The roadblock effect:
War shocks, modal shifts, and population changes

Yan Hu[*#]
CEMFI

9 November 2020

Latest version available HERE

ABSTRACT: This paper investigates a novel "roadblock effect," whereby temporary forced route changes catalyse the adoption of a new transport technology. The Taiping rebellion in 19th century China ravaged many cities, but also blocked key land routes, triggering investments facilitating steamship trade. Combining a trade model featuring modal and route choice, shipping records, and a new method to estimate historical trade costs, I show that the post-rebellion spatial variation in steamship trade was driven by blocked land routes and the feasibility of sea alternatives. This permanently moved many routes to sea and shifted population towards port cities.

Key words: city populations, path dependence, modal choice, Taiping rebellion
JEL classification: F12, J61, N95, R12, R40

[*]Centro de Estudios Monetarios y Financieros (CEMFI), Casado del Alisal 5, 28014 Madrid. E-mail: yanhu@cemfi.edu.es.

[#]I am deeply indebted to my advisor Diego Puga for his support and guidance. I am grateful to Jorge De la Roca, Mauricio Drelichman, Keith Head, Chicheng Ma, Monica Martinez-Bravo, Joan Monras, Noam Yuchtman and especially Paula Bustos and Nezih Guner for their advice. Special thanks to Shuji Cao for sharing historical population data in China. Funding from the European Research Council under the European Union's Horizon 2020 Programme (ERC Advanced Grant agreement 695107 - DYNURBAN) and from Spain's Ministerio de Economía y Competividad (Grants ECO-2016-80411-P, BES-2017-082187 and María de Maeztu Programme for Units of Excellence in R&D, MDM-2016-0684) is gratefully acknowledged.

1. Introduction

I use the Taiping rebellion in 19th century China to study the evolution of city populations when shocks hit them not just directly, through deaths and displacement, but also indirectly, by promoting some trade routes and modes over others. This highlights the complex interplay between large population shocks, shifting natural advantage, the persistence of agglomerations, and substantial changes in trade costs through technology and investment.

City locations and sizes reflect a combination of natural advantage and agglomeration economies. Natural endowments and technology differ across space, making some potential city locations much more suitable than others. At the same time, localised increasing returns also make locations that already have many firms and people more attractive to other firms and people. Such increasing returns can then reinforce or counteract comparative advantage.[1] Reinforcement happens when a location with advantageous underlying conditions sees them amplified by agglomeration economies. However, if conditions change, concentration in a location can remain very persistent even if it is no longer superior, be it because of coordination problems or sunk infrastructure costs (e.g. Bleakley and Lin, 2012; Michaels and Rauch, 2017).

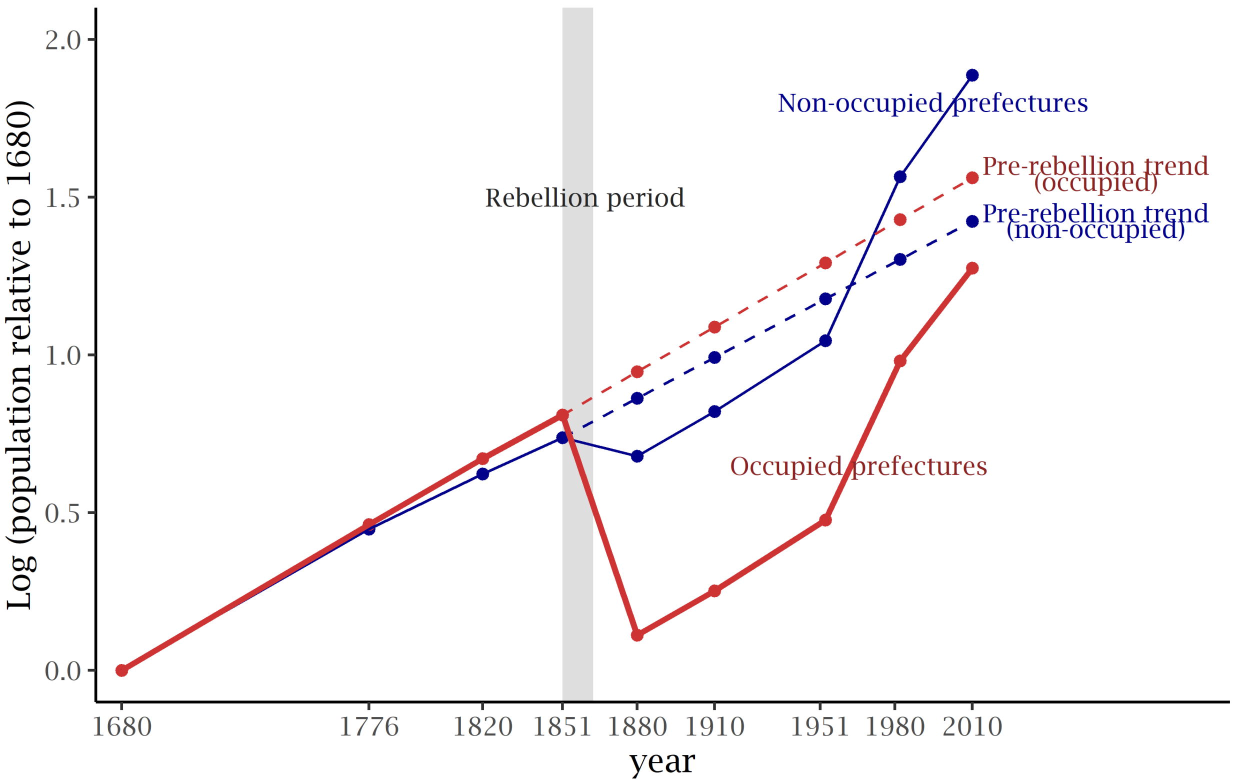
Trade costs complicate the analysis of relative city sizes further because, in addition to underlying conditions ("first nature") and the current concentration at each location ("second nature"), we have to consider the access each of them has to a larger market (Krugman, 1993). Moreover, trade costs can change over time. Normally they change because of new transport infrastructure, and a large literature examines its consequences (e.g. Faber, 2014; Donaldson, 2018). However, trade costs can also change because of new transport technologies (e.g. Tsivanidis, 2018; Heblich, Redding and Sturm, 2020). A common difficulty when investigating the effects of changing trade costs is that they are often endogenous. Here, I exploit the heterogeneous incentives to adopt steamship transport through the route and modal changes induced by shocks of civil conflict elsewhere.

The Taiping rebellion was the largest peasant revolt in the history of China. The civil war between the rebels and the Qing government led to the death of one out of every four Chinese people (Ge, Hou and Zhang, 1999). This shock was unevenly distributed, and places heavily ravaged by war, concentrated in rebellion occupied region, lost up to 80% of their population (Cao and Li, 2000). Figure 1 compares the long-term evolution of population across occupied and non-occupied prefectures. The thin blue continuous line corresponds to prefectures that were not occupied by the rebellion, while the thick red continuous line corresponds to prefectures that instead were occupied.[2] We see that the occupied prefectures had a slightly higher population growth rate than the non-occupied prefectures before the rebellion, but they experienced a much larger population drop during the rebellion. Afterwards, population in both groups grew in a parallel manner. Thus, the population in occupied prefectures did not catch up but instead suffered a permanent loss the consequences of which are still evident a century and a half later.

[1] Models examining different aspects of this trade-off include Venables, 1999; Forslid and Wooton, 2003; Amiti, 2005; Pflüger and Tabuchi, 2019.

[2] Population is expressed as the natural logarithm of population in each data year (1680, 1776, 1820, 1880, 1910, 1953, 1982 and 2010) relative to 1680 population. Thus, initial population is normalised to 0. The dashed lines plot the pre-rebellion trends for each of the two prefecture groups, based on the average growth rate between 1680 and 1851. The shaded grey rectangle marks the rebellion period, 1851-1864.

Figure 1: Evolution of population in occupied/non-occupied prefectures

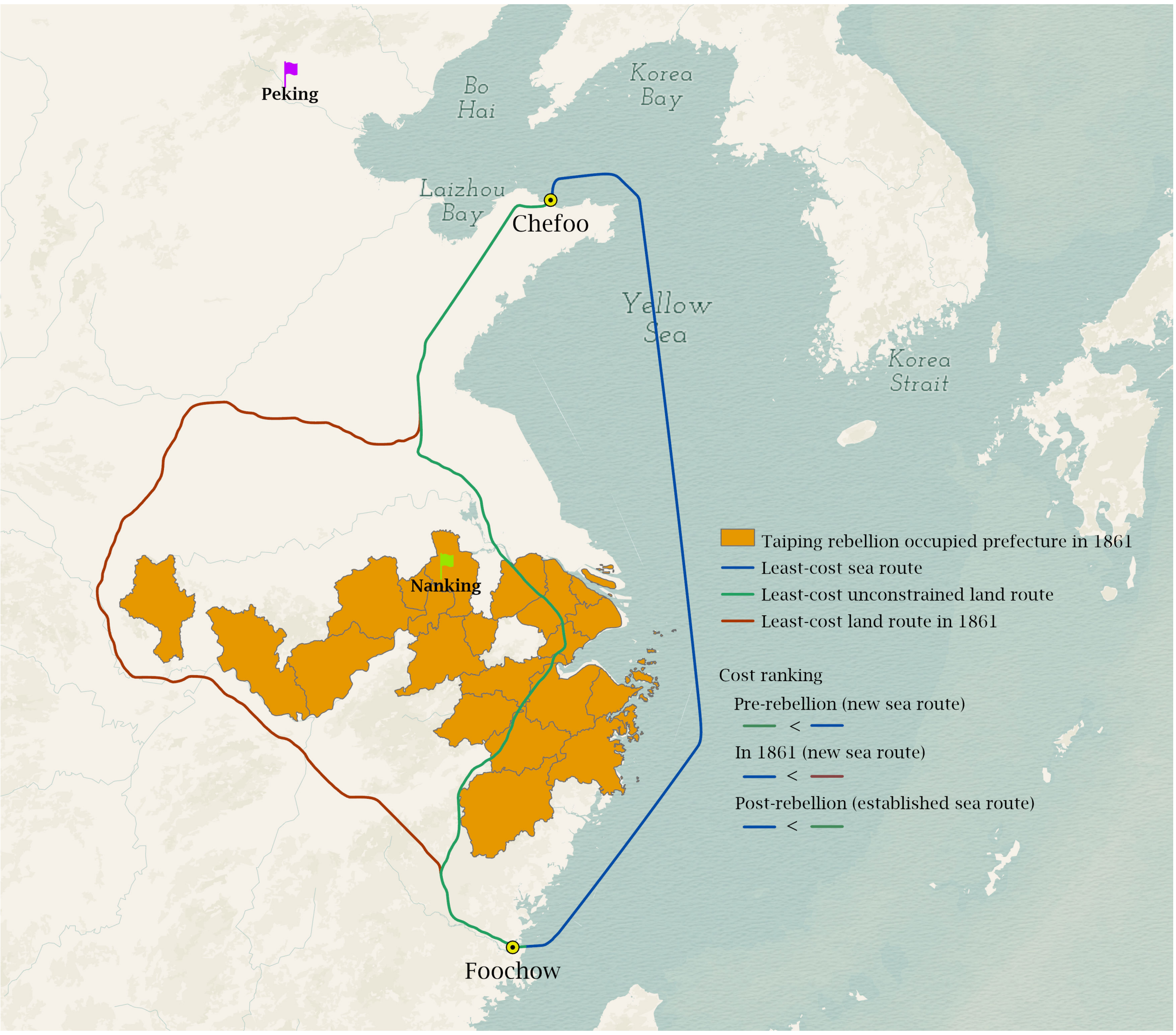


Notes: Based on the 2018 update of the prefecture population data in Cao (2000), 1982 China county census, and 2010 China township census. The pre-rebellion trends project forward the average annual growth rate of population for each group of prefectures (those that were occupied by the Taiping rebellion and those that were not) between 1680 and 1851.

There is debate about the extent to which large shocks can alter the equilibrium distribution of population across cities. Based on the Allied bombing of Japanese cities during World War II, Davis and Weinstein (2002) find that the long-run distribution of relative city sizes can be highly persistent even after large temporary shocks. Japanese cities experiencing more intense bombings quickly reverted to their pre-war relative sizes. Instead, I find strong permanent effects of war shocks. I argue this is because much of the medium-term impact on the distribution of population across Chinese cities operates through an indirect channel.

