0.010)	(0.011)	(0.011)
Cell FE	no	yes	yes	no	yes	yes
Year FE	no	no	yes	no	no	yes
Mean dep. var.	0.257	0.257	0.257	0.276	0.276	0.276
Observations	419,678	419,678	419,678	529,046	529,046	529,046

Note: The table shows cell-level regressions on the number of wells in a cell. The variable $Treat^{G30}$ ($Treat^{G1511}$) is a dummy with value 1 if the cell is in the treatment area, defined as a band with a radius of 4 km around G30 (G1511). The non-treated area includes all cells situated at least 4 km away from G1511 and from G30. The variable $Post^{G30}$ ($Post^{G1511}$) is a dummy with value 1 for years following the beginning of the construction of G30 (G1511) in 1998 (2003). The interaction of the two variables captures the DiD effect. Columns 1-3 show results for G30, and columns 4-6 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 4: Staggered DiD estimation – Effect of G30 and G1511

	Number of wells in a cell		
	(1)	(2)	(3)
Treat	0.058***		
	(0.008)		
Post^{G30}	0.138***	0.138***	
	(0.015)	(0.015)	
Post^{G1511}	0.169***	0.169***	
	(0.038)	(0.038)	
Treat*Post^{G30}	0.038***	0.038***	0.038***
	(0.010)	(0.011)	(0.011)
Treat*Post^{G1511}	0.052***	0.052***	0.052***
	(0.012)	(0.014)	(0.014)
Cell FE	no	yes	yes
Year FE	no	no	yes
Mean dep. var.	0.268	0.268	0.268
Observations	569,222	569,222	569,222

Note: The table shows cell-level regressions on the number of wells in a cell. The variable *Treat* is a dummy which equals 1 if the cell is in the treatment area, defined as a band with a radius of 2 km around G30 (G1511). The non-treated area includes all cells situated at least 2 km away from G30 and 2 km from G1511. The variable $Post^{G30}$ is a dummy which equals 1 for years following the beginning of the construction of G30 (1998). The variable $Post^{G1511}$ is a dummy which equals 1 for years following the beginning of the construction of G1511 (2003). The interaction of the two variables captures the DiD effect. 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: Effect of distance of well from G30 and G1511

	Depth of water table			
	(1)	(2)	(3)	(4)
Distance to G1511 (km)	−0.548	−0.628*	−0.624*	−0.991**
	(0.348)	(0.366)	(0.363)	(0.453)
Distance to G30 (km)	−0.442*	−0.466**	−0.463**	0.843**
	(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***	0.034*
	(0.015)	(0.017)	(0.017)	(0.018)
Distance to river		0.848**	0.850**	0.584**
		(0.350)	(0.349)	(0.277)
Distance to river squared		−0.024**	−0.024**	−0.012
		(0.011)	(0.011)	(0.008)
Village FE	yes	yes	yes	yes
Decade FE	no	no	yes	yes
Mean dep. var.	13.92	13.92	13.92	13.92
Observations	7,525	7,525	7,525	7,525

Note: This table shows the baseline results for the intensive margin analysis. In column (1) we regress the water table depth on distance from the two new highways (G30 and G1511), their interaction and village fixed effects. In column (2) we add distance to the Yellow River and its squared term, while in column (3) we add decade fixed effects. Decade fixed effects capture policy changes: in the 1980s, the introduction of the Household Responsibility System, with a rental contract period set at not less than 15 years. This led to greater land fragmentation, which may have increased well drilling. In the 1990s farmers were allowed to rent land, which may have affected profit incentives. In the 2000s, the contract period was extended to at least 30 years. Finally, in column (4), we add the square of the distance to the two highways. All distances are measured in kilometres. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Instrumental variable estimates

	Depth of 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 FE	yes	yes	yes	yes	yes	yes
Decade FE	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

Note: This table shows results for the intensive margin analysis, using instrumental variables. The instrument used is the minimum distance of the well from the virtual straight line joining the two cities connected by the highway. Columns (1), (3) and (5) show the baseline OLS results from Table 5. Columns (2), (4) and (6) show the instrumental variable results. All distances are measured in kilometres. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Estimations controlling for well density

	Depth of water table					
	Baseline	No. of wells			Distance to closest well (B)	Both (A) & (B)
		within 100m of well	within 200m of well	within 500m of well (A)		
	(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 squared (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 radius		0.038				
		(0.033)				
Number of wells in 200m radius			0.010			
			(0.021)			
Number of wells in 500m radius				0.018*		0.018*
				(0.010)		(0.009)
Distance to closest well (km)					−0.365	0.085
					(0.496)	(0.427)
Village FE	yes	yes	yes	yes	yes	yes
Decade FE	yes	yes	yes	yes	yes	yes
Observations	7,525	7,525	7,525	7,525	7,525	7,525

Note: This table shows robustness tests for the intensive margin analysis, where we control for within village variation. Column (1) shows the baseline intensive margin results. In columns 2-4 we account for local pumping externalities by controlling for the number of wells within radii 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. All distances are measured in kilometres. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Estimation without wells close to the highway

	Depth of water table					
	Baseline	Without wells			Without villages with less than	
		within 50m of highway	within 100m of highway	within 500m of highway	30 wells	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 FE	yes	yes	yes	yes	yes	yes
Decade FE	yes	yes	yes	yes	yes	yes
Observations	7,525	7,499	7,482	7,269	4,991	3,873

Note: This table shows robustness tests for the intensive margin analysis: we control for changes in the hydrology of the region and within village variation. Column (1) shows the baseline intensive margin results. In columns 2-4 we eliminate wells that are located within 50m, 100m and 500m from the highway to account for 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 40 wells. All distances are measured in kilometres. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Other roads and actual driving distance of wells to entrance/exit of G30/G1511

	Depth of water table			
	Baseline	Others	Entrance/ exit	Spatial correlation
	(1)	(2)	(3)	(4)
Distance to G1511	−0.624*	−1.416***		−0.624***
	(0.363)	(0.317)		(0.179)
Distance to G30	−0.463**	−0.456		−0.463***
	(0.215)	(0.304)		(0.113)
Distance to G220		0.882***		
		(0.236)		
Distance to G106		−0.044		
		(0.281)		
Distance to G310		−0.055		
		(0.300)		
Dist G30*Dist G1511	0.052***	0.048***		0.052***
	(0.017)	(0.013)		(0.009)
Dist G30 exit			−0.545***	
			(0.128)	
Dist G1511 exit			−0.226	
			(0.216)	
Dist G30 exit*Dist G1511 exit			0.010**	
			(0.005)	
Distance to river	0.850**	0.923***	0.284*	0.850***
	(0.349)	(0.242)	(0.171)	(0.194)
Distance to river squared	−0.024**	−0.022***	−0.005	−0.024***
	(0.011)	(0.007)	(0.004)	(0.006)
Village FE	yes	yes	yes	yes
Decade FE	yes	yes	yes	yes
Observations	7,525	7,525	7,525	7,525

Note: This table shows robustness tests for the intensive margin analysis, where we control for older highways, formal entrance/exits along the highways and spatial correlation in the error terms. Column (1) shows 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 closest entrance/exit of G30 (G1511). The distance is computed using ArcGIS and Gaode, a Chinese version of Google Maps, following the actual route between each village centroid to the closest highway entrance/exit, to which we add the straight line distance between each well and the village centroid. In column (4) we control for spatial correlation in the error term. All distances are measured in kilometres. All regressions contain a constant. Standard errors in parentheses are robust and clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Survey – plots

	Area (mu)	**Share (%)**
Winter:		
Wheat	1593.08	99.04
Vegetables	7.2	0.44
Other cash crops	8.2	0.51
Summer:		
Corn	1317.41	75.72
Peanuts	245.05	14.08
Vegetables	39.7	1.71
Melons	76.6	4.40
Other cash crops	71.1	4.10

Note: 1 ha = 15 mu. This table shows descriptive statistics for the crops grown on different plots during the two growing seasons. The data shown are from our survey of 2014. The variable *Other cash crops* represents cotton, potatoes, beans, oil seeds, apples, peaches, poplar, pears and garlic.

Table 11: Survey – descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Households					
Household size	276	4.86	1.96	1	14
Age of household head	276	53.70	10.99	25	83
Plots per household	276	3.35	1.90	1	15
Plot size (mu)	759	2.15	1.63	0.09	13
Wheat plots					
Wheat yield (kg/mu)	734	828.55	189.38	33	1300
Wheat price (yuan/kg)	593	1.14	0.13	0	1.33
Expenditure on seeds (yuan/mu)	731	156.12	127.32	0	2,196.15
Price of seed (yuan/jin)	727	4.21	4.04	0	61.33
Fertilizer use (yuan/mu)	727	283.54	128.20	22	1,077
Pesticide use (yuan/mu)	735	43.16	47.77	0	600
Herbicide use (yuan/mu)	735	25.45	17.85	0	140
Water consumption (m^3*/mu)*	736	296.92	256.78	22.5	1,642.94
Electricity expenditure	736	98.53	161.44	0	2,100
Agricultural capital stock (yuan)	704	292.87	873.91	7.5	10,040
Investment during 2013	704	188.1	1,306.83	0	15,000
Total labor use (days/mu)	730	19.19	21.58	0.75	333.33
Family labor use (days/mu)	726	19.07	21.68	0.75	333.33
Corn plots					
Corn yield (kg/mu)	631	928.27	206.85	0	1500
Corn price (yuan/kg)	590	1.05	0.09	0	1.9
Expenditure on seeds(yuan/mu)	630	152.06	113.89	0	2,196.15
Price of seed (yuan/jin)	629	4.17	3.38	0	48.8
Fertilizer use (yuan/mu)	625	282.32	124.03	24	1077
Pesticide use (yuan/mu)	632	40.56	40.37	0	360
Herbicide use (yuan/mu)	632	25.97	17.76	0	140
Water consumption (m^3*/mu)*	633	286.3	251.31	25	1,642.94
Electricity expenditure	633	102.15	169.75	0	2,100
Agricultural capital stock (yuan)	605	278.44	790.72	7.5	10,040
Investment during 2013	605	158.84	1,167.93	0	15,000
Total labor (days/mu)	627	19	21.86	0.8	333.33
Family labor (days/mu)	623	18.86	21.98	0.8	333.33
Distance to nearest highway (G30 or G1511)					
Wheat	736	5.327	4.425	0.028	17.840
Corn	633	4.942	3.883	0.028	17.533
Peanuts	133	8.126	5.553	0.047	17.84
Vegetables	16	0.816	0.453	0.250	1.858
Melons	23	1.130	1.308	0.250	6.876

Note: 1 ha = 15 mu and 1 jin = 0.5 kg. This table shows descriptive statistics for the variables collected during our survey in 2014. The sample contains 602 wells used by 276 households, which irrigate 759 plots of land in 30 villages. The distances from the nearest highway for melons, vegetables, corn and wheat are all statistically different with a confidence of 0.1%.

