Contagion, Concentration, and Commercialization: Health-Induced Markets and Land Distribution in Colonial Taiwan*

Jiadian LI[a,b,†]

[a] *School of Education and Social Policy, Northwestern University*

[b] *Department of Economics, The Chinese University of Hong Kong*

This Version: August 17, 2026

Abstract

This study investigates one origin of land inequality, focusing on the role of market access. Following Acemoglu, Johnson and Robinson (2001), our identification strategy is based on the quasi-random/exogenous variation in the prevalence of infectious diseases across villages. The extent to which a village relied on public health policies determined whether Japanese colonial authorities established public markets there. To address endogeneity, we draw on the epidemiological environment and focus on its supply side (pesthouses), combining it with evidence from its demand side (infectious mortality). We further provide complementary bioclimatic evidence and use historical spring temperatures as a source of variation in the early spread of plague. By constructing a novel cross-sectional dataset of Taiwan's villages and employing an instrumental variable approach, we provide causal evidence that market access increased land inequality. Altogether, our results highlight the role of market forces in shaping land inequality in colonial Taiwan and provide a new perspective on the economic consequences of Taiwan's postwar land reform.

Keywords: Infectious Disease, Market Access, Land Inequality, Landownership, Colonial Taiwan

JEL Codes: N35, N55, O17, Q15

*The authors would like to thank Ying Bai, James Kai-sing Kung, Yifan Zhang, and Xianqiang Zou for their constructive suggestions and useful information; conference/workshop/seminar participants from CUHK for helpful comments. No relevant financial relationships. All remaining errors are our own.

†E-mail: eric.jiadianli@gmail.com

1 Introduction

Existing literature highlights that the distribution of productive factors such as land is one crucial determinant of long-run development (Banerjee and Iyer, 2005; Cinnirella and Hornung, 2016; Galor, Moav and Vollrath, 2009). The general view that inequality hinders economic development results in the widespread adoption of redistribution policies (Alesina and Rodrik, 1994; Persson and Tabellini, 1994). However, politically driven and dogmatic redistribution policies fail to consider the fundamental causes of inequality and may therefore lead to unintended consequences. For example, Italy's land reform (*Legge Stralcio*) was primarily intended to reduce rural unrest. This helped the Christian Democratic Party, which promoted the reform, gain persistent electoral benefits (Caprettini, Casaburi and Venturini, 2021), but it also led to slow industrialization and long-term underdevelopment (Albertus, 2023; Bianchi-Vimercati, Lecce and Magnaricotte, 2026). China's redistributive policies (e.g. land reform and collectivization) were explicitly framed as class-based policies (Kung, Wu and Wu, 2012). This helped the Chinese Communist Party gain political support and loyal agents (Hao et al., 2026), but it also led to the killing of draft animals and the subsequent agricultural crisis (Chen and Lan, 2017; Lin, 1990).

The academic consensus links land reform with Taiwan's rapid economic development after World War II (Lipton, 2009; Rodrik, Grossman and Norman, 1995). Recent studies, however, suggest that the land reform in post-war Taiwan contributed very little to agricultural productivity and economic growth (Wang and Kim, 2024). In other settings, land reform may also have generated misallocation (Adamopoulos and Restuccia, 2020; Ghatak and Mookherjee, 2025). In this paper, we discuss these findings by starting from the origins of land inequality. We argue that the unequal structure of land distribution in post-war Taiwan may have been shaped by market factors. The rise of inequality most likely came from agricultural households' spontaneous land allocation and commercial transformation. Therefore, although Taiwan's land reform has also been shown to be effective in garnering electoral support, its real economic impact may have been overestimated (Chiang, Fan and Hsu, 2025).

We provide evidence showing how a public health institution established by the Japanese colonizers led to consistent changes in the relationship between market access and land inequality. Unlike most colonial settlements, whose locations reflected colonizers' choices (Acemoglu, Johnson and Robinson, 2001), Japan's colonization of Taiwan was determined by the "permanent cession" stipulated in a diplomatic treaty. Japan had no reason to leave or relinquish an island that occupied a strategic position at a maritime transportation chokepoint and on the frontier of military defense, while also generating profitable agricultural surpluses (Dell and Olken, 2020). Taiwan was an indispensable component of *the Greater East Asia Co-Prosperity Sphere*, both as an enduring ideology and policy project. Yet this fascist madness forced the Japanese colonial authorities to confront and finally address a series of challenges, including excess mortality from infectious diseases during early ruling periods.[1]

Specifically, beginning in 1895, the Japanese colonizers established the public market system in areas

[1] As early as the Japanese invasion of Taiwan, the mortality and morbidity rates exceeded half of the total military population. The year after Japan took over Taiwan, the Japanese colonial authorities promulgated regulations for the prevention and control of infectious diseases, and unofficially listed bandits and the plague, the leading infectious disease, as the two core threats.

of Taiwan where infectious diseases were prevalent.[2] Public markets accommodated sellers who had previously gathered on populous and crowded open-air streets, providing them with formal trading venues with circulating air. This cut off the major channels through which epidemic infectious diseases spread via airborne transmission, saliva, contaminated water sources, and disease vectors. As we observe, by 1918, plague had almost disappeared (although other types of infectious diseases still occurred after 1918, they were well controlled within months because the health system had already been established). By the time Japanese colonial rule ended in 1945, approximately 57% of villages received these public markets. Meanwhile, such health-induced markets objectively introduced variations in market access, which was largely driven by the nature of infectious diseases.

We manually collected and digitized a batch of historical archives and spatial information, constructing a novel village-level dataset. This covers a range of social and economic indicators. We find that spatial variation in infectious diseases appears to be quasi-random, as it is orthogonal to a wide set of predetermined and contemporaneous characteristics (including state capability, economic development, and human capital). This allows us to use the differential extent to which a village relied on public health policies to identify sources of exogenous variation in market access.

Colonial Taiwan thus provides a unique setting, which we can exploit to examine historians' narrative that transformative factors like market access matter in shaping Taiwan's land distribution. Our empirical analysis proceeds in several steps. We begin by documenting the relationship between market access and land inequality. The ordinary least squares (OLS) estimates show that villages closer to markets consistently report higher land Gini coefficients. This result remains robust after controlling for prefecture fixed effects, population, demographic variables, and initial inequality. A one-standard-deviation reduction in distance to markets increases land inequality by about 1.26% of its mean.

To address endogeneity concerns arising from measurement error and omitted variables in the OLS estimates, we apply an instrumental variables (IV) approach. We use pesthouses, which were built to quarantine and isolate patients, as the main instrument.[3] The contagious nature of infectious diseases and the medical principle of proximity allow pesthouses to capture infectious endowments more precisely. Exploiting the common spatial distribution of pesthouses and public markets, we conduct two-stage least squares (TSLS) estimation. The conclusion remains unchanged qualitatively, and the TSLS estimates are about three times as large as the OLS baseline estimates.

Three tests collectively validate the use of pesthouses as an IV. (1) We find that pesthouses show no correlation with many confounding factors. This leads us to regard their distribution as plausibly exogenous and balanced. (2) We add controls for that may jointly affect infectious diseases and agricultural land distribution in the TSLS estimation equation. These include public health indicators proxied by access to clean water and hospitals, geographical location proxied by latitude and longitude, historical agricultural shocks proxied by earthquake intensity, and agricultural endowments proxied by sugarcane suitability and rice suitability. This addresses several observable alternative channels. (3) We use the in-

[2]By contrast, public markets in Japan proper emerged only around 1917. Historians attribute this to the hyperinflation triggered by World War I, which caused sharp increases in the prices of rice and daily necessities. Therefore, the government established public markets in order to quell nationwide riots.

[3]In the term *pesthouse*, *pest* was another term used by the Japanese colonial authorities to refer to plague.

fectious mortality rate from 1917 to 1921 as a second instrument. Villages affected by infectious diseases in the early colonial period continued to report higher infectious mortality later on. When we employ the infectious mortality rate as the instrument and conduct overidentification tests, we find that infectious differences could only affect land distribution through markets. This suggests that the exclusion restriction assumption is unlikely to be violated.

People may still challenge the exogeneity of pesthouse locations and infectious mortality. In addition to examining these two measures of the epidemiological environment and rigorously testing the relevant identifying assumptions, we develop a complementary strategy based on the bioclimatic variation. Specifically, we consider the contribution of spring temperatures to plague transmission. Scientific research has shown that warmer springs are associated with greater activity among fleas and rodents. After accounting for a rich set of geographical controls, spring temperature variation across otherwise comparable villages more strongly predicts the spatial distribution of public markets, which finally predicts unequal land distribution.

After confirming the causal effect of public markets on land inequality, we turn to understanding the underlying mechanisms. We first develop a theoretical model based on historical narratives. As market access expanded, smallholders faced better outside opportunities and therefore left agriculture for non-agricultural sectors. The exit cutoff also increased conditional on market access. Larger farmers expanded operated land because the model assumes increasing returns to land over the relevant scale. Market access thus led to an upward reallocation of land. We provide several pieces of evidence consistent with these conclusions: (1) a decline in smallholder households and an increase in landlord households in agricultural landholding; (2) a increase in the non-agricultural share among smallholder households and a slight decline in the non-agricultural share among landlord households; and (3) simultaneous total population growth and a decline in total agricultural households while the total land area remained stable.

The effect of markets is possibly related to commercialization, for it changes systematically with external commercial conditions. Villages that initially had better commercial conditions (proxied by railway stations and the commercial employment share) showed weaker market effects, implying that existing commercial conditions consistent with saturation or diminishing marginal effects where commercial activity was already developed to further transformation (like barriers). We find no evidence that the market effects vary with changes in other sectors (agriculture or industry, etc.). Finally, our analysis of landownership demonstrates that only households with full ownership rights over land could reduce their holdings of self-cultivated land and report better economic performance in lower incidence of debt.

Our study contributes to the growing literature on colonial history, particularly on institutions and their consequences. Existing studies have focused on how colonizers developed different extractive or inclusive institutions under constraints imposed by settler mortality and early-stage resource accumulation, thereby explaining long-run differences in development (Acemoglu, Johnson and Robinson, 2001, 2002). The consequences of these institutions vary across cases (Heldring and Robinson, 2012). For example, the extractive cultivation system in Dutch Java contributes to long-term development (Dell and Olken,

2020); in British India, colonial rule hindered the provision of public goods (Iyer, 2010), but the railroads it built generated positive economic welfare (Donaldson, 2018). Taiwan was repeatedly colonized and decolonized throughout history, yet rigorous causal analyses of these colonial institutions remain limited (Ho, 1968). Recent research has increasingly focused on Taiwan. For example, Cheng, Fan and Wu (2022) exploit cane railroads to instrument for sugarcane cultivation and examine how cultivation helped eliminate harmful social norms; Huang and Jheng (2021) rely on distance to the residences of bank founders for instrumental identification and find that colonial banks promoted postwar industrialization. This paper focuses on market access generated by differences in infectious disease endowments and evaluates its impact on the distribution and allocation of land in agriculture, the most important productive sector in colonial and post-war Taiwan (Ho, 1968).

This study also contributes by examining the determinants of land inequality from a micro-level perspective. Cross-country studies argue that agricultural endowments can predict land concentration (Easterly, 2007), and similar explanations can also be applied to comparisons between northern and southern China (Kung, Wu and Wu, 2012). Spain's *estancias* (or *haciendas*) in Mexico (Engerman and Sokoloff, 2002) and the British system for collecting land revenue in India (Banerjee and Iyer, 2005) provide institutional explanations: regions where unequal land distribution was produced by historical institutions tend to reinforce that inequality over time (Frankema, 2010). Of course, exogenous shocks also matter (such as epidemics (Alfani, Di Tullio and Fochesato, 2026; Jedwab, Johnson and Koyama, 2022)). Our study engages with the multiple origins of land inequality in Taiwan's context and ultimately shows that the rise in land inequality was consistently related to an institutional but unintended commercialization transition. In this respect, our conclusion is highly consistent with the argument by Brandt and Sands (1990) in the context of 1930s China: non-agricultural economic opportunities led to structural selection.

The interaction between disease and colonial institutions provides a further contribution. In the classic setting of Acemoglu, Johnson and Robinson (2001), early European colonizers viewed countries with high mortality as unsuitable for long-term settlement and therefore established extractive institutions. In Taiwan, by contrast, Japanese colonizers introduced public health policies to directly eliminate the challenge posed by infectious diseases. Undoubtedly, Japan's policy-making was enabled by advances in modern medicine and technological innovation (virus detection, vaccines, and sanitary experts, etc.). Some economic historians argue that the Columbian Exchange led to the adoption of quinine, an effective treatment for malaria, which facilitated early exploration of Africa (Nunn and Qian, 2010). This comparison gives us a sense that if European colonizers had been completely unable to set foot in Africa's "death-haunted" lands, or had been able to eliminate all the disease threats, the development trajectories of many countries and regions would have been fundamentally different. The above discussion suggests that the importance of scientific development and technological change in institutional adaptation and adoption may not be limited to the mechanical level (for example, the broader effects of modern medicine may extend beyond demography and public health (Alsan and Goldin, 2019; Hollingsworth et al., 2024)) (Mokyr, Sarid and van der Beek, 2022; North, 1990).

We revisit land inequality and post-war land reform in Taiwan. Chen Cheng, the principal leader of

land reform, introduced one major reason in *A Summary of Taiwan's Land Reform*:

> *"…at that time, the situation on the mainland was deteriorating, public sentiment in Taiwan was unsettled, the economy was in disorder, and society was unstable, with the risk of upheaval at any moment. Therefore, to secure Taiwan as a base for national recovery, it was necessary first to seek stability…"*

Consistent with this ideology, many narratives characterize post-war land reform in Taiwan and worldwide as a "U.S.-backed strategy through which the capitalist bloc sought to contain communism" (Albertus, 2023; Chiang, Fan and Hsu, 2025). Politically driven redistributive policies may fail to address the underlying economic distortions. Thus, as Wang and Kim (2024) find, Taiwan's land reform may have made only a limited contribution to GDP growth. More importantly, although the second phase ("redistribution of public lands" policy) did work, it may have slowed the transition out of agriculture and even pulled manufacturing labor into agriculture (Bianchi-Vimercati, Lecce and Magnaricotte, 2026). One constraint was that land prices were set at 2.5 times total annual yield and were to be paid in kind in twenty semiannual installments. This meant that the 139,688 households that purchased land during the reform were tied to agriculture for at least ten years. Overall, we never doubt the contribution of Taiwan's land reform to reducing nominal land inequality, but its economic consequences remain to be further investigated.

The remainder of this paper is organized as follows. Section 2 introduces the relevant historical background. Section 3 describes our sample and data. Section 4 presents all empirical results. Section 5 concludes and discusses the implications.

2 Historical Background

2.1 Japanese Colonization of Taiwan

Japanese colonization of Taiwan began after the signing of the Treaty of Shimonoseki in 1895, when the Qing government ceded Taiwan to Japan. From that point on, Taiwan entered a half-century period of Japanese colonial rule (1895–1945). Prior to the outbreak of the Second Sino-Japanese War in 1937, the Japanese colonial authorities focused on infrastructure development (railways, public markets, and water systems, etc.) and agricultural production (mainly sugarcane, rice, and camphor) to establish effective control over Taiwan and increase returns from the colony. These endeavors enabled Japan to support its domestic economic development while maintaining colonial fiscal balance.

Although Japan relinquished its colonial authority in Taiwan following its unconditional surrender in 1945, the impact of its rule remained substantial. Beyond the physical developments mentioned above, its influence extended to institutional and cultural dimensions. On the one hand, Japan introduced modern institutions and facilities (e.g., statistical bureaus, police systems, formal hospitals, and formal schools), facilitating Taiwan's transition from imperial governance toward a new administrative structure. On the

other hand, during the later stage of colonial rule, Japan promoted the Kominka movement and exerted a lasting influence on both historical and contemporary Taiwan through Japanese-language education and state-led propaganda.[4]

2.2 Taiwan's Land

2.2.1 Land Survey

Japan followed a colonial economic policy of "industry in Japan and agriculture in Taiwan" and used supporting institutional arrangements to extract agricultural surplus from Taiwan. Acting as a stationary bandit (Olson, 2000), the Japanese colonial government carried out a "land survey" in 1898 (also known as the *Cadastral Survey*). This undertaking included recording Taiwan's geography, identifying land characteristics and ownership, and reforming the land tax system, etc. The land survey had two direct consequences. First, it enabled the colonial authorities to acquire detailed knowledge of and govern Taiwan. Second, parcels with mixed public and private status, unclear ownership, or no active use were identified, allocated, or developed, increasing both the amount of usable land and the returns from land (cultivation output and land taxes). To strengthen people's incentives to develop and cultivate land, the colonial government even stated that individuals could obtain permanent ownership of land as long as they "invested labor and capital" or had "sufficient factual claims" (Wu, 2017).

Once land ownership was confirmed, land was transformed from a general factor of production into a capital factor that could be formally used as collateral or traded. Historians have argued that the fundamental significance of the land survey lay in the capitalization of land. As described in Yanaihara (1988), the land survey

> *"…clarified land rights and made transactions secure… its main purpose was to attract capital and provide security for Japanese capitalists to invest in land and establish enterprises in Taiwan… in this sense, the land survey was a necessary precondition and foundational project for the capitalization of Taiwan and for Japanese capital to dominate Taiwan."*

2.2.2 Land Distribution

Taiwan entered the colonial period with a stratified system of land rights inherited from Qing rule. Rather than assigning each plot to a single legal owner, the customary system of "one field, two masters" divided claims among large-rent holders (*dazuhu*), small-rent holders (*xiaozuhu*), and actual cultivators. Large-rent holders received an upper-level rent, small-rent holders collected rent from cultivators, and cultivators retained what remained after rent and other obligations. Land rights were therefore overlapping, but the distribution of agricultural income was distinctly hierarchical: those controlling rent claims could appropriate a substantial share of the surplus despite performing little or none of the cultivation.

[4] Through ideological infiltration, Japan gradually eroded the sense of belonging that the Taiwanese people felt toward both the island of Taiwan and China. For instance, Japanese colonizers even recruited soldiers in Taiwan to fight in the Second Sino-Japanese War, pitting them against Chinese forces.

Japanese rule reorganized this hierarchy into a more unitary system of private ownership. Through the land survey and the abolition of large-rent rights, the colonial administration compensated large-rent holders, registered small-rent holders as legal owners, and generally left actual cultivators in the position of tenants. The surveys thus clarified which claims were legally enforceable, but they did not systematically redistribute land to land-poor cultivators (Tseng, 2015). Their distributive effect was consequently conservative rather than equalizing: preexisting differences in access to land were carried into the new legal regime. At the same time, clearer and more secure titles made land easier to transfer, mortgage, and use as collateral, thereby converting it more fully into a marketable capital asset (Chu, 2015). The colonial land survey, in short, formalized and capitalized an unequal agrarian structure rather than dismantling it.