Despite large potential for sea trade, trade activities and population in China remained inland before the rebellion. When the rebellion blocked some of the main land trade routes in China, this forced a look for alternatives. It triggered substantial investments to facilitate sea trade, notably using steamships, which until then had a slow take-up. After the rebellion, with sunk investment already incurred for establishing sea lanes, many trade routes permanently shifted to sea transport. This in turn catalysed a shift of population towards port cities that had effects comparable in magnitude to the huge direct shock from the rebellion. An example of this can be seen in Figure 2. The figure shows in green the least-cost land route connecting Chefoo (located in today's Yantai) and Foochow (Fuzhou) before the rebellion. These two coastal cities could also have been connected

Figure 2: An example of the “roadblock effect”



Notes: Least-cost land and sea routes between Chefoo (today’s Yantai) and Foochow (Fuzhou) based on travel costs in *Panel A* of Table 2. Land routes can combine courier and non-courier land routes with different levels of terrain ruggedness as well as river transport. The cost ranking takes into account the cost of using each mode, the cost of inter-modal transfer, and the cost of developing a new land or sea route, as estimated in *Panel B* of Table 2.

by sea, but my estimates show that the cost saving compared with the usual land route was not large enough to incur the sunk costs associated with establishing such a route. These sunk costs include learning about the current and tides along the sea lane, being warned of dangerous spots, having staging-posts ready and ensuring safe anchorage on arrival. The rebellion occupied many of the prefectures traversed by the usual land route between Chefoo and Foochow. The dark shaded area in the map shows the rebellion-occupied prefectures in 1861. To transport goods between Chefoo and Foochow one would have had to take the large detour marked in red. According to my estimates presented below, the additional cost this detour entailed was large enough to instead incur the sunk cost of establishing a steamship route between these two cities. After the rebellion, although the unconstrained regular land route became once again usable, with the sea route already in operation, steamship trade remained a less costly option. Indeed the data show that this was active steamship trade between the two ports immediately after the rebellion. In contrast, regular land route between other city pairs (e.g. Amoy and Canton) was not blocked by the rebellion and steamship trade between them did not become established. I next explain how I examine this process systematically through my analysis.

I start by providing some historical background in Section 2, then describing the data in Section 3, and characterising the impact of the rebellion on population in Section 4. After controlling for pre-existing trends and other differences across prefectures, I find that being occupied during the rebellion was associated with a permanent population loss of 40% whereas being located on the coast was associated with a 30% population increase. This increase was even greater in coastal cities that saw a surge in steamship trade.

To separate the direct effect of the rebellion on population from the indirect effect operating through the shift from land to sea trade, we need to measure how the rebellion affected land trade routes. This is feasible because regions not directly hit by the rebellion may nevertheless be substantively affected if a regular land trade route becomes unusable when the rebellion hits some intermediate point on this connection. Before sea transport was widely used, the optimal trade route could combine roads with inland navigation. The relative cost of using different land trade routes and their sea alternatives is a key driver of the differential take-up of steamships in different geographical areas. The literature on transport costs usually infers them from freight rates, so the lack of data on freight rates in 19th century China imposes a challenge.

In Section 5, I develop a new method to estimate travel costs by different transportation modes using China's 1903 postal map. The map contains the locations of postal district headquarters and the assignment of prefectures to these. Under the assumption that the assignment was made to minimise transport costs to the postal district headquarter, I am able to parameterise a rich transport cost function, including the typical speed of different transport modes, variations in speed due to road types and the terrain, as well as the fixed costs associated with transitions between modes (which can be unavailable even in modern times).

The key for my identification strategy is that, depending on their location along the regular land trade network, locations would have had their trade flows more or less affected. Maritime trade could boom broadly when the general technology of sea transport advanced. At the same time, land trade could be directly hit in areas occupied by the rebellion, since an explicit aim was to cut off supplies from this area to the capital of Qing government in Peking. However, bilateral trade flows by land to and from unoccupied locations were also affected indirectly and differently depending on roadblocks on the corresponding routes and the feasibility of sea trade alternatives. Thus, in Section 6, I proceed to show that the probability of sea trade after the rebellion is persistently higher for locality pairs where the regular land trade route connecting them becomes unusable because of the conflict.

To implement this strategy, I combine new data on bilateral historical trade flows with a trade model featuring modal and route choice. I collect a new data set of bilateral maritime trade in 19th century China by digitising archives of trade reports from the China Maritime Customs Service.[3] Given that only steamship trade is observed in the data, I need to place some additional theoretical structure into the analysis to characterise the indirect effect of the rebellion through trade. Trade models typically do not incorporate a modal choice, since trade data is generally not differentiated by which transport method was used to carry the cargoes. Instead, I add modal choice (by road, by steamship, and by sailing ship) to an otherwise standard trade model. Steamships required fixed investments, and this likely held them back until blockage of land routes made such investments a necessity. Thus, I also incorporate sunk investment costs into the model.

The model features a "roadblock effect," by which a temporary increase in land trade costs can trigger sunk investments for steamship trade, thus having permanent effects on modal choices. Given the relative transport costs via each possible mode between every pair of locations, as estimated in Section 5, the model yields the probability of steamship trade for that pair, which we observe in data. Thus, across location pairs, the heterogeneous increase in land trade costs during the rebellion can be linked to the increase in the probability of steamship trade immediately after the rebellion, mediated by the incentives to make sunk investments for steamship trade. Despite the limitations imposed by only being able to observe trade costs before, during, and after the rebellion, and steamship trade immediately after the rebellion, I am able to pin down jointly the elasticity of trade costs and sunk infrastructure costs.

The predictions of the model closely match observed modal choices. In particular, the model predicts a probability of steamship trade between "roadblocked" location pairs of 0.72 compared with 0.42 for "non-roadblocked" location pairs. The corresponding probabilities in the data are 0.71 and 0.37. The model also suggests that the "roadblock effect" was very important in practice. Under a counterfactual scenario that eliminates roadblocks between roadblocked location pairs, the model estimates that the probability of steamship trade between them would have been much lower, 0.44 instead of 0.72.[4]

Within the literature on the effect of temporary shocks on relative city sizes already mentioned above, some other papers also incorporate changing comparative advantage. Bleakley and Lin (2012) document the continuing importance of historical portage sites as city locations in the United States, although their original advantages have long since become obsolete. Michaels and Rauch (2017) also consider en episode where natural advantages shift over time, but in their case such a shift is combined with a larger population shock in one country than in another. In particular, they compare the diverging experiences of British and French cities following the collapse of the Western Roman Empire. In Britain, subject to a more intense shock, the urban network was displaced towards emerging navigable waterways. In France, where the ensuing destruction and deurbanisation was more limited, medieval towns instead re-emerged more often on their Roman-era locations. I also consider a historical episode where shifting natural advantage (from the introduction of steamships)

[3]The China Maritime Customs Service was a bureaucratic agency set up in 1853 by foreign consuls in Shanghai to collect tariffs that went unpaid when the the Qing government was overwhelmed by the Taiping Rebellion.

[4]The model also delivers a gravity equation of trade. Using the trade costs I am able to construct using the method introduced in Section 5, in AppendixE I estimate the elasticities of trade and migration with respect to distance in historical China. Reassuringly, these elasticities are similar to those estimated with modern data and for other parts of the world.

was combined with large but heterogeneous population shocks (from the Taiping rebellion). A key contribution of my framework is that, in addition to the direct channel operating through temporary population shocks, I highlight an indirect channel operating through temporary roadblocks and the incentives they provide to invest in new transport technologies.

In this respect, the paper also connects to the literature on the importance of trade costs and market access in determining the distribution of population across cities (e.g. Redding and Venables, 2004; Faber, 2014; Donaldson, 2018) and within cities (e.g. Baum-Snow, Brandt, Henderson, Turner and Zhang, 2017; Tsivanidis, 2018; Heblich et al., 2020). An important challenge when trying to evaluate the consequences of transport improvements is the endogeneity of infrastructure investments. An advantage of the setting in this paper is that the Taiping rebellion creates exogenous variation in the incentives to invest in sea transport technology.

The paper also adds to the literature on the effect of wars and conflicts on urban development (e.g. Voigtländer and Voth, 2012; Dincecco and Onorato, 2016; Glaeser and Shapiro, 2002). While the previous papers have focused on local impacts of war, this paper shows that the indirect effect on other regions via trade disruption and the ensuing incentives to adopt new technologies can be just as important.

A debate related to the persistence of city locations pertains the slow adoption of new technologies, even when they are obviously superior. This could be due to the trade-off between the sunk costs of developing and adopting new technologies and the accumulated learning by doing of incumbent technologies (Brezis and Krugman, 1997). In fact, the introduction of a new technology sometimes revitalises innovation of an incumbent technology. Ward (1967) coined the term "sailing ship effect" for this phenomenon, following the analysis by Gilfillan (1935) of advances made in sailing ships in the second half of the 19th century in response to the introduction of steamships. According to Ward, such advances led to greater improvements in sailing ships than those made in the previous three centuries.

The "roadblock effect" I document and investigate in this paper works in precisely the opposite direction: the temporary blockage of land trade routes during the Taiping rebellion triggered substantial investments in infrastructure to facilitate the use of steamships instead. This had long-term consequences for the distribution of population across Chinese cities of similar magnitude to the direct shocks of the rebellion.

2. Historical background

A *The Taiping rebellion*

The Taiping rebellion (1851-1864) was the largest peasant revolt in the history of China and one of the deadliest civil wars in human history. About one out of every four people in China died during the rebellion (Ge et al., 1999). It started as a guerrilla warfare at Kweiping, Guangxi, in southwestern China. Peasants rose up against the Qing government, which levied ever-increasing taxes after a series of natural disasters and economic crisis following its defeat at the First Opium war. The uprising escalated into to a civil war with the Qing government. The rebels swiftly moved northeast to occupy Nanking and established around it "the Taiping Heavenly Kingdom" in the mid and lower Yangtze region (Figure 3). When the rebels' attempt to besiege the imperial capital Peking was defeated, wars between the two forces receded to the area occupied by the rebellion, causing great damage and civilian deaths.[5] Some prefectures in the occupied region lost up to 80% their population and many places were razed to ground (Cao and Li, 2000).

Before the rebellion, the area around the Yangtze river, especially its lower section, had become the most prosperous part of China, contributing to half of the nation's wealth. The Qing government relied on the Grand Canal to transport tax-in-kind from this region to the imperial capital. The rebels strategically seized this area as a base to overthrow the Qing dynasty. They occupied prefectures sitting on the Grand Canal to strangle the flow of good from the lower Yangtze to the imperial capital. With their supplies through the regular routes combining land and canal transport cut off, the Qing government was forced to switch to sea transport to ship tribute grain via Shanghai and Tientsin.

This paper will show that this modal change was not unique to the transport of tribute grain nor to trade between Shanghai and Tientsin, but instead affected a much broader exchange of goods across the whole country. The military confrontation between the rebellion and the Qing government severely curtailed land transport that previously used trade routes traversing through the rebellion-occupied region. Depending on the roadblock of their usual land routes and the feasibility of sea trade alternatives, bilateral trade flows between unoccupied locations were affected differently, triggering investments facilitating steamship trade on some routes but not others.

B *The failure to adopt sea trade before the Taiping rebellion*

Since I will argue that the Taiping rebellion acted as a catalyst for sea trade, using steamships in particular, it is important to also understand the status of sea trade before the rebellion.

One reason for the weakness of sea trade prior to the Taping rebellion was the concentration of population away from coastal areas. Early during its rule, the Qing government, which ruled China from 1644 to 1912, implemented a sea ban on the mainland. The purpose was to cut off supplies to the Zheng regime in Taiwan, which was founded by loyalists of the previous Ming dynasty. As a part of the policy, coastal residents in the mainland were evacuated to areas 16-26 kilometres inland. When the Zheng family surrendered in 1683, the ban was lifted and former coastal residents were allowed to return. Thereafter, except for a short sea ban placed between 1717 and 1727, maritime trade with foreign countries was acquiesced [6]. Nevertheless, it did not recover to its full blossom in the previous dynasties and population in coastal regions continued to decline.

[5] The rebellion indirectly spurred conflicts in other regions during this period. As Qing troops from all parts of China were transferred to fight against the Taiping soldiers, the military vacuum spurred a series of uprisings in the periphery areas of China, including Yunnan, Shensi, Kansu, Szechuan and Kueichow provinces.

[6] In 1758, the Qing government established the Canton system, restricting all trade with western countries to only one port in the South, Canton. Seen by some scholars as a closed-door policy, the restriction was not strictly carried out. Foreign ship arrivals in other ports were documented. Their numbers and trade value were not restricted (e.g. merchant ships arrivals in Canton between 1757 and 1838 increased by 16 times compared to the previous 72 years and all ports combined.). Trade with non-western countries was not subject to the Canton rule.

Figure 3: The Taiping rebellion and imperial postal districts

Notes: Author's map based on the 1903 postal working map from the China Maritime Customs Postal Series and the China W dataset. Prefectures delimited in light grey and coloured if occupied by the Taiping rebellion. China's historical postal districts delimited in black with the district headquarter marked by a green pin.