Table 12: DiD estimation for the effect of the construction of G1511 on NDVI

	NDVI			
	OLS			Common trend
	(1)	(2)	(3)	(4)
Treat	0.031***			
	(0.006)			
Post	0.057	0.057		
	(0.089)	(0.089)		
Treat*Post	−0.029*	−0.029*	−0.029*	
	(0.014)	(0.014)	(0.014)	
Trend*Post				0.008
				(0.003)
Grid cell FE	no	yes	yes	yes
Year FE	no	no	yes	yes
Observations	602,470	602,470	602,470	240,988

Note: This table shows results from the regression of Normalized Difference Vegetation Index (NDVI) on the arrival of the new highways. NDVI ranges between -1 (no vegetation) and 1 (maximum vegetation cover). The variable *Treat* is a dummy with value 1 if the cell is in the treatment area, defined as a band with a radius of 4 km around G1511. The non-treated area includes all cells situated at least 4 km away from G1511 and from G30. The variable *Post* is a dummy with value 1 for years following the beginning of the construction of G1511 (2003). The interaction of the two variables captures the DiD effect. Columns 1-3 run 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: Yield and price for wheat and corn

	Wheat		Corn	
	yield	price	yield	price
	(1)	(2)	(3)	(4)
Distance to G1511 (km)	-19.599^{***}	-0.009^{**}	-5.712	0.003
	(6.043)	(0.004)	(8.891)	(0.003)
Distance to G30 (km)	-0.342	0.0001	0.418	-0.0003
	(1.516)	(0.001)	(1.798)	(0.0005)
Dist G30*Dist G1511	0.0004^{*}	0.00001	-0.00004	-0.00001
	(0.0002)	(0.00001)	(0.0003)	(0.00001)
Controls	yes	yes	yes	yes
Observations	585	463	543	510

Note: This table shows prices and yield for wheat and corn with distance as the plots from the highway. Wheat is a winter crop, grown between October and May; while corn is a summer crop, grown June-September. Yield is expressed in jin (1 jin equals 0.5 kg). Prices are expressed in yuan/jin. Distances are expressed in kilometres. Controls include soil quality, plot slope, plot type and plot quality. Farmers were asked about the quantity of output they keep, the quantity they sell and the total quantity produced. For 263 observations, the numbers reported were not consistent, they were dropped. Standard errors in parentheses are clustered by well. *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Input costs per mu

Winter wheat

	Cost of Seeds	Seed quantity	Expenditure on Fertilizer	Expenditure on Pesticide	Expenditure on Herbicide
	(1)	(2)	(3)	(4)	(5)
Distance to G1511 (km)	0.090	5.047	6.987*	0.591	0.376
	(0.221)	(7.017)	(4.101)	(1.288)	(0.688)
Distance to G30 (km)	0.030	0.719	2.438**	0.468	0.199
	(0.036)	(1.147)	(0.977)	(0.394)	(0.174)
Dist G30*Dist G1511	−0.00001	−0.00022	−0.00032**	−0.00008	−0.00004
	(0.00001)	(0.00022)	(0.00015)	(0.00005)	(0.00002)
Controls	yes	yes	yes	yes	yes
Observations	727	731	727	735	735

Summer corn

	Cost of Seeds	Seed quantity	Expenditure on Fertilizer	Expenditure on Pesticide	Expenditure on Herbicide
	(6)	(7)	(8)	(9)	(10)
Distance to G1511 (km)	0.041	4.142	6.169	0.593	0.496
	(0.261)	(8.238)	(4.397)	(1.429)	(0.798)
Distance to G30 (km)	−0.004	0.276	1.8269*	0.259	0.287
	(0.040)	(1.263)	(1.021)	(0.372)	(0.197)
Dist G30*Dist G1511	−0.00001	−0.00015	−0.00029*	−0.00007	−0.00005*
	(0.00001)	(0.00025)	(0.00016)	(0.00005)	(0.00003)
Controls	yes	yes	yes	yes	yes
Observations	629	630	625	632	632

Note: Costs are expressed in yuan per mu (1 ha = 15 mu). This table shows seed expenditures and quantity, expenditure on fertilizer, pesticide and herbicide for wheat and corn, with distance from the highway. Wheat is a winter crop, grown October-May; corn is a summer crop, grown June-September. Controls include soil quality, plot slope, plot type and plot quality. Distances are expressed in kilometers. Standard errors in parentheses are clustered by well. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 15: Water use, electricity expenditure, agricultural capital stock and labor use

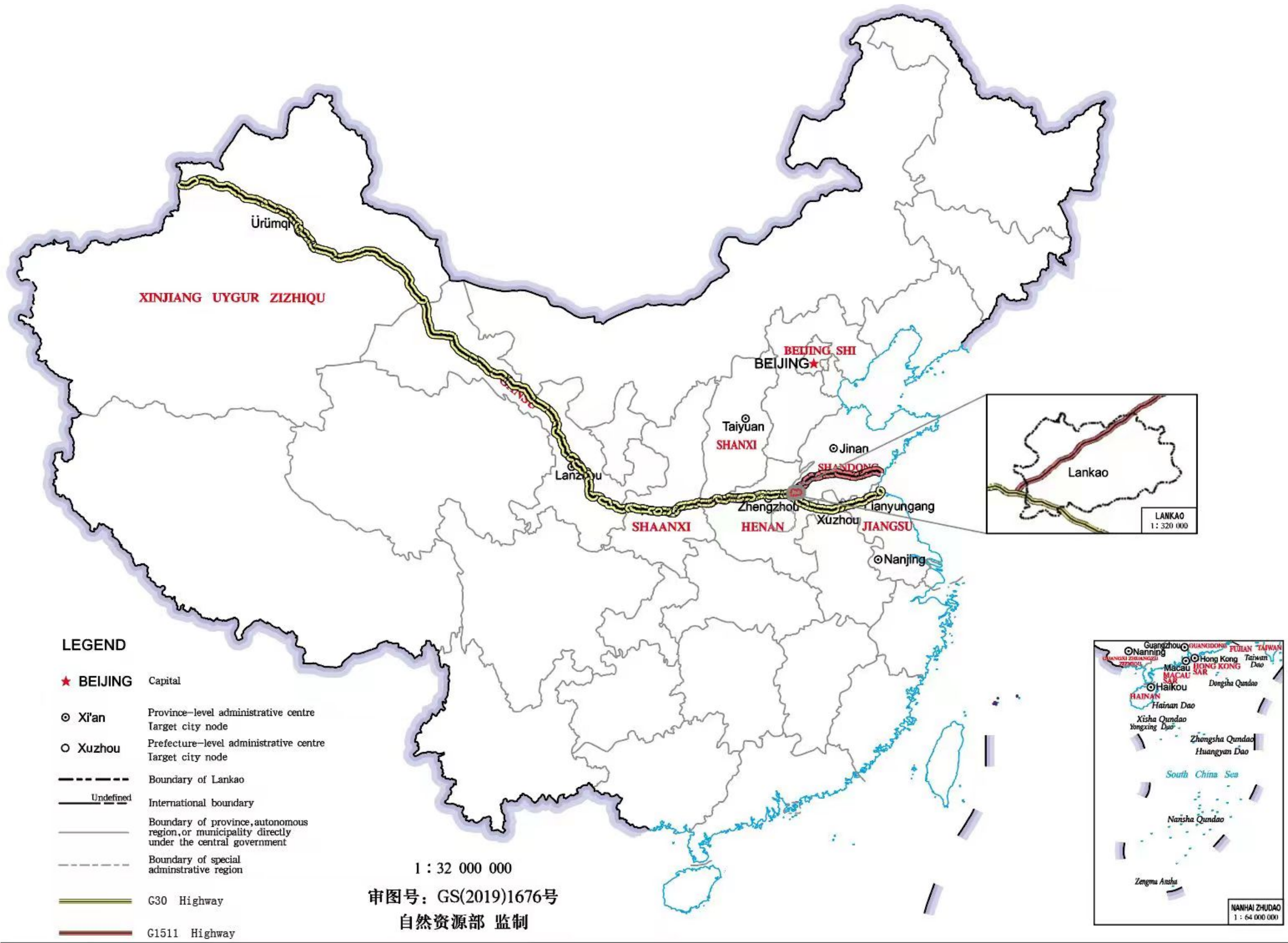
	Winter wheat					
	Water consumption	Electricity cost	Capital Stock	Investment 2013	Total labor	Family labor
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to G1511 (km)	−13.44	−12.08**	−34.59***	−25.52	−1.88*	−1.89*
	(9.37)	(5.43)	(12.74)	(28.93)	(1.04)	(1.04)
Distance to G30 (km)	−10.34***	−3.36***	−15.99**	−13.88*	−0.29	−0.29
	(2.26)	(0.90)	(7.33)	(8.09)	(0.23)	(0.23)
Dist G30*Dist G1511	0.0005	0.0002	0.001***	0.0006	0.00005	0.00005
	(0.00032)	(0.00015)	(0.00042)	(0.00088)	(0.00003)	(0.00003)
Controls	yes	yes	yes	yes	yes	yes
Observations	736	736	704	704	730	726

	Summer corn					
	Water consumption	Electricity cost	Capital Stock	Investment 2013	Total labor	Family labor
	(7)	(8)	(9)	(10)	(11)	(12)
Distance to G1511 (km)	−5.03	−10.98*	−37.64***	−39.46	−2.64**	−2.66**
	(10.41)	(6.30)	(12.60)	(28.16)	(1.21)	(1.21)
Distance to G30 (km)	−7.86***	−2.70***	−10.88	−9.88	−0.48*	−0.49*
	(2.38)	(0.99)	(6.93)	(7.54)	(0.26)	(0.26)
Dist G30*Dist G1511	0.00003	0.00014	0.00122***	0.00091	0.00008*	0.00008*
	(0.00036)	(0.00017)	(0.00044)	(0.00077)	(0.00004)	(0.00004)
Controls	yes	yes	yes	yes	yes	yes
Observations	633	633	605	605	627	623

Note: Costs are expressed in yuan per mu (1 ha = 15 mu). This table shows water consumption, spending on electricity, capital stock and investment, total labor and family labor use for wheat and corn with distance as the plots from the highway. Wheat is a winter crop, grown during October-May; while corn is a summer crop, grown during June - September. Controls include soil quality, plot slope, plot type and plot quality. Distances are expressed in kilometers. Standard errors in parentheses are clustered by well. *** p<0.01, ** p<0.05, * p<0.1.

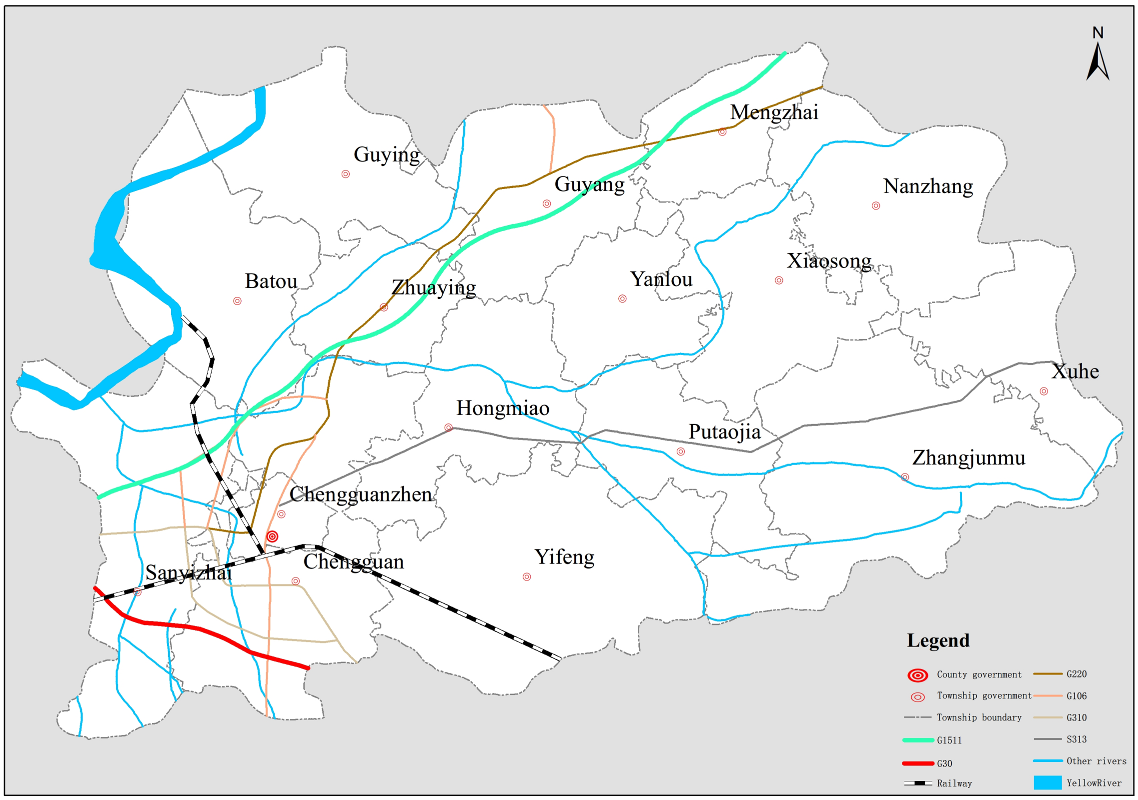
Figures

Figure 1: Location of Lankao county



Note: The two highways of interest, G30 (in red) and G1511 (in green) are shown, with the detailed county map on the right. Source: The base map of China was obtained from the Ministry of Natural Resources of the People's Republic of China website (http://bzdt.ch.mnr.gov.cn/browse.html?picId=%224o28b0625501ad13015501ad2bfc0480%22 accessed on 20 June 2022), Drawing Review No. GS(2020)4619. No modification has been made to the base map.

Figure 2: Map of Lankao county



Note: G30 is shown in red and G1511 in green. The Yellow River is seen in the North-West corner of the county. The canals, train lines and the three older highways are seen as well.

Figure 3: Informal access to G1511

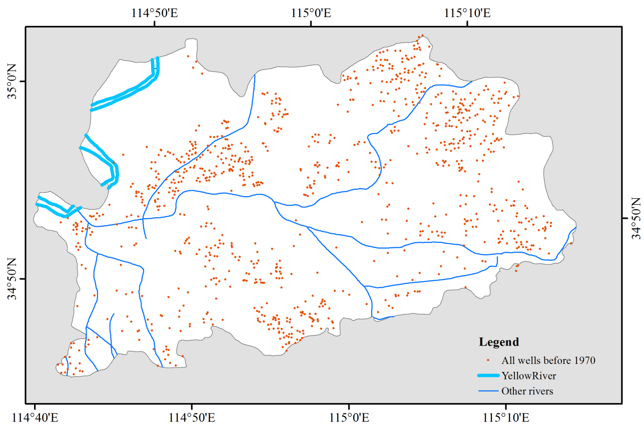
Note: This picture (taken by the authors) shows a typical informal entrance/exit into the highway. There is clear evidence of vehicular traffic through these access points.

Figure 4: Informal access to G30

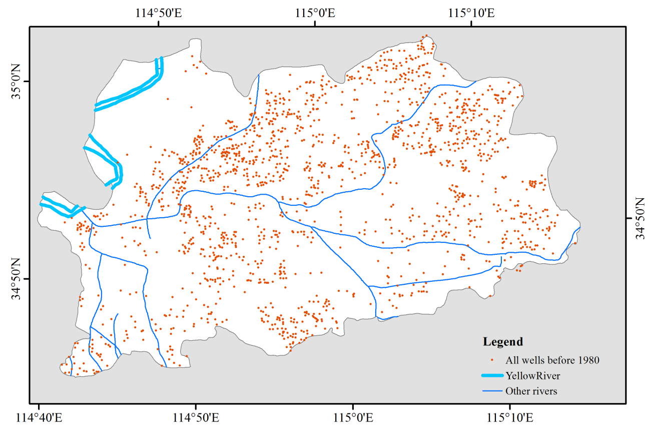
Note: This Google Maps photo shows another example of an informal entrance/exit point. The white arrow (added by the authors) suggests vehicular traffic.

Figure 5: Growth of tubewells (by decade), 1955-2011

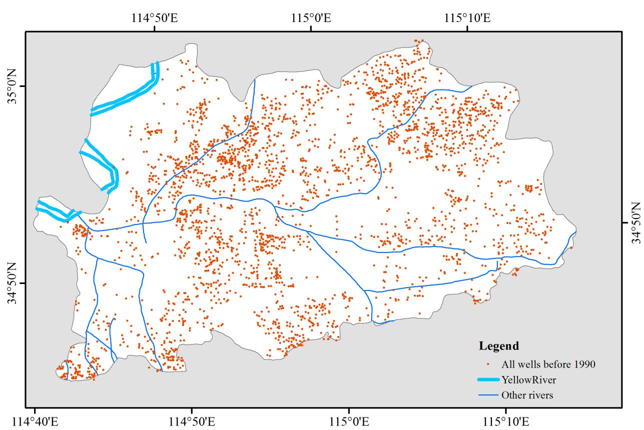
(a) Until 1970



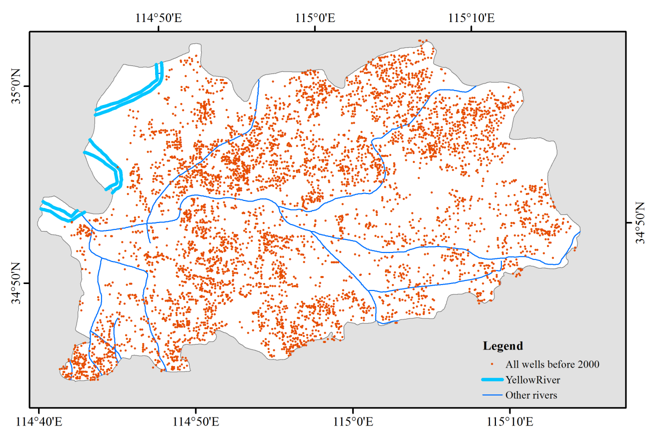
(b) Until 1980



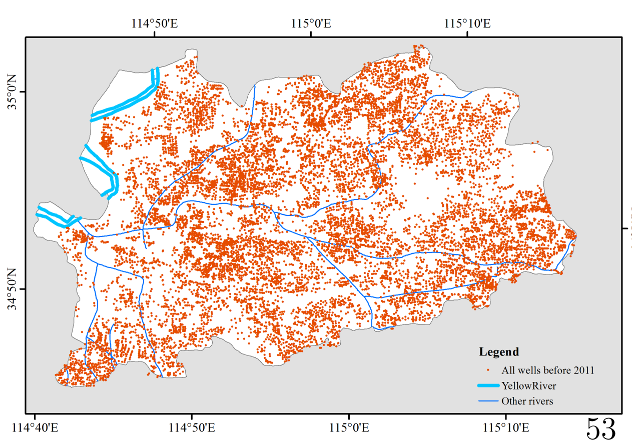
(c) Until 1990



(d) Until 2000

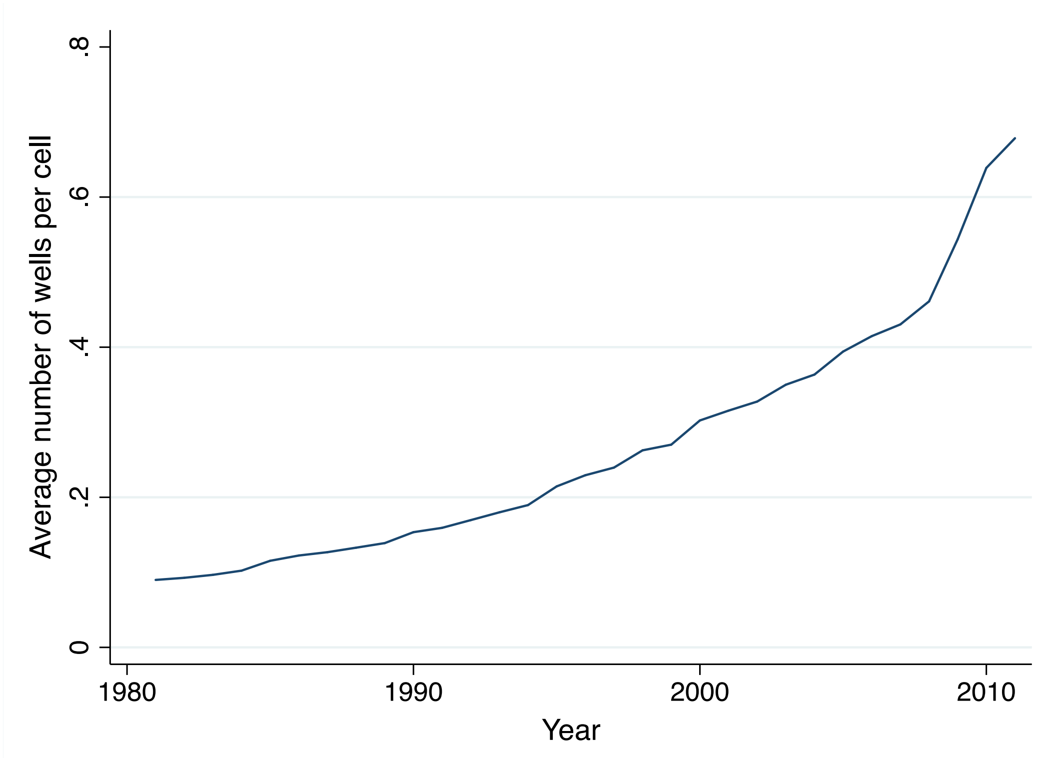


(e) Until 2011



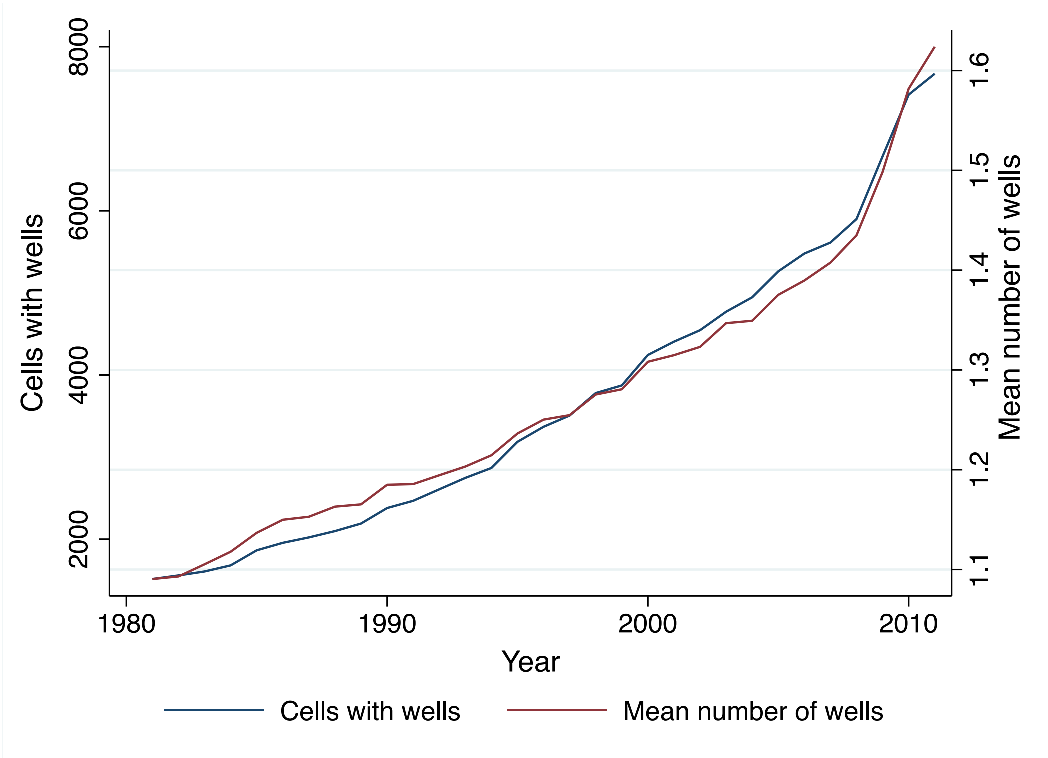
Note: The growth of tubewells in Lankao county (each orange dot represent one tubewell). Panel (a) shows all tubewells that were dug until 1970, panel (b) until 1980, panel (c) until 1990, panel (d) until 2000, and finally panel (e) until 2011, i.e. all the tubewells included in our dataset.

Figure 6: Mean number of wells per cell, 1980-2011



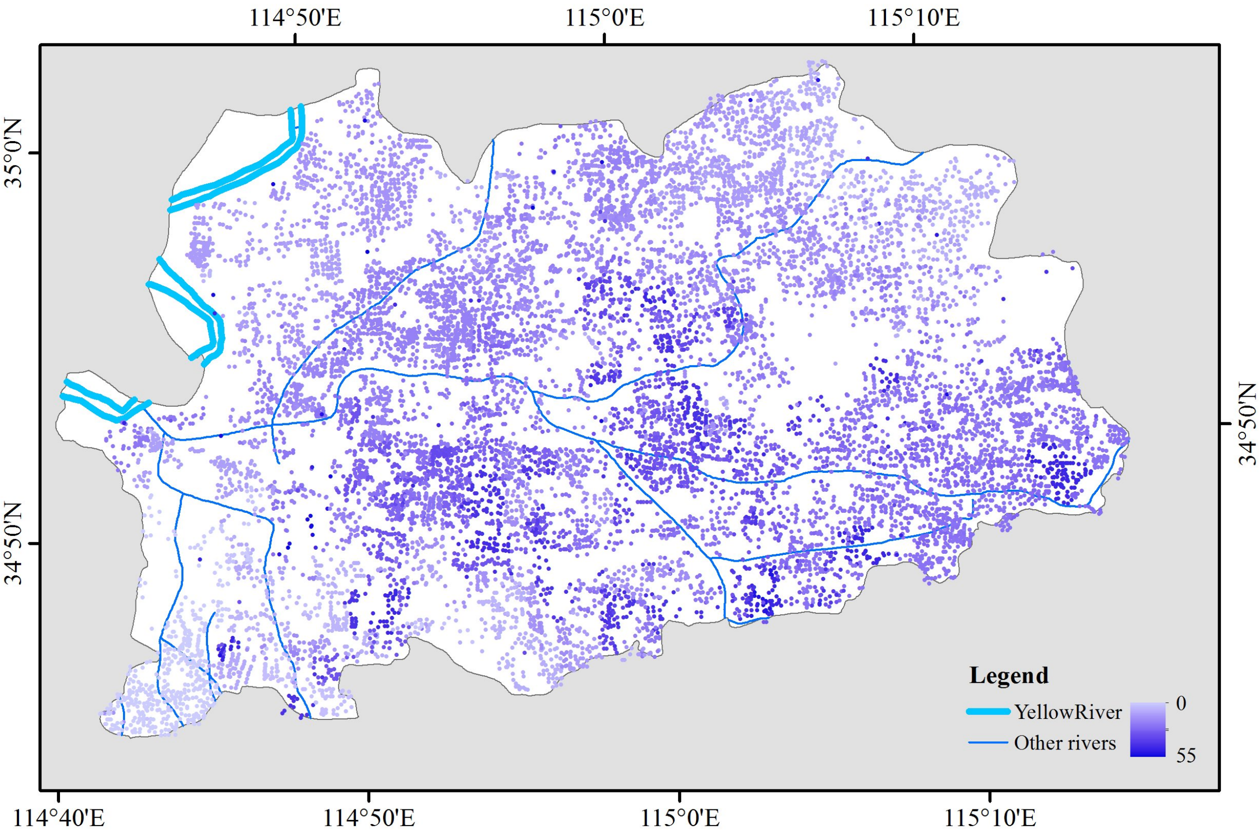
Note: Mean number of wells in the 18,362 cells (250 x 250 meters) in Lankao county by year, between 1981 and 2011. The mean increases from 0.1 in 1980 to about 0.7 in 2011.

Figure 7: Number of cells with wells, 1980-2011



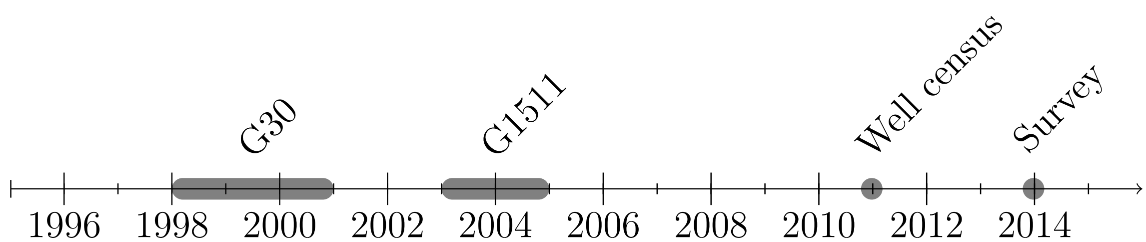
Note: The blue line shows the number of cells with wells by year, while the red line shows mean number of wells in cells conditional on having a well.

Figure 8: Depth of water table in tubewells in 2011



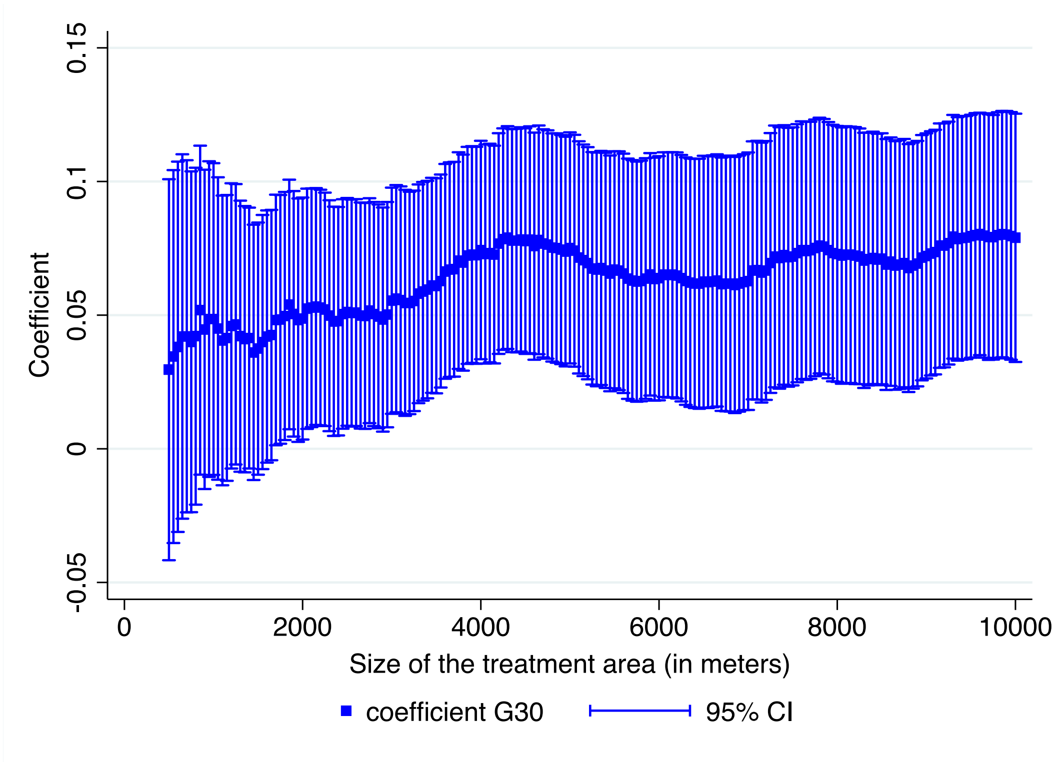
Note: This map shows the level of the water table in each of the 12,160 wells in 2011. A lighter shade of blue indicates that the water table is higher (i.e. closer to the surface), while a darker shade denotes a deeper water table.

Figure 9: Timeline of highway construction and measurement of the water table



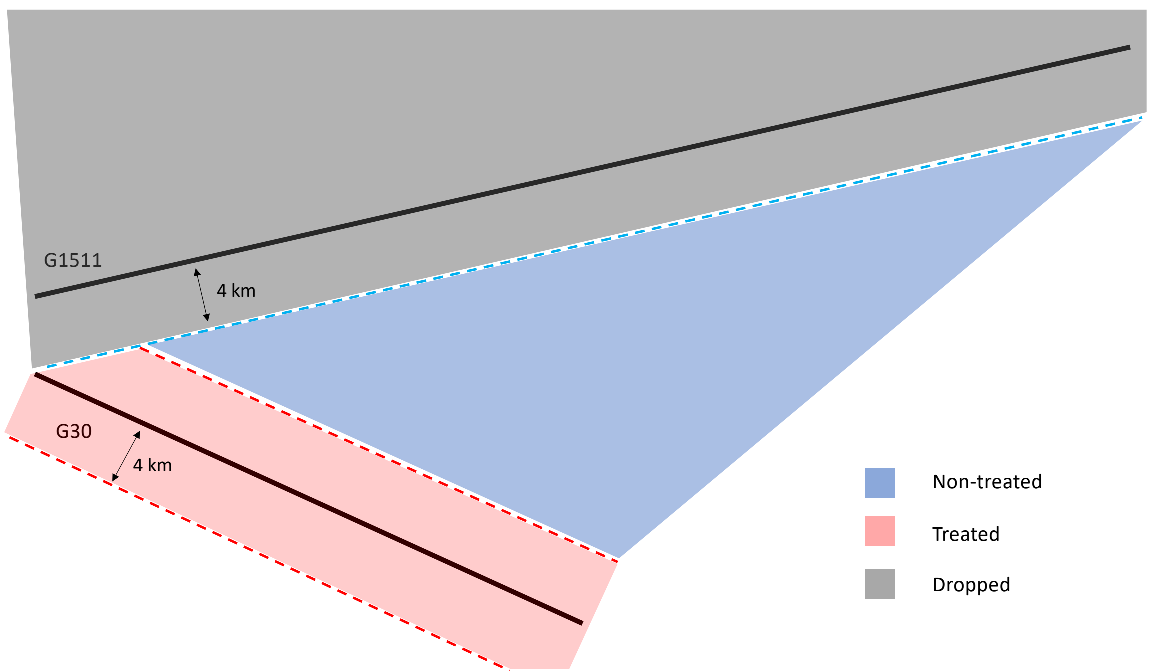
Note: "G30" denotes the length of construction of highway G30, between 1998 and 2001. "G1511" is the length of construction of highway G1511, between 2003 and 2005. Finally, "Well census" represents the year when the government measured the depth of the water table in the county and finally, "Survey" is the year in which we interviewed 300 resident households.

Figure 10: Variation of the DID coefficient as the boundary of the treatment moves away from G30



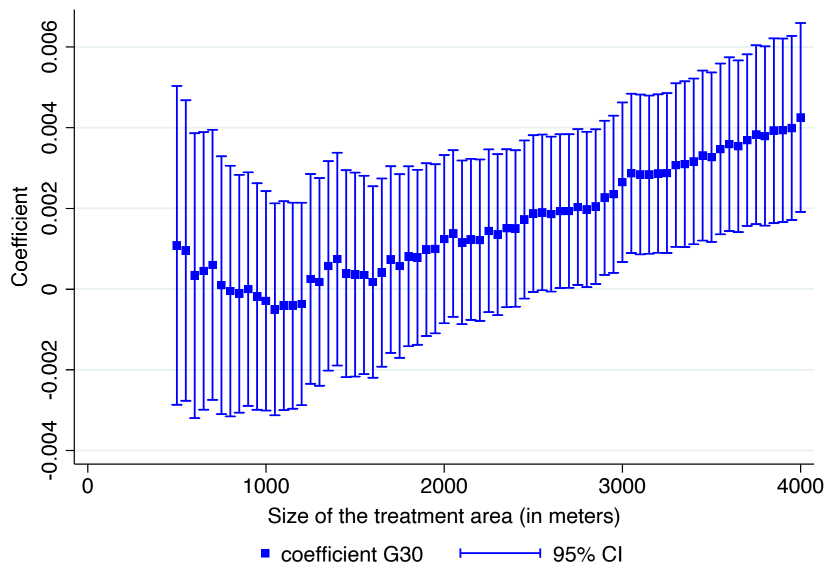
Note: Each dot represents one estimated DiD coefficient with its 95% confidence band. In the first regression the treatment is defined as a 500 meter radius around the highway. The treatment band is progressively increased until it covers a radius of 10 km on each side of the highway. The coefficients on the left of the graph are estimated less precisely because only a small number of cells is within a treated area with a radius of 500 m around the highway. At 500 m the number of treated cells is 238, representing only the 1.36% of the total, when we increase the treated area to 4000 meters the number of cells in the treated area increases to 1,540, the 11.38% of the cells in the estimation.

Figure 11: G30 – treated and non-treated areas



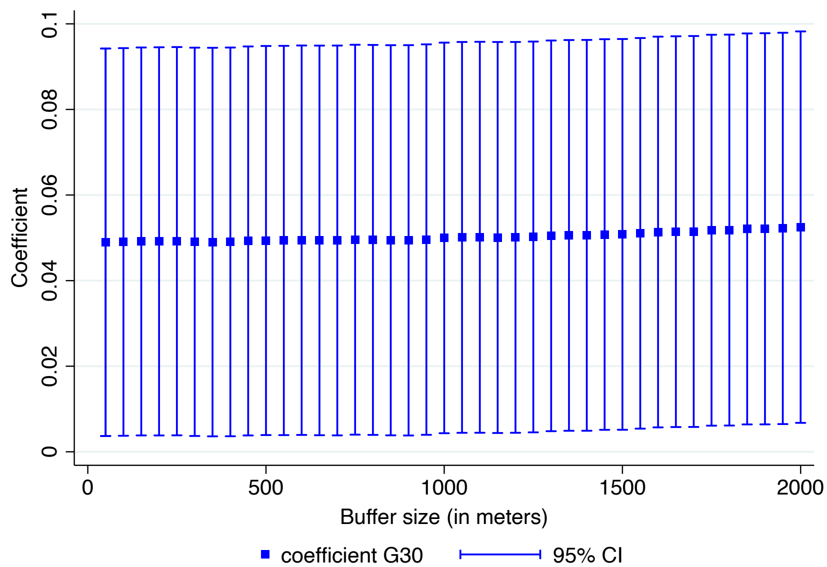
Note: Schematic of the areas defined as treated, non-treated and excluded from the regression. The red and the blue areas change as we change the size of the treatment area.

Figure 12: Variation of the falsification coefficient as the boundary of the treatment moves away from G30



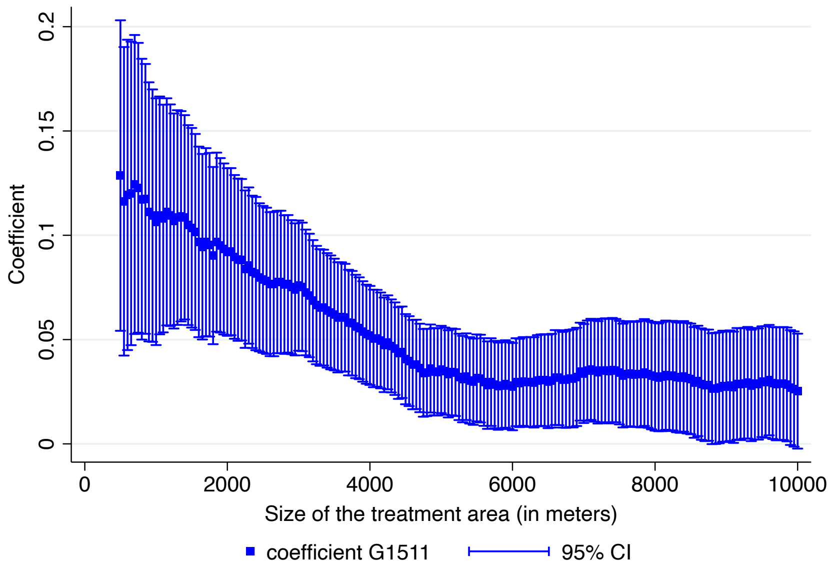
Note: Each dot represents one estimated coefficient for the falsification test with its 95% confidence band. In the first regression the treatment is defined as a 500 meter radius around the highway, which is then progressively increased.

Figure 13: Variation of the DID coefficient as the buffer between treatment and control increases – G30



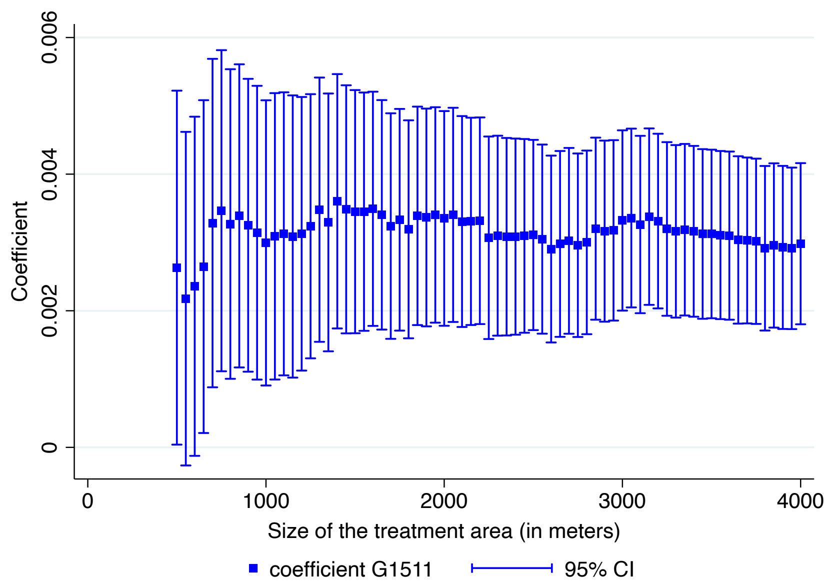
Note: Each dot represents one estimated DiD coefficient with its 95% confidence band. In the first regression we introduce a 50 meter buffer between the treated and the non-treated area on each side of the road, which is then progressively increased to 2 km on each side.

Figure 14: Variation of the DID coefficient as the boundary of the treatment moves away from G1511



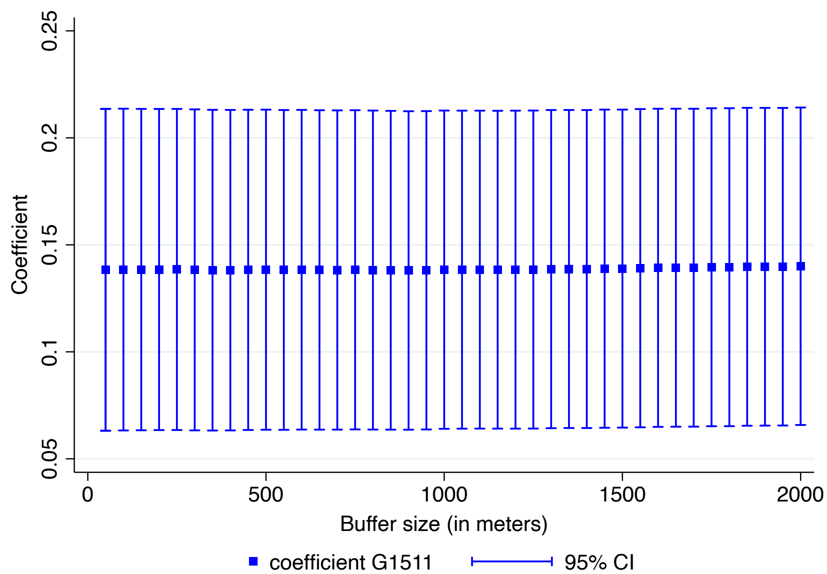
Note: Each dot represents one estimated DiD coefficient with its 95% confidence band. In the first regression the treatment is defined as a 500 meters radius around the highways, we then progressively increase the treatment and, for the last estimation, the treatment area covers 10 km on each side of the highway. The coefficients on the left of the graph are estimated less precisely because only a small number of cells is within a treated area with a radius of 500 m around the highway. At 500 m the number of treated cells is 866, representing only the 4.78% of the total, when we increase the treated area to 4000 meters the number of cells in the treated area increases to 5,068, the 29.7% of the cells in the estimation.

Figure 15: Variation of the falsification coefficient as the boundary of the treatment moves away from G1511



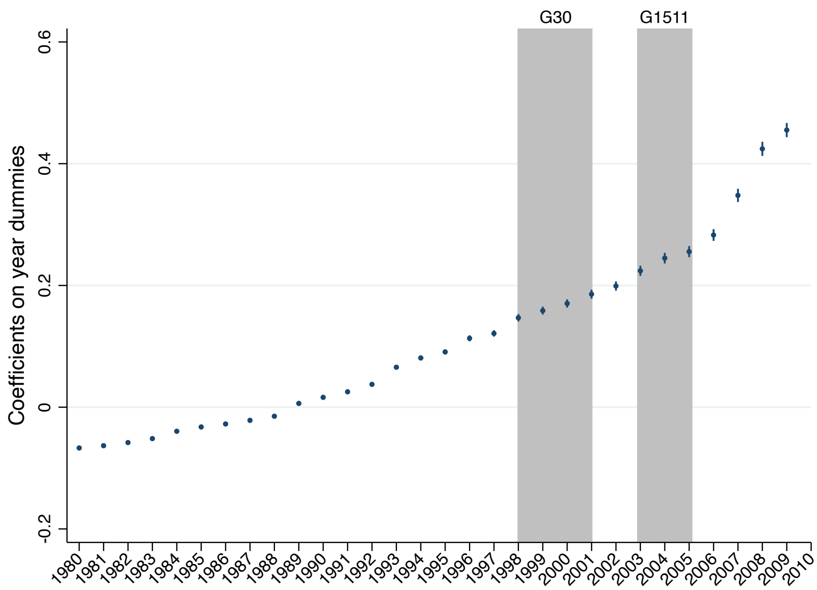
Note: Each dot represents one estimated coefficient for the falsification test with its 95% confidence band. In the first regression the treatment is defined as a 500 meter radius around the highway, which is then gradually increased to 4 km.

Figure 16: Variation of the DID coefficient as the buffer between treatment and control increases – G1511



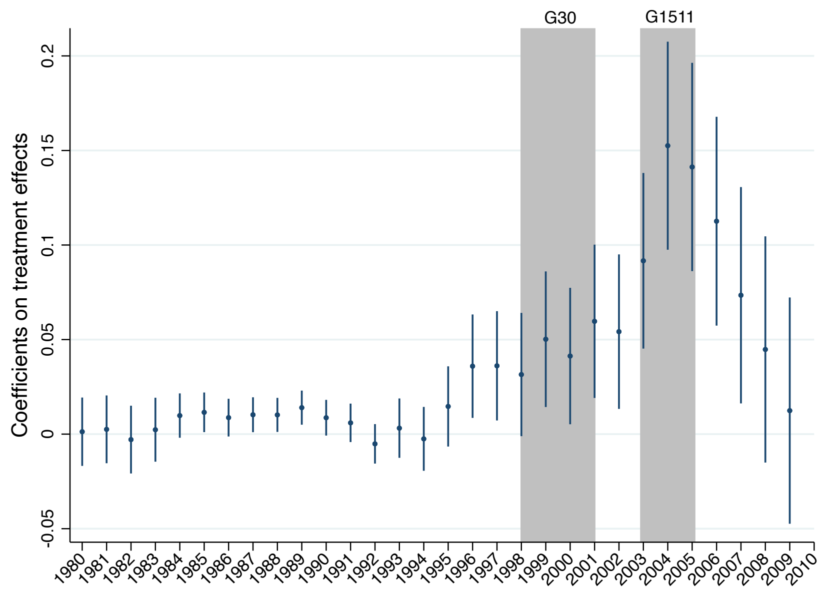
Note: Each dot represents one estimated DiD coefficient with its 95% confidence band. In the first regression we introduce a 50 meter buffer between the treated and the non-treated areas on each side of the roads, increasing the size of the buffer to 2 km on each side.

Figure 17: Coefficients from number of wells regressed on year dummies



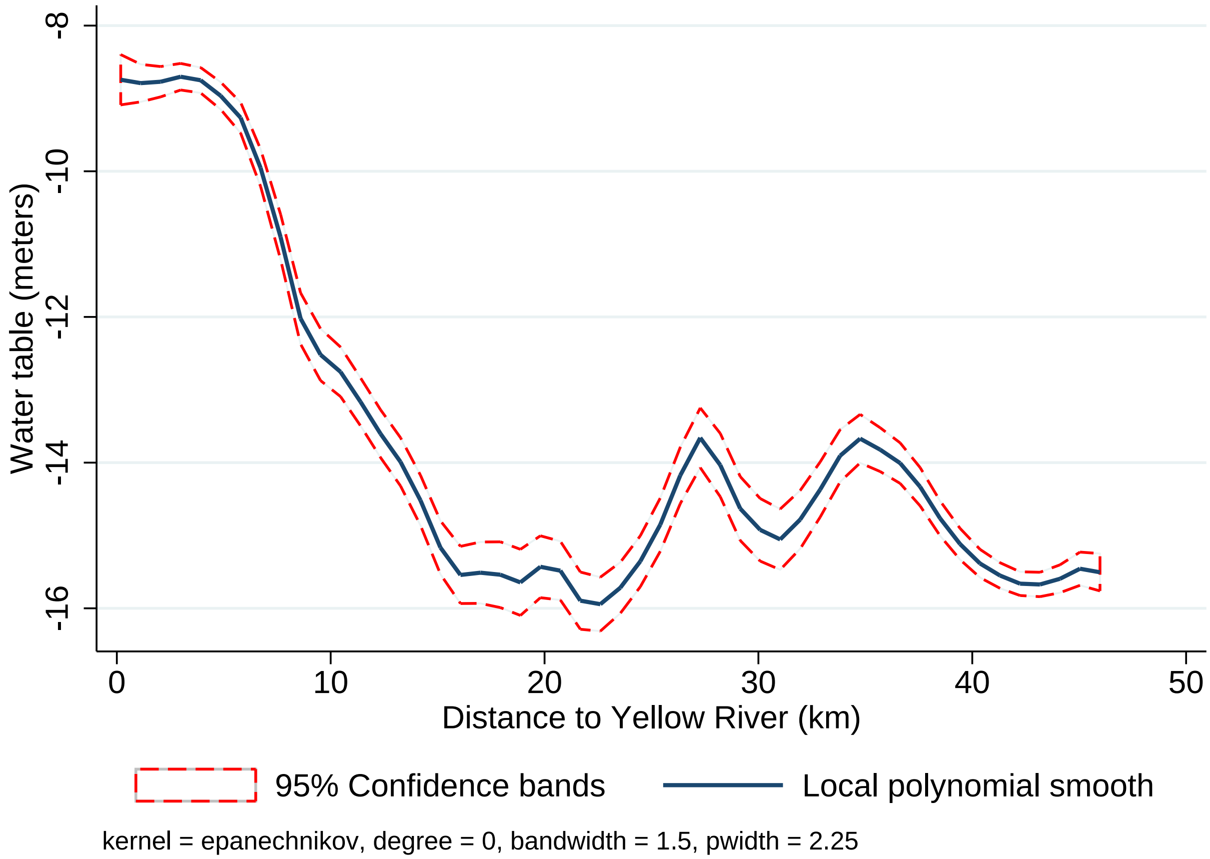
Note: The blue dots represent the year fixed effects coefficients (with their 95% confidence bands) for the estimation of a regression of the number of wells per cell over cell and year fixed effects. The two shaded areas represent the construction periods of G30 and G1511, respectively.