The resulting inequality was economically consequential. Contemporary surveys commonly placed rents at roughly 40–50 percent of the harvest, while accounts later invoked in support of post-war survey reported ordinary rents of about one-half of output and, in some cases, as much as 70 percent (Chu, 2015; Ho, 1968). Tenants also faced insecure or short-term contracts, deposits, credit obligations, and the risk of eviction. Our village-level data provide direct evidence on the distribution of holdings. Using eleven land-size categories reported for approximately 280 villages, we reconstruct village-level land Gini coefficients for 1921 and 1932. In 1932, the distribution is centered at approximately 0.616 and reaches 0.788 in the most unequal villages.[5] Moreover, the entire distribution shifts to the right between 1921 and 1932. Land inequality was therefore not merely a static remnant of Qing rule: it persisted after the early land surveys and became more pronounced during the 1920s and early 1930s.[6]

This increase coincided with the deepening commercialization and market integration of colonial agriculture. Taiwan was incorporated into the Japanese imperial economy as a supplier of sugar and rice. Preferential access to the Japanese market, together with railways and feeder roads, expanded the market facing rural producers and reduced transportation costs (Ho, 1968). From the 1920s onward, irrigation, fertilizer, improved seeds, and agricultural extension raised productivity, while the diffusion of higher-yielding and higher-value Ponlai rice after 1922 made rice cultivation an increasingly profitable commercial alternative to sugarcane.

The gains from commercialization, however, were unlikely to be distributed evenly. In combination, secure titles, improved market access, and rising agricultural returns increased both the value of productive land and the incentive to accumulate it. Households and corporations with better access to capital could purchase land, extend credit against it, or secure control through leases. Tenants and smallholders, by contrast, remained more exposed to harvest risk, indebtedness, credit constraints, and rising rents. Market integration therefore did not create the initial agrarian hierarchy, but it could amplify that hierarchy by allowing better-capitalized actors to capture a disproportionate share of the gains from agricultural growth.

[5] During the same period, the average land Gini coefficient in mainland China was approximately 0.406 (Kung, Wu and Wu, 2012).

[6] The Gini coefficients are conservatively reconstructed from eleven grouped landholding categories. Because holdings within each category are not directly observed, their levels should be regarded as approximations. Comparisons across years are more informative because an identical construction procedure is applied to both years. Details are provided in Section 3.

The sugar industry provides the clearest example. As profitable rice cultivation intensified competition for land, Japanese sugar corporations increasingly purchased land directly or secured it through long-term leases. By 1925, Japanese sugar companies reportedly owned approximately 63,000 hectares; by 1939, the six largest companies owned more than one-eighth of Taiwan's cultivated land, in addition to land controlled through leasing (Ho, 1971). These firms consequently came to occupy not only the position of purchasers of sugarcane but also, increasingly, that of large landlords.

The historical narrative therefore suggests that the observed rise in village-level inequality could have emerged from the interaction among an inherited hierarchy, a colonial property regime that formalized unequal claims, and growing commercialization.

2.3 From Fish Market to Public Market

2.3.1 Fish Market

The market is the manifestation of people's preferences for exchange within a spontaneous order. During the late nineteenth and early twentieth centuries, numerous fish markets (FMs) had already formed throughout Taiwan.[7] The primary categories of goods traded at FMs included fish, meat, vegetables, fruit, and general groceries. Sellers at FMs often set up their stalls along bustling, high-traffic streets, such as those near temples and stations (Sunagawa, 2018). The original and core objective of these FMs was to connect sellers directly with consumers, thereby facilitating rapid and convenient exchange. Initially, FMs operated through barter or self-sufficient exchange, but they gradually evolved into centralized trading hubs driven by a monetary economy (Kusunose, 1915).

However, FMs had clear disadvantages. First, the government could not manage all sellers in these markets. According to historical materials, FMs were usually located on open streets or in long, narrow buildings. The government could manage only roadside stalls, while other stalls remained almost entirely beyond its control. Second, incomplete management impeded the collection of market fees, creating fiscal pressure. Finally, and most importantly, FMs often contributed to the spread of infectious diseases because of their loose organization and chaotic locations (Taiwan Provincial Literature Committee, 1972).

2.3.2 Public Market as One Solution

In Taiwan, epidemic infectious diseases (plague, cholera, smallpox, tuberculosis, etc.) frequently broke out, posing a severe challenge to the colonizers during the early period of Japanese rule.[8] One notable example was that the number of deaths caused by infectious disease was even greater than the number of soldiers who died in conflicts with anti-Japanese Taiwanese.[9] Infectious diseases threatened colonial rule; thus, the colonial authorities introduced a series of actions to control them, including recruiting doctors, introducing vaccination, establishing hospitals, and building water-supply and sewage systems,

[7] Notably, "fish market" is merely one representative example among a variety of names.

[8] In contrast to Diamond (1997), who emphasized that the deaths of European colonizers in the Americas were mainly due to the introduction of external diseases, the relationship between Japanese colonizers and Taiwan was closer to what Acemoglu, Johnson and Robinson (2001) described for European colonizers in Africa.

[9] For example, in the 1895 war to occupy Taiwan, fewer than 700 soldiers were killed or injured in battle, but 4,624 died from disease and 26,094 were infected (Hsu, 2020).

etc. To prevent the spread of infectious diseases in densely populated areas, FMs were officially abolished in 1904 and replaced by public markets (Sunagawa, 2018).[10]

Unlike FMs, public markets operated under a more comprehensive institutional framework. Most importantly, their management shifted from private operators to the government. As a result, the government could effectively collect administrative fees as fiscal revenue. Although public markets were also located in densely populated areas, they were all located in formal buildings equipped with ventilation systems. These facilities limited bacterial growth and improved indoor air quality. These narrative accounts suggest that the direct objective of establishing public markets was to improve public health. William Kinninmond Burton, a teacher of sanitary engineering at the Imperial University, proposed street sanitation improvement plans. To stop the spread of infectious diseases, he:

> *"... promoted the relocation, demolition, and reconstruction of urban streets and compelled scattered open-air street vendors to move into designated locations. Through government designation, land acquisition, and the use of idle houses or temples, vendors were gathered into fixed places within one-story wooden or bamboo buildings that could protect food."*

In all the historical narratives and archives, the public market had never appeared as a standalone commercial facility, and it had always been subordinate to entries related to public health.

In Figure G.4, we show the distribution of the years in which public markets were established. Among the 208 recorded public markets in Taiwan, the period of expansion was concentrated in the 20 years between 1905 and 1925. As public markets spread as public-health facilities, the large-scale infectious diseases that had previously been common in Taiwan almost disappeared after the 1920s. In Appendix A, panel (a) shows the decline in the mortality rate of infectious diseases (case fatality rate), while panels (b)–(d) show the reduction in the number of cases of three major infectious diseases (plague, cholera, and smallpox). We also provide three public health indicators: the share of the population covered by piped water, the number of hospitals, and the number of health-related staff (shown in panels (e)–(g)). These facts collectively imply that (1) after 1921, both the outbreaks and the effects of infectious diseases were relatively stable, and (2) public health spending expanded. In summary, the establishment of the public market (as one of the crucial measures in public health policy rather than mere commercial infrastructure) and the control of infectious diseases in Taiwan follow a consistent pattern.

2.3.3 Objective Spillovers of Public Markets

At the same time, public markets, as part of the formal sector, also provided nearby residents with stable venues for trade and consumption. People living near public markets could buy and sell their agricultural products more regularly. In addition, standardized operating conditions encouraged other

[10] Although a small number of public markets had already been established before 1904, their growth was very slow, increasing by only 0.91 p.p. per year on average. After the 1904 act promoting the establishment of public markets was passed, the growth rate doubled to 1.99 p.p. and continued at this pace for nearly 20 years. After 1921, the growth rate became very close to its pre-1904 level (see Figure G.4 of Appendix G.4).

forms of commercial activity. For example, people could trade modern daily goods and imported products or set up luxury clothing shops, etc.[11] These activities had not been possible in the original FMs.

3 Data

The historical account above provides an ideal setting: a public health policy that inevitably granted some villages greater market access. This setting allows us to leverage such variation to examine historians' argument that market access, consistent with commercialization, contributed to the unequal distribution of land in Taiwan. Although historians have offered extensive and intuitive accounts, a quantitative exercise remains valuable for deepening our understanding of this process. In this section, we provide information about the data used in the empirical analysis. Additional information about the core variables we are interested in is provided in Appendix B.

3.1 Sample

Our benchmark sample consists of Taiwan's 285 villages in 1932. These villages covered almost all populated areas in Taiwan, except for aboriginal territories. In 1932, these villages were third-level administrative units and provided the finest level of available data, including agricultural survey, land survey, and census information, etc. Another advantage is that these villages experienced very few changes around the 1930s. In fact, 284 of them had already been defined in 1921. By examining records of their names, we are able to link different types of data across this period with minimal error.

For many other statistical indicators, the data were recorded annually for first-level administrative units (prefectures). Therefore, in the empirical analysis, we control for prefecture fixed effects. Another concern is that seven *shi* were recorded at the same statistical level as the villages. Subsequent empirical tests and robustness checks show that the influence of these administrative centers on our research design and conclusions is negligible. We also provide a classification and description of Taiwan's administrative divisions during the Japanese colonial period in Appendix C.

3.2 Land Distribution

The data on land distribution come from agricultural surveys conducted by the colonial authorities. In particular, the *Survey of Cultivated Land Distribution and Management* (SCLDM), conducted only in 1921 and 1932, provided village-level information on land distribution. These two SCLDMs classified landholdings into 12 groups by size. Except for the last group, which had no upper bound, the remaining 11 groups recorded the distribution of landholdings from 0 to 100 *Jia* within each village.[12] The surveys recorded the number of households in each land size group, allowing us to infer the land distribution in each village. Information on the data sources appears in Figure B.2.

One main challenge is that each land size group has a fixed range in every village. Thus, a Gini coefficient based only on fixed interval values would not capture variation in average holdings within

[11] Special goods such as alcohol, salt, and opium were usually sold by state-run monopoly agencies.
[12] One *Jia* is a unit of land area equivalent to about 0.97 acres.

groups across villages. Fortunately, the 1921 SCLDM reported the total land area for each land size group at the village level. For example, in 1921, Taibei Shi had 2,315 households in the 0–0.5 *Jia* group, with a total land area of 422 *Jia* (implying an average of 0.182 *Jia* per household in this group) [13]. Therefore, when calculating the land Gini coefficients for both 1921 and 1932, we use each village's 1921 average land area in each land size group as the weight. For land size groups missing in 1921, we use the average land area of the same group in other villages as an adjustment. We also correct a few obvious statistical typos using original information in the SCLDM.[14] The detailed formula for calculating the land Gini coefficient is provided in Appendix D.

[Figure 1 is about here]

We plot the primary results in Figure 1. Figure 1a shows the distribution of land Gini coefficients in 1921 and 1932. The approximately normal distributions support the validity and reliability of the data source and construction. Another pattern is that the kernel density distribution of the land Gini coefficient shifted to the right in 1932 compared to 1921. In Figure G.1 of Appendix G.1, we also plotted the Lorenz curve calculated for all of Taiwan, with the 1932 Lorenz curve always occupying a larger area. Overall, land distribution in Taiwan was unequal and became more so over time. Figure 1b shows a strong village-level correlation between the land Gini coefficients in 1921 and 1932 (with an R-squared of 0.76), which means that land inequality in 1921 is a good predictor of inequality in 1932. In the empirical design, we control for the 1921 Gini coefficient to capture factors that could have affected inequality around the 1920s (mainly the outbreak of epidemic infectious diseases).

In addition, the SCLDMs provide information on land area by land type and ownership for different villages. In terms of land type, we identify wet and non-wet land, which were used mainly to grow rice and sugarcane. In terms of ownership, we distinguish between owner-operated land (land owned and cultivated by the same household) and tenant-operated land within each village. These variables will be introduced in the subsequent analysis. Notably, the 1932 SCLDM also recorded the distribution of households across agricultural and non-agricultural sectors. Moreover, the agricultural surveys include the *Survey on Agricultural Finance*, conducted in 1933, from which we obtain information on village-level debt conditions. We are interested in the debt positions of households with different ownership types.

3.3 Locations of Public Markets

The source for the public market data is Academia Sinica.[15] Based on historical materials, it records 208 public markets established between 1895 and 1945, maps their GIS locations, and provides their names and years of establishment. Public markets on islands and in remote areas are also carefully recorded, suggesting that measurement error due to missing public markets is unlikely to be a major concern. Figure G.4 shows the establishment of public markets by year. Most public markets were established between 1905 and 1925, and only five new public markets were established after 1932. Our benchmark

[13] See Figure B.2.

[14] For example, the calculated average land area may far exceed the applicable range, or the reported figures may not correspond to the proportions.

[15] Data could be found in https://thcts.sinica.edu.tw/, accessed on February 20 2026.

analysis uses the 203 public markets established before 1932.

[Figure 2 is about here]

To measure each village's access to public markets, we calculate the distance from the village centroid to each public market and use the logarithm of the minimum distance as the indicator. Intuitively, villages located closer to markets enjoyed lower transaction costs and more non-agricultural employment opportunities, which were objective effects of public markets beyond their role as public-health infrastructure. We show the spatial pattern of each village's market distance in Figure 2d. In the robustness checks, we consider alternative measures and use remote public markets as a placebo.

In addition to the locations of public markets, Academia Sinica provides extensive geocoded data on other institutions and facilities, including pesthouses, monopoly offices, military installations, police stations, Japanese shrines, ports, coastlines, banks, courts, prisons, ancestral halls, churches, Confucian academies, public schools, water sources, hospitals, and railway stations. We gradually introduce these variables in the empirical analysis to provide further support for our findings.

3.4 Other Variables

We also collect data from the following sources.[16] (1) The 1921 and 1931 general censuses. Taiwan's population censuses were well documented, allowing us to calculate important village-level demographic indicators, including population size, sex ratio, migration rate, birth rate, and death rate. (2) The 1926 survey of population by mainland ancestral origin. We use this survey to calculate the proportion of mainland-origin residents in each village. (3) Public-health surveys from 1917 to 1921. These surveys provide mortality statistics for 41 causes of death. We focus particularly on deaths from infectious diseases. (4) The 1915 and 1920 temporary surveys of Taiwan's *Hukou*. These surveys provide additional information on the village-level distribution of occupations, the literacy rate, the proportion of women with unbound feet, and the proportion of the population able to speak Japanese, etc. (5) Natural geography data. These data include digitized historical maps of Taiwan that provide locations (longitude and latitude coordinates) and boundaries. We also collect data on historical earthquakes,[17] crop suitability (for sugarcane and rice),[18] monthly historical temperature (since 1895),[19], river density,[20] and elevation[21] by administrative unit.

[16] Data could be partly found in https://tieha.org.tw/ (Taiwan Historical Statistics), accessed on March 1 2026. Historical archives (including SCLDMs mentioned above) could be found in online libraries like https://das.nlpi.edu.tw/ and https://www.digital.archives.go.jp/.

[17] Data could be found in https://doi.org/10.5281/zenodo.7492391, accessed on February 27 2026.

[18] Data could be found in https://www.gaez.iiasa.ac.at/, accessed on July 14 2025.

[19] Data could be found in https://www.psl.noaa.gov/data/gridded/data.20thC_ReanV3.html and https://www.chelsa-climate.org/datasets/chelsa_climatologies, accessed on August 15-17 2026.

[20] Data could be found in http://www.webmap.cn, accessed on May 31 2025.

[21] Data could be found in https://registry.opendata.aws/copernicus-dem/, accessed on August 17 2026.

4 Empirical Strategy and Results

4.1 OLS Estimates

We begin with the following ordinary least squares (OLS) equation to examine the quantitative relationship between market access and land inequality:

$$Land_Gini_{i,1932} = \alpha + \beta Log_Market_Distance_i + \mathbf{X}_i'\boldsymbol{\gamma} + \rho Land_Gini_{i,1921} + \lambda_p + \varepsilon_i, \quad (1)$$

where i indexes villages; $Land_Gini_{i,1932}$ denotes land inequality in village i in 1932, with higher values indicating greater inequality; $Log_Market_Distance_i$ is the logarithm of the distance from village i to the nearest market, and we take it as an inverse proxy for market access and take its logarithm to reduce right skewness; $\mathbf{X}_i$ is a vector of baseline characteristics of village i, including log population in 1931, the female share in 1931, the net migration rate in 1931, and the share of mainland-origin residents in 1926; $Land_Gini_{i,1921}$ controls for the initial level of land inequality in village i in 1921; λ_p denotes prefecture fixed effects accounting for unobserved characteristics that vary little within prefecture p; α is the intercept term and ε_i is the error term. By clustering standard errors at the level of 61 cities, we allow errors to be heteroskedastic and spatially correlated within cities (second-level administrative units).[22] Each village is weighted by its population size.

Table 1 presents our baseline regression results. Column (1) directly examines the relationship between the log distance to the nearest public market—hereafter, market distance—and the 1932 land Gini coefficient—hereafter, land inequality—without adding any controls. The negative point estimate of -0.0214 indicates that villages closer to public markets had higher land inequality. Columns (2) and (3) sequentially add the log population control and population weights. The estimated coefficient on market distance becomes slightly smaller but remains statistically significant. Column (4) further controls for other demographic characteristics. We find little evidence that these important demographic variables are systematically related to land inequality. Column (5) adds the 1921 land Gini coefficient to control for initial inequality. This is a bad control variable because if our assumption (market access is related to higher land inequality) holds, adding the initial Gini coefficient from 1921 would cause our estimated coefficient to be closer to 0 (due to the existence of path dependence). Although this control does not qualitatively change our result, it is important for our subsequent instrumental variable design. We discuss this point in detail in Section 4.2.3.

[Table 1 is about here]

Column (5) is our preferred specification. The point estimate on market distance is -0.0078 and is significant at the 99% confidence level. A one-standard-deviation decrease in market distance, equal to 0.996, is associated with an increase in land inequality equivalent to approximately 1.26% of its sample mean of 0.616. Meanwhile, land inequality in 1921 strongly predicts land inequality in 1932,

[22] Clustering at the prefecture level does not change our results, but the small number of clusters (8 prefectures) may lead to incorrect statistical inference.

consistent with the interpretation of path dependence. The coefficient on population becomes statistically indistinguishable from zero in this specification. Among the other demographic variables, only the female share has a weak statistical relationship with land inequality. Overall, we find a strong statistical association that villages located closer to markets reported higher land inequality.

We also perform the following robustness checks. First, we adjust the way land inequality is measured. Of all these indicators, the 1932 measure of inequality is consistently higher than the 1921 measurement of inequality, and the multiple measurements of inequality are highly correlated (see Table G.1 and Table G.2 of Appendix G.2). Furthermore, these indicators exhibit a consistent and significant pattern in our preferred regression specification (see Table G.3 of Appendix G.2). We also consider using the difference in land inequality, and a binary dummy based on whether the difference is greater than 0, as shown in Table G.4 of Appendix G.2.

Secondly, we adjust the measurement of the core independent variables, using different functional forms and positive alternative indicators as proxies for market access (see Table G.5 of Appendix G.3). Figure G.2 of Appendix G.3 demonstrates that our estimation results are highly sensitive to the choice of specific markets. We obtain the largest estimated coefficients if and only if the distance is calculated using the nearest market. The actual estimate lies in the extreme tail of the permutation distribution generated by randomly reassigning market distance (see Figure G.3 of Appendix G.3).

Finally, we also consider flexible adjustments to the model specification to allow for nonlinearity and small variations in market accessibility between 1921 and 1932 (see Table G.6 and Table G.7 of Appendix G.4). We also utilized the addition of spatial fixed effects and regional dummy functions, and employed spatially robust standard errors (see Table G.8 and Table G.9 of Appendix G.4). All the results remain largely unchanged.

4.2 Instrumental Variable Design

The previous OLS estimates present stylized facts about the relationship between market access and land inequality. However, the locations of public markets were not randomly assigned. For instance, villages that established public markets earlier may also have had higher levels of economic development, which cannot be fully captured by the population size included in the baseline specifications. At the same time, the relationship between economic development and inequality across different dimensions of society is complex.

In this section, we examine the historical setting and introduce the presence of pesthouses as the major instrumental variable (IV) to reduce endogeneity concerns. This captures the supply side of the epidemiological environment. We further consider the demand side, using infectious mortality to assess the exclusion restriction underlying the pesthouse instrument. Finally, we introduce an exogenous source of bioclimatic variation, using spring temperature as an additional, albeit relatively weak, instrument to provide complementary evidence. The spatial distributions of all three instruments are presented in Figure 3. Although we acknowledge that this design cannot fully resolve endogeneity, the unique historical setting and extensive checks suggest that the correlations we identify are unlikely to be driven by chance.

[Figure 3 is about here]

4.2.1 Infectious Diseases and Pesthouses

Section 2.3.2 documents that public health concerns drove the establishment of public markets and that this policy was closely related to patterns of infectious disease contagion. This implies that public market locations responded to regional infectious disease burdens and health conditions. Therefore, we use pesthouses as an IV in this unique setting to help interpret simple correlations as causal relationships. The term "pesthouse" refers to a facility used to quarantine and isolate individuals at risk of potential exposure to infectious viruses.[23] Due to the infectious nature of the diseases, transporting affected individuals across regions to hospitals with more comprehensive medical facilities (e.g., in political centers) was unlikely to be feasible. Consequently, the location of a pesthouse was largely determined by a region's unobserved characteristics related to infectious disease outbreaks. We therefore utilize the quasi-random spatial distribution of pesthouses to capture the extent to which a village relied on public health policies. Using digitized electronic maps constructed by Academia Sinica, we obtained the precise spatial distribution of pesthouses in colonial Taiwan dating from 1899.

Another advantage of using pesthouses as opposed to the public markets of interest is their relative scarcity. Pesthouses were established specifically to handle infectious diseases and typically accommodated 10 to 40 people, which was generally sufficient to meet village-level demand. Pesthouses also required specialized personnel, mainly scarce Japanese doctors and nurses holding medical degrees from Japanese universities (Taiwan Provincial Literature Committee, 1972). Our research sample includes 203 public markets but only 45 pesthouses, implying less spatial clustering and making the captured variation more representative. Following the logic of measurement of market access, we calculate the distance from each village centroid to the nearest pesthouse and use the logarithm of this distance as our primary IV. Figure 4a shows village-level proximity to both pesthouses and public markets. Although pesthouses were fewer, distance to the nearest pesthouse strongly predicts market distance.

[Figure 4 is about here]

4.2.2 Balance Tests

Although no evidence has been found to date indicating that pesthouses were built for non-health-related reasons, concerns about their validity as an IV persist. Our IV design assumes that (1) the instrument is not correlated with potentially omitted unobservable factors and (2) the instrument does not directly affect land inequality. Given the importance of the exogeneity and exclusion restriction to our IV design, we first construct four sets of balance tests. Specifically, we examine the simple relationship between pesthouses and predetermined/contemporaneous indicators for (1) population structure, (2)

[23] "Quarantine" generally applies to individuals who have not yet manifested signs of illness, whereas "isolation" targets those who are already symptomatic (Gensini, Yacoub and Conti, 2004). In colonial Taiwan, the two modes were combined; for instance, pesthouses were used both to monitor individuals forced to leave their homes because of an infectious-disease outbreak and to house confirmed patients, who were also required to undergo free blood testing during the isolation period (Taiwan Provincial Literature Committee, 1972).

state capability, (3) economic development, and (4) human capital.

[Table 2 is about here]

The main concern is the potential mechanical impact of infectious diseases on population dynamics, which could directly change our measure of land distribution (Alfani, 2015; Alfani and Murphy, 2017; Huang and Jheng, 2021). Consequently, we examine the relationship between pesthouses and population structure. Column (1) of Table 2 shows no correlation between pesthouses and changes in birth rates. Unsurprisingly, the presence of pesthouses is correlated with reductions in death rates at the 10% significance level. However, as shown in column (3), pesthouses do not produce systematic differences in total population. Given these inconclusive results, both our baseline OLS and subsequent estimates control for the 1921 land Gini coefficient to capture inequality characteristics inherited from the epidemic period (plague was no longer a major concern by 1920, but sporadic outbreaks of other diseases reinforced the need to consider the impact of infectious diseases around 1921). While this approach may reduce the magnitude of the estimated effect, it is necessary for qualitative validation. We also include a set of relevant controls in the extended IV specifications.

Second, because pesthouses functioned as colonial institutions, their presence may merely reflect colonial penetration or broader state capability (e.g., the neglect of villages with higher infectious disease burdens), providing an alternative explanation. This possibility implies that colonists may have selected settlement sites based on disease endemicity, thereby introducing different characteristics and trajectories (Acemoglu, Johnson and Robinson, 2001).[24] Columns (4)–(10) of Table 2 examine the relationship between pesthouses and various measures of state capability, including political center dummies, the number of state-run monopoly institutions, the number of military regions, the number and normalized density of police stations, the number of Japanese shrines, and the logarithm of the foreign population. The results do not support the interpretation that pesthouse locations merely proxy for colonial penetration.

Another concern arises from the complex relationship between economic growth and inequality (Alesina and Rodrik, 1994; Kuznets, 1955), although we believe that pesthouses were largely exogenous to it. Since we cannot construct precise measures of village-level economic development for this period, we employ alternative variables as proxies for economic development (in lieu of per capita GDP). Population is a commonly used indicator in historical contexts, and we have already controlled for it in the baseline specification (Acemoglu, Johnson and Robinson, 2002; Ashraf and Galor, 2011). In columns (11)–(17) of Table 2, we also examine other potential determinants/outcomes of economic development in Taiwan. We first consider ports just before the colonial era (column (11)) and banks during the early colonial era (column (12)), as the extensive literature highlights the contributions of trade access and financial markets to economic growth (Acemoglu, Johnson and Robinson, 2005; Guiso, Sapienza and Zingales, 2004). We find no correlation between these variables and the presence of pesthouses. Greif,

[24] The geographical concentration of Dutch rule in southern Taiwan was rooted in the maritime expansion of the Age of Discovery. Dutch voyages to Asia followed the Cape of Good Hope route and entered East Asian waters through the trading networks of Southeast Asia. After an unsuccessful attempt to seize Macao in 1622, the Dutch East India Company shifted first to the Penghu Islands and established its main base at Tayouan on Taiwan's southwestern coast in 1624 (Andrade, 2006).

Mokyr and Tabellini (2025) emphasize the contributions of contractual corporations and clans to growth trajectories. Taiwan had both contractual institutions and clan structures. Therefore, we use the numbers of courts and prisons as proxies for contractual and legal development and the number of ancestral halls as a proxy for clan culture. The results remain insignificant (columns (13)–(15)). Given that Taiwan was under Dutch colonial rule from 1624 to 1662, the religious origins of economic development offer an alternative perspective (Becker and Woessmann, 2009; Weber, 1904). Column (16) shows no relationship between the distance to the nearest pesthouse and the number of churches. Finally, leveraging variation in the gender norm of foot-binding across labor markets, we find no evidence in column (17) that pesthouses are associated with higher rates of foot unbinding (Cheng, Fan and Wu, 2022; Fan and Wu, 2023). In summary, we find no evidence that the presence of pesthouses is potentially correlated with economic development or its determinants.

We further consider the interaction between human capital and the epidemic. Existing evidence suggests that positive public-health interventions can promote human capital accumulation (Bleakley, 2007). Therefore, regions that historically placed greater emphasis on public health (for complex reasons) may simultaneously exhibit lower rates of infectious disease and higher levels of human capital. In columns (18)-(20), we examine three village-level proxies for human capital: the historical number of academies, the number of public schools, and the literacy rate.[25] The results show that our instrument is uncorrelated with these factors.

To conclude, apart from one marginal concern (mechanical effects in population structure) that can be addressed with explicit controls or additional tests, we believe that pesthouse distance qualifies as a valid IV because it is statistically orthogonal to a wide range of potential confounding factors.

4.2.3 IV Estimates

We next estimate the causal effect of market access on land inequality using a two-stage least squares (TSLS) strategy. Specifically, we use the distance from each village centroid to the nearest pesthouse as an IV. The first-stage equation is

$$\begin{aligned} Log_Market_Distance_i = \pi_0 + \pi_1 Log_Pesthouse_Distance_i \\ + \mathbf{X}_i'\boldsymbol{\pi} + \theta Land_Gini_{i,1921} + \lambda_p + u_i, \end{aligned} \tag{2}$$

and the corresponding second-stage equation is

$$Land_Gini_{i,1932} = \alpha + \beta \widehat{Log_Market_Distance_i} + \mathbf{X}_i'\boldsymbol{\gamma} + \rho Land_Gini_{i,1921} + \lambda_p + \varepsilon_i, \tag{3}$$

where $Log_Pesthouse_Distance_i$ is the logarithm of the distance from village i to the nearest

[25] The available illiteracy measure does not distinguish literacy by language or population group and may therefore conflate general human capital with linguistic composition and exposure to Japanese-language education. As a diagnostic check, column (20) additionally controls for the share of the population able to speak Japanese in 1920. Once this variable is included, the estimated association between pesthouse presence and illiteracy becomes smaller and statistically indistinguishable from zero. This attenuation is consistent with the possibility that the baseline correlation partly reflects differences in Japanese-language acquisition or population composition rather than broader human-capital differences. We interpret this exercise cautiously.

pesthouse. The fitted value $\widehat{Log_Market_Distance_i}$ is obtained from the first-stage regression. Including the 1921 Gini coefficient directly as a regressor in the estimation equation has two advantages: First, it can compensate for some historically persistent inequalities, such as land distribution inherited from the Qing Dynasty. Second, since the mortality rate of infectious diseases stabilized regionally after 1920, outbreaks of infectious diseases could be effectively controlled. We assume that public health factors did not affect the changes in inequality from 1921 to 1932. We allow that the 1921 land Gini coefficient is partially compensated by public health factors. This mechanically severs the residual impact of past infectious diseases on current land distribution.

[Table 3 is about here]

We present the IV results in Table 3. Before reporting the TSLS results, we first report the reduced-form (RF) specification. Column (1) includes pesthouse distance directly as a regressor and uses the same controls and the prefecture fixed effect as the baseline specification. The negative estimate shows that greater distance from a pesthouse is correlated with lower subsequent land inequality. The baseline TSLS results appear in column (2). Panel B reports the first-stage estimates; consistent with our hypothesis, the spatial distributions of pesthouses and public markets are strongly consistent. The first-stage F-statistic is 45.74, exceeding the conventional threshold of 10 for weak IV concerns. Thus, although pesthouses number only about one-quarter as many as public markets, they have strong predictive power for the local infectious endowment that influenced public market locations. Panel A reports the second-stage estimates of interest. Using log distance to the nearest pesthouse as an instrument for log distance to the nearest public market, we find that market distance is a significant negative predictor of land inequality (coefficient estimate = -0.0257; standard error = 0.0067).

4.2.4 Accounting for Additional Controls

In this section, we directly add controls to the TSLS estimation equations. Existing literature and historical narratives suggest that these variables could plausibly be correlated with both infectious diseases and agricultural land distribution. These investigations may also help us better understand the actual processes of land distribution and inequality formation in Taiwan. This exercise is our second type of test for IV validity, and its results leave our previous conclusions virtually unchanged.

First, we incorporate a set of variables related to public health. In the previous exclusion restriction test (specifically column (2) of Table 2), we failed to reject the existence of a link—both correlational and intuitively causal—between pesthouses and declining death rates. Public health factors, particularly infectious diseases, have significant effects on the distribution of property and broader economic outcomes (Alfani and Ammannati, 2017; Jedwab, Johnson and Koyama, 2022). Therefore, we attempt to control for channels through which public health in Taiwan might directly affect land inequality. We consider (1) log distance from the village to a clean-water source (relevant to the construction of piped-water systems) and (2) log distance from the village to the nearest hospital (Alsan and Goldin, 2019; Hollingsworth et al., 2024). In column (3) of Table 3, the coefficients on these two variables are statistically indistinguishable from zero. Meanwhile, the estimated market effect decreases slightly to -0.0224 but remains highly

statistically significant. The reason may be the scale. Between 1917 and 1921, the average annual mortality rate from dozens of infectious diseases in Taiwan was 1.68 per 1,000 people. In contrast, the average annual mortality rate in Europe during the Black Death was approximately 78–120 per 1,000 people, roughly 46 to 71 times the average level in Taiwan. Thus, although infectious diseases in Taiwan represented a significant public health burden, the overall scale of mortality was far below that of catastrophic epidemics like the Black Death (Jedwab, Johnson and Koyama, 2022).

Some literature on geographical determinism suggests that variations in factors such as heat, agriculture, and disease are driven by spatial distribution, specifically latitude (Sachs, 2001). Therefore, we directly control for the latitude and longitude of each village to capture these differences. This approach also allows us to account for more flexible spatial trends. The results in column (4) indicate that geographical location is unlikely to alter our interpretation.

Because Taiwan lies along the Pacific Ring of Fire, frequent seismic activity may have shaped regional economic development and generated spatial differences in agriculture by affecting expectations of future disaster risk (Cameron and Shah, 2015; Felbermayr and Gröschl, 2014). Anecdotal evidence also suggests a link between earthquakes and infectious diseases as secondary disasters. To address this concern, we construct a village-level measure of historical exposure to major earthquakes and include it as an additional control.[26] Controlling for historical earthquake intensity leaves our results virtually unchanged in column (5).

Finally, we include a set of indicators for agricultural suitability as controls. During the land surveys, the Japanese colonial authorities encouraged residents to develop uncultivated land and granted them ownership automatically. We are concerned that villages situated in regions with superior natural conditions for agriculture might differ in terms of land allocation (Easterly, 2007; Engerman and Sokoloff, 2002). However, as McNeill (1998) notes, excessive environmental degradation and development can lead to imbalances, potentially increasing the risk of natural disasters and disease. Consequently, in column (6), we control for the suitability of two major crops (sugarcane and rice), and in column (7), we incorporate all four sets of control variables. Overall, our results remain robust across all these additional checks.

4.2.5 Infectious Mortality and Overidentification Tests

We further examine the validity of our IV strategy using an overidentification test (see Acemoglu, Johnson and Robinson (2001); Bai and Jia (2016) for similar approaches). According to our proposed channel, a village's reliance on public health policies, as captured by pesthouse location, affected the location of public markets, which thus predicted land inequality in 1932. We always assume that pesthouse location does not directly affect land inequality. Therefore, we construct another proxy for infectious disease burden: the infectious mortality rate. It will increase our confidence in the IV identification through the

[26] For village i, earthquake exposure is defined as

$$\text{Earthquake_Intensity}_i = \sum_{e \in \mathcal{E}} \frac{M_e}{d_{ie}^2}.$$

Here, M_e is the magnitude of earthquake e, and d_{ie} is the distance, measured in kilometers, from village i to its epicenter.

interaction between the two measures.

Specifically, we retrieve information on the average number of deaths by cause between 1917 and 1921 from Japanese government documents. We focus on legally notifiable infectious diseases and those transmitted primarily through airborne or droplet-borne person-to-person contact, including plague, cholera, smallpox, tuberculosis, typhoid and paratyphoid fever, dysentery, diphtheria, scarlet fever, epidemic cerebrospinal meningitis, and epidemic encephalitis.[27] We standardize the total number of deaths from these diseases by the 1921 population (per 1,000 people). Because these are the only available village-level mortality data and cannot adequately capture variation during the early colonial period, we use this imperfect indicator as a supplementary instrument for assessing the validity of pesthouse distance.

[Table 4 is about here]

In Table 4, we first confirm the relationship between pesthouses and the infectious mortality rate using a simple OLS estimator in column (1). The estimated coefficient is negative and significant at the 5% level. Villages located farther from a pesthouse reported lower mortality rates, implying that villages previously affected by infectious diseases (those closer to a pesthouse) experienced higher subsequent infectious-disease mortality. The presence of a pesthouse did not reverse the extent to which a village relied on public health policies. Columns (2) and (3) present RF and TSLS estimates using the infectious mortality rate as an IV. The results are consistent with those obtained using pesthouse distance as the IV. Furthermore, the estimated coefficient (-0.0276) is similar to that obtained using pesthouse distance (-0.0257), and the first-stage F-statistic of 24.29 also indicates that the infectious mortality rate is not a weak instrument.

If our hypothesis holds, once we use one of the IVs in a just-identified TSLS model, we should not observe the other IV exerting a direct effect on land inequality. This implies that the correlation observed in the RF specifications is driven by the public market channel. Consequently, in columns (4) and (5), we employ one IV as the instrument while including the other as a regressor in the estimating equations. The results demonstrate that, once we control for the channel through which public health affects land inequality via the market, the direct effect of public health becomes statistically indistinguishable from zero. This serves as the most direct validation of the exclusion restriction assumption.[28] Column (6) combines the two IVs and the p-value of the overidentification test is 0.8677.

4.2.6 Climatic Origins of Plague: Temperature Variation

Our preceding evidence suggests that the epidemiological environment, proxied by pesthouses and infectious mortality, provides a credible source of instrumental variation. These proxies, however, offer limited direct evidence regarding the exogeneity of the underlying epidemiological conditions. To provide further causal evidence, we therefore examine how bioclimatic factors (represented by spring tempera-

[27] Epidemiological evidence indicates that infectious diseases can also be transmitted via the fecal-oral route (e.g., cholera and typhoid). Our subsequent results remain robust when using a measure of infectious disease mortality that excludes cholera. However, because typhoid statistics are aggregated with those of other infectious diseases (such as plague), we are unable to isolate and exclude this type of impact.

[28] Even if we allow for a slight direct effect of instrumental variables on land inequality, our estimates of instrumental variables are unlikely to be reversed (see Appendix G.6).

tures) contributed to the prevalence of plague–the dominant infectious disease during the early colonial period.

Plague is a zoonotic disease caused by *Yersinia pestis*, a bacterium harbored by rodents and transmitted primarily through fleas (Xu et al., 2019). Existing research finds that a 1°C increase in spring temperature is associated with an approximately 59% increase in the average prevalence of plague during the spring (Stenseth et al., 2006). Warmer temperatures cause fleas to become active earlier and increase the availability of food and habitat resources for rodents, thereby raising the risk of plague transmission (Mora et al., 2022; Stenseth et al., 2006). This bioclimatic mechanism is particularly relevant to Taiwan for two reasons. First, the role of fleas in transmitting plague was first identified in Taiwan (Ogata, 1897). Second, spring temperatures in Taiwan fluctuate around 20°C, placing the island within a relatively warm, but not excessively hot, temperature range conducive to plague transmission.

[Table 5 is about here]

Following the approach of Stenseth et al. (2006), we calculate the average spring temperature in March and April over the plague epidemic period beginning in 1895. Appendix E describes the construction of the historical temperature measure. According to the proposed bioclimatic mechanism, villages experiencing warmer springs, and thus longer exposure to favorable transmission conditions, should be located closer to public markets. Panel B of Table 5 presents the first-stage estimates using spring temperature as the IV. Column (1) reports a specification consistent with our baseline model. Although the first-stage relationship is statistically significant, the corresponding second-stage estimate is not. We attribute this discrepancy to other geographical determinants of temperature. We therefore control for latitude, longitude, and their interaction, as well as the logarithmic distances to the coast and the nearest port, river density, and elevation. Conditional on these geographical characteristics, temperature differences across nearby and otherwise similar villages become more comparable. Consistent with this interpretation, the absolute magnitude of the first-stage estimate increases after these geographical controls are introduced.

Three additional findings are noteworthy. First, controlling for the initial Gini coefficient affects the magnitude of the estimated effect but does not alter the statistical inference. Second, the results are sensitive to slight changes in the definition of spring: extending the spring window to include either February or May produces noticeable differences in statistical significance. Third, the estimates are insensitive to the precise definition of the plague period, even when more than one-third of the later years are excluded.

Using spring temperature as the IV offers two advantages. First, existing bioclimatic evidence provides a scientific basis for the relationship between seasonal temperature variation and plague transmission. Second, systematic human manipulation of weather did not become technically feasible until 1946, making historical spring temperatures unlikely to have been influenced by local economic or institutional conditions. Nevertheless, measurement limitations and potentially local nonlinear effects make spring temperature less suitable for precisely estimating the magnitude of the causal effect and may generate a weak-IV concern. Despite these limitations, our qualitative conclusions still remain largely unchanged.

These results provide additional support for the causal effect of health-induced public markets on agricultural land distribution.

4.2.7 OLS versus IV

In the baseline model, the OLS coefficient for the effect of market distance on land inequality is approximately -0.0078, whereas the IV coefficient is approximately -0.0257 (-0.0276 when estimated using mortality rates) [29]. The IV estimate is roughly three times the size of the OLS estimate. The discrepancy between these estimates is unlikely to stem from a weak instrument, as the first-stage F-statistics across all specifications are well above 10. Weak instrumental variables are unlikely to be the main reason for the differences in our coefficients (see Appendix G.5). Although the exclusion of certain top-tier groups (where land area ranges could not be defined) and the calculation of distance based on coordinates (without accounting for transport conditions) suggest the presence of measurement error in our key independent and dependent variables, such error, and the resulting attenuation bias toward 0, cannot adequately explain a threefold difference.

A larger IV coefficient may also arise because IV identifies the effect for villages whose market access was changed by the spatial distribution of pesthouses, and this local effect need not equal the average relationship captured by OLS. To assess whether the gap instead signals an invalid instrument, we apply the Bayesian procedure of DiTraglia and García-Jimeno (2021). The procedure asks whether two beliefs suggested by our setting (market distance contains some proxy error and omitted factors may attenuate the negative OLS estimate) are jointly consistent with the observed effect size and an exogenous instrument. We find that, across different specifications, the estimated correlation between the instrument and the second-stage error includes zero, while the corrected effect of market distance remains negative and is closer to the IV estimate than to the OLS estimate (see Table G.11 of Appendix G.6).

4.3 Exploration of Underlying Mechanisms

Having provided credible evidence of causality, we turn to the mechanisms underlying the positive relationship between market access and land inequality. Section 2 presents substantial circumstantial evidence from historians and historical documents indicating that the introduction of public markets significantly increased commercial activity in nearby areas. This effect most likely arose because public markets, as part of the formal sector, provided commercial employment and access to trade.

4.3.1 A Conceptual Framework

We first develop a simple theoretical framework for better conceptualization. We assume that all farmers make trade-offs in allocating resources among three factors of production: agricultural landholdings (or leases), agricultural labor, and non-agricultural labor. The market plays a crucial role. It determines both the returns individuals derive from agricultural output (through iceberg trade costs) and the payoffs

[29] In terms of weighted standardized coefficients, these figures are -0.096 and -0.314, respectively.

(wages and land rents) they can obtain in the non-agricultural sector. We argue that, because of scale effects in land, market access generates heterogeneous effects on farmers with different land endowments. Farmers who initially hold less land face better outside options, which raises the threshold for remaining in agriculture. Landlords with more land consolidate their holdings and remain in the agricultural sector. This upward reallocation increases land inequality. We provide the detailed setup and derivation in Appendix F.

4.3.2 Non-agricultural Transformation

Our assessment regarding non-agricultural transformation is based on two pieces of quantitative evidence: the decline in the number of households owning agricultural land and the increase in the non-agricultural share.

[Table 6 is about here]

First, we present descriptive data supporting the divergence proposed above. We begin with calculations for Taiwan as a whole. We calculate the total number of households across 11 land size groups in 1921 and 1932. Table 6 reports the results. Negative contributions come from land size groups 1 through 5, whereas groups 6 through 11 make positive contributions. When calculating the total decrease and increase in land area using 1921 average weights, we find the net difference to be negligible relative to the benchmark. Based on this estimate, and in the absence of precise land transaction data, we infer that land lost by smallholders may have shifted upward.

One natural concern is that the number of smallholders decreased because they became larger landowners by acquiring more land rather than shifting to non-agricultural sectors. Table 6 shows that, although the total population increased from 1921 to 1932, the number of households classified as agricultural decreased by 62,242. Thus, some households were definitely no longer engaged in agriculture.

[Figure 5 is about here]

Figure 5a presents the more convincing results of village-level regressions for changes in each land size group from 1921 to 1932. In both the OLS and IV specifications, the point estimates for groups 2–3 are greater than zero and statistically significant. For groups 4–5, by contrast, the point estimates in the IV specification are less than zero. Although the wide confidence intervals prevent us from making any credible statistical inferences, this pattern indicates that some smallholders in groups 1–3 acquired more land and moved into groups 4–5 because of the heterogeneous effects of market access. This suggests that the "exit cutoff" in our model was likely between 2 and 3 *Jia* in 1932. Overall, both the descriptive and quantitative evidence indicates that the number of smallholders declines with market access.

Meanwhile, the 1932 SCLDM provides unique information on the number of households that owned land but were no longer engaged in agriculture, by land size group. We calculate the proportion of non-agricultural households among all landowning households and apply our OLS and IV specifications. Figure 5b reports the results. Surprisingly, for land size groups 1–8, market access consistently predicts a statistically significant increase in the non-agricultural share. For landowners with particularly large

holdings, however, this effect becomes negligible. For land size group 11 in particular, a positive estimate implies that greater market distance increased the non-agricultural share. The scale effect in our theoretical model can explain this pattern.

4.3.3 Commercialization

We believe that the rise in land inequality stems from the upward reallocation of land following the exit of smallholders. Our previous interpretation draws on historical narratives and describes the role of market access in the transition to non-agricultural production. Commercialization is the most intuitive explanation, although alternative explanations remain possible. However, we lack micro-level data on farm households and data on land transactions. Therefore, one alternative way to investigate the commercialization mechanism is to examine whether the effect of market access changes systematically with changing commercial conditions. During periods of transition, a better institutional environment may foster new commercial activities, while existing commercial barriers and a lack of unfilled market niches may limit them (Brandt and Sands, 1990; McMillan and Woodruff, 2002).

[Table 7 is about here]

Railways play a significant role in reducing transaction costs, facilitating market integration and other forms of commercial activity (Donaldson, 2018). Therefore, we use log distance to the nearest railway station as an inverse proxy for commercial conditions. In column (1) of Table 7, we examine the interaction between market distance and railway station distance. The results indicate that stronger existing commercial conditions weaken the market effect. However, given the strong endogeneity of railways, we interpret this result cautiously.

Another strategy is to examine the labor market. We use the distribution of occupations reported in the 1915 and 1920 censuses, the only two available sources. We calculate the share of the total population that was engaged in commerce and earned a living from it as the region's commercial intensity. Columns (2) and (3) report the estimates. Consistent with column (1), the results show that preexisting commercial conditions dampen the effect of market access. This pattern implies that households transitioning from agriculture to commerce faced commercial barriers, although the underlying mechanisms remain to be explored. In addition, the intercept term of commercial intensity is positively correlated with land inequality, providing indirect evidence for our conclusion.

For other forms of occupational intensity (agriculture and industry, etc.), we find no systematic differences in their interactions with market access (see Table G.12 of Appendix G.7). In conclusion, we believe that commercialization is a crucial mechanism underlying non-agricultural transformation.

4.3.4 Role of Landownership

Rising land inequality also depends on an important assumption: households could trade their land. The prevalence of tenancy in Taiwan provides a useful setting for a placebo test. We should not observe market access inducing transactions in land operated under tenancy.

[Table 8 is about here]

We focus on two types of outcomes that can be distinguished by ownership. First, we examine reductions in owner-cultivated and tenant-cultivated land. Columns (1) and (2) of Table 8 show that market access leads only to reductions in owner-cultivated land and has no significant effect on tenant-cultivated land. Second, we examine subsequent economic performance. Columns (3)–(5) report the shares of owner, mixed, and tenant households in debt. Market access is associated with a lower debt share among households with full landownership. This effect becomes statistically insignificant as the tenancy share increases. In conclusion, landownership is important to this transformation.

4.4 Additional Robustness Checks

In addition to the robustness results mentioned above, we conducted two additional checks to support our conclusions and identification strategies.

4.4.1 Falsification Tests

We conduct three falsification tests. First, following the spirit of Acemoglu et al. (2022), we conduct a falsification exercise in a subsample in which the proposed mechanism should be relatively weak. Specifically, we restrict the sample to villages in the bottom quartile of the infectious mortality distribution and reestimate our baseline specification.

Second, we use distance to public markets that became operational only after 1932 as a temporal placebo. Specifically, we calculate the log distance from each village to the nearest of the five public markets established between 1939 and 1945. We do not expect pesthouses to affect land inequality in 1932 through markets established after 1932.

Finally, we construct a non-infectious mortality measure using causes of death that are clearly not transmissible from person to person in the supplied classification: cancer and other malignant neoplasms; cerebral hemorrhage and cerebral softening; organic heart disease; chronic bronchitis; appendicitis and typhlitis; hernia and intestinal obstruction; cirrhosis of the liver; pregnancy- and childbirth-related conditions, excluding puerperal fever, which is classified separately; senility; deaths from external causes; suicide; beriberi, a nutritional-deficiency disease; and infantile convulsions and eclampsia. As with our infectious mortality measure, we normalize these deaths by the 1921 population. We then use this non-infectious mortality rate as an instrument in a TSLS regression. Because public market locations should be unrelated to non-infectious mortality, we do not expect this instrument to produce a credible estimate of the effect on land inequality.

All the TSLS point estimates from the falsification tests reported in Table G.13 of Appendix G.8 are statistically indistinguishable from zero (with small F-statistics).

4.4.2 Subsample Analysis

A potential concern is that a small number of villages with extreme characteristics may exert disproportionate influence on our estimates. We first assess this concern using a leave-one-village-out exercise.

Specifically, we sequentially exclude each of the 282 villages in the baseline estimation sample and re-estimate both the OLS and IV baseline specifications. Figure G.5 of Appendix G.9 reports the resulting coefficients and 90 percent confidence intervals. The estimates remain stable when any single village is omitted, indicating that the baseline results are not driven by one particularly influential observation.

We next impose more stringent and systematic sample restrictions designed to remove villages whose characteristics may make them especially influential. In columns (1)–(2) of Table G.14 of Appendix G.9, we exclude the administrative center of each prefecture. In columns (3)–(4), we exclude villages in the top decile of the 1931 population distribution (that is, observations above the 90th percentile). Finally, in columns (5)–(6), we remove villages in either tail of the market-distance distribution and retain only observations between the 10th and 90th percentiles of the market distance. The estimated effect of market distance remains negative and statistically distinguishable from zero across all six specifications. Thus, the qualitative conclusion is robust to each of these systematic sample restrictions.

5 Conclusion and Discussion

This paper studies how health-induced market access shaped the distribution of agricultural land in colonial Taiwan. Combining village-level landholding data for 1921 and 1932 with the locations of public markets, our identification strategy for addressing endogeneity relies on quasi-random exposure to infectious diseases and the exogenous bioclimatic variations. Our identification strategy employs three instruments: two capturing the epidemiological environment and one based on bioclimatic variation. We find that villages closer to a public market experienced greater land inequality by 1932.

One channel that we propose, but have not yet tested sufficiently directly, is that improved market access induced smallholders to sell their land and exit the agricultural sector. This mechanism is reflected in the conditional decline in landholdings and the conditional increase in the non-agricultural employment share among smallholders. Our heterogeneity analysis further shows that the distributive effects of markets were weaker in areas with more favorable preexisting commercial conditions. This pattern suggests that a transition from agriculture to commerce may provide a plausible explanation for our findings. Interestingly concluded, we may summarize that contagion led to the concentration of both people and land during the commercialization process.

Another issue that remains insufficiently investigated in our setting is land titling. Stable land rights can provide agricultural workers with greater security and better outside options, as reflected in their ability to migrate and transition into non-agricultural employment (Adamopoulos et al., 2024; De Janvry et al., 2015). Because of data limitations, we are unable to directly quantify the contribution of land titling to land inequality in colonial Taiwan. The market access mechanism examined in this study may therefore be understood as one intermediate link in the broader process connecting secure land rights to sectoral transformation (Fajgelbaum and Redding, 2022).

One further question that remains unexplored concerns the dimensions of inequality. As we observe, landholders optimized their decisions and consequently achieved better economic performance. This pattern suggests that a more unequal distribution of factors of production is not necessarily monotonically

associated with greater income or wealth inequality. One possible—but as yet untested—implication is that land reform may generate more favorable economic consequences when it corrects institutionally induced distortions in the distribution of land. By contrast, redistributive policies applied indiscriminately to distributional patterns may have more limited effects.

References

Acemoglu, Daron, Giuseppe De Feo, Giacomo De Luca, and Gianluca Russo. 2022. "War, Socialism, and the Rise of Fascism: An Empirical Exploration." *Quarterly Journal of Economics*, 137(2): 1233–1296.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369–1401.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, 117(4): 1231–1294.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review*, 95(3): 546–579.

Adamopoulos, Tasso, and Diego Restuccia. 2020. "Land Reform and Productivity: A Quantitative Analysis with Micro Data." *American Economic Journal: Macroeconomics*, 12(3): 1–39.

Adamopoulos, Tasso, Loren Brandt, Chaoran Chen, Diego Restuccia, and Xiaoyun Wei. 2024. "Land Security and Mobility Frictions." *Quarterly Journal of Economics*, 139(3): 1941–1987.

Albertus, Michael. 2023. "The Persistence of Rural Underdevelopment: Evidence from Land Reform in Italy." *Comparative Political Studies*, 56(1): 65–100.

Alesina, Alberto, and Dani Rodrik. 1994. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics*, 109(2): 465–490.

Alfani, Guido. 2015. "Economic Inequality in Northwestern Italy: A Long-Term View (Fourteenth to Eighteenth Centuries)." *Journal of Economic History*, 75(4): 1058–1096.

Alfani, Guido, and Francesco Ammannati. 2017. "Long-Term Trends in Economic Inequality: The Case of the Florentine State, c. 1300–1800." *Economic History Review*, 70(4): 1072–1102.

Alfani, Guido, and Tommy E. Murphy. 2017. "Plague and Lethal Epidemics in the Pre-Industrial World." *Journal of Economic History*, 77(1): 314–343.

Alfani, Guido, Matteo Di Tullio, and Mattia Fochesato. 2026. "Wealth Inequality and Epidemics in the Republic of Venice (1400–1800)." *Economic History Review*, 79(2): 811–849.

Alsan, Marcella, and Claudia Goldin. 2019. "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880–1920." *Journal of Political Economy*, 127(2): 586–638.

Andrade, Tonio. 2006. "The Rise and Fall of Dutch Taiwan, 1624–1662: Cooperative Colonization and the Statist Model of European Expansion." *Journal of World History*, 429–450.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101(5): 2003–2041.

Bai, Ying, and Ruixue Jia. 2016. "Elite Recruitment and Political Stability: The Impact of the Abolition of China's Civil Service Exam." *Econometrica*, 84(2): 677–733.

Banerjee, Abhijit, and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review*, 95(4): 1190–1213.

Becker, Sascha O., and Ludger Woessmann. 2009. "Was Weber Wrong?: A Human Capital Theory of Protestant Economic History." *Quarterly Journal of Economics*, 124(2): 531–596.

Bianchi-Vimercati, Riccardo, Giampaolo Lecce, and Matteo Magnaricotte. 2026. "Persistent Specialization and Growth: The Italian Land Reform." *Explorations in Economic History*, 100: 101739.

Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122(1): 73–117.

Brandt, Loren, and Barbara Sands. 1990. "Beyond Malthus and Ricardo: Economic Growth, Land Concentration, and Income Distribution in Early Twentieth-Century Rural China." *Journal of Economic History*, 50(4): 807–827.

Cameron, Lisa, and Manisha Shah. 2015. "Risk-Taking Behavior in the Wake of Natural Disasters." *Journal of Human Resources*, 50(2): 484–515.

Caprettini, Bruno, Lorenzo Casaburi, and Miriam Venturini. 2021. "Redistribution, Voting and Clientelism: Evidence from the Italian Land Reform." Centre for Economic Policy Research CEPR Discussion Paper 15679.

Cheng, Nora, Elliott Fan, and Tsong-Min Wu. 2022. "Sweet Unbinding: Sugarcane Cultivation and the Demise of Foot-Binding." *Journal of Development Economics*, 157: 102876.

Chen, Shuo, and Xiaohuan Lan. 2017. "There Will Be Killing: Collectivization and Death of Draft Animals." *American Economic Journal: Applied Economics*, 9(4): 58–77.

Chiang, Daniel Minghan, Elliott Fan, and Dexter Hsu. 2025. "A Taste of Communists' Own Medicine: The Political Consequences of Land Reforms in Japan and Taiwan." IZA Institute of Labor Economics IZA Discussion Paper 18095.

Chu, Wan-Wen. 2015. "Land Reform in Postwar Taiwan: Causes and Effects (in Chinese)." *Taiwan: A Radical Quarterly in Social Studies (in Chinese)*, , (98): 11–67. Article in Chinese.

Cinnirella, Francesco, and Erik Hornung. 2016. "Landownership Concentration and the Expansion of Education." *Journal of Development Economics*, 121: 135–152.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics*, 94(1): 260–272.

De Janvry, Alain, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet. 2015. "Delinking Land Rights from Land Use: Certification and Migration in Mexico." *American Economic Review*, 105(10): 3125–3149.

Dell, Melissa, and Benjamin A Olken. 2020. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java." *Review of Economic Studies*, 87(1): 164–203.

Diamond, Jared M. 1997. *Guns, Germs, and Steel: The Fate of Human Societies.* New York:W.W. Norton & Co.

DiTraglia, Francis J., and Camilo García-Jimeno. 2021. "A Framework for Eliciting, Incorporating, and Disciplining Identification Beliefs in Linear Models." *Journal of Business & Economic Statistics*, 39(4): 1038–1053.

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

Easterly, William. 2007. "Inequality Does Cause Underdevelopment: Insights from a New Instrument." *Journal of Development Economics*, 84(2): 755–776.

Engerman, Stanley L, and Kenneth L Sokoloff. 2002. "Factor Endowments, Inequality, and Paths of Development Among New World Economies." National Bureau of Economic Research NBER Working Paper 9259. NBER Working Paper No. 9259.

Fajgelbaum, Pablo, and Stephen J Redding. 2022. "Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914." *Journal of political economy*, 130(5): 1249–1318.

Fan, Xinyu, and Lingwei Wu. 2023. "The Shaping of a Gender Norm: Marriage, Labor, and Foot-Binding in Historical China." *International Economic Review (Philadelphia)*, 64(4): 1819–1850.

Felbermayr, Gabriel, and Jasmin Gröschl. 2014. "Naturally Negative: The Growth Effects of Natural Disasters." *Journal of Development Economics*, 111: 92–106.

Frankema, Ewout. 2010. "The Colonial Roots of Land Inequality: Geography, Factor Endowments, or Institutions?" *Economic History Review*, 63(2): 418–451.

Galor, Oded, Omer Moav, and Dietrich Vollrath. 2009. "Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence." *Review of Economic Studies*, 76(1): 143–179.

Gensini, Gian Franco, Magdi H. Yacoub, and Andrea A. Conti. 2004. "The Concept of Quarantine in History: From Plague to SARS." *Journal of Infection*, 49(4): 257–261.

Ghatak, Maitreesh, and Dilip Mookherjee. 2025. "Misallocating Misallocation?" *Annual Review of Economics*, 17.

Greif, Avner, Joel Mokyr, and Guido Tabellini. 2025. *Two Paths to Prosperity: Culture and Institutions in Europe and China, 1000–2000.* Princeton, NJ:Princeton University Press.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "Does Local Financial Development Matter?" *Quarterly Journal of Economics*, 119(3): 929–969.

Hao, Yu, Ningning Ma, Tianyang Xi, and Wuyue You. 2026. "Selecting Loyal Agents: Land Reform as a Party-Building Strategy in Post-Revolutionary China." SSRN Working Paper No. 6144010.

Heldring, Leander, and James A. Robinson. 2012. "Colonialism and Economic Development in Africa." National Bureau of Economic Research NBER Working Paper 18566. NBER Working Paper No. 18566.

Hollingsworth, Alex, Krzysztof Karbownik, Melissa A Thomasson, and Anthony Wray. 2024. "The Gift of a Lifetime: The Hospital, Modern Medicine, and Mortality." *American Economic Review*, 114(7): 2201–2238.

Ho, Samuel Pao-San. 1968. "Agricultural Transformation under Colonialism: The Case of Taiwan." *Journal of Economic History*, 28(3): 313–340.

Ho, Yhi-Min. 1971. "On Taiwan's Agricultural Transformation under Colonialism: A Critique." *Journal of Economic History*, 31(3): 672–681.

Hsu, Chiung-feng. 2020. "The Battle of Suppression and Epidemics: Infectious Diseases and Public Health in Taiwan, 1895–1945 (in Chinese)." *National Library of Public Information Website*, Accessed: April 29, 2026.

Huang, Ching-I, and Shao-Yu Jheng. 2021. "Inclusive Origins of Rapid Industrialization: The Persistent Effects of the Colonial Bank Networks on Taiwan's Economic Miracle." SSRN Working Paper No. 3982154.

Iyer, Lakshmi. 2010. "Direct Versus Indirect Colonial Rule in India: Long-Term Consequences." *Review of Economics and Statistics*, 92(4): 693–713.

Jedwab, Remi, Noel D. Johnson, and Mark Koyama. 2022. "The Economic Impact of the Black Death." *Journal of Economic Literature*, 60(1): 132–178.

Kung, James Kai-sing, Xiaogang Wu, and Yuxiao Wu. 2012. "Inequality of Land Tenure and Revolutionary Outcome: An Economic Analysis of China's Land Reform of 1946–1952." *Explorations in Economic History*, 49(4): 482–497.

Kusunose, Noboru. 1915. *Fish and Vegetable Markets in Taiwan (in Japanese).* Taihoku (Taipei):Taiwan Sōtokufu Minseibu [Bureau of Productive Industries, Civil Affairs Department, Office of the Governor-General of Taiwan]. Digitized by Keio University, June 4, 2009.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review*, 45(1): 1–28.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter. 2022. "Valid T-Ratio Inference for IV." *American Economic Review*, 112(10): 3260–3290.

Lin, Justin Yifu. 1990. "Collectivization and China's Agricultural Crisis in 1959–1961." *Journal of Political Economy*, 98(6): 1228–1252.

Lipton, Michael. 2009. *Land Reform in Developing Countries: Property Rights and Property Wrongs. Routledge priorities in development economics*, London;:Routledge.

McMillan, John, and Christopher Woodruff. 2002. "The Central Role of Entrepreneurs in Transition Economies." *Journal of Economic Perspectives*, 16(3): 153–170.

McNeill, William Hardy. 1998. *Plagues and Peoples.* New York:Anchor Books.

Mokyr, Joel, Assaf Sarid, and Karine van der Beek. 2022. "The Wheels of Change: Technology Adoption, Millwrights, and the Persistence of Britain's Industrialization." *Economic Journal (London)*, 132(645): 1894–1926.

Mora, Camilo, Tristan McKenzie, Isabella M. Gaw, Jacqueline M. Dean, Hannah von Hammerstein, Tabatha A. Knudson, Renee O. Setter, Charlotte Z. Smith, Kira M. Webster, Jonathan A. Patz, et al. 2022. "Over Half of Known Human Pathogenic Diseases Can Be Aggravated by Climate Change." *Nature Climate Change*, 12(9): 869–875.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance. Political economy of institutions and decisions.*, Cambridge:Cambridge University Press.

Nunn, Nathan, and Nancy Qian. 2010. "The Columbian Exchange: A History of Disease, Food, and Ideas." *Journal of Economic Perspectives*, 24(2): 163–188.

Ogata, Masanori. 1897. "Ueber die Pestepidemie in Formosa." *Zbl. Bakt.*, 21: 769–777.

Olson, Mancur. 2000. *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships.* New York:Basic Books.

Persson, Torsten, and Guido Tabellini. 1994. "Is Inequality Harmful for Growth?" *American Economic Review*, 84(3): 600–621.

Rodrik, Dani, Gene Grossman, and Victor Norman. 1995. "Getting Interventions Right: How South Korea and Taiwan Grew Rich." *Economic Policy*, 10(20): 53–108.

Sachs, Jeffrey D. 2001. "Tropical Underdevelopment." National Bureau of Economic Research NBER Working Paper 8119. NBER Working Paper No. 8119.

Stenseth, Nils Chr., Noelle I. Samia, Hildegunn Viljugrein, Kyrre Linné Kausrud, Mike Begon, Stephen Davis, Herwig Leirs, V. M. Dubyanskiy, Jan Esper, Vladimir S. Ageyev, et al. 2006. "Plague Dynamics Are Driven by Climate Variation." *Proceedings of the National Academy of Sciences*, 103(35): 13110–13115.

Sunagawa, Haruhiko. 2018. "Study on the Planning and Development of Modern Markets in Taiwan during the Japanese Colonial Period: Focusing on Public Markets in Local Cities in Southern Taiwan (in Japanese)." *Journal of Architecture and Planning (Transactions of AIJ)*, 83(746): 805–812.

Taiwan Provincial Literature Committee. 1972. *Gazetteer of Taiwan Province: Volume 3, Political Affairs, Part 1, Health (in Chinese).* Taichung:Taiwan Provincial Literature Committee.

Tseng, Wen-Liang. 2015. "The Arrangement of Land Relations and Its Impact on Taiwan under Japanese Rule, 1895–1905 (in Chinese)." *Cheng Kung Journal of Historical Studies (in Chinese)*, 49: 257–314. Article in Chinese.

Wang, Jen-Kuan, and Oliver Kim. 2024. "Land Reform in Taiwan, 1950–1961: Effects on Agriculture and Structural Change." SSRN Working Paper No. 4951831.

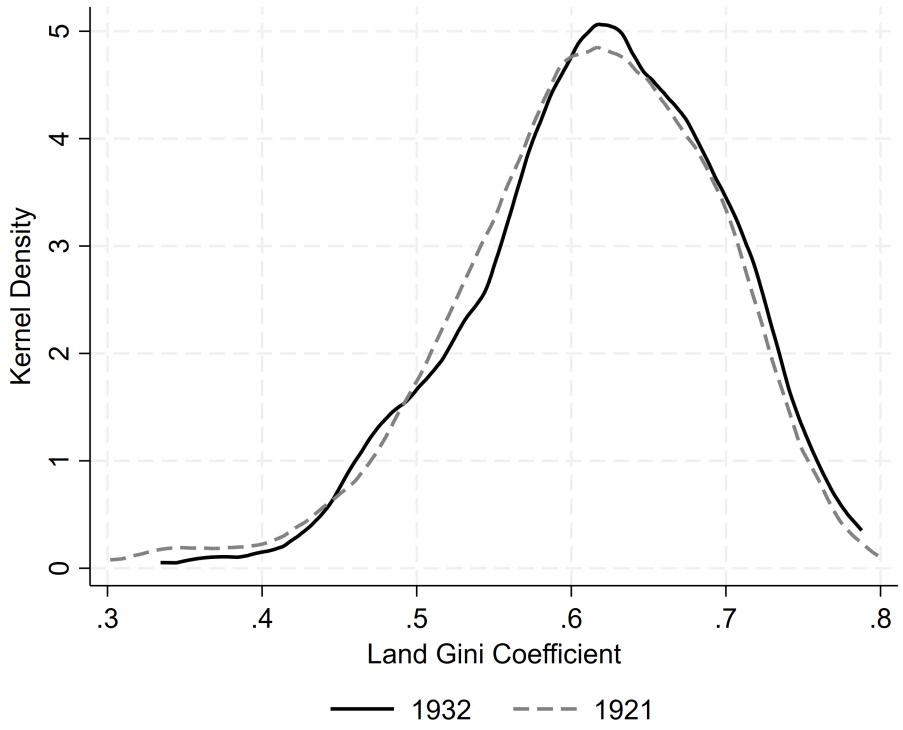
Weber, Max. 1904. "The Protestant Ethic and the Spirit of Capitalism (in German)." *Archive for Social Science and Social Policy (in German)*, 20: 1–54. Continued in volume 21 (1905), pp. 1–110.

Wu, Mi-cha. 2017. "The Launch and Significance of the Taiwan Governor-General's "Land Survey Project (1895–1905)" (in Chinese)." *Journal of Taiwan History (in Chinese)*, 10: 5–35.

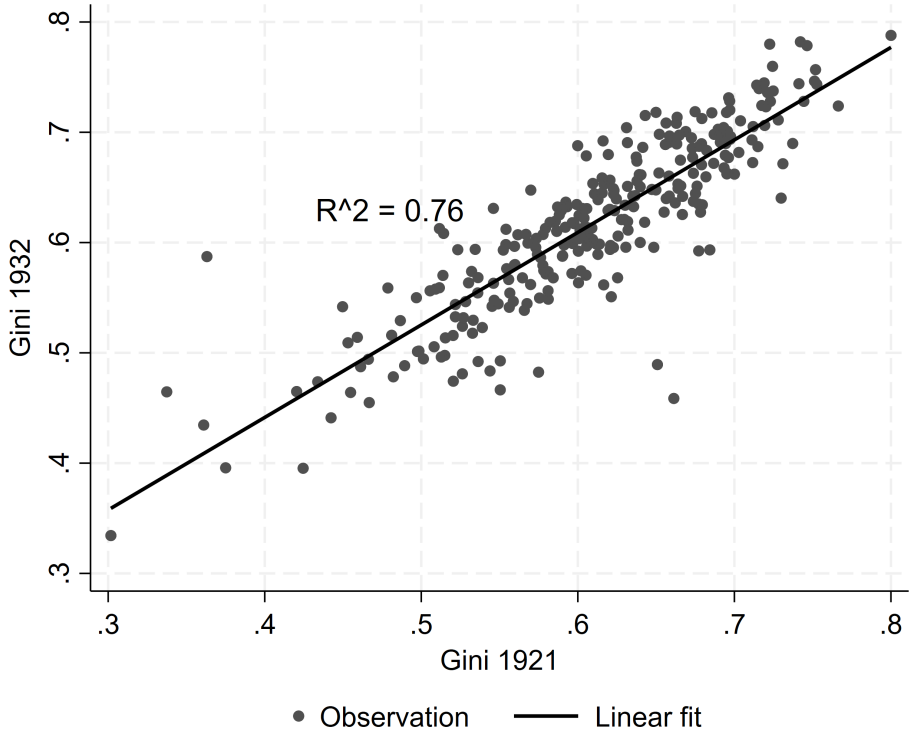
Xu, Lei, Leif C. Stige, Herwig Leirs, Simon Neerinckx, Kenneth L. Gage, Ruifu Yang, Qiyong Liu, Barbara Bramanti, Katharine R. Dean, Hui Tang, et al. 2019. "Historical and Genomic Data Reveal the Influencing Factors on Global Transmission Velocity of Plague during the Third Pandemic." *Proceedings of the National Academy of Sciences*, 116(24): 11833–11838.

Yanaihara, Tadao. 1988. *Taiwan under Imperialism (in Japanese).* . Reprint ed., Tokyo:Iwanami Shoten. Originally published in 1929.

6 Tables and Figures

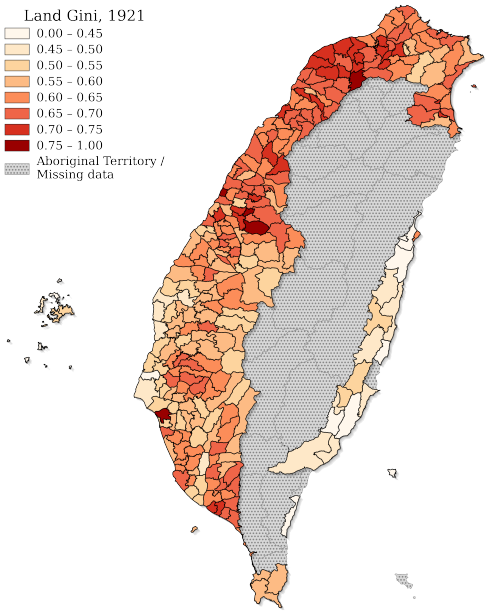


(a) Kernel Density of Land Gini (1932–1921)

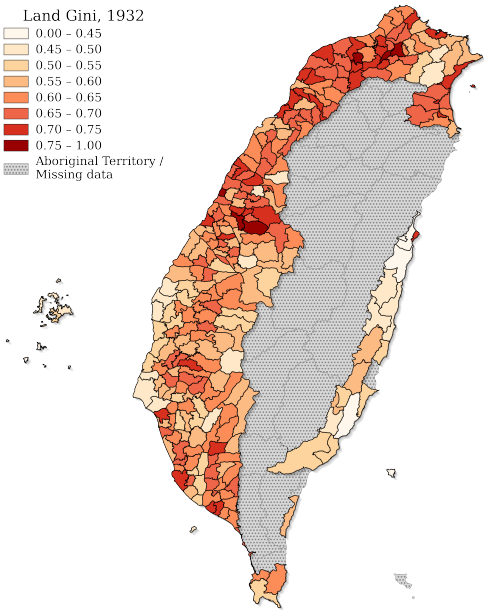


(b) Correlation between the 1932 and 1921 Gini Coefficients

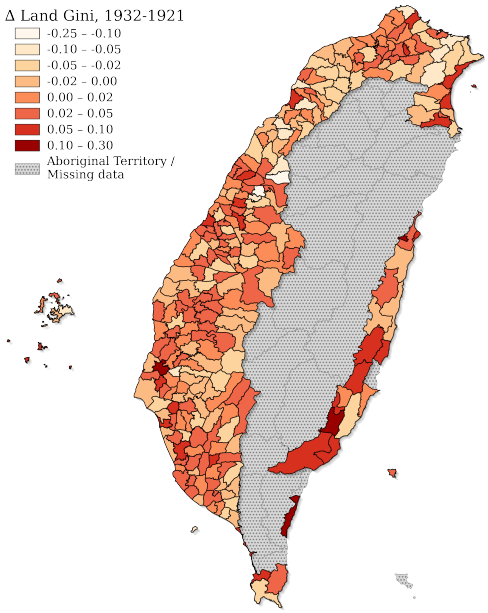
Figure 1: Kernel Density and Path Dependence of Land Inequality



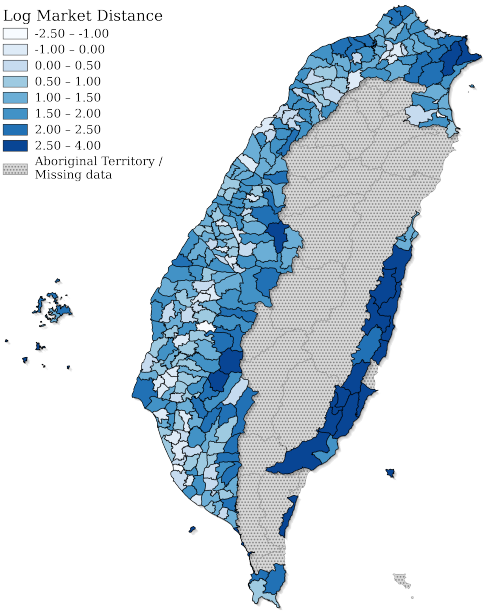
(a) Land Gini, 1921



(b) Land Gini, 1932

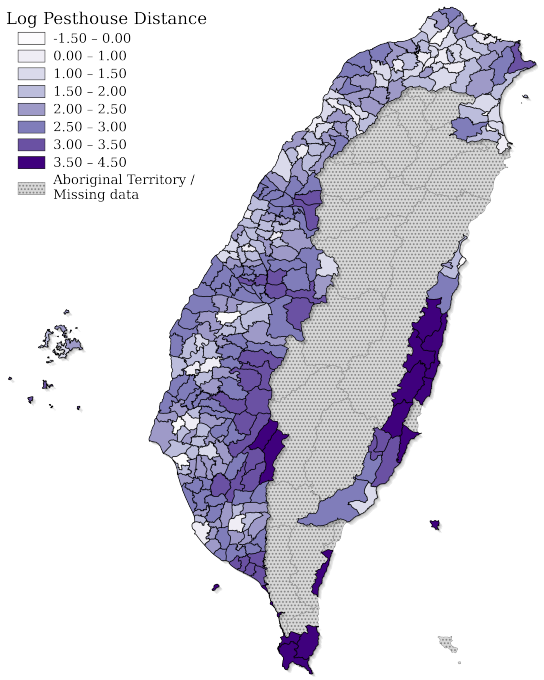


(c) Δ Land Gini, 1932–1921

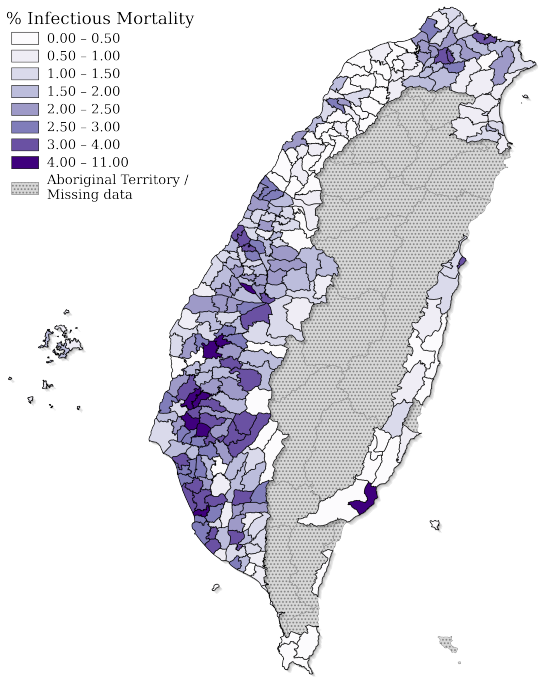


(d) Log Market Distance

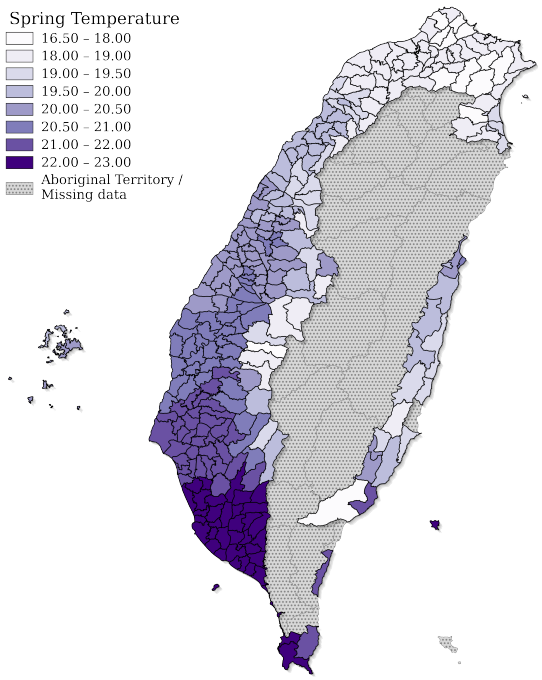
Figure 2: Spatial Distribution of Land Inequality and Market Distance



(a1) Log Pesthouse Distance



(a2) % Infectious Mortality



(b) Spring Temperature

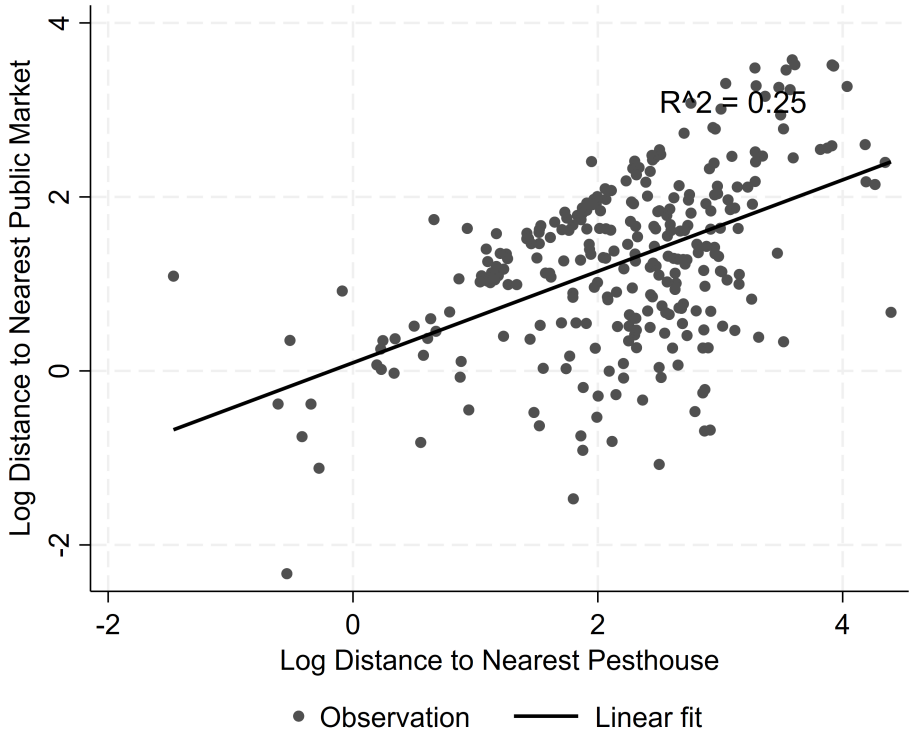
Figure 3: Spatial Distribution of Instrumental Variables

Note: Panels (a1) and (a2) represent the *epidemiological environment*, measured by log distance to the nearest pesthouse and the percentage of infectious mortality, respectively. Panel (b) represents the *bioclimatic condition*, measured by spring temperature.

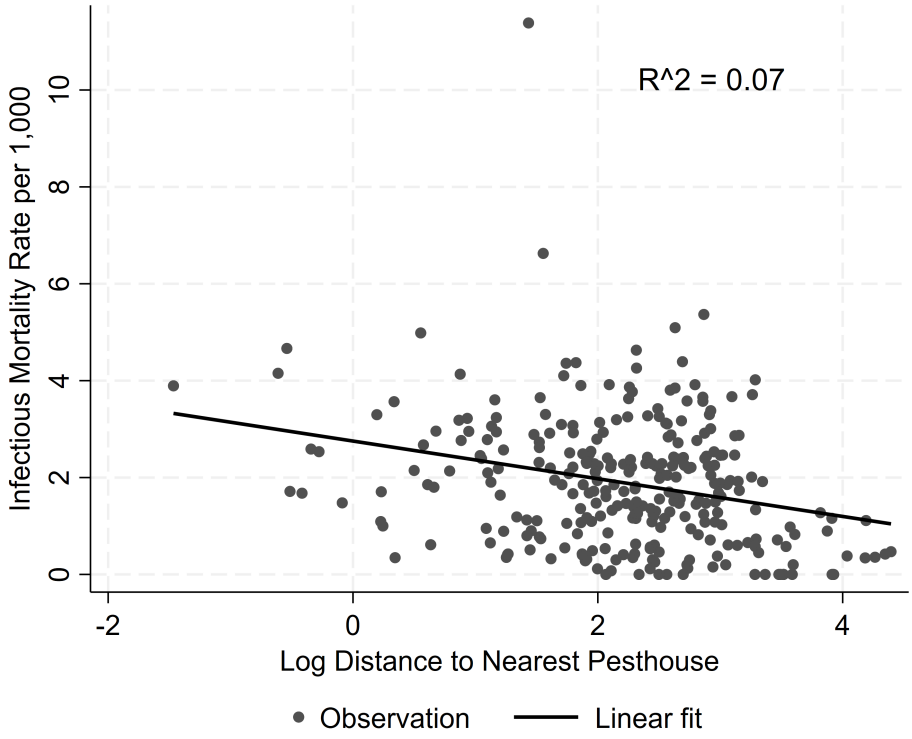
Table 1: OLS Evidence: Market Distance and Land Inequality

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Land Gini, 1932				
Log Market Distance	-0.0214***	-0.0169***	-0.0174***	-0.0167***	-0.0078**
	(0.0047)	(0.0046)	(0.0045)	(0.0046)	(0.0030)
Land Gini, 1921					0.8174***
					(0.0359)
Log Population, 1931		0.0174**	0.0186**	0.0146*	-0.0017
		(0.0080)	(0.0078)	(0.0085)	(0.0047)
% Female, 1931				-0.1636	-0.2073*
				(0.1111)	(0.1196)
% Mainland Origin, 1926				-0.0380	-0.0024
				(0.0507)	(0.0314)
% Migration, 1931				0.0180	-0.0620
				(0.0775)	(0.0445)
Constant	0.6425***	0.4726***	0.4624***	0.6089***	0.2469**
	(0.0074)	(0.0782)	(0.0758)	(0.1310)	(0.0930)
Prefecture FE	Y	Y	Y	Y	Y
Population Weighted			Y	Y	Y
Baseline Controls				Y	Y
Initial Gini					Y
Adjusted R-squared	0.3728	0.3820	0.3801	0.3798	0.7790
Observations	284	284	284	284	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.



(a) Public Markets and Pesthouses



(b) Infectious Mortality and Pesthouses

Figure 4: Pesthouses and Infectious Mortality as Measures

Table 2: Balance Tests of the Pesthouse

Balanced Type of	Population Structure			State Capability	
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Δ % Birth, 1921-1931	Δ % Death, 1921-1931	Log Δ Population, 1921-1931	Political Center Dummy	# Monopoly Institution, 1921
Log Pesthouse Distance	0.0001	0.0009*	-0.0049	0.0095	0.0272
	(0.0004)	(0.0005)	(0.0061)	(0.0113)	(0.0436)

Balanced Type of	State Capability				
	(6)	(7)	(8)	(9)	(10)
Dependent Variable	# Military Region, 1921	# Police Station, 1921	# Police Station per Taiwanese, 1921	# Japanese Shrine in 1900s	Log Foreigner Population, 1920
Log Pesthouse Distance	-0.0579	0.0817	0.0023	-0.0205	-0.0979
	(0.1523)	(0.1402)	(0.0055)	(0.0751)	(0.0859)

Balanced Type of	Economic Development				
	(11)	(12)	(13)	(14)	(15)
Dependent Variable	# Port, 1871-1895	# Bank, 1921	# Court, 1904	# Prison, 1904	# Ancestral Hall, 1904
Log Pesthouse Distance	0.1002	-0.0135	0.0084	-0.0025	-1.0698
	(0.0654)	(0.0183)	(0.0211)	(0.0162)	(0.6613)

Balanced Type of	Economic Development		Human Capital		
	(16)	(17)	(18)	(19)	(20)
Dependent Variable	# Church, 1904	% Unbound Women, 1920	# Confucian Academy, 1693-1895	# Public School, 1898-1911	% Literacy, 1920
Log Pesthouse Distance	-0.0417	-0.0204	-0.0589	-0.1275	-0.0011
	(0.0404)	(0.0161)	(0.0786)	(0.0964)	(0.0009)
Prefecture FE	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: IV Evidence: Estimation by Pesthouse with Additional Controls

Controlled for	Baseline	Baseline	Public Health	Geographical Location	Historical Agricultural Shock	Agricultural Endowment	Concluding Controls
Estimator	RF	TSLS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	Panel A: Land Gini, 1932						
Log Market Distance		-0.0257***	-0.0224***	-0.0244***	-0.0255***	-0.0270***	-0.0212**
		(0.0067)	(0.0079)	(0.0073)	(0.0069)	(0.0077)	(0.0094)
Log Pesthouse Distance	-0.0107***						
	(0.0027)						
Log Watersource Distance			-0.0009				-0.0002
			(0.0037)				(0.0039)
Log Hospital Distance			-0.0032				-0.0040
			(0.0036)				(0.0035)
Longitude				-0.0063			-0.0054
				(0.0231)			(0.0230)
Latitude				0.0126			0.0103
				(0.0115)			(0.0116)
Earthquake Intensity					0.0003		0.0007
					(0.0009)		(0.0010)
Sugarcane Suitability						0.0001	0.0001
						(0.0001)	(0.0002)
Rice Suitability						0.0002	0.0001
						(0.0005)	(0.0005)
Dependent Variable	Panel B: Log Market Distance (First-stage)						
Log Pesthouse Distance		0.4152***	0.3359***	0.3904***	0.4073***	0.3932***	0.3055***
		(0.0727)	(0.0681)	(0.0706)	(0.0733)	(0.0726)	(0.0670)
Prefecture FE	Y	Y	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y	Y	Y	Y
First-stage F-stat.		45.74	26.22	38.62	43.01	39.57	20.07
Adjusted R-squared	0.7826	0.6196	0.6347	0.6251	0.6195	0.6109	0.6363
Observations	282	282	282	282	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Infectious Mortality Rate as an Alternative Instrument and Overidentification Test

Estimator	OLS	RF	IV: Mortality	IV: Mortality	IV: Pesthouse	IV: Pesthouse & Mortality
			TSLS			
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	% Infectious Mortality	Panel A: Land Gini, 1932				
Log Market Distance			-0.0276**	-0.0282*	-0.0254***	-0.0263***
			(0.0117)	(0.0148)	(0.0075)	(0.0067)
% Infectious Mortality		0.0059**			0.0005	
		(0.0025)			(0.0029)	
Log Pesthouse Distance	-0.2528**			0.0010		
	(0.0966)			(0.0064)		
Dependent Variable		Panel B: Log Market Distance (First-stage)				
% Infectious Mortality			-0.2155***	-0.1711***	-0.1711***	-0.1711***
			(0.0547)	(0.0475)	(0.0475)	(0.0475)
Log Pesthouse Distance				0.3720***	0.3720***	0.3720***
				(0.0676)	(0.0676)	(0.0676)
Prefecture FE	Y	Y	Y	Y	Y	Y
Population Weighted		Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
Initial Gini		Y	Y	Y	Y	Y
First-stage F-stat.			24.29	16.89	37.72	32.67
Hansen J p-value						0.8677
Adjusted R-squared	0.3130	0.7792	0.6087	0.6039	0.6198	0.6161
Observations	283	282	282	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

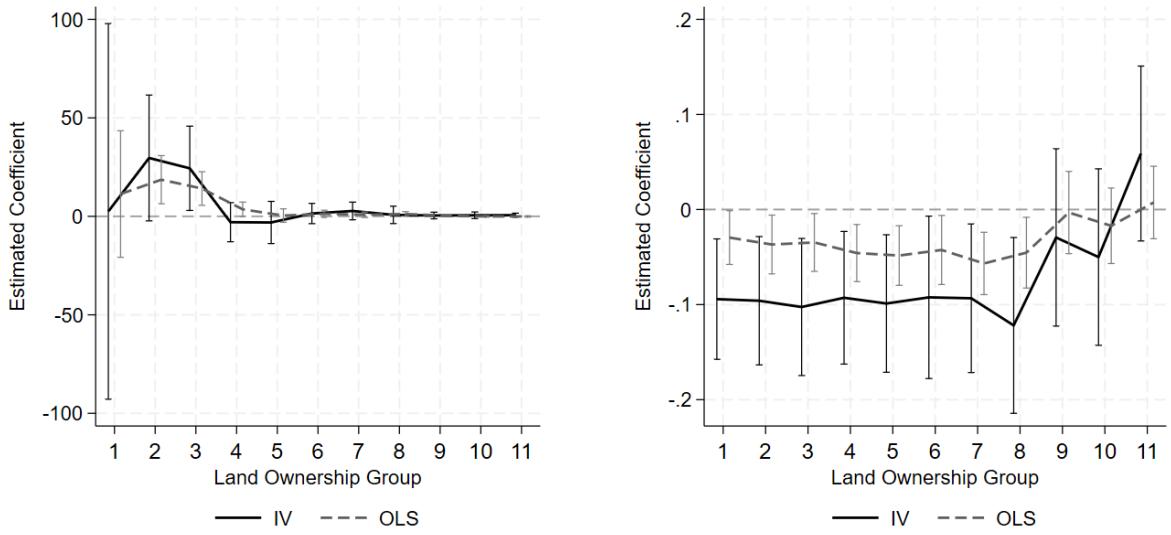
Table 5: Spring Temperature Differences during the Plague Period as an Exogenous Variation

Instrument Period	1895–1918					1895–1910			
Avg. Temperature during	Mar.–Apr.			Feb.–Apr.	Mar.–May	Mar.–Apr.		Feb.–Apr.	Mar.–May
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	Panel A: Land Gini, 1932								
Log Market Distance	-0.018	-0.064*	-0.037*	-0.037	-0.040**	-0.063*	-0.036*	-0.036	-0.039**
	(0.026)	(0.036)	(0.020)	(0.022)	(0.019)	(0.035)	(0.020)	(0.022)	(0.018)
	[0.500]	[0.078]	[0.071]	[0.101]	[0.036]	[0.081]	[0.073]	[0.102]	[0.037]
Longitude		0.969	0.324	0.324	0.336	0.964	0.321	0.323	0.333
		(0.933)	(0.586)	(0.587)	(0.600)	(0.927)	(0.581)	(0.585)	(0.596)
Latitude		4.424	1.583	1.585	1.634	4.406	1.570	1.578	1.620
		(4.703)	(2.952)	(2.956)	(3.025)	(4.674)	(2.932)	(2.947)	(3.003)
Longitude × Latitude		-0.036	-0.013	-0.013	-0.013	-0.036	-0.013	-0.013	-0.013
		(0.039)	(0.024)	(0.024)	(0.025)	(0.039)	(0.024)	(0.024)	(0.025)
Log Coast Distance		-0.008	-0.005	-0.005	-0.006	-0.007	-0.005	-0.005	-0.006
		(0.012)	(0.007)	(0.008)	(0.007)	(0.012)	(0.007)	(0.008)	(0.007)
Log Port Distance		0.004	-0.003	-0.003	-0.002	0.004	-0.003	-0.003	-0.002
		(0.011)	(0.007)	(0.008)	(0.007)	(0.011)	(0.007)	(0.007)	(0.007)
River Density		-0.885	-0.689	-0.692	-0.790	-0.837	-0.663	-0.680	-0.763
		(1.107)	(0.575)	(0.631)	(0.550)	(1.093)	(0.565)	(0.625)	(0.540)
Elevation		-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dependent Variable	Panel B: Log Market Distance (First-stage)								
Spring Temperature	-0.301***	-0.398***	-0.375***	-0.350***	-0.406***	-0.400***	-0.377***	-0.351***	-0.409***
	(0.081)	(0.127)	(0.134)	(0.131)	(0.137)	(0.127)	(0.134)	(0.131)	(0.137)
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Initial Gini			Y	Y	Y		Y	Y	Y
Geographical Controls		Y	Y	Y	Y	Y	Y	Y	Y
First-stage F-stat.	13.94	9.86	7.78	7.16	8.75	9.99	7.93	7.23	8.92
Observations	284	283	282	282	282	283	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Changes in Total Households by Type, 1921–1932

		# Total Households		Contribution to Aggregate Change			
Type	Range	1921	1932	# Change	Weight	# Change×Weight	Subtotal
Negative contributions							
1	0-0.5	172,605	131,869	-40,736	0.237	-9,673.788	
2	0.5-1	86,681	71,353	-15,328	0.720	-11,038.933	
3	1-2	70,540	64,077	-6,463	1.412	-9,126.628	
4	2-3	28,364	27,639	-725	2.460	-1,783.565	
5	3-5	23,270	22,767	-503	3.801	-1,912.113	
							-33,535.027
Positive contributions							
6	5-7	8,980	9,172	192	5.806	1,114.712	
7	7-10	5,903	6,201	298	8.264	2,462.804	
8	10-20	5,447	5,849	402	13.484	5,420.764	
9	20-30	1,354	1,593	239	24.360	5,822.139	
10	30-50	837	1,067	230	37.854	8,706.476	
11	50-100	377	529	152	67.737	10,296.085	
							33,822.980
Total Households		404,358	342,116	-62,242	*Total Land Area Change*		287.953
Total Population		(3,835,811)	(4,929,962)				



(a) Δ Households with Land, 1921–1932

(b) % Non-Agricultural, 1932

Figure 5: Market Effects on Economic Structure by Landholding Group

Note: These figures report group-level estimates for changes in households with land between 1921 and 1932 and the non-agricultural household share in 1932.

Table 7: Commercial Conditions and Barriers to Transformation

	(1)	(2)	(3)
Dependent Variable		Land Gini, 1932	
Log Market Distance	0.0021	-0.0082**	-0.0084**
	(0.0044)	(0.0038)	(0.0040)
Log Market Distance × Log Railway Station Distance	-0.0043**		
	(0.0019)		
Log Market Distance × Commercial Intensity, 1915		0.0367*	
		(0.0184)	
Log Market Distance × Commercial Intensity, 1920			0.0489*
			(0.0267)
Log Railway Station Distance	-0.0033		
	(0.0031)		
Commercial Intensity, 1915		0.1178**	
		(0.0457)	
Commercial Intensity, 1920			0.1385**
			(0.0637)
Prefecture FE	Y	Y	Y
Population Weighted	Y	Y	Y
Baseline Controls	Y	Y	Y
Initial Gini	Y	Y	Y

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Role of Landownership

	Reduction in Landholdings		Economic Performance		
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Δ Owner Cultivated Land, 1921-1932	Δ Tenant Cultivated Land, 1921-1932	% Owner Households in Debt	% Mixed Households in Debt	% Tenant Households in Debt
Log Market Distance	0.2105**	0.0816	0.0790**	0.0376	-0.0040
	(0.0896)	(0.1087)	(0.0390)	(0.0302)	(0.0309)
Prefecture FE	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A Infectious Diseases and Public Health

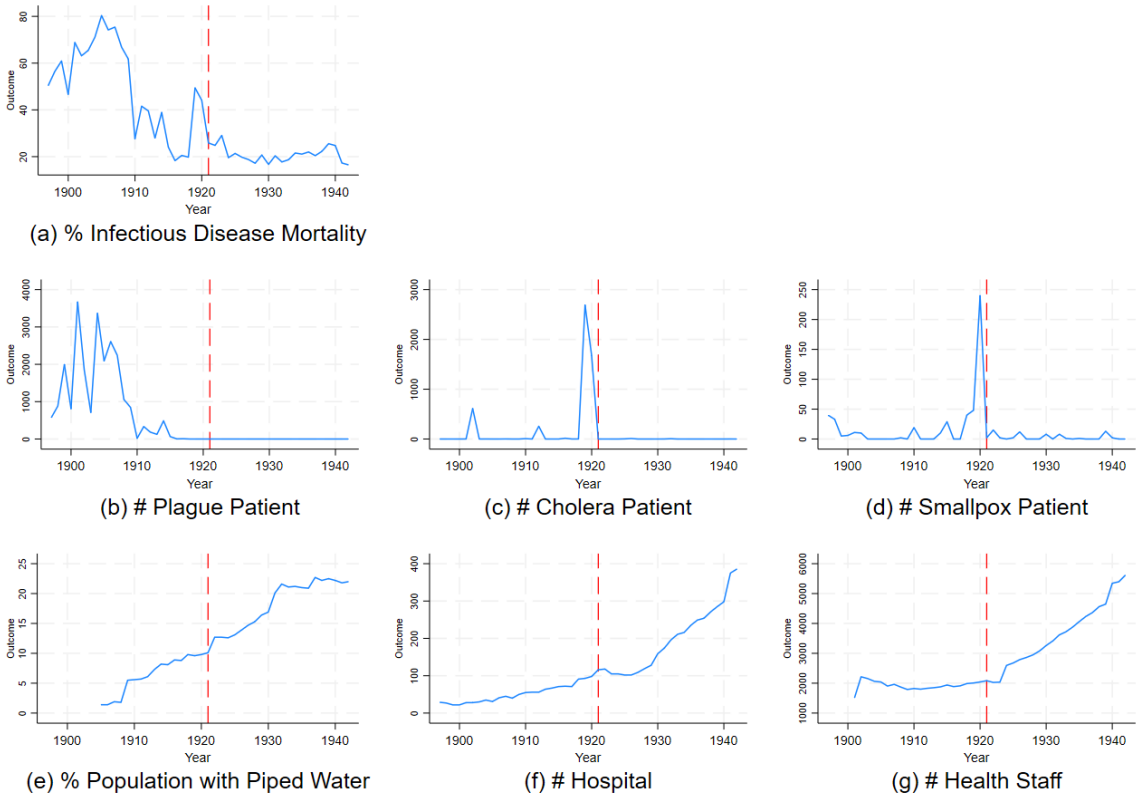


Figure A.1: Public Health Indicators and Infectious Diseases

Source: Taiwan Historical Statistics.

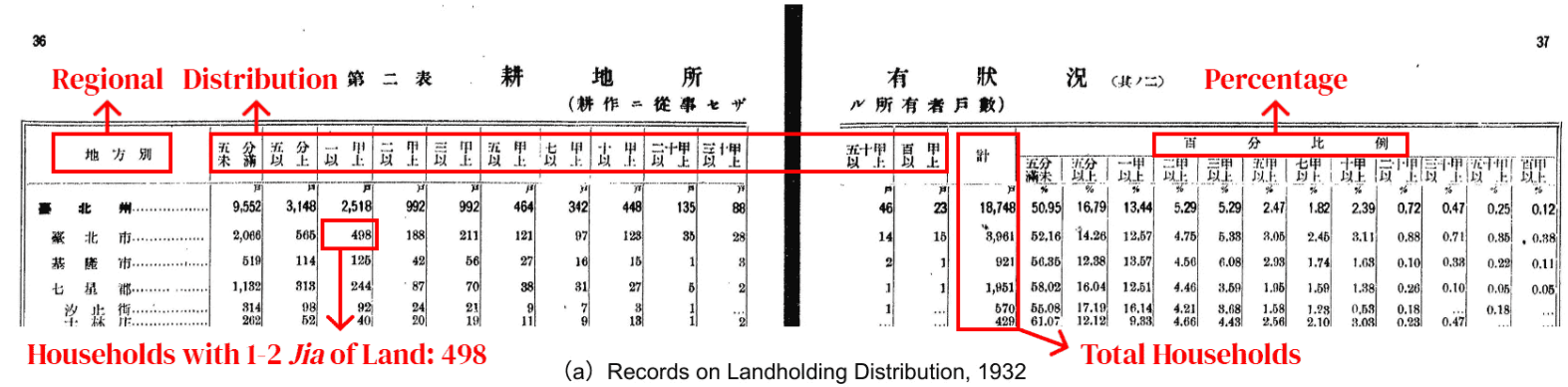
Appendix B Data Information



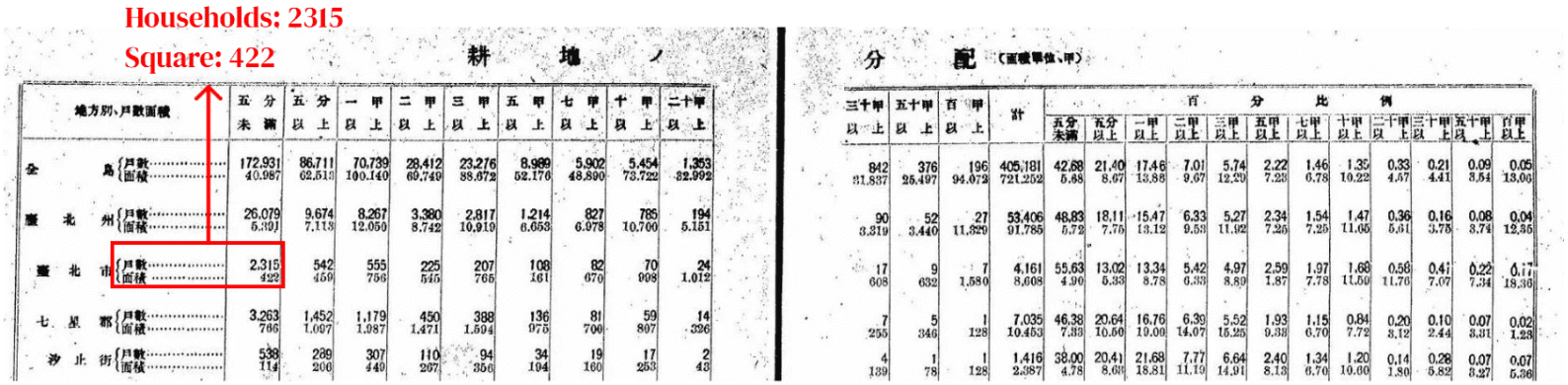
Figure B.1: A Sample Public-Market Location

Note: This figure shows the location of the Lugang public market established in 1910.

Source: Academia Sinica



(a) Records on Landholding Distribution, 1932



(b) Records on Landholding Distribution, 1921

Figure B.2: Source of Land Inequality

Note: This figure shows the source of land inequality. Information regarding the range of land area owned by households is available for the first 11 categories.

Source: Bureau of Productive Industries, Government-General of Taiwan

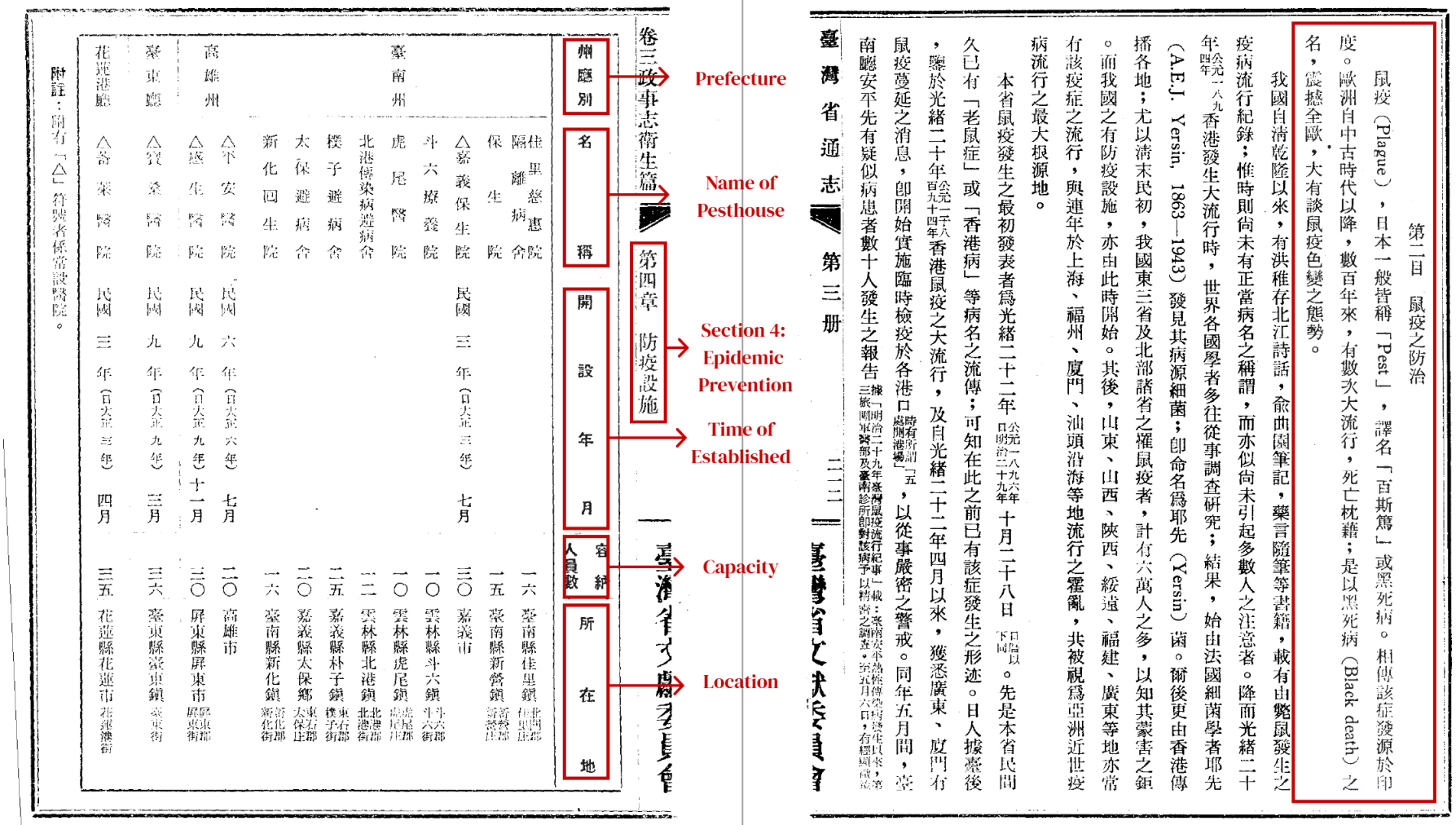


Figure B.3: Details about Pesthouses

Note: This figure provides details about pesthouses.

Source: Taiwan Provincial Literature Committee (1972)

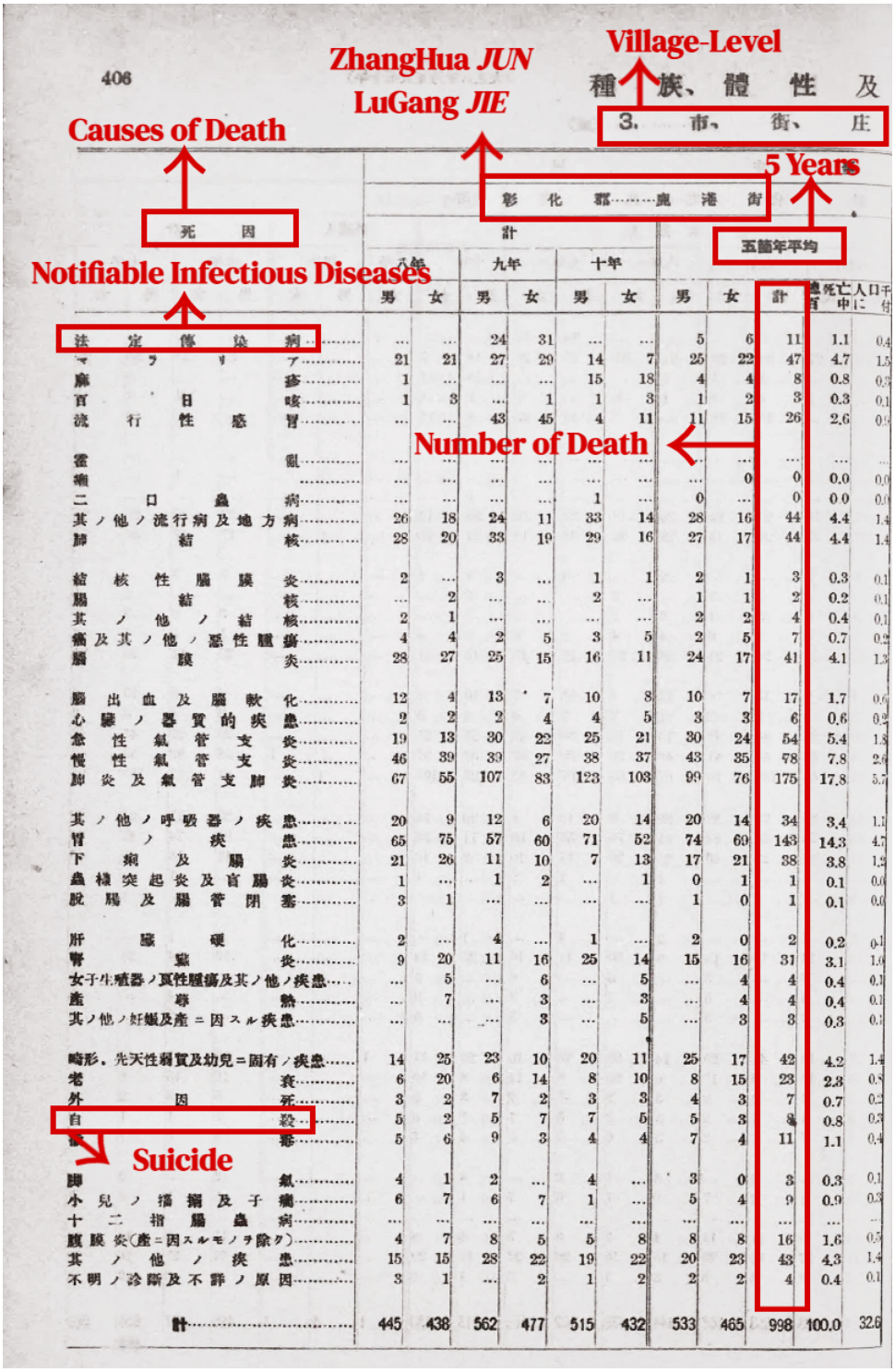


Figure B.4: Source of Infectious Mortality

Note: This figure shows how we extracted the infectious mortality data from the historical archives.
Source: Online Library

Appendix C Administrative Divisions of Taiwan

Although Taiwan was a prefecture within Imperial China, its administrative divisions were defined as "states." During the Japanese colonial period, Taiwan's administrative divisions underwent numerous adjustments. These adjustments were frequent in the early years but became relatively stable after 1920.

Based on historical census data and other surveys, we can compile a comprehensive record of these divisions in Table C.1. Broadly speaking, the administrative structure was organized into four tiers: first-, second-, third-, and fourth-level administrative units. We designate "prefectures" and "villages" to represent the first-level administrative units (8 in total) and the third-level administrative units (278 + 7 in total), respectively. The second- and fourth-level administrative units are defined as "cities" and "streets".

Table C.1: Administrative Units in Taiwan under Japanese Rule (1906–1940)

Pre-1920 System					Post-1920 System			
Unit Name	1906	1908	1909	1919	Unit Name	1920	1930	1940
ting	20	20	12	12	*zhou*	5	5	5
					ting	2	3	3
zhiting	83	83	87	85	*zhiting*	6	10	
					shi	3	7	11
					jun	47	45	51
bao	109	109	109	109	*jie*	35	34	56
li	75	75	75	75	*zhuang*	228	225	211
yu	13	13	13	13	*qu*	18	19	
xiang	5	5	5	5				
jie	76	76	73	69	*dazi*	2932		2947
zhuang	2736	2738	2734	2700	*ding*	31		221
she	118	118	114	53				
xiang	79	79	79	79				
cun				19				

Notes: All Chinese administrative terms are reported in italicized pinyin to preserve the original institutional hierarchy.

The statistical data we use are readily available at the village level. Very few observations are recorded for fourth-level administrative units, while most statistical indicators are aggregated at the prefecture level.

Appendix D Construction of Village-Level Land Inequality

We construct village-level land inequality measures for years $t \in \{1921, 1932\}$ using grouped landholding data. Let H_{ijt} denote the number of households in village i belonging to landholding category j in year t, where landholding size is divided into eleven ordered groups. Let $\bar{m}_{ij}^{1921}$ denote the adjusted average landholding size for group j in village i, which is fixed at its 1921 value and applied to both years.

Total households and total landholdings in village i at time t are given by:

$$P_{it} = \sum_{j=1}^{11} H_{ijt}, \tag{(D.1)}$$

$$L_{it} = \sum_{j=1}^{11} H_{ijt} \bar{m}_{ij}^{1921}. \tag{(D.2)}$$

We then compute the cumulative population and land shares:

$$C_{ijt}^{P} = \sum_{k=1}^{j} \frac{H_{ikt}}{P_{it}}, \tag{(D.3)}$$

$$C_{ijt}^{L} = \sum_{k=1}^{j} \frac{H_{ikt} \bar{m}_{ik}^{1921}}{L_{it}}. \tag{(D.4)}$$

The area under the Lorenz curve is approximated using the trapezoidal rule:

$$A_{it} = \frac{1}{2} \sum_{j=1}^{11} \left(C_{ijt}^{P} - C_{i,j-1,t}^{P} \right) \left(C_{ijt}^{L} + C_{i,j-1,t}^{L} \right), \tag{(D.5)}$$

where $C_{i0t}^{P} = C_{i0t}^{L} = 0$.

The village-level Gini coefficient in year t is defined as:

$$G_{it} = 1 - 2A_{it}. \tag{(D.6)}$$

Because landholding data are available only in grouped form, inequality measures are reconstructed using a discrete Lorenz-curve approximation. Fixing $\bar{m}_{ij}$ at its 1921 value ensures comparability across years by holding the within-group land distribution constant over time. This procedure implicitly assumes within-group homogeneity and therefore yields a lower-bound estimate of true inequality.

Appendix E Temperature Construction

We build monthly historical temperature for Taiwanese villages from two climate datasets. The first dataset is the NOAA–CIRES–DOE Twentieth Century Reanalysis, version 3 (20CRv3). It reports ensemble-mean monthly air temperature at 2 meters above the surface. Its grid has a resolution of $1^\circ \times 1^\circ$. We download the cells around Taiwan from January 1895 to December 2010.[30] We use 1895–1918 as the historical study period. We use 1981–2010 to calculate the monthly reference temperature.

The 20CRv3 grid is too coarse for village-level analysis. Many villages are located in the same 20CRv3 cell. These villages can still have different temperatures because they have different elevations and terrain. We therefore add the CHELSA version 2.1 monthly temperature climatology. CHELSA provides monthly means for 1981–2010 on a grid of about 1 km.[31] CHELSA provides the fine spatial baseline. It does not provide the historical variation for 1895–1918.

E.1 Anomaly-Based Spatial Downscaling

Let T_{gym}^{20CR} be the 20CRv3 temperature in grid cell g, year y, and month m. We first calculate the historical temperature anomaly. We subtract the 1981–2010 mean for the same calendar month:

$$A_{gym}^{20CR} = T_{gym}^{20CR} - \frac{1}{30} \sum_{s=1981}^{2010} T_{gsm}^{20CR}. \tag{(E.1)}$$

The anomaly measures how much a historical month differs from the 1981–2010 mean for that month. We do not assign the coarse 20CRv3 value directly to a village. Let p denote a CHELSA pixel. We use bilinear interpolation, denoted by $\mathcal{I}(\cdot)$, to move the anomaly from the coarse grid to pixel p:

$$\widetilde{A}_{pym}^{20CR} = \mathcal{I}\left(A_{gym}^{20CR}; p\right). \tag{(E.2)}$$

Let C_{pm}^{CHELSA} be the CHELSA 1981–2010 mean for pixel p and month m. We then add the interpolated anomaly to the CHELSA baseline:

$$\widehat{T}_{pym} = C_{pm}^{CHELSA} + \widetilde{A}_{pym}^{20CR}. \tag{(E.3)}$$

This calculation keeps the fine spatial pattern from CHELSA. The historical change comes from 20CRv3. For example, a CHELSA value of 18°C becomes 19.2°C when the interpolated 20CRv3 anomaly is +1.2°C.

E.2 Aggregation from Fine-Grid Pixels to Villages

We overlay the reconstructed temperature grid with the historical village polygons. Let a_{vp} be the area shared by village v and pixel p. We calculate the village temperature as an area-weighted mean:

[30] Data could be found in https://www.psl.noaa.gov/data/gridded/data.20thC_ReanV3.html, accessed on August 15 2026.

[31] Data could be found in https://www.chelsa-climate.org/datasets/chelsa_climatologies, accessed on August 17 2026.

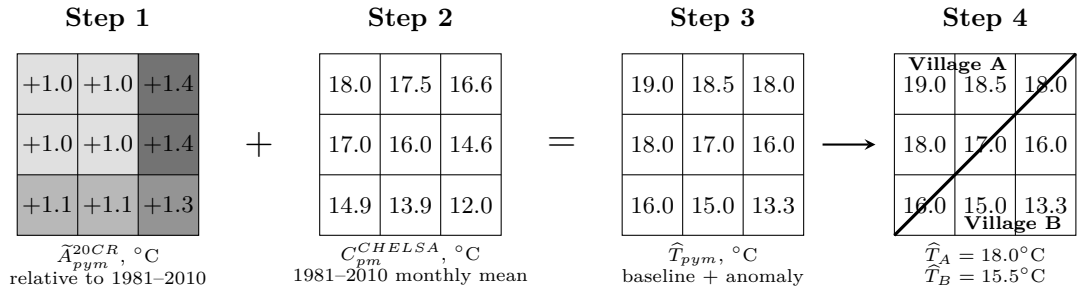


Figure E.1: Example of Construction of Village-Level Temperature

$$\widehat{T}_{vym} = \frac{\sum_p a_{vp} \widehat{T}_{pym}}{\sum_p a_{vp}}. \tag{(E.4)}$$

Step 4 in Figure E.1 shows a simple example. The diagonal divides the 3×3 grid into two villages. Village A is above the diagonal. It contains three full pixels and three half pixels. Its temperature is

$$\widehat{T}_A = \frac{19.0 + 18.5 + 18.0 + 0.5(18.0 + 17.0 + 16.0)}{3 + 0.5 + 0.5 + 0.5} = 18.0°\text{C}. \tag{(E.5)}$$

Village B contains the remaining area. It also contains three full pixels and three half pixels. Its temperature is

$$\widehat{T}_B = \frac{16.0 + 15.0 + 13.3 + 0.5(18.0 + 17.0 + 16.0)}{3 + 0.5 + 0.5 + 0.5} = 15.5°\text{C}. \tag{(E.6)}$$

This method produces one temperature value for each village, year, and month from 1895 to 1918. These values are downscaled estimates. They are not direct 1-km observations from the historical period. The method assumes that the fine spatial pattern in the 1981–2010 CHELSA climatology is a useful baseline for earlier years. It also assumes that 20CRv3 captures the historical monthly changes.

Appendix F Model

Consider a household i endowed with one unit of labor and an initial landholding $s_i > 0$. Let m denote market access. Better market access raises the non-agricultural wage $w(m)$, with $w'(m) > 0$, and lowers the iceberg trade cost $\tau(m)$, with $\tau'(m) < 0$. The farm-gate output price is therefore

$$P(m) \equiv p\big[1 - \tau(m)\big], \qquad P'(m) > 0. \tag{(F.1)}$$

To keep the model transparent, take the land rental rate $\bar{R} > 0$ as exogenously given.

If the household remains active in agriculture, it chooses operated land $T_i \geq 0$ and agricultural labor $\ell_i \in [0, 1]$. Its period payoff is

$$\begin{aligned} V_i^{\text{Agr}}(T_i, \ell_i; s_i, m) = \underbrace{P(m) A T_i^{\eta} \ell_i^{\alpha}}_{\text{farm income}} + \underbrace{w(m)(1 - \ell_i)}_{\text{non-ag earnings}} + \underbrace{\bar{R}(s_i - T_i)}_{\text{net rental income}} \\ - \underbrace{F}_{\text{fixed operating cost}} - \underbrace{\frac{\kappa}{2}\big(T_i - s_i\big)^2}_{\text{adjustment cost}}, \end{aligned} \tag{(F.2)}$$

where $0 < \alpha < 1$, $1 < \eta < 2$, $F > 0$, and $\kappa > 0$. A household that exits agriculture supplies all labor to the non-agricultural sector and rents out all land, yielding

$$V_i^{\text{Nonagr}}(s_i; m) = w(m) + \bar{R} s_i. \tag{(F.3)}$$

Subtracting the common outside-option terms, the gain from remaining in agriculture is

$$\begin{aligned} \Delta_i(T_i, \ell_i; s_i, m) &\equiv V_i^{\text{Agr}}(T_i, \ell_i; s_i, m) - V_i^{\text{Nonagr}}(s_i; m) \\ &= P(m) A T_i^{\eta} \ell_i^{\alpha} - w(m)\ell_i - \bar{R} T_i - F - \frac{\kappa}{2}\big(T_i - s_i\big)^2. \end{aligned} \tag{(F.4)}$$

This makes clear that the decision to stay in agriculture depends on whether farm operating surplus is large enough to cover the opportunity cost of labor, the rental cost of operated land, and the fixed cost of remaining active.

For a given operated area T_i, the interior labor allocation solves

$$\ell_i^*(T_i, m) = \left[\frac{\alpha P(m) A T_i^{\eta}}{w(m)}\right]^{\frac{1}{1-\alpha}}. \tag{(F.5)}$$

Hence the threshold operated scale above which the household becomes a full-time farmer is

$$\bar{T}(m) = \left[\frac{w(m)}{\alpha P(m) A}\right]^{\frac{1}{\eta}}. \tag{(F.6)}$$

When $T_i \geq \bar{T}(m)$, the unconstrained optimum would imply $\ell_i^* \geq 1$, so the household chooses $\ell_i = 1$.

Result 1 (Smallholder exit). To isolate the occupation margin from the land-reallocation margin, first consider a household that continues to operate its initial landholding, so $T_i = s_i$. Define

$$G(s_i, m) \equiv \max_{\ell_i \in [0,1]} \left\{ P(m) A s_i^{\eta} \ell_i^{\alpha} - w(m)\ell_i - \bar{R} s_i - F \right\}. \tag{(F.7)}$$

At the optimum, the derivative of the objective with respect to ℓ_i is zero, so a marginal change in s_i affects G only through the direct effect on farm revenue; the induced change in the optimal labor choice is second-order and drops out. Therefore,

$$G_s(s_i, m) = \eta P(m) A s_i^{\eta-1} \ell_i^{\alpha} - \bar{R}. \tag{(F.8)}$$

Thus, for households near the exit cutoff, larger initial holdings make remaining in agriculture more attractive whenever $G_s(s_i, m) > 0$.

If labor supply is interior at $T_i = s_i$, then

$$\ell_i^*(s_i, m) = \left[\frac{\alpha P(m) A s_i^{\eta}}{w(m)} \right]^{\frac{1}{1-\alpha}}, \tag{(F.9)}$$

and exactly the same logic applies when differentiating with respect to market access: we differentiate the objective while holding ℓ_i at its optimum because the indirect effect through $d\ell_i^*/dm$ vanishes by the first-order condition. Hence

$$\begin{aligned} G_m(s_i, m) &= P'(m) A s_i^{\eta} \left[\ell_i^*(s_i, m) \right]^{\alpha} - w'(m) \ell_i^*(s_i, m) \\ &= \ell_i^*(s_i, m) \left[\frac{w(m)}{\alpha} \frac{P'(m)}{P(m)} - w'(m) \right]. \end{aligned} \tag{(F.10)}$$

Therefore, a sufficient condition for improved market access to reduce the continuation value of marginal smallholders is

$$\frac{w'(m)}{w(m)} > \frac{1}{\alpha} \frac{P'(m)}{P(m)}. \tag{(F.11)}$$

When the growth rate of non-agricultural wages exceeds $1/\alpha$ times the growth rate of farm-gate prices, smallholders tend to exit agriculture.

For sufficiently large holdings, the household is full-time in agriculture and $\ell_i = 1$, so

$$G(s_i, m) = P(m) A s_i^{\eta} - w(m) - \bar{R} s_i - F, \qquad G_m(s_i, m) = A s_i^{\eta} P'(m) - w'(m). \tag{(F.12)}$$

Hence market access hurts small farms through the outside-option channel, but it benefits sufficiently large farms once the scale-dependent revenue effect dominates the wage effect. If the cutoff $s^*(m)$ is defined by

$$G\big(s^*(m), m\big) = 0, \tag{(F.13)}$$

then total differentiation implies

$$\frac{ds^*(m)}{dm} = -\frac{G_m\big(s^*(m), m\big)}{G_s\big(s^*(m), m\big)}. \tag{(F.14)}$$

Whenever the marginal household satisfies $G_m\big(s^*(m), m\big) < 0$, the cutoff rises with market access. Better market access therefore induces additional smallholder exit.

Result 2 (Land demand among surviving farmers). Conditional on remaining active and choosing full-time farming, the operated-land problem is

$$\max_{T_i \geq 0} \left\{ P(m) A T_i^{\eta} - \bar{R} T_i - F - \frac{\kappa}{2}\big(T_i - s_i\big)^2 \right\}. \tag{(F.15)}$$

For an interior solution, the first-order condition is

$$\eta P(m) A \big(T_i^*\big)^{\eta-1} = \bar{R} + \kappa\big(T_i^* - s_i\big). \tag{(F.16)}$$

Differentiating this condition with respect to market access gives

$$\frac{\partial T_i^*}{\partial m} = \frac{\eta A \big(T_i^*\big)^{\eta-1} P'(m)}{\kappa - \eta(\eta-1) P(m) A \big(T_i^*\big)^{\eta-2}}, \tag{(F.17)}$$

Whenever the denominator is positive, $\partial T_i^*/\partial m > 0$: better market access raises the return to operated land and stimulates land demand among larger farmers.

The same first-order condition also implies

$$\frac{\partial T_i^*}{\partial s_i} = \frac{\kappa}{\kappa - \eta(\eta-1) P(m) A \big(T_i^*\big)^{\eta-2}} > 0. \tag{(F.18)}$$

Conditional on remaining active, households with larger initial holdings optimally operate more land.

Result 3 (Upward land reallocation). Taken together, the model delivers a partial equilibrium reallocation result. Better market access raises the exit cutoff for marginal smallholders and increases land demand among surviving large farmers. Because total land is fixed, land released by smaller households is absorbed by larger producers.

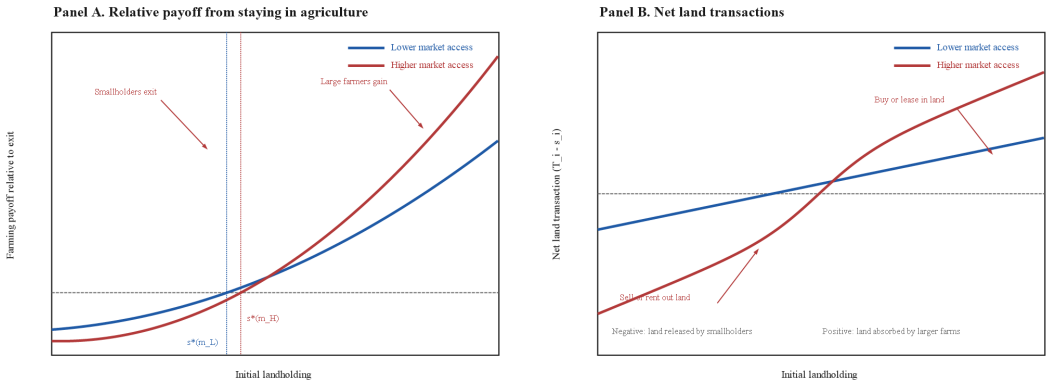


Figure F.1: Conceptual Illustration of Market Access and Upward Land Reallocation

Notes: Panel A plots the gain from remaining in agriculture, $G(s_i, m)$, against initial landholdings. Better market access shifts the curve downward for marginal smallholders because the outside-wage effect dominates there, but upward for sufficiently large farmers because the scale-dependent revenue effect becomes stronger at larger operated scales. Panel B plots net land transactions, $T_i^* - s_i$. Negative values indicate land rented out or sold; positive values indicate land rented in or purchased. Better market access therefore implies both more smallholder exit and stronger land accumulation among the larger surviving farmers.

Appendix G Supplementary Results

G.1 Lorenz Curve

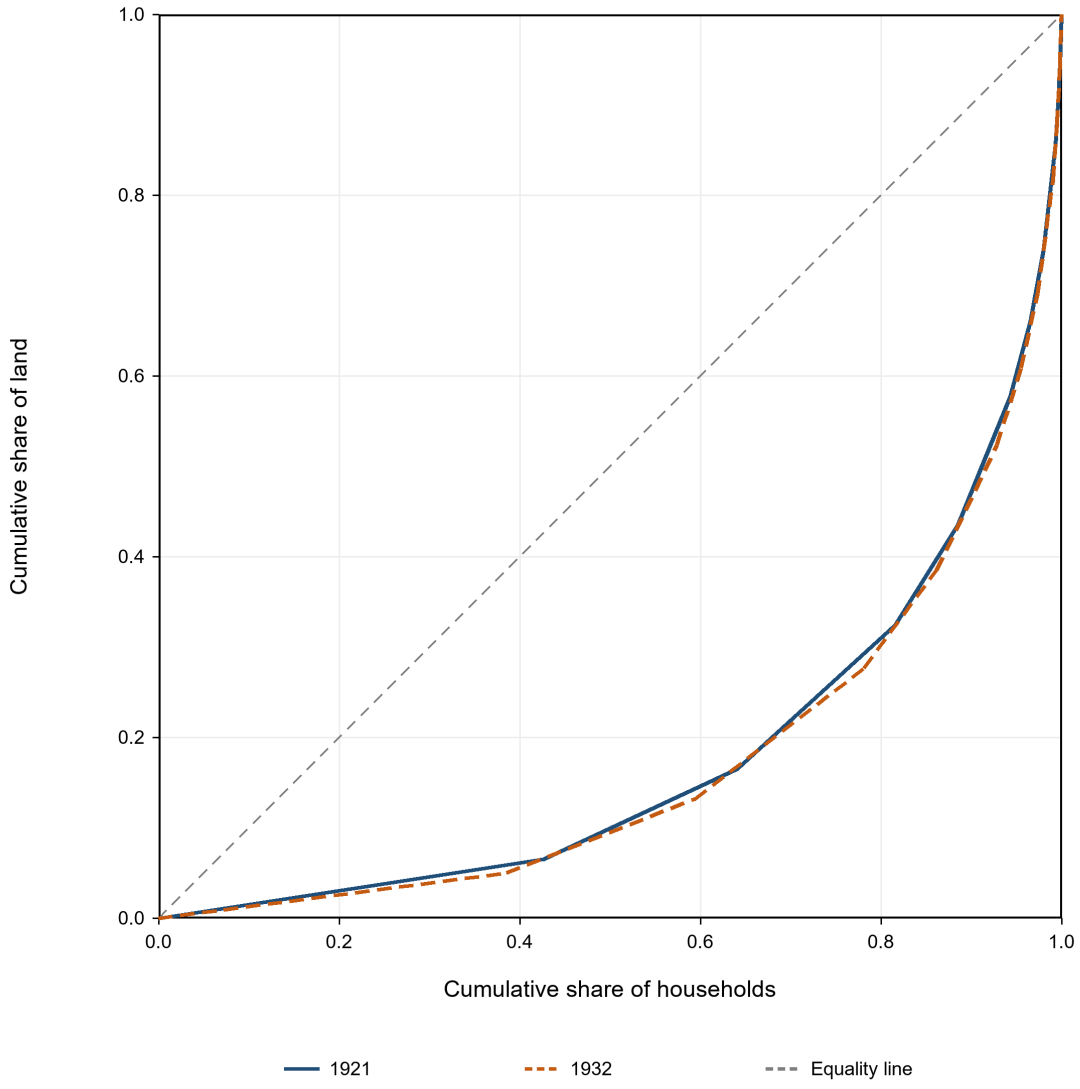


Figure G.1: Lorenz Curve

G.2 Alternative Measures of Land Inequality

We use the adjusted-weight Gini coefficient as our baseline measure of land distribution and inequality. We further consider other inequality measures to provide robustness checks and additional evidence. First, we use the mean of each landholding-size interval as a fixed weight for all villages, labeled "raw." We also consider three alternative inequality measures: the Theil, