# Railways, Telegraph and Technology Adoption: The Introduction of American Cotton in Early 20$^{\text{th}}$ Century China[†]

Ting Chen[‡] Han Qi[§] Jin Wang[¶]

This version, August 2022

**Abstract**

We present a new source of gains from infrastructure, technology adoption. Through the lens of a heterogeneous agent model with trade and information frictions, we explore how infrastructure increases welfare through the endogenous adoption of a new agricultural technology—a high-quality variety of American cotton—in early 20th century China. Using historical data, our empirical tests reveal that counties that had railway and telegraph access saw an increase in adoption of American cotton, since these infrastructures decreased the take-up thresholds. The contribution of the additional technology adoption margin to the overall welfare gains from infrastructure is quantitatively large.

**Keywords:** Technology Adoption, American Cotton, Telegraph, Railway, China
**JEL Classification Codes:** F14, H54, L92, L96, N75, O18, O33, Q16

[†]We thank seminar participants at the Joint APEBH 2019, All-UC Group in Economic History Conference at the Caltech, the 2019 CEIBS Economics Symposium hosted by the ShanghaiTech University and the Chinese University of Hong Kong, the University of Hong Kong, and Hong Kong University of Science and Technology for helpful comments and suggestions. Chen acknowledges the financial support of a research grant from Hong Kong's Research Grants Council (GRF No. 22502819). All errors remain our own.

[‡]Ting Chen, Department of Economics, Hong Kong Baptist University, 34 Renfrew Road, Hong Kong. Email: tingchen@hkbu.edu.hk.

[§]Han Qi, Shenzhen Audencia Business School, WeBank Institute of Fintech, Shenzhen University, Nanshan District, Shenzhen, China. Email: steffan@szu.edu.cn.

[¶]Jin Wang, Division of Social Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Email: sojinwang@ust.hk.

# 1 Introduction

In the last decade, economists have increasingly sought to quantify and explain the benefits of building infrastructures, one of the world's primary development strategies. In a quest to understand how investing in infrastructure can increase welfare, seminal works, most notably Donaldson (2018) for railways and Steinwender (2018) for the telegraph, highlight the improved trading environment that underpins these gains, which primarily operate by reducing trade costs and information friction. Despite these contributions, a potentially crucial margin—technology and quality upgrading generated by the development of infrastructure—remains largely unexplored. Recent discussions in the trade and development literature demonstrate the importance of this margin. Several studies have attributed substantial income and productivity gains to technology adoption (Gollin et al., 2021; Juhász et al., 2020).[1] New field experiments have also begun to explore various barriers to technology diffusion and quality upgrading by exogenously varying access to markets for premium products (Bold et al., forthcoming) or information (Beaman et al., 2021). This paper provides the first systematic evidence on technology adoption as a new source of gains from infrastructure, highlighting its key role in market access and information transmission.

We theoretically and empirically explore how railways and the telegraph affected the adoption of a new agricultural technology—high-quality American cotton—in early 20$^{\text{th}}$ century China. China's republican era represents an exceptional period in the development of modern infrastructure. In an early attempt to foster industrialization, the government increased the expansion of the railway and telegraph coverage across the country. A total of 4,335 km of railways and 87,182 km of telegraph lines were built from 1920 to 1937; by the end of this period 44% and 66% of counties were connected, respectively. Rapid global and domestic industrial development, particularly the fast-growing textile industry concentrated in coastal areas, caused inland Chinese farmers to experience sharp increases in the demand for agricultural raw materials such as cotton. The infrastructure development substantially reduced the trade costs and information delays, thus greatly improving the environment for inter-regional trade—especially long-distance trade between remote rural counties and manufacturing hubs in coastal cities. It is unlikely to be a coincidence that during almost the same period there was a surge in the cultivation of high-quality American cotton, a variety that is considered more suitable than the native Chinese variety for machine-based textile production. In a unique historical setting, the "big push" of railroad and

[1] Focusing on the Green Revolution, an agricultural technology milestone, Gollin et al. (2021) examined the impact of the global diffusion of high-yield varieties on crop productivity and per capita income in the developing world.

telegraph networks during China's early industrialization period provides a rare opportunity to assess the impact of transportation and ICT (information and communication technology) infrastructures on farmers' technology transition from low- to high-quality cotton and to identify the mechanisms behind their decision-making.

China is now the biggest cotton producer in the world. Almost all (98%) of its 6.5 million metric tonnes output in 2017 was the American upland cotton variety. Introduced in 1920, American cotton surpassed its native counterpart within decades as the dominant species in cotton-growing areas, arguably the most rapidly adopted and one of the most important agricultural technologies in modern Chinese history. Its cultivation also paved the way for China to become the world's leading textile producer and cotton exporter. Given its historical importance, the diffusion of American cotton across the country is worthy of study in its own right. Our inquiry also has broader implications for the well-being of many in developing countries. According to the World Bank (Fuglie et al., 2020), agricultural innovation and technology are key to increasing productivity and decreasing poverty. Studying how different infrastructures and possible complementarities between them facilitated agricultural technology adoption and in turn improved farmers' welfare can shed new light on this critical policy issue.

We first build a simple heterogeneous agent model that incorporates trade and information frictions. We seek to illustrate the mechanisms through which rail and telegraph networks encouraged Chinese farmers to grow the high-quality American cotton variety. We derive the equilibrium threshold for adoption denoted by the cultivation suitability and land endowment in different scenarios of infrastructure access. The key insights from the model are that when building infrastructure reduced the transportation costs and price uncertainty of American cotton, this made interregional trade possible and allowed local farmers to access markets for premium quality cotton. This lowers the adoption thresholds considerably, which leads more farmers to upgrade their cotton variety in the face of profits.

To test our main prediction about how infrastructure affects technology adoption, we assemble rich new data on 1,660 Chinese counties between 1920 and 1937 from digitized historical maps, archives, gazetteers and annual statistics of the cotton industry. This data records railway access, telegraph connections, and planting areas and yields of both American and native Chinese cotton in remarkable detail. We document the causal impact of infrastructure access on technology adoption using a difference-in-differences (DID) design, comparing adoption outcomes over time and across counties that were reached by railway and telegraph networks at different times. We show that they increase both the extensive and intensive margins of American cotton, measured as the choice to adopt the new variety and the area of land planted with the American species. We find an even larger effect for counties that have access to both infrastructures, indicating a complementarity between railway (transportation) and telegraph (information). In line with the DID identification assumptions, event studies present evidence of parallel pre-trends in adoption outcomes, which supports the causal interpretation of our estimates.

The endogenous placement of infrastructure still challenges our empirical identification. We employ an instrumental variable (IV) approach to address this issue by leveraging two important historical facts. First, during the the early years of development, the new railways were often built in straight lines between key node cities to save on construction costs. Second, the Qing government developed the telegraph network to closely match the existing postal network because local postal employees provided the new services. We therefore use whether a county is located along a straight-line corridor between major cities covered by the railway network as an instrumental variable for railway access, and whether it had a post office as an instrumental variable for telegraph connection. These instruments are strongly and significantly associated with a county's infrastructure access, but not with pre-existing socio-economic determinants of technology adoption before the expansion of the infrastructure networks. In particular, exploiting these plausibly exogenous geographical variations in infrastructure development yields a larger estimated effect of railway access on decisions to adopt (and how many hectares to plant with) American cotton, a premium product that was traded in non-local markets. Reassuringly, we do not observe such effects for the low-quality Chinese cotton. We interpret these results as strong evidence in support of the technology and quality upgrading response to infrastructure access, as predicted by the model.

To shed further light on the mechanisms at play, we compile and examine comprehensive cross-section data on the domestic trade of commodities and thresholds of adoption. We show that county pairs that had access to railway and telegraph experienced a significant increase in their trade volumes and a narrowed price spread of cotton. Similar patterns hold for all commodities aggregated. These results thus establish the premise of the model that infrastructure helped reduce trade costs and information frictions, overcoming two important hurdles to domestic trade. We then show that railway and telegraph access caused a larger increase in the adoption of American cotton, especially among farmers in counties that were less suitable for growing cotton and those with smaller farms. Our findings are highly consistent with the model's predictions about adoption thresholds, which provides robust evidence of how infrastructures facilitate cotton quality upgrading in China.

Finally, we construct the key proxy for welfare by computing the real agricultural income of Chinese counties. We then estimate a new source of welfare gains (the increase in real income) from infrastructures, which operates through the adoption of technologies. Employ-ing causal mediation analysis, we show that railway access raised farmers' welfare by a total of 8.4%, 4.1% of which can be attributed to the adoption of American cotton. The telegraph connection alone does not significantly affect farmers' welfare. The results are closely related to the context we study: the long-distance trade of American cotton is only possible with enabling transportation infrastructure, which plays a more important role here.

Our study contributes to, and bridges, two strands of scholarly work. First, it advances a burgeoning infrastructure literature that focuses on assessing the gains generated by railway (Hornung, 2015; Donaldson and Hornbeck, 2016; Donaldson, 2018) and telegraph networks (Ejrnes and Persson, 2010; Steinwender, 2018; Gao and Lei, 2021) that arise from increased trade. In historical China, our study shows that endogenous technology adoption could be the important missing piece in the welfare analysis of infrastructure development.[2] This newly discovered mechanism explains a considerable proportion of the benefits from infrastructure, which market integration and trade could not previously account for. To the best of our knowledge, our study is the first to comprehensively assess the railroad and telegraph networks within a single research context. We demonstrate that transportation infrastructure has a first-order effect on technology adoption, while ICT infrastructure appears to exhibit strong complementarity.

This paper also connects to a significant body of trade literature on technology and quality upgrading. Predominantly within the manufacturing sector, this line of work has employed trade liberalization reforms or field experiments to demonstrate that better market access can promote quality upgrading (Verhoogen, 2008; Bustos, 2011; Amiti and Khandelwal, 2013; Atkin et al., 2017). Prior research has measured quality in terms of unit values (shipment values divided by quantities), industry-level quality ladders inferred from market shares, and firm spending on technology. In this paper, we identify specific benefits of infrastructure associated with the adoption of technologies in the context of agricultural markets, which many developing countries rely on for employment and income generation. Our results suggest that by improving market access and information transmission, infrastructure is a powerful policy tool to overcome technological adoption barriers. Methodologically, we utilize a large-scale historical natural experiment with a high degree of external validity. We also deploy an accurate and fine-grained measure of quality upgrading and technology adoption: the planting of high-quality American versus low-quality Chinese cotton species. Both aspects are rarely available in other contemporary settings, which we view as important contributions of the paper.

[2]In investigating the welfare impact of railroads, Donaldson (2018) takes technology as exogenous and driven by rainfall and finds that about half of the gains are attributable to enhanced opportunities to trade ("trade share").

Notable exceptions include the most recent experimental studies of Bold et al. (forthcoming) and Beaman et al. (2021).[3] A demand-side force similar to that described in Bold et al. (forthcoming) is at work in our paper: Chinese farmers rely on infrastructure (railway) to ship their products to the national market, which pays a higher price for the high-quality cotton, thus giving them an incentive to adopt the American variety. Broadly related to Beaman et al. (2021), we highlight information friction as a hurdle to technology adoption; in our research setting, the telegraph can deliver price and market information in a timely manner to farmers to reduce their uncertainty of adopting the new variety of cotton. We push the analysis further than in these two prior papers by exploring the role of heterogeneity in how building infrastructure affects quality upgrading. Most notably, we believe our study is the first to quantify the welfare gains operating through the technology adoption margin.

The remainder of the paper is organized as follows. The next section describes the historical background. Section 3 sets out a simple theoretical model. Section 4 introduces the data. Section 5 investigates how infrastructure affects technology adoption. Section 6 explores the underlying mechanisms and quantifies the welfare gains. Section 7 concludes.

# 2 Historical Background

In this section, we discuss the history of the infrastructure development in late Qing and Republican-era China. We then describe the diffusion of American cotton, a high-quality foreign variety.

## 2.1 Infrastructure Expansion

The first telegraph and railway lines were built in China in the $19^{\text{th}}$ century under the influence of the western powers, particularly their modern knowledge of infrastructure. Li Hongzhang, a prominent Qing official and a reformist, recognized the utility advantage of telecommunications in transmitting information, particularly for military and administrative purposes; this motivated him to promote the development of modern infrastructures. He was instrumental in the installation of China's first independent telegraph line in 1881, which connected Tianjin (a major city in northern China adjacent to Beijing) and Shanghai. Li also spearheaded the construction of the country's first independent railway in the same

[3] Bold et al. (forthcoming) provide new experimental evidence of the importance of market access in quality upgrading among maize farmers in Uganda. They show that a major barrier to upgrading is the absence of a price premium for higher-quality products in the local market. Beaman et al. (2021) conducted a field experiment in Malawi and identified a specific constraint to agricultural technology adoption—the absence of proper information targeting to central farmers for the effective diffusion of farming practices.

year—the "Tangxu" line in Hebei—to transport coal from the mines to the marketplace.

The large-scale rollout of both infrastructures soon gained momentum. Two major waves of construction occurred in 1903–1920 and 1927–1937, as shown in the left panel of Figure 1, which depicts the kilometers of railway (upper) and telegraph lines (lower) completed annually during this period. A total of 70,674 km of telegraph lines were installed between 1881 and 1920, in 592 counties (36% of the counties in China at that time). While the expansion of the telegraph importantly could rely on the pre-existing postal network, railway construction had no prior foundation and thus required much more capital investment. As a result, the development of the railway (denoted by maroon lines in the lower left panel of Figure 1) appeared to lag behind that of the telegraph (dark blue line). A total of 5,427 km of railway lines were built and put into operation, connecting 501 counties (30% of Chinese counties) located within a 50 km radius of the railway.

The telegraph and railway networks experienced a second wave of mass expansion during the study period (1920–1937), in which an additional 87,182 km of telegraph lines and 4,335 km of railroad tracks were constructed and completed. Before the Sino–Japanese War, 1,094 counties (66%) were connected to the telegraph network and 728 (44%) could be reached by rail, covering vast areas of the nation's territory. By 1937, the telegraph network (light blue lines in the right panel of Figure 1) extensively covered China Proper (the 18 major provinces in late Qing and early Republican China that covered the majority of the inland territories), while railways (red lines) had reached major cities and ports along the coasts and the Yangtze River.

[Figure 1 about here]

By significantly reducing the costs, railway development greatly facilitated freight transportation. Dirt roads, river transport and coastal shipping were previously the country's dominant freight transportation modes. For inland transport, freighters used horses, mules or camels to carry bulk commodities (mainly in the North-west) along the road network, which was riddled with many difficulties. On well-worn road surfaces and in good weather conditions, animal-powered transport could carry a cart of goods 20—30 km per day. Yet good-quality roads were rare and typically monopolized by the government. Thus the shipments were often charged high rates. Although water transport was economically superior to roads, it was possible only in the Yangtze River basin that spreads across the South. Due to limited river networks in the North, it was extremely difficult to ship goods over waterways from inland areas to the coast. For example, shipping wool from the northwestern regions to the coastal city of Tianjin (Chen, 2010) began with camel caravans. After 10–30 days the wool would arrive in the regional transportation hub of Zhangjiakou. From there, the wool would be carried by horses to the next stop, Tongzhou (near Beijing), which took another 5–10 days. In the final stage, the goods would be moved to Tianjin by water for several days. If there was a shortage of horses, cargo may have to wait for months or even half a year to be shipped out of Zhangjiakou. The completion of the Jing Zhang railway in 1909 dramatically improved the country's transport network. The shipping times between the two connected hubs—Tianjin and Zhangjiakou—were substantially shortened to 2–3 days and the risk of delays and transport costs decreased considerably. In 1906, an estimated 48% of the shipments between Tianjin and the inland regions were transported through railways; 45% were transported via waterways and 7% over dirt roads (Mi, 1963). By 1920, the corresponding share of railway transport had increased to 71%, and water transport had decreased to 25% and road shipments to 4%.

The introduction of the telegraph system also dramatically reduced the costs of information transmission in China. Before it was developed, an imperial postal relay system transmitted information by horse over land or through trade routes along waterways. In the late 19$^{th}$ century, it took at least 3 days to transmit information in this way from the capital, Peking, to Nanjing—a major southeastern city along the Yangtze River at a speed of 365 km per day. The movement and circulation of information from Peking to Hangzhou—another city on the Yangtze River—took at least 6 days at a speed of 255 km per day (Fairbank and Teng, 1972). The telegraph network transmitted messages over long distances almost instantaneously. By avoiding lengthy information time lags, telegraphs provided a new means of extending commerce.

## 2.2 The Diffusion of American Cotton in China

The American cotton species (*Gossypium hirsutum* in Latin), also known as upland cotton, originated in Mexico and Central America. During the Industrial Revolution, it quickly spread to the U.S. and the rest of the world because the booming textile industry in the United Kingdom and other early industrializers increased the demand for this high-quality cotton variety (Beckert, 2014). Today around 90% of global cotton production is based on cultivars developed from these species. Over 95% of the cotton cultivated in China, the largest grower in the world, is the American variety.

China had a long history of cotton cultivation before the introduction of American cotton. Starting with the domestication of Asian cotton (also known as Indian cotton, *Gossypium arboreum* and *Gossypium hardense* in Latin), there was a surge in cotton planting in the 16$^{th}$ century in southern China along the Yangtze River to the North near the Yellow and Huai rivers. From the late 19$^{th}$ to the early 20$^{th}$ century, China's attempts to adopt new and high-quality American cotton came in two waves. The first was led by local officials from the Qing government who bought sample seeds from the U.S. and experimented with cotton domestication in Shanghai and Wuhan around 1890. The local farmers who received the sample seeds were not given appropriate guidance or sufficient incentives for cultivation, which led to adoption failures.