Figure 18: Coefficients from number of wells regressed on treatment times year dummies



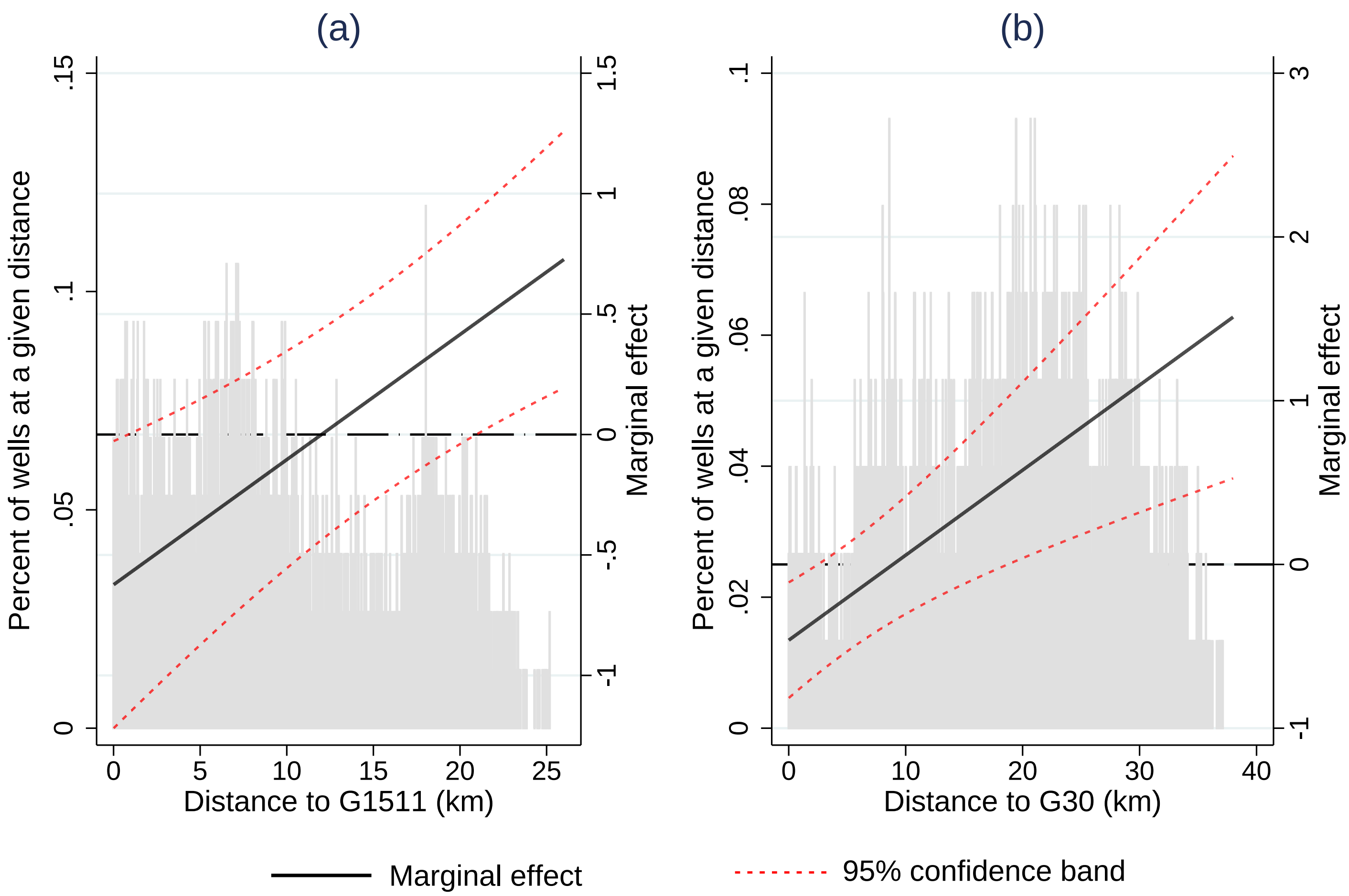
Note: The blue dots represent the year fixed effects coefficients (with their 95% confidence bands) for the estimation of a regression of the number of wells per cell over cell and year fixed effects interacted with the treatment dummy. The two shaded areas represent the construction periods of G30 and G1511, respectively.

Figure 19: Depth of water table in wells with distance from the Yellow River



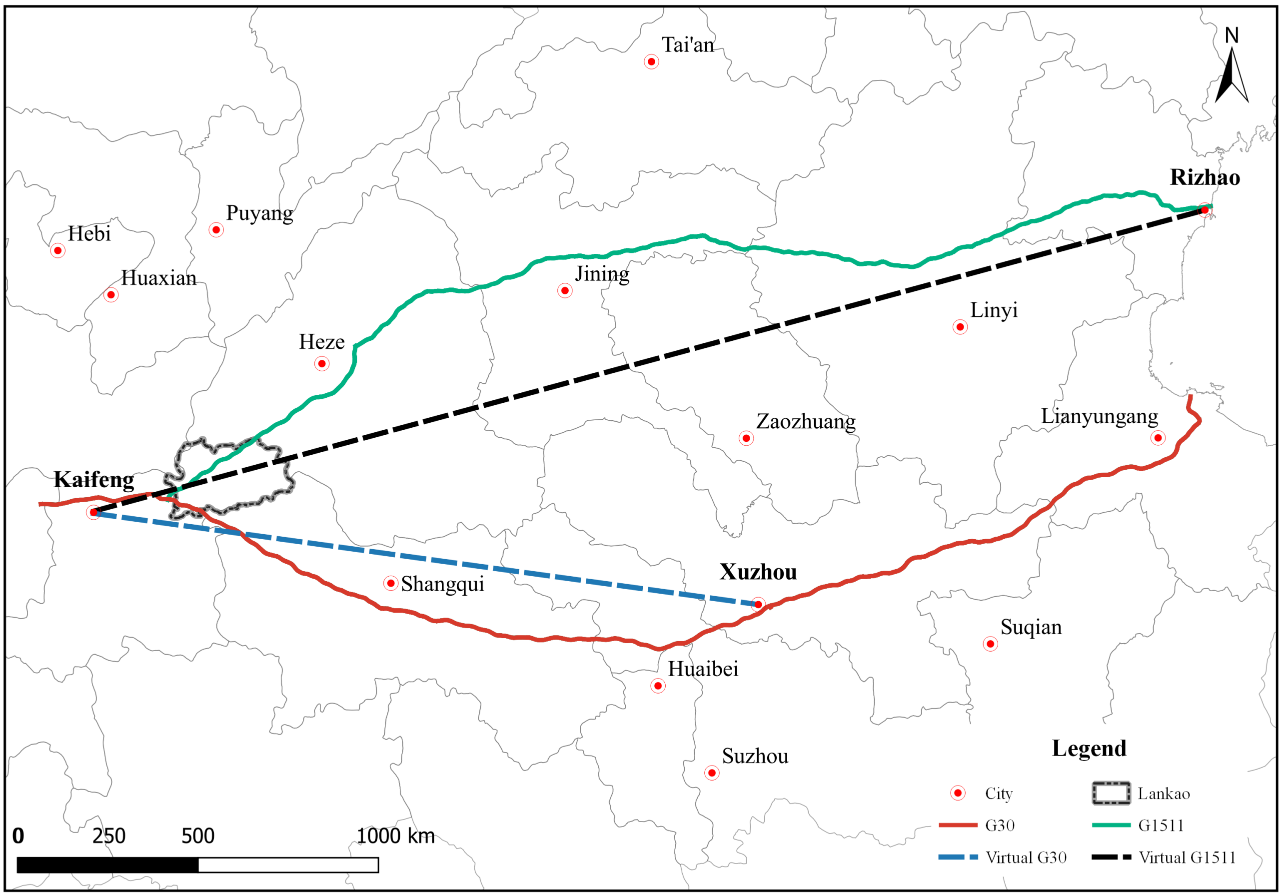
Note: This graph shows a local polynomial smoothing for the depth of the water table regressed on the distance to the Yellow River.

Figure 20: Marginal effect of the distance from G1511 and G30



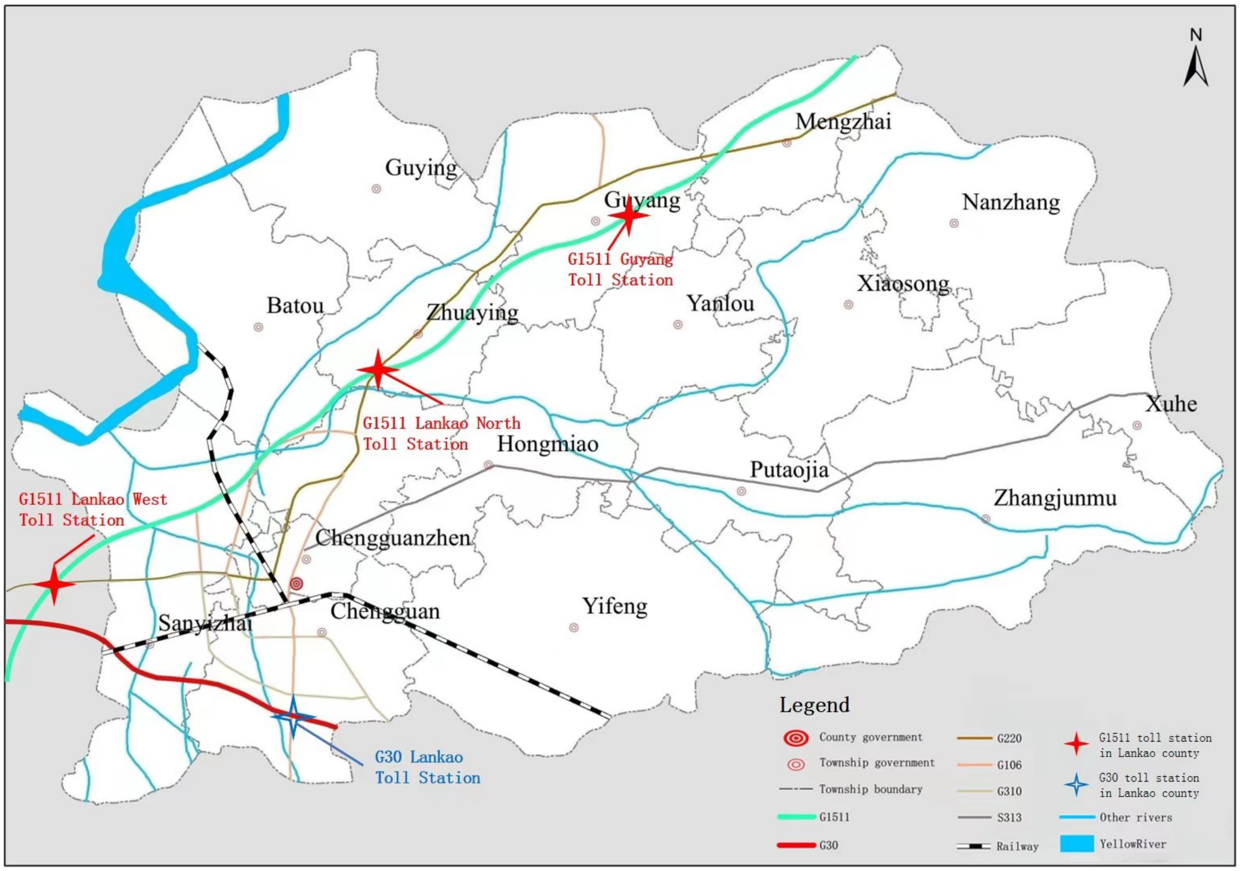
Note: We plot the marginal effect on the water table of distance to G30 and G1511, and the histogram of the number of wells at each distance from the roads. In panel (a) the distance to G30 is fixed at its mean value. We show the depth of the water table when the distance to G1511 changes. Panel (b) shows a similar plot for G30.

Figure 21: Real and virtual highways: G1511 and G30



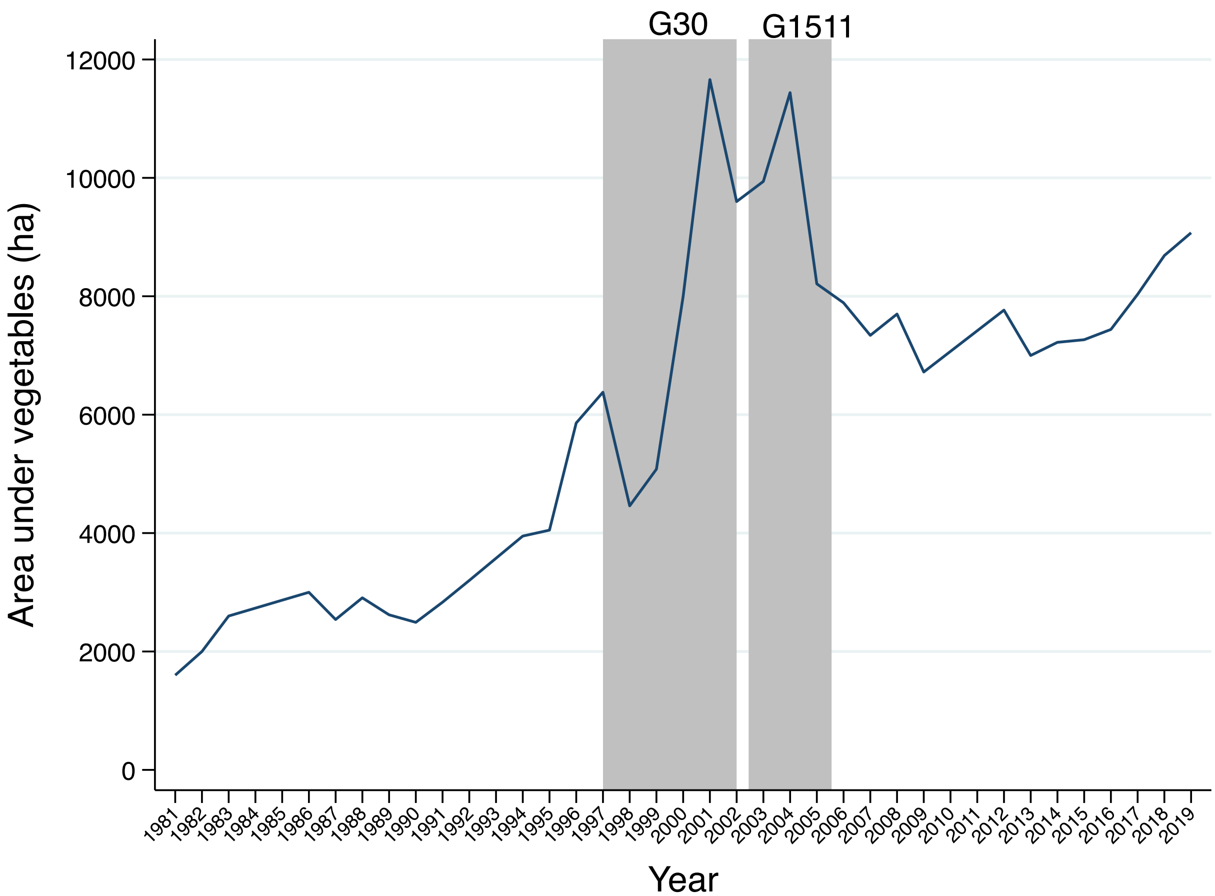
Note: The dashed straight lines show the position of the virtual highways connecting major cities, in relation to the actual roads: G30 is in red and G1511 in green.

Figure 22: Entry/exit ramps for G30 and G1511



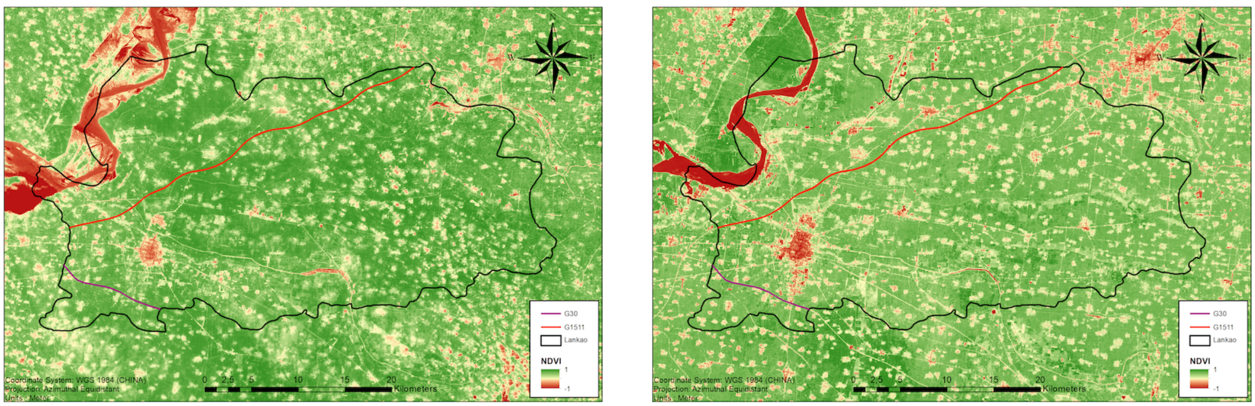
Note: The blue and red stars show the entrance/exit ramps to G30 and G1511 respectively. The entrance and exit ramps are too close to each other to be shown separately. Note that the Lankao West ramp falls just outside the county, however we include it in the analysis since using this ramp may be a least cost option for some residents.

Figure 23: Total area under vegetables in the county (1981-2019)



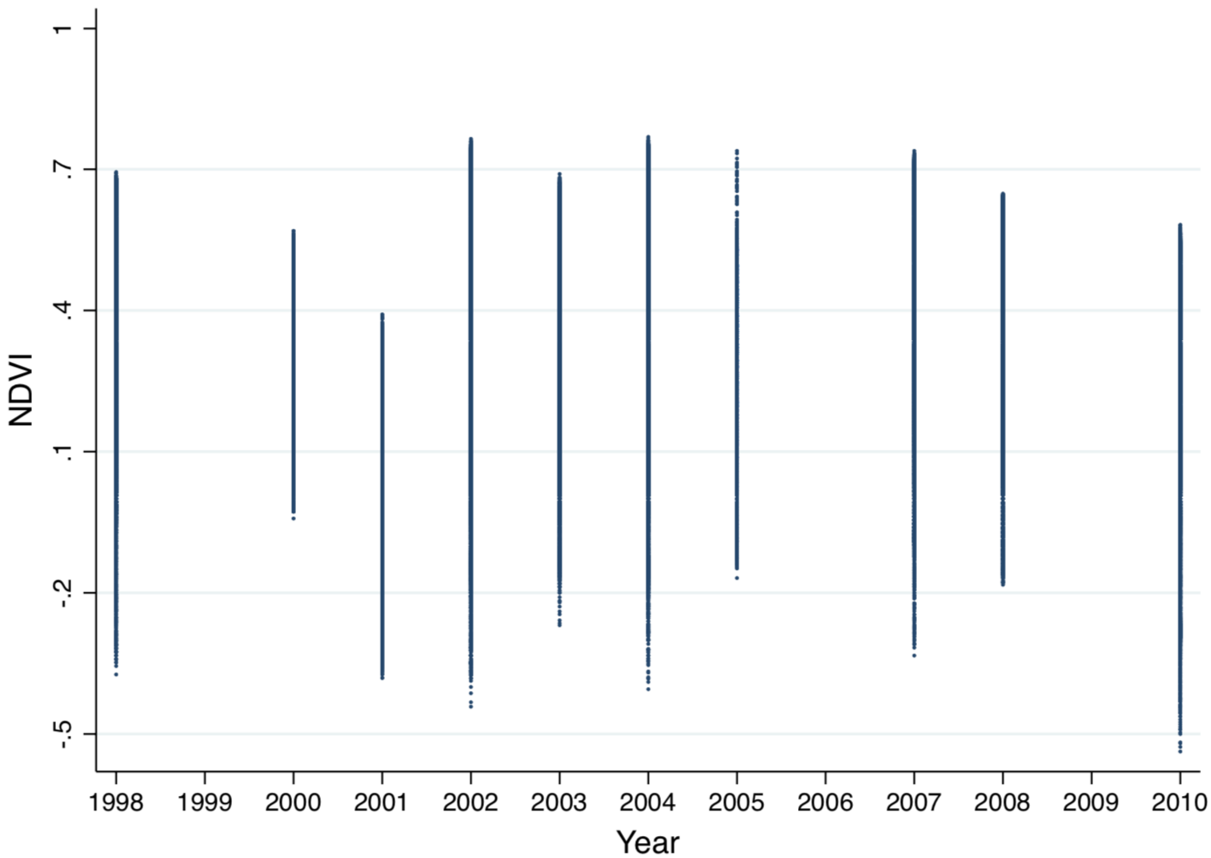
Note: The shaded area on the left shows the construction period of G30, the one on the right is for G1511. Source: Chinese National Bureau of Statistics.

Figure 24: Comparison of NDVI (1998 and 2010)



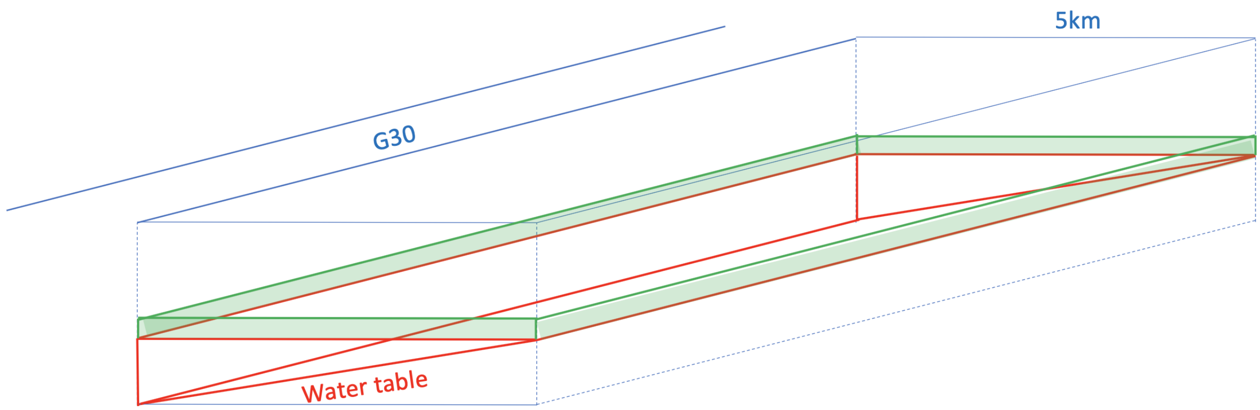
Note: The left panel shows NDVI for Lankao county during the summer of 1998, and the right panel while the right panel for the summer of 2010. These images are constructed using satellite data from the United States Geological Survey (USGS). NDVI values are computed for each of the 75,828 (150 x 150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

Figure 25: NDVI by cell and year for Lankao (1998-2010)



Note: NDVI by year for the 75,828 cells (150x150 meters) in Lankao county. There is no clear trend over the whole time period. Data for 1999, 2006 and 2009 are missing because quality images were not available. The data used to compute NDVI are from the United States Geological Survey. NDVI varies between -1 – indicating no vegetation in the cell, to 1, indicating full cover.

Figure 26: Depletion of the water table relative to G30.



Note: The green shaded area shows mean depletion measured by satellites (Feng et al., 2013), and the red triangle represents the estimated additional depletion generated by the highway.

8 Appendix A: Model Details

We can now differentiate the above equations with respect to x to get:

$$(p_1 - tx)\left[\frac{\partial^2 y_1}{\partial w_1^2}\frac{dw_1}{dx} + \frac{\partial^2 y_1}{\partial w_1 \partial L_1}\frac{dL_1}{dx}\right] - t\frac{\partial y_1}{\partial w_1} = 0 \quad (14)$$

$$p_2\left[\frac{\partial^2 y_2}{\partial w_2^2}\frac{dw_2}{dx} + \frac{\partial^2 y_2}{\partial w_2 \partial L_2}\frac{dL_2}{dx}\right] = 0, \ \textit{and} \quad (15)$$

$$(p_1 - tx)\left[\frac{\partial^2 y_1}{\partial w_1 \partial L_1}\frac{dw_1}{dx} + \frac{\partial^2 y_1}{\partial L_1^2}\frac{dL_1}{dx}\right]$$
$$-p_2\left[\frac{\partial^2 y_2}{\partial w_2 \partial L_2}\frac{dw_2}{dx} + \frac{\partial^2 y_2}{\partial L_2^2}\frac{dL_2}{dx}\right] - t\frac{\partial y_1}{\partial L_1} = 0. \quad (16)$$

From (15), we get

$$\frac{dw_2}{dx} = \frac{dL_2}{dx}\left[\frac{-\frac{\partial^2 y_2}{\partial w_2 \partial L_2}}{\frac{\partial^2 y_2}{\partial w_2^2}}\right]. \quad (17)$$

Substituting (17) into (16) and writing $\text{sgn}\left(\frac{dw_2}{dx}\right) = \text{sgn}\left(\frac{dL_2}{dx}\right) = -\text{sgn}\left(\frac{dL_1}{dx}\right)$ yields

$$(p_1 - tx)\frac{\partial^2 y_1}{\partial w_1 \partial L_1}\frac{dw_1}{dx} + \frac{dL_1}{dx}\left[(p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2} + p_2\frac{\partial^2 y_2}{\partial L_2^2} - p_2\frac{(\frac{\partial^2 y_2}{\partial L_2 \partial w_2})^2}{\frac{\partial^2 y_2}{\partial w_2^2}}\right] - t\frac{\partial y_1}{\partial L_1} = 0. \quad (18)$$

From (14) and (18) we have two linear equations with two unknowns, $\frac{dw_1}{dx}$ and $\frac{dL_1}{dx}$. We can write them as

$$A_1\frac{dw_1}{dx} + B_1\frac{dL_1}{dx} + C_1 = 0$$
$$A_2\frac{dw_1}{dx} + B_2\frac{dL_1}{dx} + C_2 = 0$$

where

$$A_1 = (p_1 - tx)\frac{\partial^2 y_1}{\partial w_1^2} \tag{19}$$

$$B_1 = (p_1 - tx)\frac{\partial^2 y_1}{\partial w_1 \partial L_1} \tag{20}$$

$$C_1 = -t\frac{\partial y_1}{\partial w_1} \tag{21}$$

$$A_2 = (p_1 - tx)\frac{\partial^2 y_1}{\partial L_1 \partial w_1} \tag{22}$$

$$B_2 = (p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2} + p_2\frac{\partial^2 y_2}{\partial L_2^2} - p_2\frac{\left(\frac{\partial^2 y_2}{\partial L_2 \partial w_2}\right)^2}{\frac{\partial^2 y_2}{\partial w_2^2}} \tag{23}$$

$$C_2 = -t\frac{\partial y_1}{\partial L_1}. \tag{24}$$

Solving the system of equations for $\frac{dw_1}{dx}$ and $\frac{dL_1}{dx}$, we get

$$\frac{dw_1}{dx} = \frac{B_1C_2 - B_2C_1}{A_1B_2 - A_2B_1} = \frac{F}{G}$$

where $F = B_1C_2 - B_2C_1$ and $G = A_1B_2 - A_2B_1$. Now we can determine the sign of F and G. Starting with F:

$$F = -(p_1 - tx)\frac{\partial^2 y_1}{\partial w_1 \partial L_1}t\frac{\partial y_1}{\partial L_1} + t\frac{\partial y_1}{\partial w_1}\left[(p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2} + p_2\frac{\partial^2 y_2}{\partial L_2^2} - p_2\frac{\left(\frac{\partial^2 y_2}{\partial L_2 \partial w_2}\right)^2}{\frac{\partial^2 y_2}{\partial w_2^2}}\right]$$

We know that $-(p_1 - tx)\frac{\partial^2 y_1}{\partial w_1 \partial L_1} < 0$, $(p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2} < 0$, and that

$$p_2\frac{\left[\frac{\partial^2 y_2}{\partial L_2^2}\frac{\partial^2 y_2}{\partial w_2^2} - \left(\frac{\partial^2 y_2}{\partial L_2 \partial w_2}\right)^2\right]}{\frac{\partial^2 y_2}{\partial w_2^2}} < 0 \tag{25}$$

and, therefore, $F < 0$. G can be re-written as:

$$G = (p_1 - tx)^2 \left[\frac{\partial^2 y_1}{\partial w_1^2} \frac{\partial^2 y_1}{\partial L_1^2} - \left(\frac{\partial^2 y_1}{\partial L_1 \partial w_1} \right)^2 \right] + p_2 \frac{\left[\frac{\partial^2 y_2}{\partial L_2^2} \frac{\partial^2 y_2}{\partial w_2^2} - \left(\frac{\partial^2 y_2}{\partial L_2 \partial w_2} \right)^2 \right]}{\frac{\partial^2 y_2}{\partial w_2^2}} (p_1 - tx) \frac{\partial^2 y_1}{\partial w_1^2} \quad (26)$$

Since $\left[\frac{\partial^2 y_1}{\partial w_1^2} \frac{\partial^2 y_1}{\partial L_1^2} - \left(\frac{\partial^2 y_1}{\partial L_1 \partial w_1} \right)^2 \right] > 0$, $(p_1 - tx) \frac{\partial^2 y_1}{\partial w_1^2}$, and using (25), $G > 0$. This allows us to conclude that

$$\frac{dw_1}{dx} < 0 \quad (27)$$

We have

$$\begin{aligned} \frac{dL_1}{dx} &= \frac{-A_2 \frac{dw_1}{dx} - C_2}{B_2} \\ &= \frac{-(p_1 - tx) \frac{\partial^2 y_1}{\partial L_1 \partial w_1} \frac{dw_1}{dx} - C_2}{B_2} \end{aligned} \quad (28)$$

Since $B_2 < 0$, $C_2 < 0$, $\frac{dw_1}{dx} < 0$, and $-(p_1 - tx) \frac{\partial^2 y_1}{\partial L_1 \partial w_1} < 0$, we can conclude that

$$\frac{dL_1}{dx} < 0. \quad (29)$$

This yields $\frac{dL_2}{dx} > 0$ and from (17), $\frac{dw_2}{dx} > 0$.

For the aggregate impact on water usage at each location, we have:

$$\frac{dw_1}{dx} + \frac{dw_2}{dx} = \frac{B_1 C_2 - B_2 C_1}{A_1 B_2 - A_2 B_1} + \frac{\frac{\partial^2 y_2}{\partial w_2 \partial L_2}}{\frac{\partial^2 y_2}{\partial w_2^2}} \frac{dL_1}{dx} \quad (30)$$

Let us define $\theta \equiv \frac{\frac{\partial^2 y_2}{\partial w_2 \partial L_2}}{\frac{\partial^2 y_2}{\partial w_2^2}}$ and replace $\frac{dL_1}{dx}$ with $\frac{-A_2 \frac{dw_1}{dx} - C_2}{B_2}$, obtaining:

$$\frac{dw_1}{dx} + \frac{dw_2}{dx} = \frac{\frac{dw_1}{dx} (B_2 - \theta A_2) - \theta C_2}{B_2} \quad (31)$$

We now need to determine its sign, so for B_2:

$$B_2 = \frac{(p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2}\frac{\partial^2 y_2}{\partial w_2^2} + p_2\left[\frac{\partial^2 y_2}{\partial L_2^2}\frac{\partial^2 y_2}{\partial w_2^2} - \left(\frac{\partial^2 y_2}{\partial w_1 \partial w_2}\right)^2\right]}{\frac{\partial^2 y_2}{\partial w_2^2}} \tag{32}$$

We know that $(p_1 - tx)\frac{\partial^2 y_1}{\partial L_1^2}\frac{\partial^2 y_2}{\partial w_2^2} > 0$, $\left[\frac{\partial^2 y_2}{\partial L_2^2}\frac{\partial^2 y_2}{\partial w_2^2} - \left(\frac{\partial^2 y_2}{\partial w_1 \partial w_2}\right)^2\right] > 0$, and that $\frac{\partial^2 y_2}{\partial w_2^2} < 0$, therefore

$$B_2 < 0$$

The numerator $\frac{dw_1}{dx}(B_2 - \theta A_2) - \theta C_2$ can be written as:

$$\frac{dw_1}{dx}B_2 - \frac{dw_1}{dx}\theta A_2 - \theta C_2 > 0 \tag{33}$$

$$\left|\frac{dw_1}{dx}B_2\right| > \left|\theta\left(\frac{dw_1}{dx}A_2 + C_2\right)\right| \tag{34}$$

This is true if $\frac{\partial^2 y_2}{\partial w_2 \partial L_2}$ is sufficiently small. Thus, aggregate water use in plots declines away from the highway.

9 Appendix B: Details of the 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 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.

Appendix Tables

Table 16: Robustness – DiD estimation, 300m cells

	Number of wells in a cell					
	G30			G1511		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat^{G30}	−0.005					
	(0.015)					
Post^{G30}	0.346***	0.346∗∗∗				
	(0.051)	(0.051)				
Treat^{G30}*Post^{G30}	0.077***	0.077**	0.077**			
	(0.028)	(0.029)	(0.029)			
Treat^{G1511}				0.117***		
				(0.013)		
Post^{G1511}				0.392***	0.392***	
				(0.059)	(0.060)	
Treat^{G1511}*Post^{G1511}				0.068***	0.068***	0.068***
				(0.015)	(0.016)	(0.016)
Cell FE	no	yes	yes	no	yes	yes
Year FE	no	no	yes	no	no	yes
Mean dep. var.	0.344	0.344	0.344	0.382	0.382	0.382
Observations	291,400	291,400	291,400	368,280	368,280	368,280

Note: The table shows cell-level regressions on the number of wells in a cell. The variable $Treat^{G30}$ ($Treat^{G1511}$) is a dummy with value 1 if the cell is in the treatment area, defined as a band with a radius of 4 km around G30 (G1511). The non-treated area includes all cells situated at least 4 km away from G1511 and from G30. The variable $Post^{G30}$ ($Post^{G1511}$) is a dummy with value 1 for years following the beginning of the construction of G30 (G1511) in 1998 (2003). The interaction of the two variables captures the DiD effect. Columns 1-3 show results for G30, and columns 4-6 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 17: Robustness – DiD estimation, 500m cells

	Number of wells in a cell					
	G30			G1511		
	(1)	(2)	(3)	(4)	(5)	(6)
$Treat^{G30}$	-0.019					
	(0.045)					
$Post^{G30}$	0.940***	0.940***				
	(0.139)	(0.139)				
$Treat^{G30}$*$Post^{G30}$	0.192**	0.192**	0.192**			
	(0.084)	(0.087)	(0.087)			
$Treat^{G1511}$				0.321***		
				(0.041)		
$Post^{G1511}$				1.067***	1.067***	
				(0.162)	(0.162)	
$Treat^{G1511}$*$Post^{G1511}$				0.189***	0.189***	0.189***
				(0.046)	(0.051)	(0.052)
Cell FE	no	yes	yes	no	yes	yes
Year FE	no	no	yes	no	no	yes
Mean dep. var.	0.933	0.933	0.933	1.040	1.040	1.040
Observations	106,764	106,764	106,764	134,974	134,974	134,974

Note: The table shows cell-level regressions on the number of wells in a cell. The variable $Treat^{G30}$ ($Treat^{G1511}$) is a dummy with value 1 if the cell is in the treatment area, defined as a band with a radius of 4 km around G30 (G1511). The non-treated area includes all cells situated at least 4 km away from G1511 and from G30. The variable $Post^{G30}$ ($Post^{G1511}$) is a dummy with value 1 for years following the beginning of the construction of G30 (G1511) in 1998 (2003). The interaction of the two variables captures the DiD effect. Columns 1-3 show results for G30, and columns 4-6 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 18: Robustness *Post* – G30

	Number of wells in a cell					
	Post taking value 1 starting in					
	1996	1997	1998	1999	2000	2001
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	0.060***	0.058***	0.055***	0.052**	0.048**	0.045*
	(0.018)	(0.018)	(0.019)	(0.020)	(0.021)	(0.022)
Cell FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Mean dep. var.	0.257	0.257	0.257	0.257	0.257	0.257
Observations	419,678	419,678	419,678	419,678	419,678	419,678

Note: The variable *Treat* is a dummy with value 1 if the cell is in the treatment area, defined as 4000 m of G30. The non-treated area includes all cells situated at least 4000 m away from G30 and from G1511. The variable *Post* defines the period after the construction of the highway, which changes from 2 years before the beginning of construction (1998), in column (1), to the last year of the construction (2001), in column (6). The interaction of the two variables captures the DiD effect. Standard errors in parentheses are two-way clustered at the cell and year level. *** p<0.01, ** p<0.05, * p<0.1.

Table 19: Robustness *Post* – G1511

	Number of wells in a cell				
	Post taking value 1 starting in				
	2001	2002	2003	2004	2005
	(1)	(2)	(3)	(4)	(5)
Treat*Post	0.048***	0.047***	0.047***	0.045***	0.046**
	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)
Cell FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Mean dep. var.	0.276	0.276	0.276	0.276	0.276
Observations	529,046	529,046	529,046	529,046	529,046

Note: The variable *Treat* is a dummy with value 1 if the cell is in the treatment area, defined as 4000 m of G1511. The non-treated area includes all cells situated at least 4000 m away from G1511 and from G30. The variable *Post* defines the period after the construction of the highway, which changes from 2 years before the beginning of construction (2001), in column (1), to the last year of construction (2005), in column (5). The interaction of the two variables captures the DiD effect. Standard errors in parentheses are two-way clustered at the cell and year level. *** p<0.01, ** p<0.05, * p<0.1.