Traditionally, domestic trade used the combination of land transport and inland navigation.[7] By the beginning of 19th century, canal shipping had become increasingly costly, yet flows of good between different regions continued to rely heavily on the Grand Canal.[8] At the time, the canal was constantly flooded by the Yellow River, which called for regular and costly maintenance through damming up the river and dredging the canal. The problem became so severe that in 1826 the emperor Daoguang launched an initiative to transport tax-in-kind by sea from the lower Yangtze region to the imperial capital. The trial was promising. Sea transport appeared to be one-third as costly and much faster than canal shipping. [9] Yet the initiative met strong political opposition and was quickly abandoned.

In 1842, after the defeat at the First Opium War, China was forced to further integrate into the global market by opening up treaty ports, in addition to Canton, to trade with western countries. Foreign steamships started to appear in large numbers by Chinese coast.[10] Compared to sailing ships, they were faster and less reliant on wind and currents. Despite these new changes, trade routes remained persistent. Foochow was one of four ports signed to open, but its celebrated black tea continued to take a lengthy detour overland to Canton for export, which took ten times longer than a direct export from Foochow. Not until the outbreak of Taiping rebellion, when the regular land route to Canton became blocked, was the first batch of tea shipped directly from Foochow.

The inertia in the use of established land and river routes could be due to a large sunk cost usually associated with establishing new trade routes. Especially for sea trade, sailors need to harness the current and tides along the way, be warned of dangerous spots, have supplies when needed and find safe anchorage on arrival. The large investment needed for safe voyage became evident when concerted efforts were made by China Maritime Custom Service in late 19th century drawing on expertise of mechanics, engineers, surveyors and constructors from home and abroad to build a chain of lighthouses and a system of navigation aid to safeguard trips along the Chinese coast and the Yangtze River.[11] The scope of the investment was something that no individual firm before 20th century could undertake without generous support from the state. While collectively, firms could benefit from investment to facilitate sea trade, in the absence of coordinated efforts, the gain from engaging in sea trade for any individual firm would not be large enough to compensate it for the large investment upfront, which resulted in inefficient modal choice.

[7]Domestic sea trade was used for some specific commodities (e.g. export of soy bean from the northeast to Shanghai), but it was never the major transport method. It also varied substantially from year to year: there could be 85 ship arrivals in one year and only 2 in the next (**?**). Being unpredictable and marginal, there was barely any infrastructure developed for sea trade.

[8]The Grand Canal is a 1,776 km canal that connects Peking, the capital, to Hangchow in the lower Yangzte region, passing through from the north to the south Tientsin, Hopeh, Shantung, Kiangsu and Chekiang. It formed the inland navigation network of China connecting the Yellow River and the Yangtze River.

[9]It took sea transport ten to twenty days to transport tribute grain compared to two months by canal (Wei, 1826).

[10]Steamships first appeared in China sea as early as in late 1820s and early 1830s.

[11]These efforts include surveys by British Royal Navy since 1842, a series of updates on *Notice to Mariners* and construction of lighthouses, beacons and buoys (**?**).

3. Data

To study the "roadblock effect" on sea trade, I collect a new data set of bilateral domestic maritime trade between treaty ports by digitising archival reports of *Returns of Trade at Treaty Ports* published by China Maritime Customs Service (Henceforth CMCS). CMCS was a bureaucratic agency established in 1853 by foreign consuls in Shanghai to collect tariffs that went unpaid when the Qing government became overwhelmed by the Taiping rebellion.[12] CMCS supervised tariff collection of foreign export and import, as well as domestic maritime trade that went through treaty ports. Its statistic department published trade reports following strict statistics standards that are consistent across ports and over time. The trade data from CMCS is believed to be "the only reliable and systematic material" to study China's trade in history (Cheng et al., 1956). Keller, Li and Shiue (2012) made a thorough introduction of the trade data from CMCS.

The earliest data on bilateral domestic trade flow is available for 1867 when more accurate and consistent accounting practice was established, which is 3 years after the Taiping rebellion ended. In addition, I digitised bilateral domestic trade for 1873, 1882, 1887, 1892 and 1897 to look at the long-term effect of rebellion.[13] The nice feature of CMCS data is that it differentiates between exporting and re-exporting. I use direct export to measure trade between treaty ports. The limitation of the CMCS data is that it only includes domestic maritime trade between treaty ports, which were ports opened to trade with western countries, and only the part of maritime trade carried by steamship. Maritime trade by traditional sailing ships in China (i.e. junks) went missing for most years and ports.

To deal with this data limitation, I build a trade model featuring transport modal and route choice, which helps to infer the missing trade by other transportation methods. To estimate the model however we need some additional inputs, which are the trade costs of using different transport methods in 19th century China.

While in the literature transport costs are usually derived from freight rates or travel speed, this information is not available in historical China. In Section 5, I develop a novel method to estimate transport cost in history based on the index to the postal working map in 1903, which lists each prefecture in China under a post district. Post districts were regional divisions of the imperial post office that was established by CMCS in 1896 to provide mail delivery service to the general public. The service expanded from the post district headquarter, usually a treaty port, to inland area

[12]CMCS is a Qing government agency, but its staff is a mix of Chinese usually in lower rank positions and foreigners performing managerial roles. Set up to assess taxes of foreign import and export, overtime its role extended to the collection of domestic tariffs, the maintenance of harbours and lighthouses, weather inspection, the payment of foreign loans and the establishment of the modern postal system in China.

[13]Trade was recorded in Customs Taels, a silver based monetary unit uniform across all ports and for all years except for 1867. In 1867, the southern ports reported their trade value in Mexican dollars and I converted to the Customs Taels using the conversion rate at the time. In 1901, CMCS took over native customs within 25 kilometres of the treaty port, trade data may start to include trade by sailing ships that went through native customs. Therefore I use domestic trade data until 1897.

through its extensive delivery network (See Figure 3 for post districts and their headquarters. [14]). [15]

The imperial postal office developed a system of transliterating Chinese places names, referred to as postal romanisation. In this paper, I use postal romanised place names as they appear in trade reports to indicate the reference to the historical places. The mapping from the postal place names to their current locations can be found in Table A.1.

Transport cost by land is modelled to vary with road types (i.e. courier routes and navigable rivers) and terrain. I obtain the locations of Qing courier route (1800-1900) from China W dataset, locations of rivers from CHGIS V4 dataset of coded river in 1820 and terrain ruggedness from Nunn and Puga (2012).[16]

Transport by sea can use both steamships and junk sailing ships. The relative unit transport cost of the two is taken from Pascali (2017), which covers the same period. To get the transition cost from land to steamship, I digitised CMCS *Returns of Trade at Treaty Ports* in 1904, which has information about the usage of steamships and sailing ships at two treaty ports, Newchwang and Kiaochow, in their interregional trade with locations over various distances.

To look at the effect rebellion on population, I use the 2018 update of historical prefecture population data from Cao (2000) [17]. The data is constructed based on government administered census, cross-checked with literary accounts, notes and more than 3000 local gazettes. Prefectures are the second level administrative divisions in China. They are subdivisions of provinces, followed by counties and towns. Historical prefecture population is available for the pre-rebellion period in 1680, 1776, 1820, 1851 and post-rebellion period in 1880, 1910 and 1953. I supplement with population from 1982 county census and from 2010 township census and map modern population to the historical prefecture boundaries in 1820 taken from China Historical Geographic Information System (CHGIS V4), constructed by Harvard Yenching Institute and Fudan Center for Historical Geography.

I obtain the information about occupied prefectures by the rebellion on a monthly basis from **?** and Hua (1991). To measure roadblock, I construct a measure of yearly occupation by the rebellion from 1853 to 1864 by aggregating occupied prefectures within one year.[18]

[14]The headquarter of each post district was usually a treaty port and custom commissioner the postmaster, with the exception of Peking and Chinan districts.

[15]There was no major change in land transportation technology in 19th century China. Railway was a nascent technology and its wider application in China need to wait until 20th century. In 1903, only regional railways operated in the northern part of China surrounding the capital. They were rail lines connecting Chengting to Peking and Peking to Tientsin and Chinchow. Therefore the planning of post districts that could be affected by the railway network are dropped from the estimation (i.e. Peking, Tientsin and Newchwang districts). The Yellow River changed its course in 1855 and could potentially affect the part of courier routes where the new river intersected. The post district in charge of the region was Chinan and is also excluded from the estimation.

[16]Qing courier route was a transportation network of roads, canal and smaller rivers. The rivers are coded with values between 1 to 6. I keep rivers of order higher than 4, which corresponds to navigable rivers.

[17]The previous version of the data is used by Jia (2014) to study the long-run effect of treaty ports on population and economic development in China.

[18]Although the earliest uprising took place in Kweiping, Kuangsi in 1851, the rebels advanced to the middle and lower Yangtze region and occupied the first prefecture, Chinkiang, in January, 1853. The last occupied prefectures fell to Qing government in August, 1864.

4. Population Evolution

In this section, I estimate a reduced-form relation between population, rebellion occupation, coastal location and sea trade using Equation 1:

$$\log pop_{it} = \alpha post_t + \beta(post_t \times X_i) + \sum_{I=prov} \gamma_I t + \psi_i + \epsilon_{it} \tag{1}$$

Each observation is one of the 179 prefectures from 13 provinces in the core region of China at each data year [19]. I regress the natural logarithm of population at prefecture i in each data year t on an indicator whether the observation is post rebellion ($post_t$) and interact post dummy with a set of characteristics of the location (X_i). The coefficient on the interaction tells us whether after the rebellion, the level of population in prefectures with those characteristics increases more than others. I include prefecture fixed effect (ψ_i) to control for any time invariant differences across prefectures and allow for provincial time trend (γ_I) to account for long-term population trends. Standard errors are clustered at the province level to adjust for heteroskedasticity and within-province correlation over time.

The results are shown in Table 1. In column (1), the post-rebellion dummy indicates that there is a drop in the level of population of about 38% for all prefectures after the rebellion. The interaction between post dummy and indicator for rebellion occupation shows that the civil war has a much larger effect on rebellion occupied prefectures. The decline of population in occupied prefectures is 47% larger, confirming the pattern found in Figure 1.

In addition to the direct impact of rebellion on war-ravaged region, it could affect relative attractiveness of locations through diverting trade routes to sea. Places that stand to benefit from sea trade were coastal prefectures, and I investigate this possibility in column (2). I include an interaction between post and coastal dummies. The coefficient on the interaction is positive and significant, indicating that the level of population in coastal prefectures increases by 25% after the rebellion, after controlling for the direct effect of rebellion on occupied region.

We may think this is driven by other characteristics of coastal areas. While prefecture fixed effects take care of the differences across prefectures in level, those characteristics may have a differential impact on population growth after a large shock to population sizes. For instance, coastal areas may have better agricultural productivity, which may attract migration after the rebellion. Therefore in column (3), I include a variety of locational traits interacted with the post dummy. These are the initial level of human capital and a set of geographic controls including terrain ruggedness, agricultural suitability for major crops in historical China (i.e. rice, foxtail millet and sweet potato) and whether the Grand Canal went through the prefecture. The positive coefficient estimate of the interaction on post and coast dummies becomes even larger after controlling for differential impacts of other locational characteristics on population after the rebellion, indicating a 30 % increase in the level of population in coastal prefectures after the rebellion.

[19] The Taiping rebellion spurred uprisings in Yunan, Shensi, Kansu, Szechuen and Kueichow at the same time. These were periphery regions inhabited also by non-Han minority groups, so they are dropped from the analysis. Nomadic regions of Qing Dynasty such as the Northeast provinces (Kirin, Heilungchiang and Shengching), Outer Mongolia, Inner Mongolia, Xinjiang, and Tibet are also excluded.

Table 1: Prefecture population evolution 1680–2010

	Dep. var.: Log (population)				
	(1)	(2)	(3)	(4)	(5)
Post-rebellion	-0.376***	-0.430***	-0.195	-0.434***	-0.139
	(0.0886)	(0.0933)	(0.209)	(0.0976)	(0.250)
Post-reb. × occupied	-0.468***	-0.448***	-0.400***	-0.430***	-0.394***
	(0.0784)	(0.0797)	(0.0965)	(0.0887)	(0.105)
Post-reb. × coastal		0.253***	0.300***	0.181**	0.422**
		(0.0600)	(0.0873)	(0.0740)	(0.156)
Post-reb. × coastal × trade val.				0.184***	0.138**
				(0.0365)	(0.0615)
Post-reb. × initial human capital			-0.0476		-0.0683*
			(0.0274)		(0.0337)
Post-reb. × geographic controls			Yes		Yes
Prefecture fixed effects	Yes	Yes	Yes	Yes	Yes
Provincial trend	Yes	Yes	Yes	Yes	Yes
Observations	1,610	1,610	1,565	1,368	1,323
Number of prefectures	179	179	174	152	147
Number of coastal prefectures	35	35	35	8	8
R^2	0.740	0.756	0.780	0.745	0.775

Notes: In columns (4) and (5), trade val. is the natural logarithm of domestic trade value by steamship in 1867, which is only available for 8 out of 35 treaty port prefectures (and is 0 for all 144 non-coastal prefectures). In columns (3) and (5), initial human capital is proxied by the natural logarithm number of graduates from the Imperial Exam in the 1840s, which is not available for 5 prefectures. The geographic controls are an indicator for whether the Grand Canal passes through, terrain ruggedness, and categorical variables for agricultural suitability for major crops in historical China. Robust standard errors clustered by province are reported in parentheses. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

To provide suggestive evidence that population relocated because trade activities were diverted to coast as the rebellion blocked regular land trade routes, in column (4) in addition to the interaction between coastal and post dummies, I further interact them with the natural logarithm of the value of domestic trade by steamship in 1867 to see whether port cities with more sea trade had a larger increase in population after the rebellion. Using the port with the lowest value of trade as a benchmark and normalising its trade value to one, we can reinterpret the coefficient on the post and coast dummies interaction as the population increase of the benchmark port. The coefficient on the triple interaction represents additional population growth in coastal prefectures associated with the increase in value of sea trade relative to the benchmark. The result suggests an increase in sea trade by 1% is associated with a 0.19% increase in the level of population after the rebellion. On the basis of column (4), column (5) further controls for locational characteristics, interacted with the post dummy. The general picture remains the same, indicating that the positive association between population and trade is not confounded by locational characteristics that can at the same time affect sea trade.

However, results on column (4) and (5) need to be viewed with some caveats. The data on the value of domestic trade is only available for a subset of port cities, which were treaty ports opened for trade with western countries. Also the value of sea trade only includes the part that was carried by steamship, therefore can be sensitive to availability of using sailing ships as an alternative. Since trade data comes in more details (i.e. we know who trade with whom by sea), we can identify the "roadblock effect" on bilateral trade by exploiting the regional variation in incentives to use sea trade depending on whether their regular land routes were affected by the rebellion, which explains population relocation to coastal regions with more steamship trade after the rebellion. For this closer look at the effect of rebellion on trade routes and modes, we need an intermediate step of estimating transport costs during this period, which is detailed in the next section.

5. Estimation of Historical Transport Cost

The relative cost of using different land trade routes and their sea alternatives is a key determinant of the differential impact of rebellion on steamship take-up. While previous literature on transport costs usually infers them from freight rates, lack of data on freight rates in 19th century China imposes a challenge.

To deal with this problem, I develop a novel method of estimating transport costs in history based on China's postal map in 1903 and data on sea trade in 1904. I allow typical speed of different transport modes (e.g. by land, by sea), variations in speed due to road types, vehicle and terrain, as well as fixed cost associated with transitions between modes (which can be unavailable even in modern times). I divide this section into two parts: estimation on land transport cost based on postal map, followed by estimation on sea transport cost using trade data on usage of steamships and sailing ships.

A *Land Transport Cost*

The index to the postal working map in 1903 lists each prefecture in China under a post district. The post districts are regional divisions of imperial post office, which expanded the mail delivery service from the post headquarter, usually a treaty port, to inland prefectures. Under the assumption that the allocation of prefectures to post districts is made to minimise travel costs to their post headquarters, we can infer the cost of using different land transport infrastructure available along the route.

I divide the surface of China into 0.1°by 0.1°cells (553$\times$ 826 cells in total, each covering about $90km^2$ area at 30°latitude) and overlay with courier routes, rivers and terrain (See Figure B.1). Each cell is assigned a (relative) cost to traverse based on whether it falls into river, courier route or point of transition between modes and road types. Given a set of transport cost for courier land route cell, non-courier land route cell, river cell, cell involving land to river modal change and courier to non-courier route transition, and penalty for ruggedness, we can use Fast Matching Method (FMM) to find the least-cost travelled post headquarter for each prefecture. The optimal transport cost is chosen by maximising the number of matches between actual and FMM allocations of prefectures to postal headquarters.

In the end, 167 prefectures out of 199 can be matched to the actual post districts compared to 111 when using the distance alone. The majority of the unmatched prefectures (25 out of 29) are very close to be matched to their actual headquarter as their neighbour is matched to that headquarter.

B *Sea Transport Cost*

From the estimation based on post map we derive transport cost by land. We also need to know the transport cost by sea compared to land, as the relative advantage of using sea transport and its change during the rebellion is a potential driver for adoption of sea trade.

Sea transport in 19th century of China can use two technologies: sail and steam ships. Sailing ships are also used in river, therefore we can assign the unit transport cost by river estimated from the post map to sailing ships. The same applies to the cost of transition from land to sailing ship, which is set by the cost of modal change from land to river. For steamships, I use the relative cost of steam compared to sail ships from Pascali (2017), which covers the same historical period, and set the unit transport cost by steamship half the cost by sailing ship. Steamships have lower cost per distance travelled than sailing ships, but because of its greater size, a larger fixed cost would be needed for transshipment, therefore is more suitable for long-distance trade. To estimate the fixed cost associated with transition from land to steamship, I exploit trade data on two treaty ports (Newchwang and Kiaochow) in 1904, which details the use of steam and sail ships for their trade to/from other ports with varying distance.

The probability of using steamship is an increasing function of the relative cost of sail to steam ship. The relative cost of using sail compared to steam ships depends on their transport cost by distance and the cost of transition from land to sailing ship, which we already know; and the modal transition from land to steamship, which needs to be estimated. The probability of having steamship trade varies in a positive direction with distance between two ports, as steamships have lower cost per distance travelled, but potentially a larger fixed cost associated with modal transition. Therefore we can estimate the transition cost from land to steamship by maximising the probability of observing the use of steam and sail ship for sea trade with varying distance. The detailed estimation can be found in AppendixC.

The estimation results on transport cost are shown in *Panel A* of Table 2. I normalise the cost of travelling by land without any road infrastructure or ruggedness to 1 so that the unit transport cost by other modes is relative to that of plain land. The cost of using courier route is about a quarter of the cost of non-courier land route. Transport cost by sailing ship is about 13% the cost of non-courier land transport and by steamship about 7%. An increase in terrain ruggedness by one standard deviation is associated with 20% increase in transport cost. Regarding transition, I convert the cost into equivalence to distance travelled by land. The transition cost from non-courier to courier land route amounts to the cost of travelling by flat land for 1.1 kilometres. The transition between different modes is much more costly. Going from land to sailing ship, there is a fixed cost equivalent to travelling 11 kilometres by land. The transition from land route to steamship is about three times the transition to sailing ship, which is reasonable given steamships come larger in sizes therefore require longer time for transshipment.

6. The Roadblock Effect

In this section, I exploit the spatial variation in steamship trade, using data on bilateral trade digitised from CMCS, and the heterogeneous incentives to adopt sea trade through the route and modal changes induced by the rebellion, estimated using the novel method developed in the previous section, to identify the "roadblock effect". Given only trade by steamship is observed, I need to place some additional theoretical structure into the analysis. I add transport modal and route choice to an otherwise standard trade model, which is usually not incorporated as trade data is generally not differentiated by which transport method was used to carry the cargoes.

We consider a country with J cities (prefectures in historical China), indexed by $j = 1,2....J$. A continuum of symmetric firms in each city produces a distinct product, which can be consumed locally as well as in other locations subject to iceberg trade cost.[20] If we assume that consumers have constant elasticity of substitution utility and firms face monopolistic competition, we can write firm profit function of exporting merchandise from location j to i at time t:

$$\pi_{ijt} = (1-\alpha)(\frac{\tau^m_{ijt} c_{jt}}{\alpha P_{it}})^{1-\epsilon} Y_{it} - c_{jt} f. \tag{2}$$

The elasticity of substitution is denoted as α, which falls between 0 and 1. To keep the notation concise, I introduce $\epsilon = 1/(1-\alpha)$. As standard in trade models, the profit increases with income Y_i and Dixit-Stiglitz price index P_i in the destination $(1-\epsilon < 1)$ and decreases with the cost of production c_j, which is common to all firms in origin j. There is a fixed cost of entry f , which can be thought of as the number of workers needed to build the firm.

Departing from the standard setting, I allow the iceberg trade cost τ^m_{ijt} to vary with transport mode m. There are three prevailing transport modes: by land, by steamship and by junk (Chinese sailing ships), which I denote l, st, jk. In addition to make the trade cost a function of transport cost D^m_{ijt}, which takes into account optimal route choice within each mode, I incorporate intertemporal investment for establishing new trade routes for sea transport I^s_t and land transport I^l_t. The modal-dependent iceberg trade costs are:

$$(\tau^l_{ijt})^{1-\epsilon} = (D^l_{ijt} I^l_t))^{-\gamma} e^{\eta_{l,ijt}}, \quad (\tau^{st}_{ijt})^{1-\epsilon} = (D^{st}_{ijt} I^s_t)^{-\gamma} e^{\eta_{s,ijt}}, \quad (\tau^{jk}_{ijt})^{1-\epsilon} = (D^{jk}_{ijt} I^s_t)^{-\gamma} e^{\eta_{s,ijt}} e^{\eta_{j,ijt}} \tag{3}$$

for land, steamship and junk respectively. Investment cost is modelled here as a multiplier to transport cost, so that we can interpret it as the cost ratio of establishing a new route over using an existing route.[21] The benefit and cost of investment in establishing new sea routes are shared by sail and steam ships. Its investment cost I^s depends on whether a sea route has already been

[20] Recent trade models have incorporated heterogeneous firms (e.g. Helpman, Melitz and Rubinstein, 2008; ?), since only aggregate but not firm level trade data is available in 19th century of China, there is no gain from modelling heterogeneous firms here.

[21] The implicit assumption here is that sunk investment is proportional to transport cost. This is reasonable assumption given establishing a new sea route involves acquainting with tides, current, dangerous spot and installing staging posts along the way. These costs all increase with distance.

established in the previous period. Define C_{ijt} as the indicator whether sea trade happens in the previous period $t-1$. We have:

$$I_t^s = \begin{cases} I^s, & \text{if } C_{ijt-1} = 0. \\ 1, & \text{if } C_{ijt-1} = 1. \end{cases} \tag{4}$$

If sea route was established in the previous period, there is no extra cost for using sea transport in addition to the transport cost. If instead sea transport was not used previously, firms need to incur the sunk investment cost for establishing a new sea route in order to use sea transport. The investment cost for new land routes I_t^l is defined analogously.

Trade costs are also subject to shocks. There are three types of shocks in the model. The shocks to land and sea transport $\eta_{l,ijt}$ and $\eta_{s,ijt}$ are modelled to be realised annually, and could be thought of as medium-term shocks to transport cost (e.g. road damage, port maintenance). In addition to common shocks to sea transport for both sail and steam ships, sailing ships are subject to an additional shock $\eta_{j,ijt}$. As sailing ships are vulnerable to changes in current and wind, they are susceptible to weather shocks that are more short-lived (e.g. storms). Therefore I assume their shocks to be realised for T times within a year.[22] All shocks are assumed to be log normal and independent:

$$\eta_{l,ijt} \sim \mathcal{N}(0,\sigma_l^2), \quad \eta_{s,ijt} \sim \mathcal{N}(0,\sigma_s^2), \quad \eta_{j,ijt} \sim \mathcal{N}(0,\sigma_j^2). \tag{5}$$

The timing of the events are the following: at the beginning of each year, firms observe origin-destination specific shocks to land and sea transport and choose either land or sea transport for that year; Afterwards, shocks to sailing ships are realised for multiple times within the year. After each realisation, firms that chose sea transport in the first stage choose between steamships and junks.

From the model set-up, we can write the probability of sea trade in any given year. As sea transport was generally not available before the rebellion and sunk investment cost had not been incurred, we can ignore the pre-rebellion period as it does not affect decision on adopting of sea trade after the rebellion. The period during the rebellion matters. When land transport was curtailed by the rebellion, it forced some locations to look for alternative ways to transport goods, which can have a long-term impact on modal choices through investment in establishing sea trade. This can be seen from the probability of using sea trade after the rebellion:

$$Pr(C_{ij,post} = 1) = \Phi\left(\frac{\gamma}{\sigma} \ln \frac{D^l_{ij,post}}{D^s_{ij,post}} - \mathbb{1}(C_{ij,reb} = 0) i^s\right), \tag{6}$$

which depends first on the post-rebellion transport cost by land relative to sea ($D^l_{ij,post}/D^s_{ij,post}$).[23] The elasticity of trade with respect to transport cost γ, normalised by the variation of transport shocks $\sigma = \sqrt{\sigma_l^2 + \sigma_s^2}$, governs how responsive usage of sea trade is to the relative cost of land transport. Second, it depends on the sunk investment cost for sea transport, which may or may not has been incurred during the rebellion. To keep the notation concise, I introduce $i^s = \gamma/\sigma \ln I^s$,

[22]The discrete number of shocks is modelled because of zero trade by sailing ship in the data.

[23]The bilateral sea transport cost $D^s_{ij,post}$ is the cost of using steamship or junks, depending on whose cost is expected to be lower for locality pairs.

which measures the effect of sunk investment cost on the probability of trade. If during the rebellion, sea trade was used and therefore sea route established ($\mathbb{1}(C_{ij,reb} = 0) = 0$), in post-rebellion period, firms do not need to incur the sunk investment cost, therefore the probability of using sea trade increases.[24] The probability of using sea trade during the rebellion period is:

$$Pr(C_{ij,reb} = 1) = \Phi(\frac{\gamma}{\sigma} \ln \frac{D^l_{ij,reb}}{D^s_{ij,reb}} + \mathbb{1}(D^l_{ij,reb} \neq D^l_{ij,pre}) i^l - i^s). \tag{7}$$

As trade routes by sea have not been established, using sea transport requires investment i^s during the rebellion. The rebellion could incentivise adoption of sea trade through two channels. First, for locality pairs whose regular land route was blocked as the rebellion occupied prefectures that it previously traversed, finding alternative land route entailed an investment cost i^l. Second, as the new route was not the previous unconstrained optimal one, the relative cost of land transport ($D^l_{ij,reb}/D^s_{ij,reb}$) increased compared to sea. Both raised the relative desirability of sea transport during the rebellion, which could incentivise roadblocked localities to incur the sunk cost for establishing sea trade. Here I measure roadblock and their resulting increase in land transport cost by assuming that the prefectures under occupation were impossible to pass. I calculate the least cost routes for each locality pair in unoccupied region for each year during the rebellion, and use the year in which land transport of the pair was affected the worst. As the decision for using sea trade is linked intertemporally through investment, the temporary increase in land transport cost has a long-lasting effect on transport modal choices (See an example of "roadblock effect" in Figure 2). I provide reduced-form evidence on the persistent effect of rebellion on steamship trade from 1867 to 1897 in AppendixD.

With the probability of sea trade, we can estimate the parameters of the model with observed modal choices. However, the data comes with some limitation. We only observe trade by one transport method (i.e. steamship) and we only observe it immediately after the rebellion. Therefore I write the probability of using steamship after the rebellion first conditional on the post-rebellion probability of having sea trade:

$$\begin{aligned} Pr(S_{ij,post} = 1) &= Pr(C_{ij,post} = 1) Pr(S_{ij,post} = 1 | C_{ij,post} = 1) \\ &= Pr(C_{ij,post} = 1)\Big(1 - \big(1 - \Phi(\frac{\gamma}{\sigma_j} \ln \frac{D^{jk}_{ij,post}}{D^{st}_{ij,post}})^T\Big), \end{aligned} \tag{8}$$

which depends on the relative cost of sail to steam ship and the number of times (T) shocks to sailing ships are realised. To proceed, I write the probability of sea trade after the rebellion conditional on the probability of sea trade during the rebellion:

$$Pr(C_{ij,post} = 1) = Pr(C_{ij,reb} = 1)\Phi(\frac{\gamma}{\sigma} \ln \frac{D^l_{ij,post}}{D^s_{ij,post}}) + \big(1 - Pr(C_{ij,reb} = 1)\big)\Phi(\frac{\gamma}{\sigma} \ln \frac{D^l_{ij,post}}{D^s_{ij,post}} - i^s). \tag{9}$$

We can then write the post-rebellion probability of steamship trade, which is observed in data, by

[24] Here I assume that after the rebellion, the land route reverted to the one used before the rebellion. I also consider the possibility that the rebellion damaged road therefore we need to incur investment cost for a new land route for pairs whose old route passes through the rebellion controlled area. The results are very similar.

Table 2: Estimation of roadblock effect: trade cost, sunk cost and model prediction

Panel A: Estimation results on transport costs

Unit cost relative to flat land		Transition cost equivalence to km of flat land	
Flat land (without courier route)	1.00		
Courier route	0.24	Non-courier to courier land route	1.1 km
Sailing ship	0.13	Land route to sailing ship	11.0 km
Steamship	0.07	Land route to steamship	29.3 km
One std. increase in ruggedness	1.20		

Panel B: Estimation results on trade model parameters

	Parameter	Value
Elasticity of probability of sea trade w.r.t. land/sea cost ratio	γ/σ	0.979
		(0.330)
Sunk cost of new sea route	I^s	7.297
(cost ratio of new/established sea route)		(2.293)
Sunk investment for new land routes	I^l	2.773
(cost ratio of new/established land route)		(2.224)
Observations		82

Panel C: Mean probability of steamship trade after the rebellion

	Observations	Data	Model prediction	
			With roadblock	W/o roadblock
	(1)	(2)	(3)	(4)
Roadblocked pairs	52	0.712	0.720	0.435
			(0.021)	(0.021)
Unaffected pairs	30	0.367	0.420	0.420
			(0.034)	(0.034)

Notes: In *Panel C*, the model prediction with roadblock in column (3) provides the mean estimated probability for city prefecture pairs for which the least costly land route was roadblocked by the rebellion, taking into account whether the required detour was costly enough to incur the sunk cost of starting to trade through steamships between those locations, and whether the cost saving once that sunk cost was incurred was large enough to keep trading by steamship after the rebellion. Column (2) provides the actual mean probability observed in the data. Column (4) estimates the same mean probability under a counterfactual where the route had not been roadblocked by the rebellion. Standard errors in parentheses in *Panel B*. Bootstrap standard errors using 100 replicates in parentheses in *Panel C*.

combining Equation 7, 8 and 9:

$$
\begin{aligned}
Pr(S_{ij,post}=1) =& \Big(1-\big(1-\Phi(\frac{\gamma}{\sigma_j}\ln\frac{D^{jk}_{ij,post}}{D^{st}_{ij,post}})^T\Big)\times \\
&\Big(\Phi(\frac{\gamma}{\sigma}\ln\frac{D^l_{ij,reb}}{D^s_{ij,reb}}+\mathbb{1}(D^l_{ij,reb}\neq D^l_{ij,pre})i^l-i^s)\Phi(\frac{\gamma}{\sigma}\ln\frac{D^l_{ij,post}}{D^s_{ij,post}}) \\
&+\Big(1-\Phi(\frac{\gamma}{\sigma}\ln\frac{D^l_{ij,reb}}{D^s_{ij,reb}}+\mathbb{1}(D^l_{ij,reb}\neq D^l_{ij,pre})i^l-i^s)\Big)\Phi(\frac{\gamma}{\sigma}\ln\frac{D^l_{ij,post}}{D^s_{ij,post}}-i^s)\Big).
\end{aligned}
\tag{10}
$$

It is a function of (1) land transport cost before and during the rebellion, the relative cost of land transport compared to sea during and after the rebellion, the relative cost of steamship compared to sailing ship after the rebellion and the number of times shocks to sailing ships are realised within one year; (2) the parameters of the model (γ/σ, I^s and I^l). We have already obtained all elements of (1), estimated in Section 5. Therefore we can estimate the parameters of the model by maximising the likelihood of observing steamship trade in data.

The results are presented in *Panel B* of Table 2. [25] The elasticity of using sea transport with respect to land to sea transport cost ratio is positive as expected, indicating that the probability of sea trade increases when the cost of using alternative mode of land transport in comparison is high. There is a substantial cost associated with initialising sea trade. The cost of using a new sea route is more than 7 times the cost of an existing one. Compared to sea transport, the investment cost for land transport is modest. The cost ratio of using a new land route over an existing route is less than 3.

We can use the parameter estimates in *Panel B* to predict the mean probability of steamship after the rebellion, which is shown in *Panel C*. The observations are divided into two groups: locality pairs that experienced roadblock during the rebellion and ones whose usual land routes remained intact. In the data, shown in column (2), the mean probability of steamship trade immediately after the rebellion in 1867 is 0.712 for roadblocked pairs and 0.367 for unaffected pairs. Column (3) shows in comparison the model predicted mean probability of steamship. The prediction on the mean probability of using steamship for roadblocked pairs is 0.72 and is very close what we observe in the data. The prediction on the unaffected pairs is slightly higher, but overall the model matches well with observed trade. In column (4), I calculate the probability of steamship trade under a counterfactual where roadblock is eliminated during the rebellion. As unaffected pairs did not have roadblock to start with, their counterfactual probability is exactly the same as in column (3). For the roadblocked pairs, their counterfactual probability of steamship would be only 0.435 without roadblock, which is very similar to that of the unaffected pairs. The counterfactual estimation result suggests that if it were not for the rebellion, the two groups would act very similarly. With roadblock, however, the probability of steamship take-up increased by 66%.

7. Conclusion

Concentration of population in a location can remain very persistent even when it is no longer superior. Trade costs play a key role here because they affect market access, consequently affecting the distribution of economic activities and population across space. Trade costs can change over time because of new transport infrastructures as well as new technologies. Like other newborn technologies, new transport technology typically required large sunk investment cost upfront, which could set it back. Failure to adopt the superior technology, however, could trap population in sub-optimal locations.

This paper uses the Taiping rebellion in 19th century China to study population changes after a large but temporary shock to city sizes. In contrast to a large literature that finds persistence

[25] Estimation excludes trade from/to Shanghai. For more than one year, because it was surrounded by occupied prefectures, the land transport to Shanghai was impossible ($D^l_{ij,reb}/D_{ij,reb}$ would be positive infinite). Shanghai was also the only port that had steamship trade with all other ports in 1867. To show that the roadblock effect is not entirely driven by this exceptional example, it is dropped from the estimation.

of relative city sizes after temporary shocks, I find a permanent loss in the level of population of about 40% in war-ravaged cities and an increase of 30% in coastal cities after the rebellion. I provide evidence that this is because in addition to the direct impact of war through death and displacement, the rebellion affected population indirectly through trade routes and transport modes. Before the rebellion, as the adoption of sea trade was lagging far behind its potential, population continued to decline in coast region. This was because the scope of investment for establishing sea routes and building a system of navigation aid was beyond the capacity of any individuals. Without coordinated efforts or state intervention, wide participation in sea trade was prohibitive. The rebellion blocked regular land trade routes, forcing experiment with steamships, which triggered substantial investment to facilitate maritime trade and shifted permanently many trade routes to sea. As trade activities relocated, this catalysed a shift of population to port cities.

Faced with limitation of data in 19th century China (i.e. only observing trade by steamship after the rebellion), I develop a simple trade model with transport modal and route choices to identify the heterogeneous effect of the rebellion on regional trade through facilitating investment for sea trade depending on whether regular land routes became blocked. The estimation of the model requires an additional input: transport cost in history, which I obtain by developing a novel method based on the postal map. The model delivers via the investment channel an intertemporal link between temporary increase in land trade costs and permanent shift in transport modes, consistent with observed trade patterns. The effect was quantitatively large, facilitating transport technology to upgrade, trade patterns to adjust, and population trapped in sub-optimal locations to relocate.

References

Amiti, Mary, "Location of vertically linked industries: agglomeration versus comparative advantage," *European Economic Review*, 2005, *49* (4), 809–832.

Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang, "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 2017, *99* (3), 435–448.

Bleakley, Hoyt and Jeffrey Lin, "Portage and path dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Brezis, Elise S and Paul R Krugman, "Technology and the life cycle of cities," *Journal of Economic Growth*, 1997, *2* (4), 369–383.

Cao, Shuji, "Zhongguo yi min shi [The migration history of China]," *Fujian Education Press*, 1997.

_ , "Zhongguo ren kou shi [The demographic history of China]," *Shanghai: Fudan University Press. Chaliand, G.(2005). Nomadic Empires. New Brunswick: Transaction Books. Translated by AM Berrett*, 2000.

_ **and Yushang Li**, "The effect of taiping rebellion on population in Zhejiang," *Fudan Journal (Social Sciences Edition)*, 2000, *5*, 005.

Cheng, Yukuei et al., "Foreign trade and industrial development of China," 1956.

Davis, Donald R and David E Weinstein, "Bones, bombs, and break points: the geography of economic activity," *American Economic Review*, 2002, *92* (5), 1269–1289.

Dincecco, Mark and Massimiliano Gaetano Onorato, "Military conflict and the rise of urban Europe," *Journal of Economic Growth*, 2016, *21* (3), 259–282.

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

Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Forslid, Rikard and Ian Wooton, "Comparative advantage and the location of production," *Review of International Economics*, 2003, *11* (4), 588–603.

Ge, J, Y Hou, and G Zhang, "Population and the modernization of China: since 1850," *Xuelin Press*, 1999.

Gilfillan, S Colum, *Inventing the ship: a study of the inventions made in her history between floating log and rotorship; a self-contained but companion volume to the author's" Sociology of invention"; with 80 illustrations, bibliographies, notes and index*, Follett, 1935.

Glaeser, Edward L and Jesse M Shapiro, "Cities and warfare: The impact of terrorism on urban form," *Journal of Urban Economics*, 2002, *51* (2), 205–224.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm, "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *The quarterly journal of economics*, 2008, *123* (2), 441–487.

Hua, Qiang, "The Geography of Taiping Rebellion," 1991.

Jia, Ruixue, "The legacies of forced freedom: China's treaty ports," *Review of Economics and Statistics*, 2014, *96* (4), 596–608.

Keller, Wolfgang, Ben Li, and Carol H Shiue, "The evolution of domestic trade flows when foreign trade is liberalized: Evidence from the Chinese Maritime Customs Service," in "Institutions and Comparative Economic Development," Springer, 2012, pp. 152–172.

Krugman, Paul, "First nature, second nature, and metropolitan location," *Journal of regional science*, 1993, *33* (2), 129–144.

Michaels, Guy and Ferdinand Rauch, "Resetting the urban network: 117–2012," *The Economic Journal*, 2017, *128* (608), 378–412.

Nunn, Nathan and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Pascali, Luigi, "The wind of change: Maritime technology, trade, and economic development," *American Economic Review*, 2017, *107* (9), 2821–54.

Pflüger, Michael and Takatoshi Tabuchi, "Comparative advantage, agglomeration economies and trade costs," *Journal of Urban Economics*, 2019, *109*, 1–13.

Redding, Stephen and Anthony J. Venables, "Economic geography and international inequality," *Journal of International Economics*, 2004, *62* (1), 53–82.

Tsivanidis, Nick, "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Unpublished manuscript*, 2018.

Venables, Anthony J, "The international division of industries: Clustering and comparative advantage in a multi-industry model," *Scandinavian Journal of Economics*, 1999, *101* (4), 495–513.

Voigtländer, Nico and Hans-Joachim Voth, "The three horsemen of riches: Plague, war, and urbanization in early modern Europe," *Review of Economic Studies*, 2012, *80* (2), 774–811.

Ward, WH, "The sailing ship effect," *Physics Bulletin*, 1967, *18* (6), 169.

Wei, Yuan, "Emperor Daoguang Bingyu Shipping Notes," 1826.

AppendixA. Table Appendix

Table A.1: Mapping between postal names and current locations

Postal name (Romanization)	Current location (Pinyin)	Postal name (Romanization)	Current location (Pinyin)
Amoy	Xiamen	Kiukiang	Jiujiang
Canton	Guangzhou	Kiungchow	Haikou
Chefoo	Yantai	Kweiping	Guigang
Chengting	Shijiazhuang	Newchwang	Yingkou
Chinan	Jinan	Ningpo	Ningbo
Chinchow	Jinzhou	Shanghai	Shanghai
Chinkiang	Zhenjiang	Tientsin	Tianjin
Foochow	Fuzhou	Swatow	Shantou
Hankow	Hankou	Wenchow	Wenzhou
Ichang	Ichang	Wuhu	Wuhu
Kiaochow	Qingdao		

AppendixB. Figure Appendix

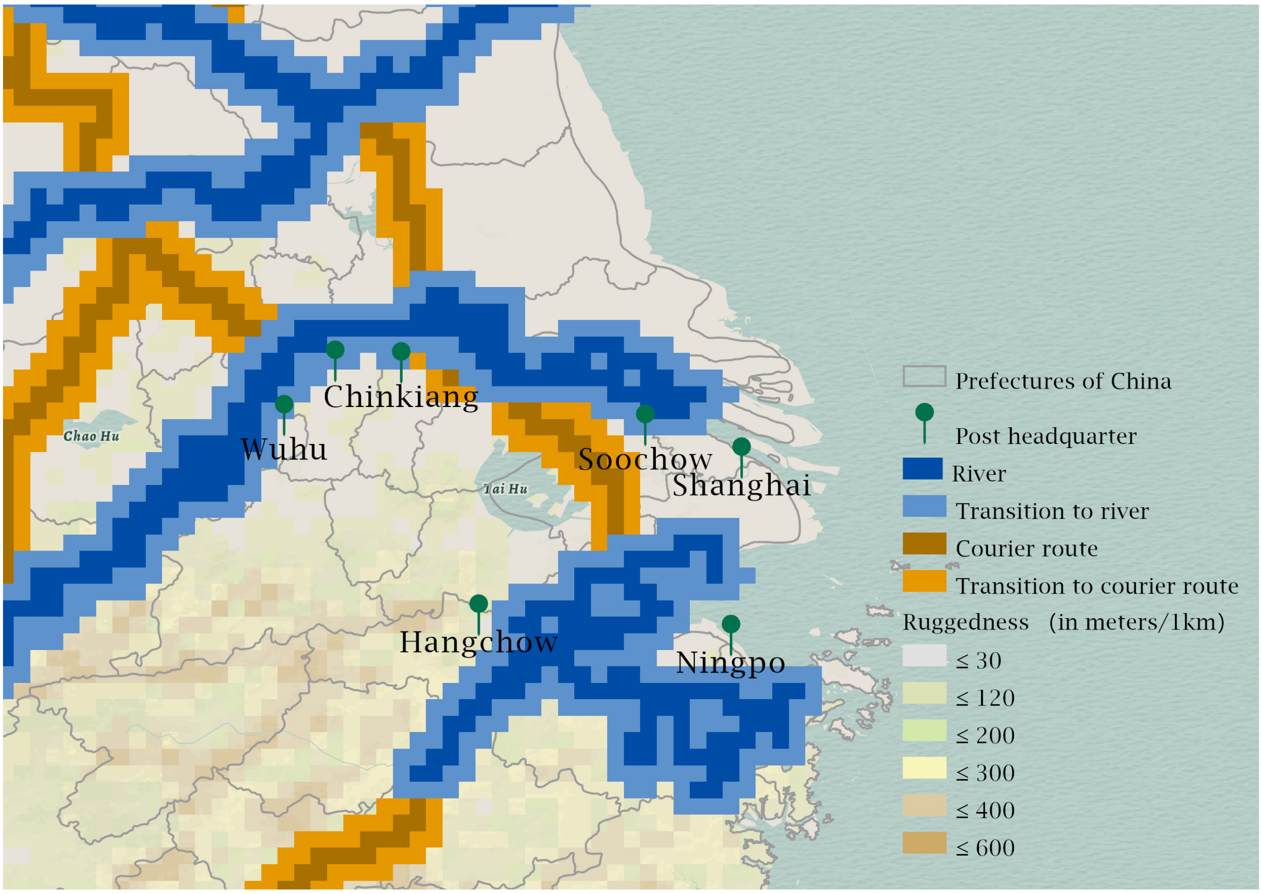
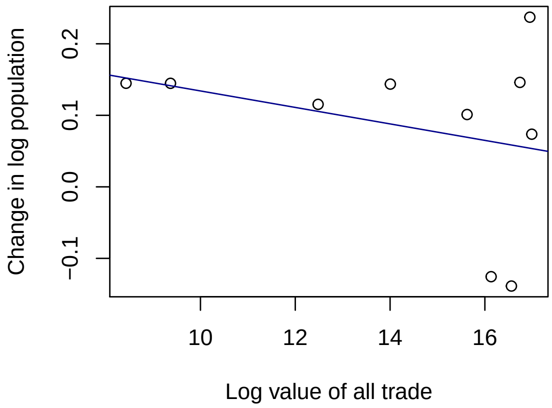
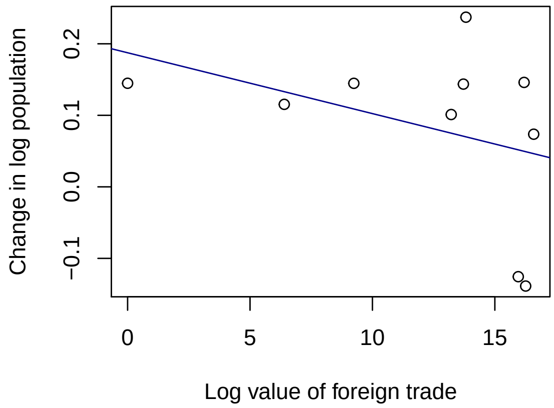


Figure B.1: Travel cost in 19th century China

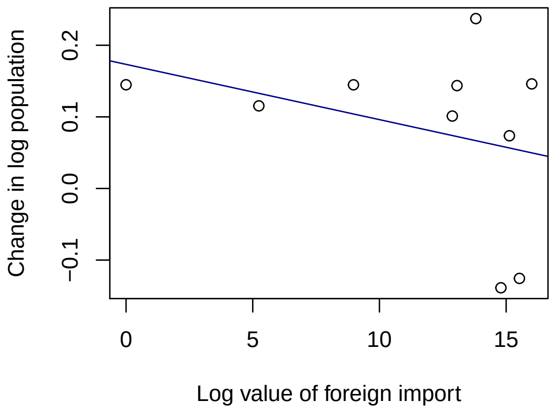
Notes: This figure shows the underlying raster (0.1°by 0.1°) used to estimate travel costs. River cells are created using CHGIS V4 dataset of coded river 1820. The Qing courier routes (1800-1900) are taken from China W dataset. Terrain ruggedness is from Nunn and Puga (2012).



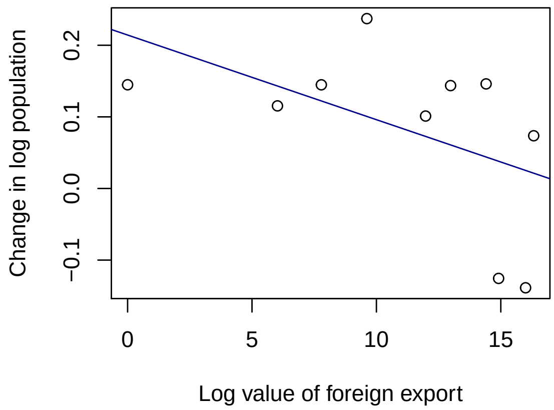
(a) All trade



(b) Foreign trade



(c) Foreign import



(d) Foreign export

Figure A.2: Foreign trade and population change

Notes: The figure plots changes in log population between 1851 and 1880 in treaty ports against (a) log value of total trade (b) log value of foreign trade (c) log value of foreign import (d) log value of foreign export. Treaty ports under the direct control of the rebels are excluded.

AppendixC. Model Appendix

Following the set-up of the model with Equation 2, 3 and 5, after each realisation of shock to sailing ship, firms that have chosen maritime trade choose steamship if:

$$\gamma \ln D^{jk}_{ijt} - \eta_j > \gamma \ln D^{st}_{ijt}$$

The shocks to sailing ships are realised for T times within a year. The discrete number of shocks are modelled to rationalise the observed zero trade by steamship and junk sailing ship. Within a year, junk trade does not happen if under the best scenario for junks, steamships are preferred. This is equivalent to max $\eta_{j,ijt}$ of T realisations is smaller than $\gamma \ln(D^{jk}_{ijt} / \ln D^{st}_{ijt})$. The normalised maximum of shocks to sailing ships $\max(\eta_{j,ijt}/\sigma_j)$ follows the distribution Φ^T, therefore we can write the probability of observing no junk trade in year t given sea trade is chosen:

$$Pr(J_{ijt} = 0|C_{ijt} = 1) = \Phi(\frac{\gamma}{\sigma_j} \ln \frac{D^{jk}_{ijt}}{D^{st}_{ijt}})^T,$$

where J_{ijt} and C_{ijt} are indicators for junk trade and sea trade respectively. The probability of observing junk trade in year t is therefore:

$$Pr(J_{ijt} = 1|C_{ijt} = 1) = 1 - \Phi(\frac{\gamma}{\sigma_j} \ln \frac{D^{jk}_{ijt}}{D^{st}_{ijt}})^T.$$

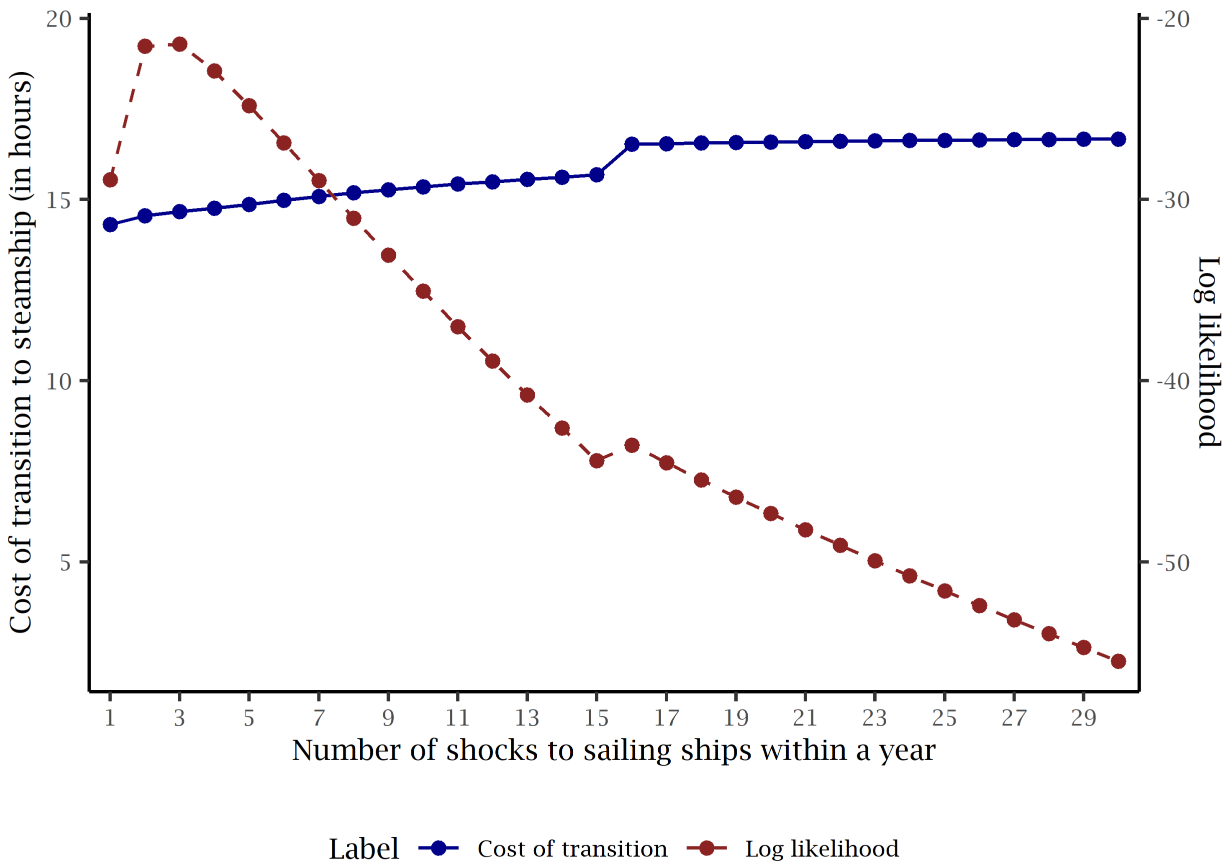
Steamship trade in any given year is observed if at least under the worst scenario for junks, steamship is preferred. This is equivalent to min $\eta_{j,ijt} < \gamma \ln(D^{jk}_{ijt} / \ln D^{st}_{ijt})$ after T realisations of $\eta_{j,ijt}$. We can write the conditional probability that steamship trade happens during year t, indicated by S_{ijt}:

$$Pr(S_{ijt} = 1|C_{ijt} = 1) = 1 - \big(1 - \Phi(\frac{\gamma}{\sigma_j} \ln \frac{D^{jk}_{ijt}}{D^{st}_{ijt}})^T.$$

The transport costs for sail and steam ship consist of two parts: the cost by distance and two-way fixed cost of transition between land and steamship. Using postal map, we get the distance cost and transition cost for sailing ships. The distance cost of steamship is set to be half the cost of sailing ship according to Pascali (2017). I estimate the fixed cost associated with transition between land to steamship, for different numbers of realisations of shocks to sailing ships within one year, by maximising the probability of observing the usage of steamship and sailing ship in two ports for their trade with other locations with varying distance.

Figure C.1 plots the estimated cost of transition between land and steamship with the corresponding value of log likelihood for different numbers of realisations of shocks to sailing ships. Shocks being realised every four month ($T = 3$) matches the data the best, where the maximised log likelihood function reaches the highest value. Table C.1 shows the maximum likelihood estimation results on cost of transition when we assume the shocks to sailing ships are realised every four months. With the speed of travelling by land without courier route normalised to 4 kilometres per hour, the time spent on transition between land and steamship is about 14.7 hours, which is about 2.67 times the time spent between land and sailing ship.

Figure C.1: Log likelihood estimation of transition costs



Notes: The estimated cost of transition (in hours) is based on normalisation of land transport speed of 4 kilometres/hour. The cost of transition shown here include both the cost of switching from land to steamship and from steamship to land.

Table C.1: Estimation of transport cost by steamship

	Parameter	Value
Elasticity of probability of steamship ship w.r.t. sail/steam cost ratio	γ/σ_j	3.444
		(0.626)
Cost of transition between land and steamship		14.664
		(1.268)
Observations		27

Notes: Shocks to sailing ship assumed to realise every four months. Land transport speed normalised to 4 kilometres/hour. The estimated cost of transition is time spent (in hours) on transition from land to steamship and steamship to land combined.

AppendixD. Reduced form evidence on roadblock

This section provides reduce-form evidence on the persistent effect of roadblock on steamship trade. I estimate the following:

$$S_{ijt} = \mathbb{1}\big(\beta \ln \frac{D^l_{ijt}}{D^s_{ijt}} + \delta roadblock_{ij} + \iota_{ijt} \geq 0\big),$$

where S_{ijt} is indicator for steamship trade between port i and j in year t. Assume ι_{ijt} follows standard normal distribution, we have:

$$Pr(S_{ijt} = 1) = \Phi\big(\beta \ln \frac{D^l_{ijt}}{D^s_{ijt}} + \delta roadblock_{ij}\big).$$

This equation looks at whether locality pairs that had roadblock during the rebellion ($roadblock_{ij} = 1$) have systematically higher probability of using steamship in post-rebellion periods controlling for the transport cost of land relative to sea. Table D.1 shows the results from 1867 to 1897. Because steamship trade was only allowed in treaty ports, to compare steamship take-up in 1867, I only use locality pairs that were both active treaty ports in 1867. To keep the sample consistent over time, I also restrict the observations to the active ports in 1867 to look at trade in later years.

Table D.1: Roadblock effect on steamship trade : baseline sample

	Dep. var.: Probability of steamship trade						
	1867	1873	1877	1882	1887	1892	1897
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log ($Dist_{land}/Dist_{sea}$)	1.032**	0.406	0.831**	0.405	0.168	0.274	0.383
	(0.435)	(0.362)	(0.355)	(0.364)	(0.362)	(0.379)	(0.417)
Roadblock	1.168***	0.889***	0.679**	0.818***	0.644**	0.827**	0.828**
	(0.307)	(0.285)	(0.286)	(0.288)	(0.298)	(0.325)	(0.353)
Constant	-1.700***	-0.696	-1.075**	-0.523	0.125	0.173	0.244
	(0.648)	(0.532)	(0.513)	(0.530)	(0.527)	(0.542)	(0.595)
Observations	98	98	98	98	98	98	98

Notes: Estimation is based on treaty ports that were functioning in 1867 (except trade between river ports), so that the sample is consistent over time. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

We can see that roadblocked pairs have a much higher probability of using steamship trade in 1867. At the means, the probability of using steamship is 0.79 for roadblocked pairs, compared to 0.36 for unaffected pairs. Although all pairs have a higher tendency to use steamship (captured by the constant term) over time, the difference between roadblocked pairs and unaffected pairs is remarkably persistent even after the three decades in 1897, which is the last year we observe the data.

An explicit aim of the rebels was to cut off supplies to the imperial capital. The flows of goods from rebellion occupied area to the imperial capital through regular land routes were curbed, which forced Qing government to find ways to transport tribute grain by sea. I argue that this also affected

Table D.2: Roadblock effect on steamship trade : exclude trade from/to rebellion region to Tienstin

	Dep. var.: Probability of steamship trade						
	1867	1873	1877	1882	1887	1892	1897
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log ($Dist_{land}/Dist_{sea}$)	0.986**	0.448	0.891**	0.428	0.280	0.382	0.481
	(0.446)	(0.372)	(0.359)	(0.372)	(0.365)	(0.380)	(0.418)
Roadblock	1.179***	0.916***	0.656**	0.838***	0.678**	0.864***	0.861**
	(0.319)	(0.296)	(0.295)	(0.299)	(0.306)	(0.332)	(0.361)
Constant	-1.594**	-0.776	-1.183**	-0.566	-0.0966	-0.0361	0.0551
	(0.674)	(0.559)	(0.529)	(0.554)	(0.547)	(0.561)	(0.613)
Observations	92	92	92	92	92	92	92

Notes: Estimation is based on treaty ports that were functioning in 1867 (except trade between river ports), but excludes bilateral trade of ports that were near occupied prefectures by rebellion (Shanghai, Hankow, Chinkiang, Kiukiang) with Tienstin. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

indirectly and differently modal choices of bilateral trade flows between unoccupied locations. I investigate this possibility by excluding trade between ports that were near occupied prefectures and Tienstin in Table D.2. The results are very similar to Table D.1, suggesting that modal change was not unique to transport of tribute grain and roadblock affected much broader exchange of goods across the whole country.

The rebellion could affect modal choice through long-term investment to facilitate sea trade triggered by roadblock, but the warfare could also increase sea trade in the long run by destroying road infrastructure. To separate the effect of the two, in Table D.3, I calculate the post-rebellion the least cost land route assuming that road infrastructure in rebellion occupied region was not functional after the rebellion. The difference between the roadblocked pairs and unaffected pairs remains as large after controlling for the possibility that the rebellion damaged some roads permanently and more so for roads used previously by roadblocked pairs.

Previous tables look at trade between treaty ports that were active in 1867. In Table D.4, however, I include trade between treaty ports as long as they were functioning ports in the corresponding year. The sample size increases over time as there were more treaty ports opened. Also we see that the sharp distinction between roadblocked location pairs and unaffected pairs was only present for the initial pairs that were opened as treaty ports before 1867. For ports that were opened after 1873, even if they had roadblock, since steamship trade was not allowed at the time, the rebellion did not lead to adoption of steamship in the long run. Although they may have turned to sailing ship when regular land routes became blocked, the advantage of sailing ship over land transport could be not large enough to shift trade routes permanently to sea. The results in Table D.4 suggest that both the advance in technology (i.e. introduction of steamship) and the impetus to change necessitated by roadblock are essential for China's final transition to sea trade in late 19th century.

Table D.3: Roadblock effect on steamship trade : broken routes

	Dep. var.: Probability of steamship trade						
	1867	1873	1877	1882	1887	1892	1897
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log ($\hat{Dist}_{land} / Dist_{sea}$)	0.987**	0.292	0.778**	0.371	0.157	0.372	0.410
	(0.437)	(0.371)	(0.364)	(0.372)	(0.371)	(0.383)	(0.421)
Roadblock	1.162***	0.889***	0.682**	0.819***	0.645**	0.832**	0.835**
	(0.307)	(0.285)	(0.286)	(0.288)	(0.298)	(0.326)	(0.354)
Constant	-1.681**	-0.562	-1.046*	-0.498	0.131	0.0288	0.187
	(0.674)	(0.557)	(0.542)	(0.554)	(0.553)	(0.560)	(0.614)
Observations	98	98	98	98	98	98	98

Notes: Estimation is based on treaty ports that were functioning in 1867 (except trade between river ports), so that the sample is consistent over time. The relative transport cost by land ($\hat{Dist}_{land} / Dist_{sea}$) takes into account that road infrastructure (i.e. courier routes) within rebellion occupied region was unusable afterwards. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table D.4: Roadblock effect on steamship trade : all pairs

	Dep. var.: Probability of steamship trade						
	1867	1873	1877	1882	1887	1892	1897
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Dist_{land} / Dist_{sea}$	1.032**	0.406	0.623**	0.376	0.296	0.634***	0.705***
	(0.435)	(0.362)	(0.254)	(0.257)	(0.255)	(0.246)	(0.207)
Roadblock	1.168***	0.889***	0.360*	0.330*	0.188	0.361*	0.0273
	(0.307)	(0.285)	(0.197)	(0.197)	(0.194)	(0.189)	(0.159)
Constant	-1.700***	-0.696	-1.239***	-0.914**	-0.607*	-0.996***	-1.104***
	(0.648)	(0.532)	(0.363)	(0.367)	(0.363)	(0.344)	(0.282)
Observations	98	98	188	188	188	202	302

Notes: Sample includes all pairs of treaty ports that were active in the corresponding year (except trade between river ports). Robust standard errors in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

AppendixE. Estimation of Trade and Migration Gravity

In this section, I estimate the distance elasticities of trade and migration in historical China using transport cost estimated in Section 5.

A. Trade gravity

Take log on both sides of Equation 2, we derive the gravity equation of trade:

$$\ln X_{ijt}^{st} = \psi + \zeta_{jt} + \xi_{it} - \gamma ln D_{ijt}^{st} + \nu_{s,ijt},$$

where $\psi = -(1-\epsilon)\ln\alpha$, $\zeta_{jt} = (1-\epsilon)\ln c_{jt} + \ln N_{jt}$, $\xi_{it} = \ln Y_{it} - (1-\epsilon)\ln P_{it}$. However, different from the standard setting where we observe bilateral trade by all transport methods, here we only observe trade by steamship. Regressing value of trade by steamship on transport cost is likely to generate a biased estimate of γ.

One source of the bias is that the shorter the distance, it is more likely that maritime trade is carried out by sailing ships. I correct the bias using the probability of using steamships conditional on sea transport ($Pr(S_{ijt} = 1|C_{ijt} = 1)$). Moreover, there could be a bias stemming from the correlation between the distance of sea transport and the size of market that uses steamship for sea trade. This is because steamships were only available at treaty ports, which were large ports allowed for foreign trade. For shorter distance trade, in which sailing ships had an advantage, inland locations could export their goods through the closest ports by sailing ship, which were not necessarily treaty ports. For longer distance trade however, it may worth the trouble transferring the merchandise to a treaty port to transport by steamship, whose average cost decreases with distance, even for inland locations that were not close by. Therefore, longer distance trade by steamship disproportionately attracts and therefore includes goods from a larger market.

To correct for this bias, I find for each pair of treaty ports, all the combinations of prefectures that would choose optimally to export and import from the treaty ports by steamship. Assume that nearby prefectures have similar wage levels, and that the income at the destination Y_{it} is proportional to its population L_{it}. Free entry condition pins down the equilibrium number of firms at the origin $N_{jt} = L_{jt}/\epsilon f$, which is also proportional to its population. I correct for the correlation between steamship trade distance and market size by summing over the product of population at the origin and destination: $\sum_{o,d\in\mathbb{S}_{i,j}} L_o \times L_d$, where $\mathbb{S}_{i,j}$ is the set of prefecture pairs o,d that optimally choose to transport goods by steamship from the treaty ports i,j. The adjusted value of trade between treaty ports by steamship is therefore:

$$\overline{X}_{ijt} = \frac{X_{ijt}}{Pr(S_{ijt} = 1|C_{ijt} = 1)\sum_{o,d\in\mathbb{S}_{i,j}} L_o \times L_d}.$$

Another source of biased comes from selection on pairs of trade that choose sea transport over land transport. Prefecture pairs that have a smaller land to sea transport cost ratio would use sea transport if they receive large positive shocks for sea transport (higher $\eta_{s,ijt}$). As the land to sea transport cost ratio can be correlated to transport costs by steamship, this results in a correlation between the error term ($\eta_{s,ijt}$) and the transport cost ($\ln D_{ijt}^{st}$) in Equation A.. This bias can be corrected using the Heckman selection framework. From the set-up of the model, we can write the first stage selection of sea transport as a function of the relative cost of land transport to sea and the sunk investment cost of sea trade:

$$C_{ijt} = \mathbb{1}\left(\frac{\gamma}{\sigma}\ln\frac{D^{l}_{ijt}}{D^{s}_{ijt}} - \mathbb{1}(C_{ij,reb}=0)i^{s} + v_{ij}\right), \qquad v_{ijt} \sim N(0,1),$$

where $v_{ijt} = (\eta_{s,ijt} - \eta_{l,ijt})/\sigma$. The expected value of trade by steamship controlling for selection bias is:

$$E[ln\overline{X}_{ijt}|C_{ijt}=1] = \psi + \zeta_{jt} + \xi_{it} - \gamma lnD^{st}_{ijt} + E[\eta_{s,ijt}|C_{ijt}=1].$$

Denote $\alpha_v = \gamma/\sigma\ln(D^{l}_{ijt}/D^{s}_{ijt}) - \mathbb{1}(C_{ijt0}=0)i^{s}$, we have:

$$\begin{aligned} E[\eta_{s,ijt}|C_{ijt}=1] &= E[\eta_{s,ijt}|v_{ijt} > -\alpha_v] \\ &= E[\sigma v_{ijt} + \eta_{l,ijt}|v_{ijt} > -\alpha_v] \\ &= \sigma E[v_{ijt}|v_{ijt} > -\alpha_v] \\ &= \sigma\lambda(\alpha_v), \end{aligned}$$

where $\lambda(\alpha_v) = \phi(\alpha_v)/\Phi(\alpha_v)$ is the inverse mills ratios. We can use the estimates from *Panel B* of Table 2 to predict $\lambda(\alpha_v)$.

Table E.1: Trade gravity

	Dep. var.: Log (value of trade by steamship)				
	(1)	(2)	(3)	(4)	(5)
Log ($Dist_{steam}$)	2.035*	0.316	-0.307	-2.027	-2.217**
	(1.134)	(1.214)	(1.141)	(1.215)	(1.084)
Inverse mills ratio					7.248***
					(1.590)
Adjust trade by sailing ship		Yes		Yes	Yes
Adjust market size			Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes
Observations	111	111	111	111	111
R-squared	0.715	0.669	0.722	0.717	0.776

Notes: Estimation based on bilateral trade by steamship between treaty ports in 1877 and 1882. Market size is calculated based on prefecture population in 1880. All specifications include time variant origin and destination fixed effects. Robust standard errors in parenthesis are clustered in origin-destination pair. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table E.1 presents estimation results on elasticity of steamship trade with respect to sea transport cost using a sample of bilateral trade by steamship in 1877 and 1882 with market size calculated using prefecture population in 1880. I regress the natural logarithm of value of trade by steamship on garithm of transport cost by steamship, controlling for time variant origin and destination fixed effects. In column (1), I present the unadjusted trade elasticity. The elasticity is positive, indicating that the value of steamship trade increases with distance. In column (2), I correct for the fraction of trade carried out by junk sailing ships. The elasticity becomes less positive. In column (3), controlling for the factor that long distance steamship trade includes exchange of goods between a larger market, the coefficient estimate on trade elasticity becomes negative. In column (4), I control both trade by sailing ship and market size, the elasticity becomes negative and similar to what has been found in the literature. When accounting for the selection bias of sea transport over land transport, in column (5), the elasticity does not change much but becomes statistically significant. The coefficient of inverse mills ratio has the expected sign and magnitude.

B. Migration gravity

In this part, I estimate the following migration gravity equation:

$$\ln L_{ijt} = \chi + \eta_{it} + \psi_{jt} - \kappa \ln D_{ijt} + u_{ijt}.$$

The dependent variable is the natural logarithm of migration flow from location i to j at time t. Historical migration flow data is rarely available. Here I construct a new data set of migration flows in historical China based on studies that traces the origin of contemporary villages. Historically, many villages in China were founded by clans from outside areas where living conditions in their place of origin worsened. Thanks to the tradition of keeping a rich family history, Cao (1997) was able to obtain the information about those migration villages regarding the the place of origin of their founders and formation time. We can use the number of villages at location i founded by individuals from location j at time t to proxy for the migration flows from i to j at t. The implicit assumption here is that the initial population when the villages were founded is similar. For a subset of villages where their recent population is available, I calculate the implied annual population growth rate to check if this is a reasonable assumption. I assume that all migration villages have initial population of 30 to start with and plot them against the actual annual population growth rate in the prefecture to which the villages belong (Figure F.1). It seems plausible to assume that their initial migration population is similar.

In Cao (1997), the destinations are counties. For the majority of villages, we only know the origin province and for some we know the prefectures or counties. In column (1) of Table E.2, I present estimation results on migration gravity using all available origin and destination pairs. The study by Cao (1997) focuses on migration to Kiangsi province and some nearby prefectures. To the extent that the prefecture destination outside of Kiangsi may have different migration patterns, in column (2), I restrict attention to only migration flows to Kiangxi province. In column (3), I restrict sample to only county to province pairs, where the location of the province is proxied by the provincial centroid. For all specifications I allow the origin and destination fixed effects to vary by time. Still the distance elasticity can vary over time [26]. Therefore I divide the sample into two periods in column (4) and column (5). The results show that over time migration flows responds more to transport cost. In Table E.3, I aggregate the destinations from counties to prefectures, which leaves us with fewer observations, but the estimated distance elasticities are similar. They are comparable to those estimated with modern data and for other parts of world.

[26] The migration time frame is between 1643-1722, 1643-1796, 1723-1796 or 1796-1912.

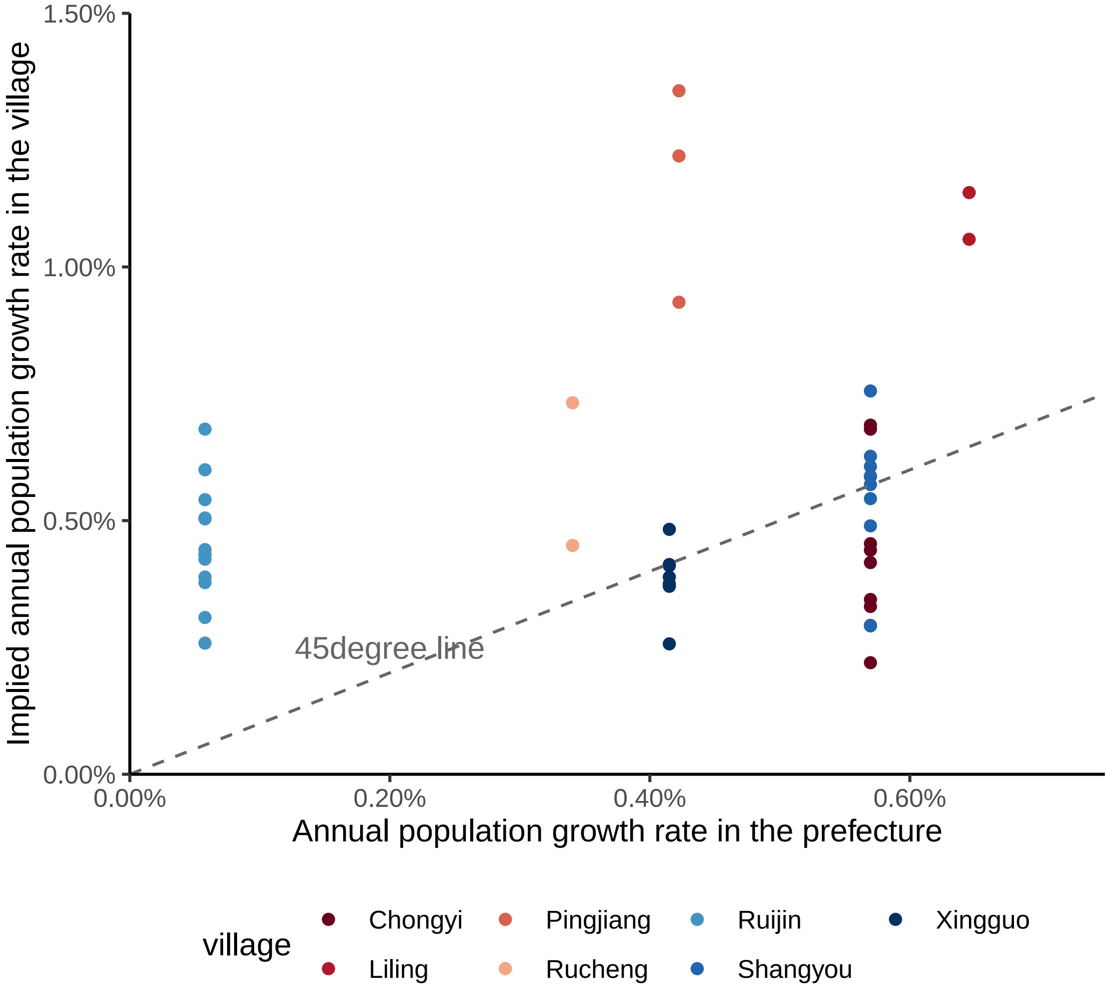


Figure F.1: Actual population growth and implied population growth of migration villages

Notes: Each colour denotes a distinct village. Based on recent population of villages and their foundation time from Cao (1997) and historical prefecture population data from Cao (2000).

Table E.2: Migration gravity: county destination

	Dep. var.: Log (number of villages)				
	All	Kiangsi only	Province only	1643-1722	1723-1796
	(1)	(2)	(3)	(4)	(5)
Log (*Dist*)	-1.485***	-1.271***	-1.796***	-0.990**	-1.552***
	(0.353)	(0.306)	(0.548)	(0.445)	(0.396)
Origin FE	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes	Yes	Yes
Observations	113	93	62	45	48
R-squared	0.894	0.837	0.942	0.823	0.854

Notes: Column (1) to (3) includes time variant origin and destination fixed effects. Robust standard errors in parenthesis. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table E.3: Migration gravity: prefecture destination

	Dep. var.: Log (number of villages)		
	All	Kiangsi only	Province only
	(1)	(2)	(3)
Log (*Dist*)	-1.598*	-1.292*	-2.128**
	(0.781)	(0.590)	(0.776)
Origin FE	Yes	Yes	Yes
Dest FE	Yes	Yes	Yes
Observations	42	27	32
R-squared	0.935	0.923	0.956

Notes: All specifications include time variant origin and destination fixed effects. Robust standard errors in parenthesis. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.