We focus here on the second wave. Between the 1920s and 1930s, having acknowledged the importance of high-quality cotton for the mechanical textile industry, the government devoted considerable effort to introducing and domesticating the American species, and played a crucial role in promoting its cultivation. In the mid-1910s, Zhang Jian, the Minister of Agriculture and Business, had selected five fields and hired foreign experts to undertake pilot studies. The scale of the experiments was expanded after the government established the Cotton Planting Committee, which was comprised of officials, entrepreneurs and scientists. In 1920, the committee set up over 26 pilot farms to experiment on eight varieties of American cotton across the North China Plain, which had favorable conditions for growing the crop (soil, topography and weather). The most viable varieties were identified and domesticated for cultivation. After extensive efforts by various associations and entrepreneurs from the textile industry, farmers gradually adopted American cotton on a larger scale.

Compared to the native Chinese species, American cotton had clear advantages in both the quality of end products (such as cotton yarns and cloth) and the suitability for machine-based textile production. First, the staples of the American-type cotton tend to be smoother and more lustrous than those of most other commercial cottons (including the native Chinese species), and thus produce superior yarns and cloth. The long-staple American cotton also outperforms the short-staple Chinese cotton in machine-based textile production and has high value in processing. It is much easier for machines to spin the American variety's longer and more tenacious fibers into yarn, and then to weave the yarn into cloth without damage.[4] With these attractive properties, textile factories were willing to pay a much higher price for American cotton. Table 1 presents more detailed quality comparisons of American vs. Chinese cotton on a host of characteristics, based on two sources—the Ministry of Agriculture's cotton production improvement program and the Shanghai Commodity Inspection Bureau's cotton quality grading experiment. Appendix Figure A.1 illustrates the differences between the two cotton species.

[Table 1 about here]

[4]To make fabric from cotton, it must first be prepared for spinning, including cleaning the fibers and sorting them into shapes. The primary processing is to remove the seeds from the raw cotton. Before the cotton gin came into use, cotton fiber could only be separated from the seeds by hand. The second step is to make cotton yarn using spindles and prepared cotton. In the final step, cotton yarn is woven into cloth.

While the costs of planting both types of cotton were similar, the profits were remarkably different. The domestic tree cotton was mainly used for handicraft spinning and weaving in rural China, and the native cloth was largely consumed by local households (Feuerwerker, 1970; Fairbank and Liu, 1980). By contrast, the American cotton was harvested and sold to local brokers, who transported it from where it was grown (mostly inland rural counties) to the coastal cities where the textile factories were located. Appendix Figure A.2 shows a map on the long-distance trade routes of American cotton along the railway lines in 1936, produced by the (Japan) East Asia Institute.[5] Although the farmers had to share their profits with brokers and pay some transportation costs, it was still profitable for them to grow American cotton. According to a 1928 survey in Shandong Province, cultivating American cotton yielded a profit of 12.25 taels per hectare, compared to 8.73 taels for Chinese cotton; other species yielded less than 5 taels per hectare.

Figure 2 depicts the rapid diffusion of American cotton across counties. Introduced in 1920, the American species was soaring after 1927. It dominated but had not necessarily replaced the native species by 1933 among local farmers (Appendix Figure A.3).

[Figure 2 about here]

Historians have linked the adoption of American cotton after 1927 to the increase in industrial demand for cotton, which was stimulated by the contemporaneous development of transportation and ICT infrastructure (Buck, 1937; Fong, 1932; Yan, 1943). Figure 3 displays the spatial variation in the adoption of American cotton over time. The two maps depict the total cropping area of American cotton by county in 1927 and 1937. The figure indicates that the newly constructed railroads and telegraph lines were mostly located in the vast northern plain and appear to coincide with regions that rapidly adopted and produced American cotton in the late 1920s and early 1930s. The strong spatial correlation between the infrastructure and cotton adoption seems to be consistent with historians' observations, and sets the stage for our empirical inquiry detailed in the proceeding sections.

[Figure 3 about here]

# 3 A Model of Infrastructure and Technology Adoption

In this section, we present a heterogeneous agent model with trade and information frictions. The model is based on the premise that when these frictions are high, the transportation

[5] We obtained the pertinent map from the Stanford Library (Source: https://searchworks.stanford.edu/view/10313137).

costs and output price uncertainty in the associated national market discourage risk-averse farmers from adopting the new technology. We use the model to illustrate the mechanisms through which the railways and telegraph system promoted the adoption of the high-quality cotton variety in China and to guide our empirical analysis. Note that we assume away the supply of information from sources such as farmers' social networks and middlemen, which we do not focus on here.

## 3.1 Settings

China has $N$ counties, indexed by $i$. $L^i$ denotes county $i$'s land endowment. In each county, individual farmers with different farm sizes can choose to grow either grain or cotton. The productivity of growing each on one unit of land is denoted as $A_g^i$ and $A_c^i$, respectively. We postulate that there is no fixed cost associated with growing grain. However, producing cotton requires an annual fixed cost $f$, and a farmer can choose from among two feasible options: to grow the low-quality native variety or introduce the high-quality American cotton.

We assume that each county functions as a small open economy. Our model takes the national prices of products as exogenously given, subject to the trade costs across counties. If a farmer chooses to grow the native cotton, he can sell it in the regional market at a constant price, $p_t$, with an iceberg transportation cost $\kappa > 1$; if he decides to upgrade the technology and grow the American variety, he can sell it in the national market at price $p_a + \varepsilon$, with an iceberg transportation cost $\tau > \kappa$, where $\varepsilon \sim N(0, \sigma^2)$ denotes the price fluctuations. For simplicity, we focus on the meaningful case where $p_a > p_t$.

A farmer's utility function is defined as:

$$E(u) = E\left(u(g^\alpha c^{1-\alpha})\right),$$

where $g^\alpha c^{1-\alpha}$ is the Cobb-Douglas demand function, and $u$ is a concave function, e.g. $u = 1 - e^{-\gamma(g^\alpha c^{1-\alpha})}$ with $\gamma > 0$, which indicates that the farmers are risk averse and they dislike uncertainty; $g$ is the consumption of grain, and $c$ is the consumption of cloth.

Recall that the domestic raw cotton was spun and woven locally into cloth for rural household consumption (Section 2.2). We assume that cloth is produced at the immediate regional market, and the production function of cloth is given by:

$$y_c = y_t,$$

where $y_c$ is the output of cloth, and $y_t$ is the input of tree cotton. For simplicity, in our model cloth can be produced from cotton with a productivity of 1. Thus, the price of cloth

$p_c = p_t$ and the price for the consumption of cloth for the farmers is $\kappa p_t$.

Each farmer's budget constraint is defined as:

$$p_g g + \kappa p_t c = I,$$

where $p_g$ is the price of grain and $I = p_g y_g + p_t y_t$ is the farmer's total income. Given the greater price volatility, a farmer's income is less certain if he chooses to upgrade to the high-quality American cotton.

Maximizing the utility function subject to the budget constraints and production technologies gives the usual marginal rate of substitution (MRS) equation:

$$\frac{g}{c} = \frac{\kappa\alpha}{1-\alpha}\frac{p_t}{p_g}.$$

## 3.2 Equilibrium

We denote the total consumption of grain and cloth in county $i$ as $G^i = \sum g$ and $C^i = \sum c$, respectively; $I^i = \sum I$ is county $i$'s total income. Aggregating the MRS condition and making use of the market clearing condition of goods, we obtain:

$$\begin{aligned} p_g G^i &= \alpha I^i, \\ \kappa p_t C^i &= (1-\alpha) I^i, \end{aligned}$$

which yields:

$$\frac{G^i}{C^i} = \frac{\kappa\alpha}{1-\alpha}\frac{p_t}{p_g}.$$

We now consider the model under different scenarios and characterize the related equilibrium conditions. We show how the equilibrium changes as the national market becomes increasingly integrated and the price is more certain with the expansion of the transportation and ICT infrastructure.

- **No Trade Between Counties**

We start with a closed economy in which there is no trade between counties. In this case, no farmer will choose to grow American cotton because they cannot trade with other regions. Therefore the total consumption of grain and cotton, $G^i$ and $C^i$, are equal to the total production of grain and cotton in county $i$. Thus:

$$\frac{G^i}{C^i} = \frac{\kappa\alpha}{1-\alpha}\frac{p_t}{p_g} = \frac{A_g^i L_g^i}{A_c^i L_c^i},$$

where $L_g^i$ and $L_c^i$ are the total land used for growing grain and cotton, respectively. In the closed economy equilibrium, the farmers are indifferent between growing grain or cotton. This condition equates the two income levels, which we use to obtain:

$$\frac{A_c^i}{A_g^i} = \frac{\kappa p_g}{p_t}.$$

Recall that the total amount of land made available for grain and cotton production is fixed:

$$L^i = L_g^i + L_c^i.$$

Given the relative price, $\frac{p_t}{p_g}$, the land allocation between grain and cotton, is stated as:

$$\begin{aligned} L_g^i &= \alpha L^i, \\ L_c^i &= (1-\alpha) L^i. \end{aligned}$$

- **Specialization With Trade**

Infrastructure development decreases the transportation costs between counties and makes trade between them possible. Each county is an open economy that is assumed to be too small to influence the national price. A county's comparative advantage in producing cotton (measured as $\frac{A_c^i}{A_g^i}$) is denoted by:

$$\frac{A_c^i}{A_g^i} > \frac{\kappa p_g}{p_t}.$$

An equilibrium in this open economy has two main features. First, farmers in counties with a greater comparative advantage in cotton production will grow more cotton than those in the closed economy case because $\frac{p_t}{p_g}$ becomes larger for these counties after they open to trade. Farmers who were indifferent between growing grain or cotton before trade is opened will switch to growing cotton after the trade opening. Second, only farmers with a land endowment $l$ that is large enough to satisfy the following condition will produce cotton:

$$\left(\frac{p_t}{\kappa} A_c^i - p_g A_g^i\right) l > f.$$

When $\kappa$ decreases, the threshold of comparative advantage in planting cotton, $\frac{\kappa p_g}{p_t}$, decreases. The threshold of minimal land endowment for producing cotton is:

$$\frac{f}{\frac{p_t}{\kappa} A_c^i - p_g A_g^i},$$

which decreases with $\frac{p_t}{p_g}$ and $\frac{A_c^i}{A_g^i}$ for a given $\frac{f}{p_g A_g^i}$. The threshold condition thus has the simple interpretation that counties with a greater comparative advantage in cotton production will plant more cotton. When the national price of cotton increases, more farmers will choose to plant cotton over grain.

- **Technology Upgrading in Cotton Production**

Cotton farmers can choose to upgrade their technology from tree cotton to American cotton for a higher expected price of $p_a > p_t$. Yet two important hurdles could make such an upgrade unprofitable. First, they need to ship the product to the national market, which incurs a transportation cost of $\tau > \kappa$. Second, information frictions create a great deal of uncertainty surrounding the price of future output (American cotton). The condition for quality upgrading is:

$$\int u\left(B\frac{p_a+\varepsilon}{\tau}\right) f(\varepsilon)\, d\varepsilon > u\left(B\frac{p_t}{\kappa}\right). \tag{$\star$}$$

where $B = \left(\frac{\alpha}{p_g}\right)^{\alpha}\left(\frac{1-\alpha}{\kappa p_t}\right)^{1-\alpha} A_c^i l$ is a constant for a given farmer.

A necessary condition of quality upgrading is:

$$\frac{p_a}{\tau} > \frac{p_t}{\kappa}.$$

Since farmers are risk averse, they will only upgrade their technology if they expect to earn more from growing American cotton than tree cotton. The results of this section are summarized in the following propositions.

**Proposition 1** *In counties with more suitable planting conditions for cotton, or where the average farm size is larger, more farmers will, holding all else constant, choose to upgrade from tree cotton to American cotton when it becomes available.*

*Proof:* $l_i^*$ denotes the threshold of land endowment that equalizes the utility of quality upgrading (American cotton) and the status quo (tree cotton). At the equilibrium cutoff, a marginal farmer is indifferent between quality upgrading and not upgrading. To solve the equality ($\star$) in county $i$, it is straightforward to show that $A_c^i l_i^*$ will be the same across counties with the same $p_a$, $p_t$, $\tau$ and $F(\varepsilon)$. Hence, $l_i^*$ decreases with $A_c^i$. This implies that counties with higher $A_c^i$ or farmers with larger farms can more easily meet the critical level for quality upgrading. Q.E.D.

Cotton tends to be cultivated by farmers with sufficient land $l_i$ or those in counties with a comparative advantage ($\frac{A_c^i}{A_g^i}$). In the absence of infrastructure development, the costs of transporting goods to the national market and information frictions are substantial. Only a subset of well-off farmers endowed with even more land or from counties that are much more suitable to cotton cultivation will upgrade to American cotton.

**Proposition 2** *When the transportation costs decrease with newly constructed railways, upgrading from tree cotton to American cotton becomes more profitable for farmers. Furthermore, the suitability threshold for American cotton cultivation decreases; the threshold level of the land endowment for adopting American cotton also falls.*

*Proof:* When $\tau$ decreases, the left-hand side of inequality ($\star$) increases. $B(l_i^*) = \left(\frac{\alpha}{p_g}\right)^{\alpha}\left(\frac{1-\alpha}{\kappa p_t}\right)^{1-\alpha} A_c^i l_i^*$ reduces accordingly to accommodate a less concave $u(\cdot)$ function. Thus $A_c^i l_i^*$ strictly increases with $\tau$. When trade costs $\tau$ decline, the suitability and land endowment thresholds of cotton cultivation decrease and more farmers will choose to upgrade. Q.E.D.

This proposition can be conveyed by an intuitive interpretation. The railway connects more counties at a lower cost to facilitate trade in the national market, where the high-quality American cotton is valued substantially more than the native tree cotton. This causes the expected profit of American cotton cultivation to increase. Adopting this technology would then become beneficial to farmers with smaller farms or in counties less suitable to growing cotton, leading to a larger increase in American cotton cultivation along the railway lines.

**Proposition 3** *When the improved ICT infrastructure (telegraph system) reduces the uncertainty associated with the national price of American cotton, the suitability threshold of American cotton cultivation decreases, as does the threshold of the average land endowment for adopting American cotton.*

*Proof:* When $\sigma$ decreases, there is less uncertainty about the price of American cotton in the national market. The left-hand side of inequality ($\star$) increases. As shown in Appendix Figure A.4, a smaller risk premium, denoted by the gap between $B\frac{p_a}{\tau}$ and $B\frac{p_t}{\kappa}$, is required to compensate farmers for the price uncertainty they face while upgrading their cotton variety. Hence, $B(l_i^*)$ decreases for the marginal farmer, causing a fall in $A_c^i l_i^*$. The adoption threshold levels of cotton cultivation suitability and land endowment will fall, causing more farmers in the county to plant American cotton. Q.E.D.

Intuitively, the development of the telegraph improved the amount of information available to the newly connected counties about the conditions in the national market for premium quality cotton. More precise and recent information greatly increases the benefits for smaller farms and counties with less suitable growing conditions of adopting the new technology (American cotton), which is derived from the concavity of the utility function—risk aversion.

# 4 Data Construction

We compiled information from several sources to create a new dataset that provides comprehensive information on infrastructure access and cotton planting in 1,660 Chinese counties from 1920 to 1937. We complement this data with measures of adoption thresholds.

## 4.1 Railway and Telegraph

We collected data from three principal sources. We first digitized and geo-referenced a map of China's railroad network from 1953, which was maintained by the United States Library of Congress. We then collected information on what year each section of the railway lines started to operate trains from *the History of Railroad Network Construction (1876–1949)* (Zhang, 1997) and *the Collection of Republican Period Archives on Railroad*, edited and published by the National Library Press (Cao, 2013). We then synthesized this information to create a collection of GIS maps on the annual development of railway lines from 1876 to 1953.

By spatially merging the railway maps with the county boundary map in ArcGIS, we constructed a key variable in our county-level panel dataset: railway access. We defined a county as being reached by the railway if it is within 50 km of the railway network in a given year.

We gathered data on each county's access to the telegraph network from two historical sources. First, we used *the Telegraph Network Map of 1938* compiled by the Transportation Bureau to construct a corresponding GIS map of telegraph lines. Second, we collected data on telegraph stations from Chinese county gazetteers published in the 1980s and 1990s. The "Telecommunication Infrastructure" section of the local gazetteers provides systematic and detailed information on the historical development of various types of telecommunication infrastructures in the county, including telegraph stations. We hand coded the year that each county in the sample received its first telegraph station.

Combining data from both sources, we then constructed another key dummy variable in our county-level panel that equals 1 if a county had established a telegraph station in a specific year, and 0 otherwise. Table 2 reports the summary statistics of the pertinent variables. About 47% of the 1,660 counties in China Proper established a telegraph station between 1920 and 1937. Around 33% of county-year observations had railway connections. But only 18% of counties had access to both types of infrastructures during the sample period.

[Table 2 about here]

## 4.2 Cotton Cultivation in China

We gathered data on the diffusion of American cotton across China from county-level statistics compiled by the Chinese Cotton Industry Statistics Association (*Zhonghua Mianye Tongjihui*). This includes detailed information on all the cotton regions in China (13 provinces and 786 counties) from 1920 to 1937, and documents each county's planting area and yields of both American and native cotton.

We use this data to organize our analysis around two dimensions of American cotton adoption: (1) the *extensive margin*, denoted by a dummy variable that equals 1 if a county was recorded as producing American cotton in a specific year and (2) the *intensive margin*, which is a county's logged planting area for American cotton in a specific year. During our sample period, 14.2% of counties had ever planted American cotton and the average area cultivated with this variety was 3.51 ($e^{1.255}$) 10,000 acres per county.

As a robustness check, we also constructed variables to measure a county's adoption of Chinese cotton, in parallel with American cotton. As illustrated in Table 2, 17.2% of counties in our sample had produced Chinese cotton and the average area of cultivation was 5.12 ($e^{1.634}$) 10,000 acres—more than that of American cotton in a naïve comparison.

## 4.3 Thresholds for Adoption: Suitability and Land Endowment

To test the model's predictions that railway and telegraph access promoted the adoption of American cotton by reducing the take-up thresholds—agricultural suitability and land endowment—we use two variables to proxy for them in cross-section. The first is a county's comparative advantage in the production of cotton relative to grain, and the second is a county's average household farm size. For the former, we collected data on Agricultural Suitability and Potential Yields from the Global Agro-Ecological Zones (Gaze) Database.[6] At a granular geographic level, Gaze data assesses the suitability for cultivating all major staple and economic crops from 0 (least suitable) to 100 (most suitable). By merging the Gaze data with the GIS map of China, we obtained county-level average suitability indices for cultivating cotton and four other main staple crops—rice, wheat, sorghum and maize. We then constructed a normalized measure of a county's relative suitability for cultivating cotton by dividing the suitability index of planting cotton by the average suitability of planting other staple crops.

While data constraints prevented us from constructing a household-level normalized mea-

[6]This database is constructed and maintained by the UN Food and Agriculture Organization. It provides crop-specific suitability indices—based on soil, slope, and climatic characteristics—in 0.5*0.5 grid cells covering all land on Earth.

sure, we were able to collect cross-sectional data on each county's average household farm size from an agricultural survey conducted by the Bureau of Statistics of the Directorate-General of Budget Accounting and Statistics (*Xingzhengyuan Zhujichu Tongjiju*). Published in the Report of Agricultural Statistics in Various Chinese Provinces (*Gesheng Nongye Gaishu Guji Zongbaogao*, 1932), the survey contains comprehensive agricultural statistics for 1,498 counties from China Proper in 1931. Similar to the previously described procedure, we digitized the data and merged it with our county panel for further empirical analyses. Table 2 provides the summary statistics for these two threshold variables for adoption, which illustrates considerable variation across counties.

# 5 Effects of Railways and Telegraph on Technology Adoption

In this section we analyze the causal relationship between infrastructure access and technology adoption. A key challenge associated with our identification strategy is that a county's connections to railway and/or telegraph networks may be correlated with unobserved determinants of American cotton adoption. We address this critical issue in multiple ways. First, we apply a DID model to county-level panel data to use as our baseline specification. Second, building on recent advances in the econometrics of DID, we present new estimators for our staggered case. Third, we introduce two instrumental variables for railway and telegraph access to address endogenous infrastructure route decisions.

## 5.1 Difference-in-differences Estimates

First, we formally test the effect of railway and/or telegraph access on the adoption of a new agricultural technology (American cotton) using a standard DID design. Of the 1,660 counties in China Proper (18 provinces) between 1920 and 1937, we exploit variation in the timing of the counties' access to the railway and/or telegraph using the following baseline specification:

$$Cotton_{it} = \alpha + \beta Railway_{it} + \gamma Telegraph_{it} + \sigma Both_{it} + \lambda_i + \mu_t + \epsilon_{it}. \quad (1)$$

where the explanatory variable $Railway_{it}$ equals 1 if county $i$ was within a 50 km radius of the rail routes in year $t$, and $Telegraph_{it}$ is a dummy variable that equals 1 if county $i$ had established a telegraph station in year $t$. $Both_{it}$ is a a dummy variable that indicates whether county $i$ was within 50 km of the railway network *and* had a telegraph station in year $t$. The coefficient $\sigma$ allows us to test whether there is any interaction effect between two types of infrastructures conditional on the individual effect of each, $\beta$ and $\gamma$, respectively. For the key outcome of interest—the adoption of—American cotton, we use two dependent variables: (1) an adoption dummy indicator to measure the extensive margin and (2) the logarithm of the area planted with American cotton in county $i$ in year $t$ to measure the intensive margin. As a robustness check, we repeat the analysis on analogous outcomes of Chinese cotton. All specifications include county and year fixed effects. To control for potential heteroskedasticity, standard errors are clustered at the county level.

Each column of Panel A in Table 3 reports the reduced-form estimates by adding variables indicating railway and telegraph access sequentially on the adoption dummy (columns (1)–(4)) and the total size of the planting area (columns (5)–(8)). Column (3) shows that railway access increased a county's probability of adopting high-quality American cotton by 11.8%, and access to the telegraph network raised the likelihood of adoption by 5.5%. Regarding the cultivation area, the corresponding impacts are 240% (i.e., $exp(1.224) - 1$) and 51.9%, as shown in column (7). Columns (4) and (8) add the dummy indicator denoting access to both infrastructures. The results suggest that beyond the beneficial effects of either type of infrastructure, counties with both railway and telegraph access were 9.4% more likely to adopt American cotton and exhibited a two-fold increase in the size of the planting area, pointing to strong complementarities between transport and ICT infrastructure.

[Table 3 about here]

Panel B reports the analogous results for Chinese cotton. Access to the telegraph network had no effect on the adoption of native cotton at either the extensive or intensive margins. Although substantially smaller than those for American cotton, railway access appears to have some positive effects on both margins at 4.1% and 34.2%. Here, we suggest caution when interpreting the ordinary least squares (OLS) estimates of railway access on Chinese cotton adoption. We discuss the details of the IV estimates (the estimated value of the coefficients falls and becomes insignificant) in the next subsection. As mentioned in Section 2, the low fiber density and quality of the Chinese cotton made it difficult to earn a sufficient profit to offset the cost of long-distance trade. This helps explain why the adoption of Chinese cotton was less sensitive to the reduction in trade costs and did not respond to diminished information frictions. Moreover, the two cotton species serve very different markets, which might explain why American cotton did not crowd out its native counterpart. All told, these findings suggest that infrastructure has large positive effects on the adoption of American cotton, and has no analogous effects on the native variety.

To verify the parallel time trend assumption for the DID estimation, we test whether a county's adoption of American cotton occurred in correspondence with getting connected to infrastructure networks. We estimate the specification as in columns (3) and (7) of Table 3, substituting $Railway_{it}$ and $Telegraph_{it}$ with a full set of year-wise dummies going from 10 years before becoming connected to the infrastructure network to 10 years after. Figure 4 displays the estimated coefficients and 95% confidence intervals.

[Figure 4 about here]

The upper panel depicts the coefficients of railway access on the adoption dummy and planting area of American cotton, respectively, while the lower panel displays the results for telegraph access. The analysis reveals no clear signs of a deterministic trend in American cotton adoption before a county had access to either type of infrastructure. However, positive and significant effects gradually emerged after a county was connected to infrastructures, particularly the railway. These results boost our confidence in the validity of our identification strategy, lending support to the causal effect of infrastructure on technology adoption.

Goodman-Bacon (2021) highlights that the standard two-way fixed effects (TWFE) estimator for a staggered DID design could be biased when treatment effects vary across groups and over time.[7] To address this potential bias, Callaway and ant'Annab (2021) proposes a non-parametric estimator, the group-time average treatment effect (ATT)—a unique ATT for a cohort of units treated at the same time. We use their method to re-estimate the effects of railway and telegraph access on the adoption of American cotton. Figure 5 plots the coefficients of both types of infrastructure on both margins of adoption for 10 years before getting access to 10 years after. The upper panel illustrates that railway access, denoted by near railway (<=50 km), has a quantitatively large and robust positive effect on a county's adoption and planting area of American cotton. Further, there was no pre-trend before a county became connected to the railway network. However, the lower panel shows no significant effect of connecting to the telegraph on either margin over time after adjusting the bias from TWFE.[8]

[Figure 5 about here]

[7] De Chaisemartin and d'Haultfoeuille (2020) investigate the pertinent bias of TWFE from a slightly different perspective, focusing on the negative weighting problem of the staggered treatment (for the later periods).

[8] To further test whether having access to both infrastructures has an additive effect, we depict the group-time average ATT on the dummy of $Both_{it}$, and find that both types of infrastructure have a robust, complementary effect.

## 5.2 IV Estimates

The DID estimates might still be biased due to omitted variables that determine both the expansion of the infrastructure network and the diffusion of American cotton. To rigorously address this issue, we employ an IV approach that identifies the causal effect of infrastructure access on technology adoption. With two key explanatory variables, we need two instrumental variables that can separately predict the spatial development pattern of each infrastructure and display no direct effect on the adoption of American cotton. Below we describe how each of the instrumental variables is constructed in ways that are deeply grounded in the historical facts. We also carefully justify both the instrument relevance and validity of the exclusion restriction.

First, for the railway infrastructure, in a similar vein to Hornung (2015) and Banerjee et al. (2020), we predict each county's actual railway access $Railway_i$ at different cross-sectional time points between 1920 and 1937 with the contemporaneous likelihood of railway connection. We use an instrumental variable that denotes whether a county is located along a straight-line corridor $StraightLineCorridor_i$ between major cities included on the railway network.

This IV analysis is motivated by historical evidence pertaining to the construction of railways. Chinese railways were initially built primarily to connect important cities. For instance, the Jianghan Line—the longest railway line constructed prior to 1949—connected two of the largest cities in central and southern China, Wuhan and Guangzhou, both of which were provincial capitals and early treaty ports (opened between 1827 and 1880).[9] To reduce the construction costs associated with acquiring land and building tracks and stations, and with future operation and maintenance in mind, new lines to further expand the rail network were generally built along straight lines between key node cities. Counties located along these straight-line corridors were thus more likely to be connected. The rare exception is when there are natural geographical obstacles such as lakes, rivers, and hills along the straight line corridor, which could introduce a random component into our instrument. However, the actual railway routes may occasionally deviate from straight lines for some endogenous reasons.[10] Nevertheless, the nearest distance from each county's centroid to the straight line between two major rail hubs was still an important factor determining its access to the railway. The geographic pattern of rail construction can provide variation

[9] We choose the end year at 1880 as the next year (1881) was the year that the first railway in China opened to use.

[10] For example, the Yuehan line—the main railway line connecting two major cities in southern China, Guangzhou and Wuhan—was initially designed to be constructed along a straight line route through Jiangxi Province. However, political elites from neighboring Hunan Province lobbied the central government to construct the line through Hunan Province instead.

in railway connections that is plausibly exogenous to underlying determinants of American cotton adoption. Below we construct this distance instrument for the actual rail connection.

Using the GIS technique, we first connect the major cities (nodes) with straight lines. Following the lead of Banerjee et al. (2020), we define all provincial capitals and treaty ports of the late Qing Dynasty as major cities (see Figure 6). We then construct a corridor by buffering the straight line by 1.5 km in each direction. Finally, we construct the $StraightLineCorridor_i$ variable by assigning a value of 1 to each county $i$ if its centroid is within any corridor, and 0 otherwise.

[Figure 6 about here]

Second, for the telegraph infrastructure, we use data about the contemporaneous availability of a post office as an instrument for telegraph connection. Tracing the history of the telegraph back to the establishment of the National Telegraph Office in 1880, we find that the expansion of the telegraph network in China relied heavily on the existing network of post offices and lines. A primary reason was that the provision of telegraph services relied on local postal workers. For instance, the postman would deliver telegram messages sent through the local telegraph office to the recipient. Another important reason was that since it greatly shortened the information transmission time between major cities, the telegraph was initially designed to complement the postal network, the only ICT infrastructure prior to the birth of the telegraph. Sir Robert Hart, in an official proposal for the Qing government, outlined a grand plan to develop a modern telecommunication service in 1898. He suggested connecting every post office in the treaty ports with telegraph lines and then expanding the telegraph network to closely match the entire postal network to save money and improve the network's efficiency (Zhang, 1985). Figure 7 maps the postal network in Guangdong Province as of 1907. As the figure demonstrates, in the early 20$^{\text{th}}$ century this network already connected most of the prefectures, treaty ports and important counties. The red symbols $\mp$ on the map depict the locations of telegraph stations at the time, which were at the important junctures along the postal network (represented by the thin black lines). Taken together, the contemporaneous availability of a post office in a county can serve as a sound predictor of whether it also has a telegraph station.

[Figure 7 about here]

We conduct a series of tests to assess the plausibility of the exclusion restriction for both instruments. We regress a number of socio-economic proxies before or in the early years of our sample period on our instrumental variables (i.e., the straight-line corridor indicator of the late Qing Dynasty; the postal station indicator in 1920). These pre-existing outcomes include county-level population density (columns (1)–(2) and (9)–(10)), total firm capital (columns (3)–(4) and (11)–(12)), total bank capital (columns (5)–(6) and (13)–(14)) in 1915 and 1920, and the average and nearest distance to ports (columns (7)–(8) and (15)–(16)). Table 4 confirms that no significant relationship is observed between the IVs and economic proxies before the infrastructure was built.[11] The results suggest that the two IVs provide plausibly "good" variations in railway and telegraph access that are exogenous to counties' pre-existing socio-economic determinants of technology adoption.

[Table 4 about here]

We then proceed to the IV estimations. Given that our instrumental variable for railway access is time-invariant, we begin by examining the effect of infrastructures on American cotton adoption for each of three cross sections separately: the early year of our sample, 1921, the year in between, 1928, and the last year of the sample period, 1937. The corresponding 2SLS model we estimate instrumenting for infrastructure access is given by:

$$\begin{aligned} Cotton_i &= \alpha \widehat{Railway}_i + \beta \widehat{Telegraph}_i + \epsilon_i, \\ Railway_i &= \eta StraightLineCorridor_i + \phi PostalOffice_i + \nu_i, \\ Telegraph_i &= \pi StraightLineCorridor_i + \rho PostalOffice_i + \mu_i. \end{aligned} \tag{2}$$

The corresponding first- and second-stage results are reported in Panels A and B of Table 5. Columns (1)–(2) and (7)–(8) show the result for 1921, while columns (3)–(4) and (9)–(10) show the result for 1928, and columns (5)–(6) and (11)–(12) for 1937. Looking across columns, Panel A confirms that counties located along the straight line corridor are indeed significantly more likely to have railway access. Likewise, counties with a postal station are more likely to have a telegraph station. Thus, instrument relevance is demonstrated. In Panel B, columns (7)–(8) show that counties with railway access are 13.1% more likely to grow American cotton and to increase the total size of the cultivation area by two-folds (i.e., $exp(1.124) - 1 = 2.08$) in 1921. In contrast, the effect of telegraph access on the extensive margin is only marginally significant with a smaller magnitude of 5.7%. Further, columns (9)–(12) present evidence that the size of the effect of railway access appears to increase over time, but the effect of telegraph access is small and statistically insignificant in later years.

Next, Panel C repeats the IV regressions using the panel data between 1920 and 1937. To accommodate the time-invariant instrumental variable for railway access, we remove county fixed effects but still include year fixed effects and cluster the standard errors at the

[11] Since each province completed the 1915 population census in different years, to control for the potential confounding effect, we control for province fixed effects in the regressions reported in columns (1) and (9).

county level as in our baseline specification. Consistent with our cross-sectional IV estimates, columns (13)–(14) show that the two instrumental variables are positively correlated with the endogenous variables of infrastructure access. Columns (15)–(16) suggest that only railway access but not telegraph connection significantly increases both the extensive and intensive margins of American cotton cultivation. In particular, the coefficients on railway access estimated from the IV approach are about three times as large as those obtained from the OLS regression (i.e., columns (3) and (7) of Table 3).

Finally, when i) both the endogenous regressor and the instrumental variable are binary variables, and ii) the effect of the instrumental variable on the key regressor is non-monotonic, Masten and Torgovitsky (2016) suggests that the 2SLS estimation "becomes an increasingly complicated weighted average of local average treatment effects". They propose an alternative approach, namely, "instrumental variable correlated random coefficients (IVCRC)". We apply their technique to our study, allowing for such treatment effect heterogeneity. As reported in columns (17)–(18) of Panel C, the IVCRC estimations produce results very similar to those from the OLS estimations: 0.156 versus 0.118 for the effect of railway access on the adoption dummy, 1.412 versus 1.224 on the cultivation area, while 0.033 versus 0.055 and 0.390 versus 0.519 for the effects of telegraph connection on analogous outcomes. Infrastructure access does show positive effects on both the extensive and intensive margins of American cotton cultivation. In term of the magnitude, a county's access to railway increases the likelihood of adoption and the area of cultivation by 15.6% and three folds, respectively. Telegraph connection increases them by 3.3% and 39%. Overall, various estimation approaches that exploit different assumptions yield qualitatively similar results.

In a similar vein to the baseline, we also estimate the impacts of both infrastructures on the cultivation of Chinese cotton as placebo tests. Appendix Table A.1 reports the results, confirming no significant effects on either the adoption or planting area of Chinese cotton.

[Table 5 about here]

# 6 Mechanisms and Welfare Implications for Technology Adoption

In this section, we elucidate the channels through which infrastructures enhance the adoption of new agricultural technologies such as American cotton. We then quantify the welfare gains from infrastructures through technology adoption, an additional margin that has not yet been explored in the trade literature.

## 6.1 Infrastructure, Trade Costs and Information Friction

An important hurdle to both international and domestic trade is the physical cost of transportation (i.e., trade costs).[12] Transportation infrastructure such as railways has been proved to be a powerful instrument that effectively reduces trade costs. Using detailed railway network and price information from colonial India, Donaldson (2018) reveals large benefits of railways including reducing trade costs, increasing trade volumes and improving market integration. Donaldson and Hornbeck (2016) quantify the aggregate impact of railway network expansion on the American agricultural sector in 1890. Railroads are shown to have played an irreplaceable role in increasing counties' market access and ultimately yielding substantial economic gains.[13]

Beyond transportation costs, Steinwender (2018) proposed information frictions as another critical factor hindering market integration and trade. Her study explored the historical establishment of the transatlantic telegraph in 1866, which connected New York, a key trading center in the US at the time, and Liverpool, a manufacturing hub in England. Access to the telegraph network generates sizable welfare gains by improving the efficiency of information transmission, reducing uncertainty in transactions and thus facilitating international trade.[14]

Motivated by these earlier works, in the heterogeneous agent model we lay out the key premises through which infrastructures induce the adoption of new technology: reducing trade costs and mitigating information frictions. Therefore, before investigating how these channels influenced technology adoption, we seek to establish whether railways and the telegraph network helped China reduce trade costs and information frictions in the early 20th-century. To do so, we digitize comprehensive domestic trade data from a commodity census conducted by the General Post Office of Ministry of Transportation (*Jiaotongbu Youzheng Zhongju*) in 1934.[15] The census contained detailed transaction-level information on more than 76 types of commodities including cotton at the origins, destinations, trade volumes and prices. We use this source to construct a cross-sectional gravity data set on

[12] The pioneering work of Ramondo et al. (2016) emphasizes the importance of domestic trade costs and their impact on scale effects. Tombe and Zhu (2019) have shown how the decline in domestic trade costs can benefit a large country like China.

[13] In addition to directly reducing trade costs, Hummels and Schaur (2013) document that infrastructure developments can also reduce trade *time*.

[14] In a similar vein, Ejrnes and Persson (2010) demonstrate that the transatlantic telegraph connection between Chicago and Liverpool in the late 19th century also improved market efficiency, proxied by the convergence of prices to the law of one price equilibrium. Along these lines, a number of previous studies exploited exogenous variation in information frictions such as the introduction of cell phone (Jensen, 2007; Acharya et al., 2016) and internet kiosks (Goyal, 2010). All confirm significant welfare improvement that results from greater market efficiency and a reduction in information frictions.

[15] This data was published in 28 volumes titled *Zhongguo Tongyou Difang Wuchanzhi*.

domestic trade. In turn, we examine the effects of a county's access to railway and telegraph networks on a set of outcomes (in logs) including trade volume, price spread (defined as the gap between the highest and lowest prices for the same commodity traded between any pair of counties) and average price.

Table 6 reports the results for all commodities aggregated (Panel A) and for cotton (Panel B). Column (1) of Panel A illustrates that access to either type of infrastructure significantly increased a county's trade volume. Column (2) shows that the effect of telegraph connections seems to be largely driven by county pairs with access to both infrastructures. Similarly, column (7) of Panel B demonstrates that railway (but not telegraph) access significantly increased the volume of cotton traded. Column (8) shows that the pair of counties that have access to both infrastructures traded much larger volumes of cotton, echoing the findings of all commodities aggregated. Taken together, we offer evidence that the railway and telegraph infrastructures had a substantial effect on reducing trade costs, reflected in the greatly improved trade flows across counties. As for the information frictions, column (3) of Panel A shows that either railway or telegraph access narrowed the price spread for all types of commodities. However, column (4) confirms that the effect is again only driven by county pairs with both types of infrastructure. Columns (9) and (10) present evidence that this is also the case for the price spread of cotton. Lastly, Columns (5)–(6) and (11)–(12) of Panel B show that infrastructure access does not display any significant effects on the average price of all commodities or for cotton specifically. Although the reduction of trade costs could potentially drive down prices, we conjecture that county pairs with infrastructure access may also be more likely to trade products of higher quality (such as American cotton). As a result, the infrastructures' net effect on price could well be ambiguous in the absence of an accurate control variable for product quality.

[Table 6 about here]

## 6.2 Mechanisms for Technology Adoption

Having verified the crucial role of both railway and telegraph infrastructures in reducing trade costs and information frictions, we now examine the mechanisms by which access to these infrastructures could induce technology adoption. We first test propositions (1)–(3) of our model on how comparative advantages in different locations are related to American cotton adoption without and with infrastructure. To do so, we use a normalized county-level indicator of the agricultural suitability of cotton relative to grain, computed as the ratio of the suitability of planting cotton to the average suitability of planting grain crops including rice, wheat, sorghum and maize. We then interact this ratio with the dummy variables of infrastructure connections sequentially and include them as regressors in the OLS estimation; the outcome variables are the two measures of American cotton adoption. Our regression specification is as follows:

$$Cotton_{it} = \phi Railway_{it} + \eta Railway_{it} * SuitabilityRatio_i + \theta Telegraph_{it} + \varphi Both_{it} + \upsilon_i + \nu_t + \varrho_{it}. \quad (3)$$

Table 7 reports the estimation results. Columns (1)–(4) present the effects of the adoption dummy indicator, while columns (5)–(8) present the effects on the size of the cultivated area. By including the time-invariant suitability ratio, columns (1) and (5) modify the specification by temporally removing county fixed effects. The results confirm that a county's comparative advantage in cotton growing is positively associated with new technology adoption at both margins. Columns (2)–(4) and (6)–(8) add the interaction terms between the suitability ratio and the infrastructure access indicator into the regression, which are the key regressors of interest underpinning the mechanism. The negative and significant coefficients on the interaction terms across columns indicate that counties with a lower comparative advantage in cotton growing experienced greater increases in the adoption of American cotton after they gained access to railway and telegraph networks. In other words, transport and ICT infrastructures greatly reduced the thresholds for counties to take up the new agricultural technology. Appendix Table A.2 shows that the results are robust to alternative measures of cotton cultivation suitability (i.e., not divided by the suitability indices of other staple crops).

[Table 7 about here]

Next we test the model's propositions on the take-up threshold of a farmer's endowment in terms of landholding. We use a county-level proxy of average farm size per household dated in 1931 as previously described. We regress the American cotton adoption outcomes on the interaction terms between the average farm size and the access of different infrastructures along with other control variables. The specification is the following:

$$Cotton_{it} = \phi Railway_{it} + \eta Railway_{it} * FarmSize_i + \theta Telegraph_{it} + \varphi Both_{it} + \upsilon_i + \nu_t + \varrho_{it}. \quad (4)$$

Table 8 reports the results with farm size as another proxy for technology adoption threshold. Columns (1) and (5) control for the main effect of farm size without including the county fixed effects and find a positive relationship between the farm size per household and the technology adoption. Reassuringly, in columns (2)–(4) and (6)–(8), the negative coefficients of the interaction terms confirm that access to either type of infrastructures boosted technology adoption particularly for those with a smaller farm size. The patterns documented here are broadly consistent with the findings in Table 7.

[Table 8 about here]

## 6.3 Welfare Gains through Technology Adoption

We employ the unified "sufficient statistics" framework developed by Arkolakis et al. (2012) to evaluate the direct welfare gains from trade cost reduction.[16] Donaldson (2018) has shown in the context of colonial India that sufficient statistics—a proxy for direct welfare gains from enhanced trade—can explain 52% of the total welfare impact of the railway network expansion, while a substantial share of benefits remains unexplained. Most notably, our paper explores a new channel of welfare gains from infrastructure development—the income effect through the adoption of new technologies. With the expansion of the railway and telegraph networks, farmers in connected counties switched from planting native cotton to adopting the high-quality American variety that consequently increased their income. We show that this mechanism can account for a large proportion of the "unexplained" welfare gains from infrastructure development.

To construct our key proxy for farmer welfare, we apply the methodology developed in Donaldson (2018) and calculate each county's real agricultural income in each year (denoted by $W_{it}$) as follows:

$$W_{it} = \frac{I_{it}}{p_{g,it}^{\alpha} p_{c,it}^{1-\alpha}}.$$

The measure of nominal income $I_{it}$ is the sum of the product of a county's agricultural outputs of cotton, rice and wheat, and their prices. We use the cultivated area of these three crops to calculate output.[17] We then compile the corresponding crop-level price information from sources including the Pre-war Government Archives on Consumer Price (*Kangzhanqian Jiage Cankao Ziliao*) and a Commodity Census Covering All Postal Accessed Counties (*Tongyou Wuchan Zhi*) conducted around 1934.[18] Finally, for the Cobb-Douglas expenditure shares of grain ($\alpha$) and cloth ($1-\alpha$), we obtain data between 1933 and 1935 on China's total production of the two goods from Yan (1955), and import and export values from the Customs statistics.

[16] This framework essentially captures the term-of-trade gains and has been widely employed by the majority of new trade models.

[17] Because data on the total cultivated area and the area growing cotton are available only at the county level, we use the caloric suitability of rice and wheat to construct the crop-specific cultivated areas.

[18] The price data documents information for a different set of counties across different years. For those without prices, we impute the missing values using those from the nearest county with available information.

We calculate the representative citizen's average expenditure share of each product during the contemporaneous period.

Using the real income construct, we examine the crude effects of either infrastructure on farmers' welfare (in logs). In Panel A of Table 9, columns (1)–(2) report the OLS estimation results for each infrastructure indicator, respectively. We obtain similar results if we include both infrastructure indicators in the regression. Columns (3)–(4) report the IV estimation results. Here we use the straight-line corridor (postal station availability) as the IV for railway (telegraph) access separately, conditioning on the other infrastructure indicator as the control variable in the 2SLS estimation. We choose this particular specification to be consistent with those of the proceeding mediation analysis.[19] Column (1) shows that a county's railway access has a significant effect on enhancing farmers' welfare: it increases income by 8.4%. In column (3), the IV estimate pertaining to the welfare gains from railway access is 46.4%, again indicating that railway access plays a crucial role in improving welfare.[20] However, access to the telegraph network alone does not have the same significant effect on farmers' welfare. Yet when combined with railway access, telegraph access can almost double the welfare gains from railway access alone (additional OLS regression results available upon request). This finding is broadly consistent with our baseline DID and IV estimation results, which illustrate that telegraph access had an insignificant effect on the adoption of American cotton. Without a substantial reduction in trade costs, it was virtually impossible for inland farmers to ship the premium-quality American cotton to textile manufacturing hubs in coastal areas for higher potential income.

[Table 9 about here]

In a mediation analysis, we estimate the welfare effect of the mediator factor—the adoption of American cotton, a possible channel through which the treatment (infrastructure access) can influence the real income. Columns (5) and (6) of Panel B regress the welfare measure on the extensive and intensive margins of adoption. Columns (7) and (8) then isolate the direct effects of the two infrastructures on welfare by excluding the mediator effect. For the latter, we adopt the two-step approach proposed by Acharya et al. (2016) to estimate the average controlled direct effect of each infrastructure.[21] Using this estimation technique, we carefully address the possibility of intermediate variable bias. In the first step, we regress

[19]As discussed in more detail later, the mediation method applies to a treatment of interest rather than multiple treatments simultaneously. So we need to modify our estimation framework by focusing on separate estimates for each infrastructure access, conditioning on the other infrastructure indicator.

[20]Note that the estimates are robust to the specification in Panel C of Table 5.

[21]See recent examples such as Bietenbeck (2019) and Brown et al. (2019). This approach represents a substantial shift away from the previous practice of simply conditioning on (or "controlling for") the post-treatment pathway variables.

the real income of farmers in each county on the treatment (i.e., railway access for column (7) and telegraph access for column (8), respectively), the mediator–a dummy variable indicating whether American cotton is grown in the county, the interactions of the mediator with the treatment and the pretreatment confounder (i.e., relative suitability for growing cotton), and other controls (i.e., the county's area of land cultivated with Chinese cotton, the other infrastructure connection, and county and year fixed effects). Online Appendix B contains more details on the specifications. On that basis, we obtain a "de-mediated" measure of real income, which is then used as the dependent variable in the second step. We regress it on the treatment, controlling for the other network connection and fixed effects. We use a bootstrap method (over 1,000 replications) to obtain the standard error of the estimated direct effect excluding mediators. The "de-mediated" results presented in columns (7) and (8) can then be directly compared with the original effects in columns (1) and (2) with the mediator effects still in force. These estimates provide important insights into how much of the welfare gains from infrastructures can be explained by the new channel of technology adoption.

Column (5) shows that the adoption of American cotton leads to a 30.6% increase in farmers' welfare. At the intensive margin in Column (6), for counties that grew the American variety, more area devoted to cotton production is significantly associated with larger welfare gains accrued to the farmers. A 10% increase in the American cotton planting area, on average, improves a farmer's welfare by over 1.2%, which is economically large. Further, a comparison of columns (7) and (1) suggests that the direct effect of the railway almost halves after controlling for the mediation pathway. After the technology adoption channel is accounted for, the welfare impact of railway access decreases from 8.4% (the crude effect in column (1)) to 4.3% (the direct controlled effect in column (7)). By contrast, comparing column (8) to column (2), the direct welfare effect of access to the telegraph is negative, while the crude effect is not significantly different from zero. This implies that the mediation effect through the technology adoption channel still has a positive effect on welfare. The results of the median analysis largely confirm that the lion's share of the welfare-enhancing effect of infrastructure operates through the adoption of American cotton.

# 7 Conclusion

This article investigates how access to railway and telegraph networks affects technology adoption, using the introduction of a high-quality variety of American cotton in early 20$^{\text{th}}$ century China as a case study. We develop a theoretical framework that generates clean insights on how infrastructure development would affect trade costs and information frictions, and thus farmers' adoption thresholds in terms of their counties' comparative advantage (cultivation suitability) and land endowment.

Combining rich historical data, we put forward empirical analyses to test the theory. We confirm that railway and telegraph access substantially increased the adoption (extensive margin) and planting area (intensive margin) of American cotton. Beyond the direct individual effects, access to both infrastructures yields additional benefits, suggesting the existence of transport-information complementarity. We further present evidence on the key mechanisms behind these results. We show that railway and telegraph access have increased the volume of inter-regional trade and decreased price uncertainty in the presence of reduced trade costs and information frictions. Greatly improved access to the national market for premium-quality cotton and information about market conditions decrease the thresholds required for adopting the new agricultural technology. Examination of heterogeneity in the impact of railway and telegraph access reveals that counties that were less suitable for cultivating cotton and had smaller average farm sizes experienced greater increases in the adoption of the high-quality American cotton after they were connected to the infrastructure network.

To better understand the significance of the technology adoption channel, we quantify the associated welfare changes by calculating the income effect of infrastructure. We find that railway connections have considerably raised farmers' welfare–real income; approximately half of this increase can be explained by the adoption of the new technology. However, mere telegraph connections do not significantly improve welfare. This comparison highlights the salience of transport infrastructure, likely through a stronger display of trade cost reduction in the present setting, where long-distance trade in high-quality cotton is facilitated.

Our results provide strong support for policies that seek to invest in infrastructure and increase access for large proportions of the population in developing countries. Beyond revolutionizing economic exchanges with the unprecedented movement of goods and information, transport and ICT infrastructure could potentially overcome barriers to the broader adoption of beneficial technologies and ultimately improve welfare. This is an important aspect of "infrastructure for economic development" that policy makers and practitioners should take into account.

# References

**Acharya, Avidit, Matthew Blackwell, and Maya Sen**, "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects," *American Political Science Review*, 2016, *110* (3), 512–529.

**Amiti, Mary and Amit Khandelwal**, "Import competition and quality upgrading," *Review of Economics and Statistics*, 2013, *95* (2), 476–490.

**Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

**Atkin, David, Amit Khandelwal, and Adam Osman**, "Exporting and Firm Performance: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

**Banerjee, Abhijit, Esther Duflo, and Nancy Qian**, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 2020, *145*, 102442.

**Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak**, "Can Network Theory-Based Targeting Increase Technology Adoption?," *American Economic Review*, 2021, *111* (6), 1918–43.

**Beckert, Sven**, *Empire of Cotton: A Global History*, New York: Vintage, 2014.

**Bietenbeck, Jan**, "The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR," *Journal of the European Economic Association*, 2019, *18* (1), 392–426.

**Bold, Tessa, Selene Ghisolfi, Frances Nsonzi, and Jakob Svensson**, "Market Access and Quality Upgrading: Evidence from Four Field Experiments," *American Economic Review*, forthcoming.

**Brown, Ryan, Verónica Montalva, Duncan Thomas, and Andrea Velásquez**, "Impact of violent crime on risk aversion: Evidence from the Mexican drug war," *Review of Economics and Statistics*, 2019, *101* (5), 892–904.

**Buck, John Lossing**, *Land Utilization in China: Atlas*, Chicago: University of Chicago Press, 1937.

**Bustos, Paula**, "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American economic review*, 2011, *101* (2), 304–40.

**Callaway, Brantly and Pedro H.C. ant'Annab**, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**Cao, Ning**, *The Collection of Republican Period Archives on Railroad*, Beijing: National

Library of China Press, 2013.

**Chaisemartin, Clément De and Xavier d'Haultfoeuille**, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–96.

**Chen, Zhengping**, "Impacts of Republican China's Railway Constructions on the Integration of the Western Market," *Zhejiang Academia*, 2010, (3), 13–19.

**Donaldson, Dave**, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

**Ejrnes, Mette and Karl Gunnar Persson**, "The Gains from Improved Market Efficiency: Trade Before and After the Transatlantic Telegraph," *European Review of Economic History*, 2010, *14* (3), 361–381.

**Fairbank, John King and Kwang-Ching Liu**, *The Cambridge History of China. Volume 11: Late Ch'ing, 1800–1911, Part 2*, Cambridge: Cambridge University Press, 1980.

_ **and Ssu yu Teng**, "The Traditional Role of Tribute," *In Chen K.(eds). The Foreign Policy of China*, 1972.

**Feuerwerker, Albert**, "Handicraft and Manufactured Cotton Textiles in China, 1871–1910," *The Journal of Economic History*, 1970, *30* (2), 338–378.

**Fong, H.D. (Fang Xianting)**, *Cotton Industry and Trade in China*, Tientsin: Nankai Institute of Economics, 1932.

**Fuglie, Keith, Madhur Gautam, Aparajita Goyal, and William F. Maloney**, "Harvesting Prosperity: Technology and Productivity Growth in Agriculture," *Washington, DC: World Bank*, 2020.

**Gao, Pei and Yu-Hsiang Lei**, "Communication Infrastructure and Stabilizing Food Prices: Evidence from the Telegraph Network in China," *American Economic Journal: Applied Economics*, 2021, *13* (3), 65–101.

**Gollin, Douglas, Casper Worm Hansen, and Asger Mose Wingender**, "Two Blades of Grass: The Impact of the Green Revolution," *Journal of Political Economy*, 2021, *129* (8), 2344–2384.

**Goodman-Bacon, Andrew**, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

**Goyal, Aparajita**, "Information, Direct Access to Farmers, and Rural Market Performance in Central India," *American Economic Journal: Applied Economics*, 2010, *2* (3), 22–45.

**Hornung, Erik**, "Railroads and Growth in Prussia," *Journal of the European Economic Association*, 2015, *13* (4), 699–736.

**Hummels, David L. and Georg Schaur**, "Time as a Trade Barrier," *American Economic Review*, 2013, *103* (7), 2935–59.

**Jensen, R.**, "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *The Quarterly Journal of Economics*, 2007, *122* (3), 879–924.

**Juhász, Réka, Mara P. Squicciarini, and Nico Voigtländer**, "Technology Adoption and Productivity Growth: Evidence from Industrialization in France," *NBER Working Paper, No. w27503*, 2020.

**Masten, Matthew A. and Alexander Torgovitsky**, "Identification of Instrumental Variable Correlated Random Coefficients Models," *The Review of Economics and Statistics*, 2016, *98* (5), 1001–1005.

**Mi, Rucheng**, *Zhongguo Jin Dai Tie Lu Shi Zi Liao: 1863-1911 (Source Materials on the History of Railways in Modern China, 1863-1911)*, Beijing: Zhonghua shu ju, 1963.

**Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez**, "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, 2016, *106* (10), 3159–84.

**Steinwender, Claudia**, "Real Effects of Information Frictions: When the States and the Kingdom Became United," *American Economic Review*, 2018, *108* (3), 657–96.

**Tombe, Trevor and Xiaodong Zhu**, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–72.

**Verhoogen, Eric A.**, "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

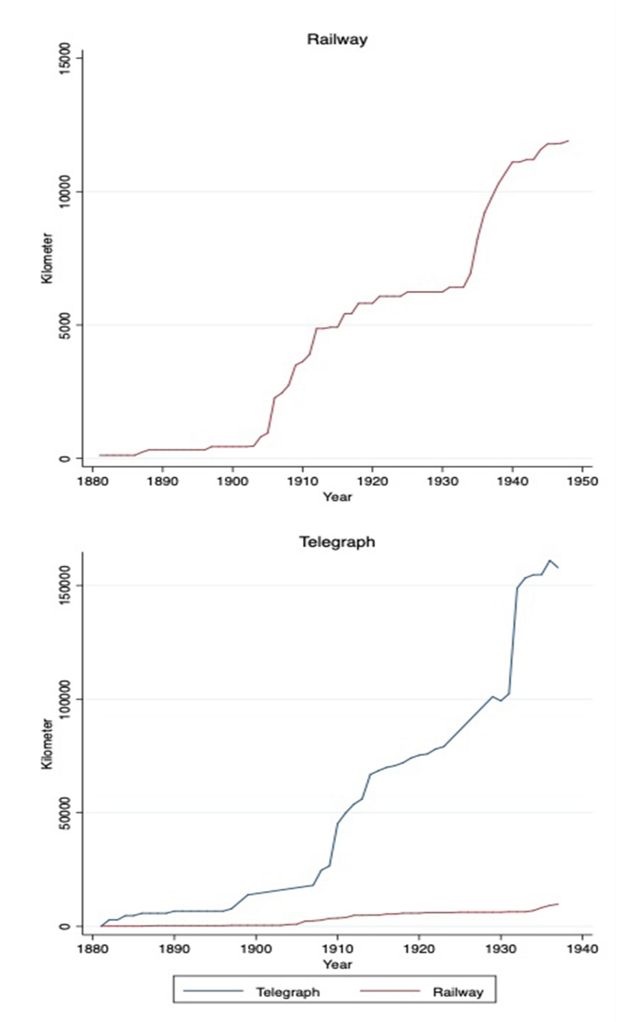
**Yan, Zhongping**, *Zhongguo Mian Ye Zhi Fa Zhan (The Development of the Chinese Cotton Industry)*, The Commercial Press, 1943.

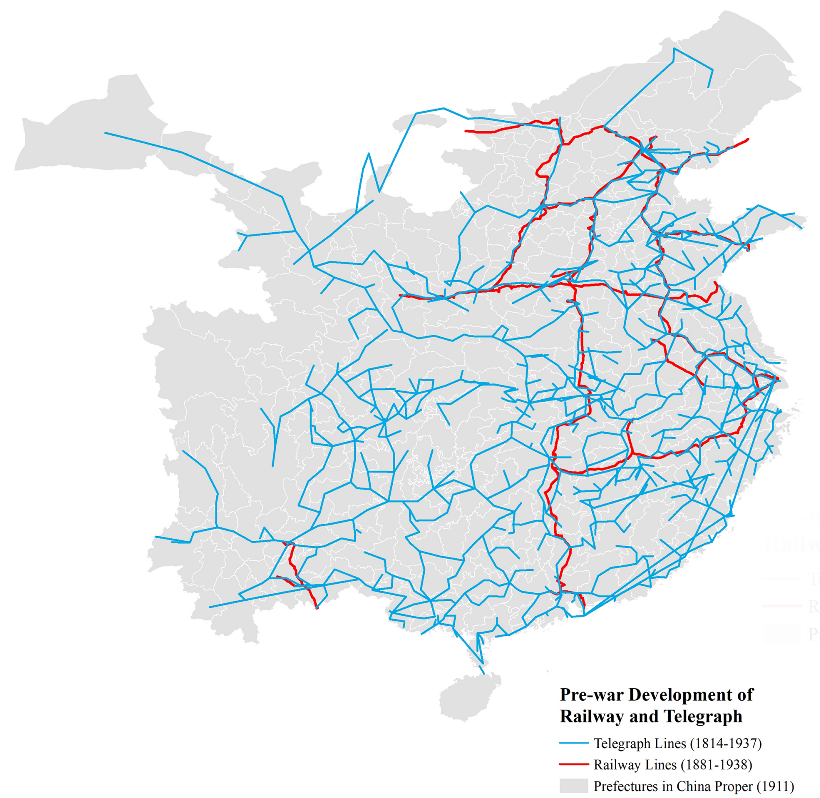
_ , *Draft History of Cotton Textiles in China*, Beijing: Kexue chubanshe, 1955.

**Zhang, Yi**, *History of China's Postal System*, Taipei: Soochow University University Press, 1985.

**Zhang, Yucai**, *Brief History of China's Railway Construction: 1876-1949*, Beijing: China Railway Press, 1997.

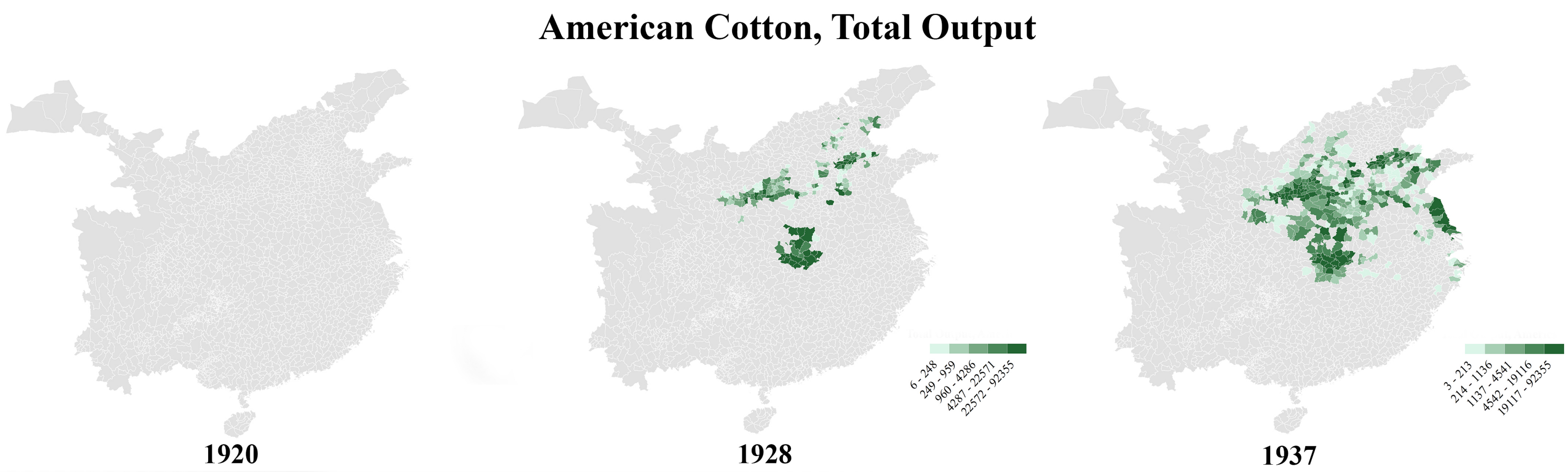
Figure 1: Expansion of Railway and Telegraph Networks over Time, 1880-1937



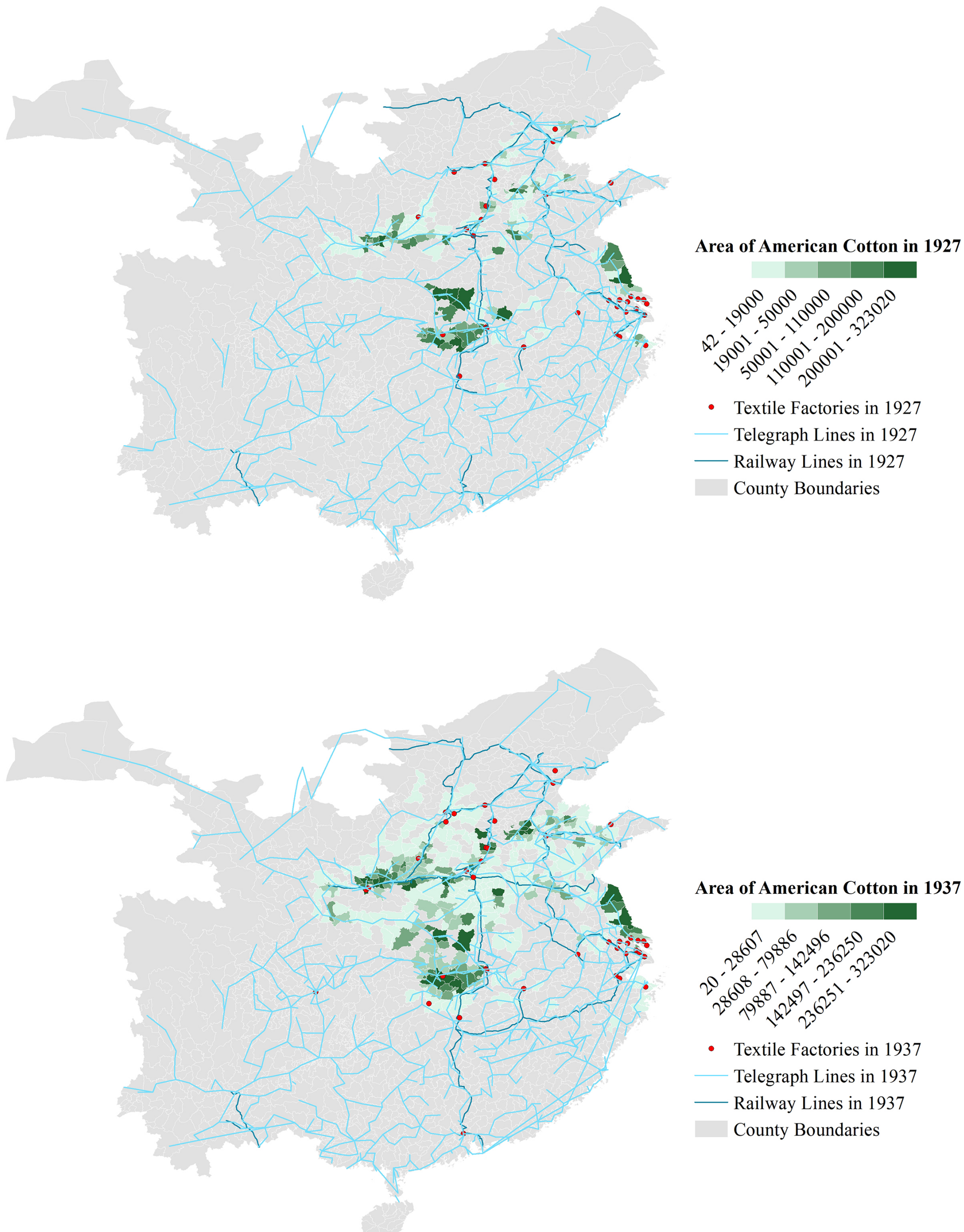

Notes: This figure shows the expansion of infrastructures over the 1880–1937 period. The left panel plots the length of railways and telegraph (in kilometers) against time, respectively. The right panel displays the geographical distribution of railway and telegraph lines across prefectures in China Proper. Source: *China's railroad network at 1953; The History of Railroad Network Construction (1876-1949); The Collection of Republican Period Archives on Railroad; The Telegraph Network Map of 1938; Contemporary Chinese county gazetteers, 1980–1990s.*

Figure 2: Diffusion of American Cotton over Time, 1920-1937



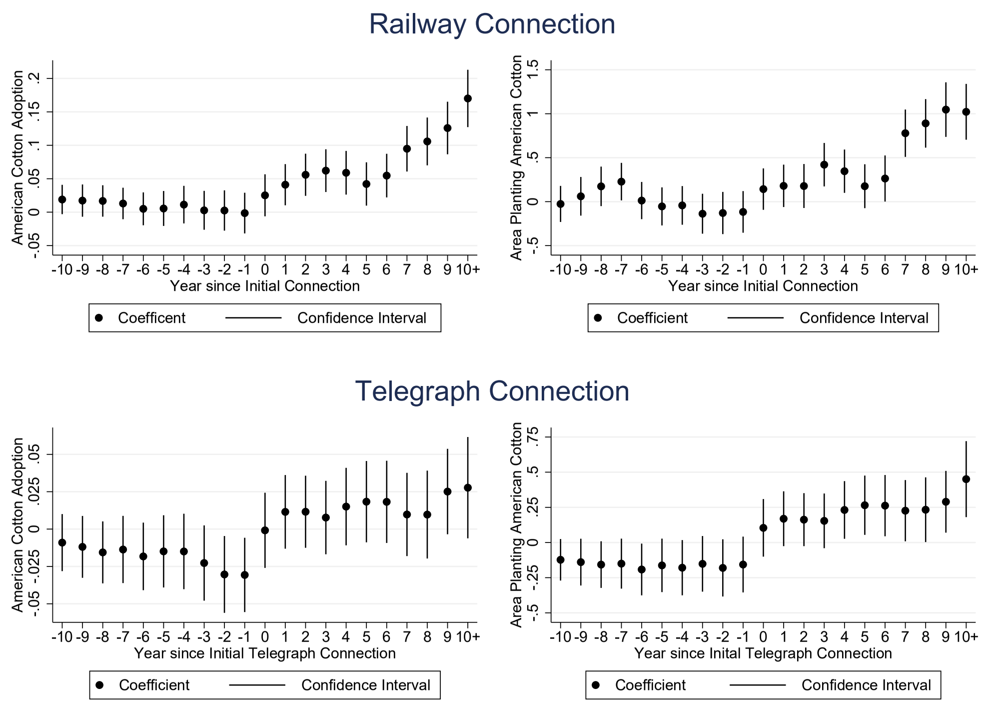

Notes: This figure shows the output of American cotton across prefectures in China Proper, 1920–1937. Source: Annual county-level statistics compiled by the Chinese Cotton Industry Statistics Association *(Zhonghua Mianye Tongjihui)*.

Figure 3: Diffusion of American Cotton and Infrastructure Expansion over Time, 1927-1937



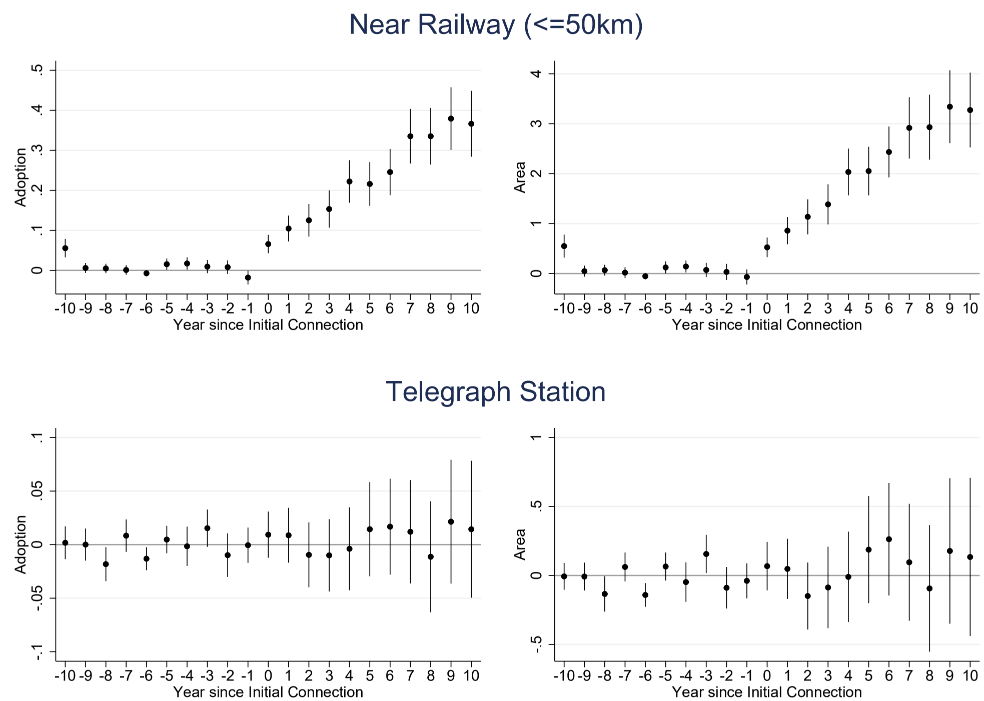

Notes: This figure shows the cropping area of American cotton across prefectures in China Proper, 1927–1937.

Figure 4: Effects of Railway and Telegraph on the Adoption of American Cotton



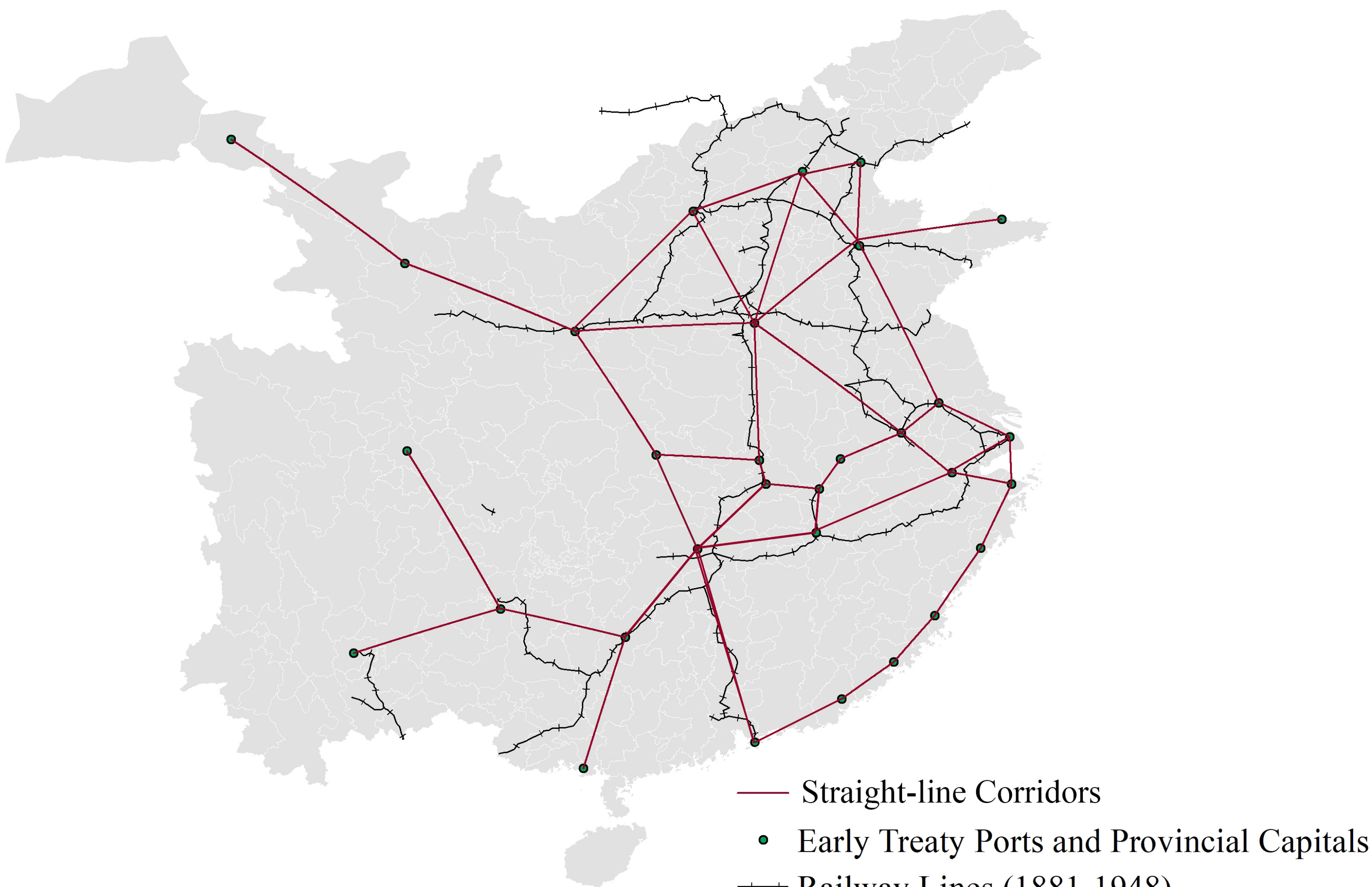

Notes: This figure reports estimated coefficients and their 95 percent confidence intervals for time periods from ten years before getting connected to the infrastructure networks to ten years after, using the specification as in columns (3) and (7) of Table 3. The upper panel depicts the coefficients of railways on the adoption dummy and plantation area of American cotton respectively, while the lower panel displays those of telegraph.

Figure 5: Effects of Railway and Telegraph on the Adoption of American Cotton: Group-Time Average Treatment Effect Estimation



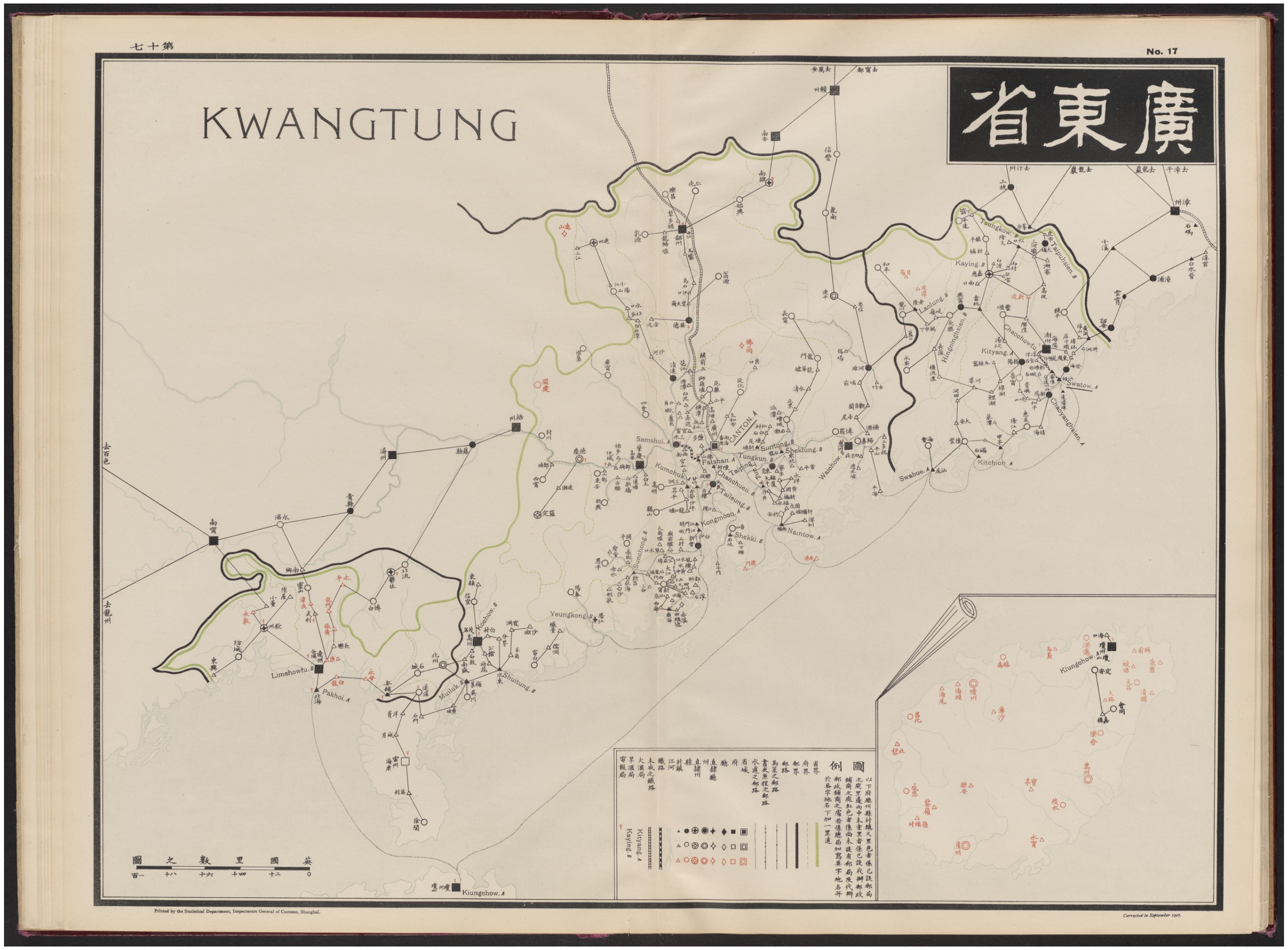

Notes: This figure reports estimated coefficients of group-time average treatment effect (Callaway and ant'Annab, 2021) and their 95 percent confidence intervals for time periods from ten years before getting connected to the infrastructure networks to ten years after, using the specification as in columns (3) and (7) of Table 3. The upper panel depicts the coefficients of railways on the adoption dummy and plantation area of American cotton respectively, while the lower panel displays those of telegraph.

Figure 6: Major Rail-hub Cities and Straight-line Corridor in the Late Qing Dynasty




Notes: This figure displays major rail-hub cities including provincial capitals and early treaty ports (1827-1881) from late Qing dynasty and the straight-line corridors connecting them.

Figure 7: Postal Network and Telegraph Stations in Guangdong Province, 1907




Notes: This figure displays a map of the postal network in Guangdong Province, 1907. The red symbols ∓ denote the locations of telegraph stations, all sitting at the important junctures along the postal network (represented by the thin black lines). Source: *Postal Map of Qing Dynasty, 1907* (*Daqing Youzheng Yutu, 1907*).

Table 1: Quality Differences between American and Chinese Cotton

| | Mean | SD | Mean | SD | Diff. | SE |
|---|---|---|---|---|---|---|
| | American Cotton | | Chinese Cotton | | | |
| | Cotton Production Improvement Program (1920-1925) | | | | | |
| | Central Cotton Production Lab of Ministry of Agriculture | | | | | |
| # of Species | 34 | | 18 | | | |
| Yield of Cotton (Kilogram) | 14.73 | 6.27 | 10.52 | 5.75 | 4.21** | 1.78 |
| Fiber Length (Inch) | 1.00 | 0.19 | 0.82 | 0.14 | 0.18*** | 0.05 |
| Price (Yuan) | 4.43 | 1.60 | 3.07 | 1.67 | 1.36*** | 0.47 |
| | Cotton Quality Grading Experiment (1940) | | | | | |
| | Cotton Research Lab of Shanghai Commodity Inspection Bureau | | | | | |
| # of Species | 7 | | 4 | | | |
| Fiber Length | 0.98 | 0.09 | 0.86 | 0.08 | 0.12** | 0.05 |
| Tenacity | 84.13 | 20.48 | 64.08 | 10.21 | 20.06* | 11.11 |
| Fineness (*1000) | 0.79 | 0.01 | 0.86 | 0.02 | -0.074*** | 0.02 |
| Strength | 5.77 | 1.05 | 6.71 | 2.34 | -0.93 | 1.00 |
| Brightness (% of Highest Rank) | 100% | | 25% | | | |
| Counts of Yarn | 40-20 | | 10-20 | | | |

Note: This table presents a comparison of quality parameters between American and Chinese cotton.

Table 2: Summary Statistics

| Variables | Number of Obs. | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|---|
| American Cotton Adoption | 31868 | 0.142 | 0.349 | 0 | 1 |
| Chinese Cotton Adoption | 31868 | 0.172 | 0.378 | 0 | 1 |
| American Cotton Area (logged) | 31868 | 1.255 | 3.202 | 0 | 12.685 |
| Chinese Cotton Area (logged) | 31868 | 1.634 | 3.693 | 0 | 13.082 |
| Near Railway ($<= 50km$) | 31868 | 0.330 | 0.470 | 0 | 1 |
| Telegraph Station | 31868 | 0.471 | 0.499 | 0 | 1 |
| Both | 31868 | 0.176 | 0.381 | 0 | 1 |
| Straight Line Corridor | 31868 | 0.123 | 0.328 | 0 | 1 |
| Postal Station | 31868 | 0.888 | 0.315 | 0 | 1 |
| Suitability Ratio of Cotton Cultivation | 1652 | 0.904 | 0.546 | 0 | 3.236 |
| Farm Size Per Household (acre) | 1498 | 2.870 | 0.638 | 0.267 | 4.965 |

Note: This table displays summary statistics for the variables described in Section 4.

Table 3: Railways, Telegraph and Adoption of American/Chinese Cotton, OLS Estimation

| | Adoption Dummy | | | | Area (logged) | | | |
|---|---|---|---|---|---|---|---|---|
| Panel A | American Cotton | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Near Railway ($<= 50km$) | 0.128*** | | 0.118*** | 0.081*** | 1.312*** | | 1.224*** | 0.814*** |
| | (0.009) | | (0.009) | (0.011) | (0.087) | | (0.087) | (0.098) |
| Telegraph Station (=1) | | 0.068*** | 0.055*** | 0.020 | | 0.649*** | 0.519*** | 0.137 |
| | | (0.009) | (0.009) | (0.010) | | (0.090) | (0.089) | (0.096) |
| Both | | | | 0.094*** | | | | 1.019*** |
| | | | | (0.016) | | | | (0.148) |
| Number of Observations | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 |
| Adj. R-squared | 0.571 | 0.567 | 0.573 | 0.574 | 0.600 | 0.596 | 0.602 | 0.604 |
| Panel B | Chinese Cotton | | | | | | | |
| | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| Near Railway ($<= 50km$) | 0.042*** | | 0.041*** | 0.030* | 0.339*** | | 0.342*** | 0.300** |
| | (0.011) | | (0.011) | (0.012) | (0.097) | | (0.099) | (0.114) |
| Telegraph Station (=1) | | 0.010 | 0.006 | -0.004 | | 0.017 | -0.020 | -0.059 |
| | | (0.009) | (0.009) | (0.009) | | (0.084) | (0.085) | (0.082) |
| Both | | | | 0.026 | | | | 0.105 |
| | | | | (0.017) | | | | (0.158) |
| Number of Observations | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 |
| Adj. R-squared | 0.635 | 0.634 | 0.635 | 0.635 | 0.671 | 0.670 | 0.671 | 0.671 |
| County Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Note: This table reports the coefficients on railway and telegraph access from estimating Equation (1). Column heading shows dependent variables: in columns (1)–(4), Panel A, the adoption dummy is an indicator variable for whether a county adopted the American cotton, denoting the extensive margin. In columns (5)–(8), Panel A, the log of the plantation area of American cotton measures the intensive margin. Panel B uses the corresponding dependent variables for the Chinese cotton. All specifications control for county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. * p¡0.05, ** p¡0.01, *** p¡0.001.

Table 4: Exclusion Restriction Tests on the Instrumental Variables

| | Population Density | | Total Firm Capital | | Total Bank Capital | | Average Distance to Ports | Nearest Distance to Port |
|---|---|---|---|---|---|---|---|---|
| | 1912–1916 | 1920 | 1915 | 1920 | 1915 | 1920 | Time invariant | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Straight Line Corridor | 0.076 | 0.116 | 0.321 | 0.566 | 0.141 | 0.151 | -0.044 | -0.833 |
| | (0.109) | (0.177) | (0.221) | (0.327) | (0.164) | (0.153) | (0.044) | (0.511) |
| Number of Observations | 1459 | 1497 | 1799 | 1799 | 1799 | 1799 | 1799 | 1799 |
| Adj. R-squared | 0.427 | 0.001 | 0.003 | 0.008 | 0.000 | 0.000 | 0.005 | 0.003 |
| | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
| Postal Station (1920) | 0.109 | 0.008 | 0.040 | 0.011 | -0.291 | -0.178 | -0.041 | 0.244 |
| | (0.088) | (0.116) | (0.136) | (0.170) | (0.206) | (0.241) | (0.029) | (0.328) |
| Number of Observations | 1459 | 1497 | 1799 | 1799 | 1799 | 1799 | 1799 | 1799 |
| Adj. R-squared | 0.428 | 0.001 | 0.000 | 0.001 | 0.002 | 0.000 | 0.005 | 0.000 |

Note: This table reports the coefficients on instrument variables from regressing county-level socio-economic proxies before or in the early years of the sample period on the Straight Line Corridor indicator of the late Qing Dynasty and the Postal Station indicator in 1920, respectively. Column heading shows dependent variables: in columns (1)–(2) and (9)–(10), the county-level population density; in columns (3)–(4) and (11)–(12), a county's total firm capital; in columns (5)–(6) and (13)–(14), total bank capital in 1915 and 1920; in columns (7)–(8) and (15)–(16), the county's average and nearest distance to ports. Columns (1) and (9) control for province fixed effects since each province completed the 1915 population census in different years. Standard errors are clustered at the province level.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 5: Railways, Telegraph and Adoption of American Cotton, IV Estimation

| Panel A | Year=1921 | | Year=1928 | | Year=1937 | |
|---|---|---|---|---|---|---|
| | | | First-Stage | | | |
| | Near Railway (1) | Telegraph Station (2) | Near Railway (3) | Telegraph Station (4) | Near Railway (5) | Telegraph Station (6) |
| Straight Line Corridor | 0.356*** | 0.043 | 0.411*** | 0.061* | 0.496*** | 0.107** |
| | (0.022) | (0.029) | (0.025) | (0.030) | (0.036) | (0.039) |
| Postal Office | 0.007 | 0.284*** | -0.011 | 0.293*** | 0.031 | 0.254*** |
| | (0.024) | (0.022) | (0.024) | (0.021) | (0.022) | (0.024) |
| Number of Observations | 1660 | 1660 | 1660 | 1660 | 1660 | 1660 |
| Adjusted R-squared | 0.057 | 0.099 | 0.068 | 0.106 | 0.085 | 0.069 |
| Panel B | Year=1921 | | Year=1928 | | Year=1937 | |
| | | | Second-Stage | | | |
| | Adoption Dummy (7) | Area (Logged) (8) | Adoption Dummy (9) | Area (Logged) (10) | Adoption Dummy (11) | Area (Logged) (12) |
| Near Railway (<=50km) | 0.131*** | 1.124*** | 0.395*** | 3.448*** | 0.578*** | 5.458*** |
| | (0.036) | (0.322) | (0.057) | (0.529) | (0.085) | (0.818) |
| Telegraph Station | 0.057* | 0.458 | 0.015 | 0.233 | -0.067 | -0.671 |
| | (0.028) | (0.263) | (0.053) | (0.477) | (0.070) | (0.674) |
| Number of Observations | 1660 | 1660 | 1660 | 1660 | 1660 | 1660 |
| Adjusted R-squared | 0.046 | 0.051 | 0.191 | 0.168 | 0.041 | 0.039 |
| Panel C | | | Panel Data (1920-1937) | | | |
| | First-Stage | | Second-Stage | | IVCRC Estimations | |
| | Near Railway (13) | Telegraph Station (14) | Adoption Dummy (15) | Area (Logged) (16) | Adoption Dummy (17) | Area (Logged) (18) |
| Straight Line Corridor | 0.403*** | 0.058* | | | | |
| | (0.021) | (0.026) | | | | |
| Postal Office | 0.015 | 0.304*** | | | | |
| | (0.020) | (0.019) | | | | |
| Near Railway (<=50km) | | | 0.377*** | 3.322*** | 0.156*** | 1.412*** |
| | | | (0.044) | (0.412) | (0.005) | (0.039) |
| Telegraph Station | | | 0.005 | 0.087 | 0.033*** | 0.390*** |
| | | | (0.042) | (0.377) | (0.004) | (0.037) |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Observations | 31868 | 31868 | 31868 | 31868 | 31868 | 31868 |
| Adjusted R-squared | 0.066 | 0.107 | 0.042 | 0.058 | | |

Note: This table reports the two-stage least squares estimates, instrumenting for railway access and telegraph connections using the straight-line corridor and postal station indicator. Column heading shows dependent variables. Panels A and B report the first- and second-stage results for each of three cross sections (1921, 1928, 1937) separately, using Equation (2). Panel C reports the IV estimation results using the panel data between 1920 and 1937, and the specifications control for year fixed effects. Columns (17)–(18) report the results of the IVCRC estimations in the spirit of Masten and Torgovitsky (2016). Standard errors are clustered at the county level. * p<0.05, ** p<0.01, *** p<0.001.

Table 6: Effects of Infrastructures on Trade Volume, Price Spread and Average Price: A Gravity Model Estimation

| | At County Pairs Level | | | | | |
|---|---|---|---|---|---|---|
| | Trade Volume (logged) | | Price Spread (logged) | | Average Price (logged) | |
| Panel A | All Commodities | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Near Railway ($<= 50km$) | 0.904*** | 0.604*** | -0.123* | -0.070 | 0.005 | 0.026 |
| | (0.177) | (0.179) | (0.058) | (0.066) | (0.098) | (0.108) |
| Telegraph Station ($= 1$) | 0.261* | -0.075 | -0.123** | -0.063 | 0.055 | 0.079 |
| | (0.113) | (0.125) | (0.041) | (0.043) | (0.072) | (0.079) |
| Both | | 0.771*** | | -0.138* | | -0.056 |
| | | (0.139) | | (0.068) | | (0.092) |
| Distance (logged) | 0.075 | 0.079 | 0.046** | 0.045* | 0.009 | -0.056 |
| | (0.042) | (0.041) | (0.018) | (0.018) | (0.027) | (0.092) |
| Number of Observations | 13905 | 13905 | 13042 | 13042 | 13127 | 13127 |
| Adj. R-squared | 0.639 | 0.640 | 0.614 | 0.614 | 0.658 | 0.658 |
| Origin County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Commodity FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Panel B | Cotton | | | | | |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| Near Railway ($<= 50km$) | 0.948* | 0.548 | -0.292 | -0.191 | -0.352 | -0.273 |
| | (0.375) | (0.371) | (0.174) | (0.172) | (0.260) | (0.269) |
| Telegraph Station ($= 1$) | 0.282 | -0.188 | -0.318*** | -0.208* | 0.194 | 0.280 |
| | (0.247) | (0.265) | (0.086) | (0.093) | (0.149) | (0.158) |
| Both | | 1.077*** | | -0.256* | | -0.199 |
| | | (0.249) | | (0.129) | | (0.184) |
| Distance (logged) | 0.124 | 0.125 | 0.051 | 0.051 | -0.024 | -0.024 |
| | (0.073) | (0.072) | (0.039) | (0.040) | (0.065) | (0.065) |
| Number of Observations | 2547 | 2547 | 2467 | 2467 | 2481 | 2481 |
| Adj. R-squared | 0.545 | 0.548 | 0.560 | 0.561 | 0.555 | 0.555 |
| Origin County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination County FE | Yes | Yes | Yes | Yes | Yes | Yes |

Note: This table reports the gravity model estimation results on a comprehensive cross-sectional domestic trade data set (Source: commodity census conducted by the General Post Office of Ministry of Transportation in 1934). Panel A reports the estimates for all commodities aggregated. Panel B reports the estimates for cotton specifically. Column heading shows dependent variables including the log of trade volume, price spread (defined as the gap between the highest and lowest price for the same commodity traded between any pair of counties), and average price. All specifications control for county fixed effects and year fixed effects. Standard errors are clustered at the county level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 7: Effects of Infrastructures on American Cotton Adoption and the Comparative Suitability Threshold

| | American Cotton | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Adoption Dummy | | | | Area (logged) | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Near Railway ($<= 50km$) | 0.080*** | 0.079*** | 0.288*** | 0.084*** | 0.692*** | 0.788*** | 2.459*** | 0.834*** |
| | (0.014) | (0.011) | (0.025) | (0.011) | (0.131) | (0.109) | (0.250) | (0.111) |
| Telegraph Station ($= 1$) | 0.038*** | 0.119*** | 0.023* | 0.024* | 0.527*** | 1.048*** | 0.301** | 0.302** |
| | (0.009) | (0.019) | (0.010) | (0.010) | (0.092) | (0.200) | (0.113) | (0.113) |
| Both | -0.031 | 0.132*** | 0.117*** | 0.318*** | -0.014 | 1.640*** | 1.529*** | 3.089*** |
| | (0.019) | (0.018) | (0.017) | (0.032) | (0.193) | (0.197) | (0.179) | (0.341) |
| Suitability Ratio | 0.110*** | | | | 1.175*** | | | |
| | (0.016) | | | | (0.164) | | | |
| Near Railway*Suitability Ratio | | -0.192*** | | | | -1.527*** | | |
| | | (0.020) | | | | (0.205) | | |
| Telegraph Station*Suitability Ratio | | | -0.116*** | | | | -0.902*** | |
| | | | (0.020) | | | | (0.223) | |
| Both*Suitability Ratio | | | | -0.186*** | | | | -1.450*** |
| | | | | (0.021) | | | | (0.229) |
| Province Fixed Effects | Yes | No | No | No | Yes | No | No | No |
| County Fixed Effects | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Observations | 31720 | 31720 | 31720 | 31720 | 31720 | 31720 | 31720 | 31720 |
| Adj. R-squared | 0.280 | 0.579 | 0.576 | 0.579 | 0.279 | 0.611 | 0.609 | 0.610 |

Note: This table reports estimates using Equation (3). A county's comparative suitability of planting cotton is normalized by the average suitability of planting various grain crops (including rice, wheat, sorghum and maize) to yield a ratio variable. Column heading shows dependent variables including the adoption dummy and planting areas of American cotton. Columns (1) and (5) include dummy indicators of infrastructure access and the suitability ratio while controlling for province fixed effects and year fixed effects. Columns (2)–(4) and columns (6)–(8) include the pairwise interactions of infrastructure access and suitability ratio while controlling for county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

Table 8: Effects of Infrastructures on American Cotton Adoption and the Farm Size Threshold

| | American Cotton | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Adoption Dummy | | | | Area (logged) | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Near Railway ($<= 50km$) | 0.080*** | 0.079*** | 0.481*** | 0.085*** | 0.718*** | 0.797*** | 4.533*** | 0.854*** |
| | (0.015) | (0.011) | (0.043) | (0.011) | (0.144) | (0.112) | (0.456) | (0.112) |
| Telegraph Station (= 1) | 0.051*** | 0.222*** | 0.019 | 0.017 | 0.678*** | 2.172*** | 0.271* | 0.255* |
| | (0.010) | (0.037) | (0.011) | (0.011) | (0.111) | (0.382) | (0.119) | (0.119) |
| Both | -0.011 | 0.089*** | 0.077*** | 0.521*** | 0.205 | 1.313*** | 1.204*** | 5.253*** |
| | (0.020) | (0.017) | (0.016) | (0.050) | (0.202) | (0.176) | (0.174) | (0.535) |
| Farm Size Per Capita | 0.017 | | | | 0.189* | | | |
| | (0.009) | | | | (0.093) | | | |
| Near Railway*Farm Size per HH | | -0.134*** | | | | -1.243*** | | |
| | | (0.014) | | | | (0.149) | | |
| Telegraph Connection*Farm Size per HH | | | -0.072*** | | | | -0.675*** | |
| | | | (0.013) | | | | (0.135) | |
| Both*Farm Size per HH | | | | -0.152*** | | | | -1.385*** |
| | | | | (0.016) | | | | (0.168) |
| Province Fixed Effects | Yes | No | No | No | Yes | No | No | No |
| County Fixed Effects | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Observations | 28838 | 28838 | 28838 | 28838 | 28838 | 28838 | 28838 | 28838 |
| Adj. R-squared | 0.270 | 0.580 | 0.577 | 0.581 | 0.267 | 0.613 | 0.611 | 0.614 |

Note: This table reports estimates using equation (4). The farm size threshold is a county-level proxy that denotes average farm size per household in 1931 (Source: the Report of Agricultural Statistics in various Chinese Provinces, 1932). Column heading shows dependent variables including the adoption dummy and planting areas of American cotton. Columns (1) and (5) include indicators of infrastructure access and farm size while controlling for province fixed effects and year fixed effects. Columns (2)–(4) and columns (6)–(8) include the pairwise interactions of infrastructure access and farm size while controlling for county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

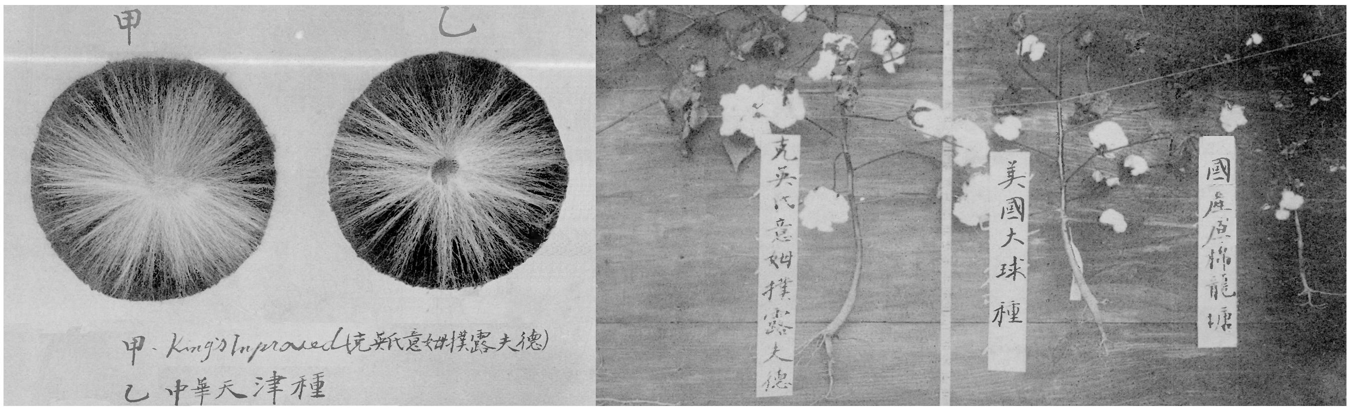
Table 9: Railways, Telegraph and Welfare of Farmers

| | Welfare: log(real agricultural income) | | | |
|---|---|---|---|---|
| Panel A | Crude Effect | | | |
| | OLS | | IV | |
| | (1) | (2) | (3) | (4) |
| Near Railway (<=50km) | 0.084*** | | 0.464*** | |
| | (0.013) | | (0.074) | |
| Telegraph Station (=1) | | 0.009 | | -0.021 |
| | | (0.012) | | (0.065) |
| Number of Observations | 15866 | 15866 | 15866 | 15866 |
| Adj. R-squared | 0.868 | 0.868 | 0.043 | 0.048 |
| Panel B | Median Analysis | | | |
| | Mediators' Effect on Welfare | | Direct Effect Excluding Mediators | |
| | (5) | (6) | (7) | (8) |
| American Cotton (=1) | 0.306*** | | | |
| | (0.012) | | | |
| American Cotton Area (logged) | | 0.125*** | | |
| | | (0.007) | | |
| Near Railway (<=50km) | | | 0.043* | |
| | | | (0.023) | |
| Telegraph Station (=1) | | | | -0.046*** |
| | | | | (0.017) |
| Number of Observations | 15866 | 4289 | 15832 | 15832 |
| Adj. R-squared | 0.548 | 0.614 | 0.530 | 0.525 |
| County Fixed Effects | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes |

Note: This table reports estimates of the infrastructure's effects on welfare measured by a county's real agricultural income (in logs). Panel A presents the crude effect. Columns (1)–(2) report the OLS estimation results with each infrastructure indicator, respectively. Columns (3)–(4) report the IV estimation results by using each instrument separately with the other infrastructure indicator conditioned on as control variables in both the first and the second stage of the 2SLS estimation. Panel B presents the mediation analysis in the spirit of Brown et al. (2019) and Acharya et al. (2016). Columns (5)–(6) report the mediator (extensive and intensive margins of American cotton adoption)'s effect on welfare. Columns (7)–(8) estimate the average controlled direct effect of each infrastructure on farmers' welfare, while excluding the mediation channel of the American cotton adoption (for further detail about the specifications, see Appendix B). All specifications control for county fixed effects and year fixed effects except for column (3), which removes county fixed effects due to the time-invariant IV (straight line corridor) used. * p<0.05, ** p<0.01, *** p<0.001.

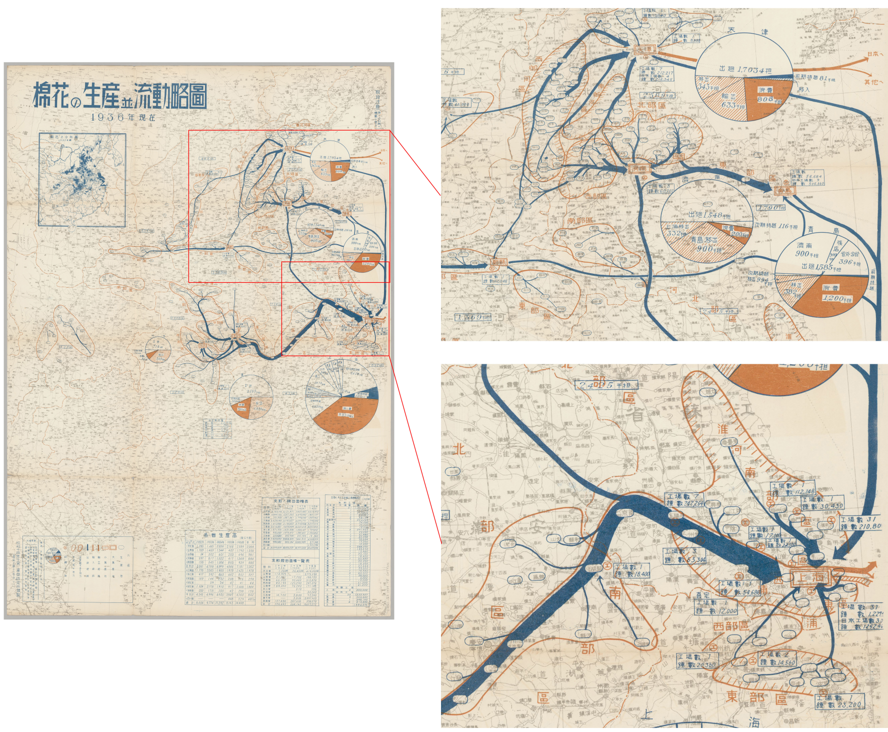
# A Appendix A

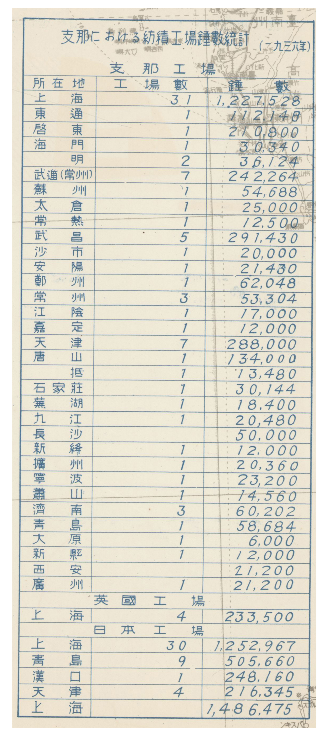
Figure A.1: A Visual Comparison of American and Chinese Cotton




*Notes:* This figure presents visual comparisons of American and Chinese cotton. The left panel shows two images of the seed and associated fibre from American (left) and Chinese (Tianjin) cotton species (right). The right panel shows images of three cotton plants: the first two belong to American species (left) and the last corresponds to Chinese species (right). These images were obtained from experimental sites for cultivation and domestication of American cotton, operated by the Ministry of Agriculture and Business. Source: *Working Report on the Experiments for Improving Cotton Production* (1934).

Figure A.2: Long-distance Trade Routes of Cotton along Railway Lines in 1936



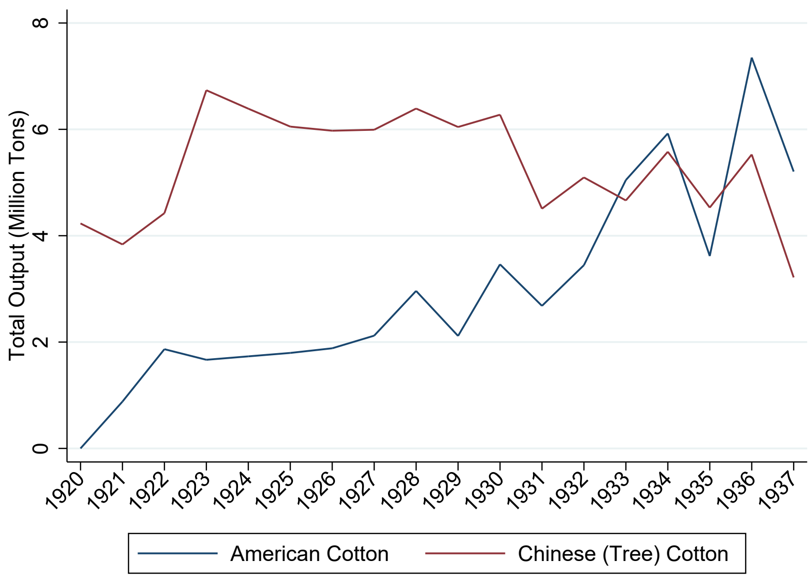



支那における紡績工場錘數統計 (一九三六年)

| 所在地 | 工場數 | 錘數 |
|---|---|---|
| 支那工場 | | |
| 上海 | 31 | 1,227,528 |
| 東通 | 1 | 112,148 |
| 啓東 | 1 | 210,800 |
| 海門 | 1 | 30,340 |
| 明 | 2 | 36,124 |
| 武進(常州) | 7 | 242,264 |
| 蘇州 | 1 | 54,688 |
| 太倉 | 1 | 25,000 |
| 常熱 | 1 | 12,500 |
| 武昌 | 5 | 291,430 |
| 沙市 | 1 | 20,000 |
| 安陽 | 1 | 21,430 |
| 鄭州 | 1 | 62,048 |
| 常州 | 3 | 53,304 |
| 江陰 | 1 | 17,000 |
| 嘉定 | 1 | 12,000 |
| 天津 | 7 | 288,000 |
| 唐山 | 1 | 134,000 |
| 抵 | 1 | 13,480 |
| 石家莊 | 1 | 30,144 |
| 蕪湖 | 1 | 18,400 |
| 九江 | 1 | 20,480 |
| 長沙 | | 50,000 |
| 新絳 | 1 | 12,000 |
| 撫州 | 1 | 20,360 |
| 寧波 | 1 | 23,200 |
| 蕭山 | 1 | 14,560 |
| 濟南 | 3 | 60,202 |
| 青島 | 1 | 58,684 |
| 大原 | 1 | 6,000 |
| 新縣 | 1 | 12,000 |
| 西安 | | 21,200 |
| 廣州 | 1 | 21,200 |
| 英國工場 | | |
| 上海 | 4 | 233,500 |
| 日本工場 | | |
| 上海 | 30 | 1,252,967 |
| 青島 | 9 | 505,660 |
| 漢口 | 1 | 248,160 |
| 天津 | 4 | 216,345 |
| 上海 | | 1,486,475 |

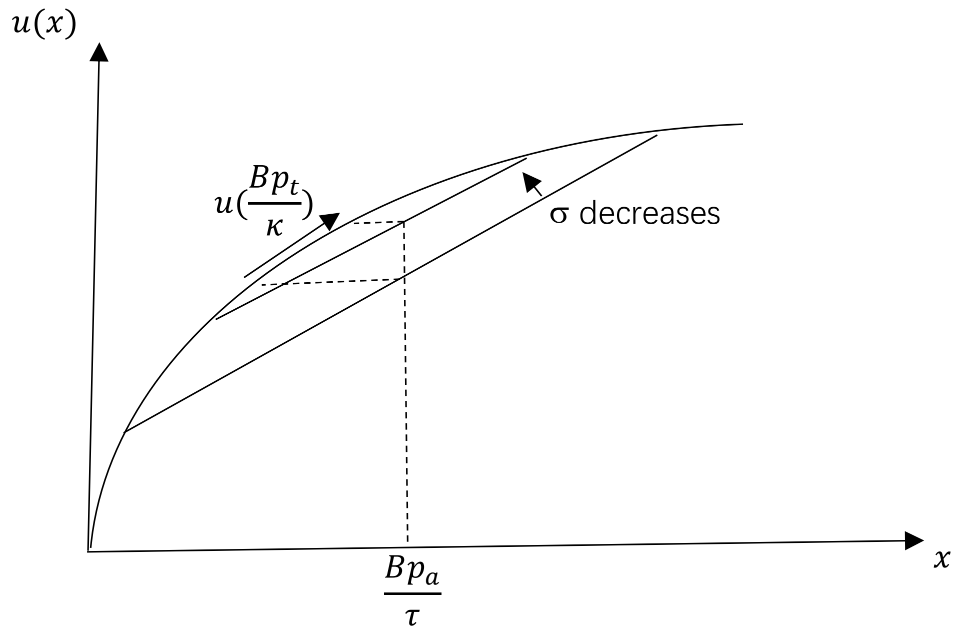
*Notes:* This figure shows the trade routes of American cotton (in dark blue) along the railway lines which help ship the raw materials from inland plantation counties to major textile firms in the coastal areas. The left panel displays the whole map. The middle panel zooms images of the Northern China—a main production area for American Cotton at that time (upper) and Shanghai—the largest centre of China's textile industry (lower). The right panel zooms to a table at the bottom right of the original map, which documents the number and production of the textile factories in various cities. Source: the (Japan) East Asia Institute, 1936.

Figure A.3: Total Output of American and Chinese Cotton (1920-1937)




*Notes:* This figures shows the total output of the native Chinese tree cotton and the newly introduced American cotton. Source: Authors' calculations based on annual county-level statistics compiled by the Chinese Cotton Industry Statistics Association (Zhonghua Mianye Tongjihui).

Figure A.4: An Illustration of Reduced Price Uncertainty in the Model




*Notes:* This figures shows the condition for technology adoption (quality upgrading) under different price uncertainty, corresponding to Proposition 3 of Section 3. With decreased uncertainty denoted by a fall in $\sigma$, the risk premium required for adopting the American cotton is lower, i.e., the gap between expected earning from American cotton plantation, $B\frac{p_a}{\tau}$ and that from Chinese cotton plantation, $B\frac{p_t}{\kappa}$ decreases for the marginal farmer. Recall that $B = \left(\frac{\alpha}{p_g}\right)^{\alpha}\left(\frac{1-\alpha}{\kappa p_t}\right)^{1-\alpha} A_c^i l$. This in turn implies a fall in the adoption threshold $A_c^i l_i^*$.

Table A.1: Railways, Telegraph and Adoption of Chinese cotton, IV Estimation

| Panel A | Year=1921 | | Year=1928 | | Year=1937 | |
|---|---|---|---|---|---|---|
| | Second-Stage | | | | | |
| | Adoption Dummy | Area (logged) | Adoption Dummy | Area (logged) | Adoption Dummy | Area (logged) |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Near Railway (<=50km) | 0.081 | 0.637 | 0.088 | 0.868 | 0.102 | 1.554 |
| | (0.061) | (0.793) | (0.486) | (0.884) | (0.090) | (1.381) |
| Telegraph Station | 0.029 | 0.318 | -0.038 | -0.469 | -0.156 | -1.529 |
| | (0.054) | (0.555) | (0.068) | (0.694) | (0.085) | (0.794) |
| Number of Observations | 1660 | 1660 | 1660 | 1660 | 1660 | 1660 |
| Adjusted R-sequared | 0.239 | 0.221 | 0.287 | 0.278 | 0.313 | 0.415 |

| Panel B | Panel Data (1920-1937) | | | |
|---|---|---|---|---|
| | Second-Stage | | IVCRC Estimations | |
| | Adoption Dummy | Area (logged) | Adoption Dummy | Area (logged) |
| | (7) | (8) | (9) | (10) |
| Near Railway (<=50km) | 0.637 | 3.136 | 0.273 | 1.734 |
| | (0.793) | (3.608) | (0.181) | (3.046) |
| Telegraph Station | -0.062 | -0.664 | 0.005 | 0.019 |
| | (0.051) | (0.508) | (0.004) | (0.047) |
| Year Fixed Effects | Yes | Yes | Yes | Yes |
| Number of Observations | 31868 | 31868 | 31868 | 31868 |
| Adjusted R-sequared | 0.187 | 0.204 | | |

Note: This table reports the two-stage least squares estimates, instrumenting for railway access and telegraph connections using the straight-line corridor and postal station indicator. Column heading shows dependent variables. Panel A reports the second-stage results for each of three cross sections (1921, 1928, 1937) separately, while Columns (7)–(8) of Panel B report the IV estimation results using the panel data between 1920 and 1937, and the specifications control for year fixed effects. Columns (9)–(10) report the results of the IVCRC estimations in the spirit of Masten and Torgovitsky (2016). Standard errors are clustered at the county level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.2: Effects of Infrastructures on American Cotton Adoption and the Cotton Suitability Threshold

| | American Cotton | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Adoption Dummy | | | | Area (logged) | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Near Railway (<=50km) | 0.079*** | 0.332*** | 0.079*** | 0.077*** | 0.717*** | 2.747*** | 0.796*** | 0.781*** |
| | (0.014) | (0.064) | (0.011) | (0.011) | (0.127) | (0.547) | (0.099) | (0.100) |
| Telegraph Station (=1) | 0.037*** | 0.021* | 0.116** | 0.023* | 0.379*** | 0.146 | 0.958** | 0.161 |
| | (0.009) | (0.010) | (0.041) | (0.010) | (0.084) | (0.097) | (0.332) | (0.097) |
| Both | -0.010 | 0.088*** | 0.099*** | 0.429*** | 0.031 | 0.974*** | 1.062*** | 3.875*** |
| | (0.019) | (0.016) | (0.016) | (0.105) | (0.171) | (0.147) | (0.151) | (0.978) |
| Suitability Index | 0.030*** | | | | 0.283*** | | | |
| | (0.003) | | | | (0.028) | | | |
| Near Railway*Suitability Index | | -0.036*** | | | | -0.275*** | | |
| | | (0.009) | | | | (0.075) | | |
| Telegraph Station*Suitability Index | | | -0.014* | | | | -0.122* | |
| | | | (0.006) | | | | (0.049) | |
| Both*Suitability Index | | | | -0.048*** | | | | -0.405** |
| | | | | (0.014) | | | | (0.133) |
| Province Fixed Effects | Yes | No | No | No | Yes | No | No | No |
| County Fixed Effects | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Observations | 31792 | 31792 | 31792 | 31792 | 31792 | 31792 | 31792 | 31792 |
| Adj. R-squared | 0.283 | 0.575 | 0.574 | 0.575 | 0.278 | 0.605 | 0.604 | 0.605 |

Note: This table reports estimates using equation (4). The key explanatory variable is the interaction terms between a county's suitability index of cultivating cotton and the infrastructure access indicators. Column heading shows dependent variables including the adoption dummy and planting areas of American cotton. Columns (1) and (5) include dummy indicators of infrastructure access and the suitability ratio while controlling for province fixed effects and year fixed effects. Columns (2)–(4) and columns (6)–(8) include the pairwise interactions of infrastructure access and suitability ratio while controlling for county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

# B Appendix B

## Mediation Analysis

We estimate the average controlled direct effect (ACDE) of each infrastructure following the two-step approach proposed by Acharya et al. (2016). Specifically, we first estimate the "de-mediated" real income measure, and then use it as the dependent variable in the second step.

Below we use the "railway access" treatment as an example to illustrate the estimation procedure.[B.1] As the first step, we estimate the following specification:

$$W_{it} = \alpha + \beta_1 Railway_{it} + \beta_2 f(M_{it}) + \beta_3 Z_{it} + \beta_4 Telegraph_{it} + \upsilon_i + \nu_t + \epsilon_{it}. \tag{B.1}$$

Here, $W_{it}$ is the welfare indicator–real agricultural income of county $i$ in year $t$. $Railway_{it}$ denotes the treatment variable; $M_{it}$, the American cotton adoption dummy, is the mediator we use to explain the relationship between the treatment variable (access to the railway network) and the dependent variable (real income) in the sequential g-estimation. $X_i$ denotes the pretreatment confounder–the suitability ratio for cotton plantation. To fully capture the mediation effect, $f(M_{it})$ includes the mediator $M_{it}$, the interaction terms between the mediator and the treatment, and the pretreatment confounder (i.e., $M_{it}Railway_{it}$, $M_{it}X_i$). $Z_{it}$ represents the intermediate confounder (other potential mediator) such as the plantation area of Chinese cotton. We also include access to the other infrastructure network (telegraph) as the control variable to be consistent with our main specifications. Note that the direct effect of pretreatment confounder is absorbed by county fixed effects. The "de-mediated" outcome variable is given by:

$$\widetilde{W_{it}} = W_{it} - \hat{\beta}_2 f(M_{it}) \tag{B.2}$$

The "demediated" dependent variable is then used to estimate the following equation:

$$\widetilde{W_{it}} = \phi + \varphi Railway_{it} + \psi Telegraph_{it} + \rho_i + \sigma_t + \omega_{it}. \tag{B.3}$$

where $\omega_{it}$ represents bootstrapped clustered standard errors based on 1,000 replications. We then report the "de-mediated" effect of railway connection $\hat{\varphi}$ in the column (7) of Table 9. The procedure to obtain the "de-mediated" effect of telegraph connection in column (8) of Table 9 is analogous.

[B.1] The mediation method applies to a treatment of interest rather than multiple treatments simultaneously. So we modify our estimation framework by focusing on separate estimates for each infrastructure access, conditioning on the other infrastructure indicator.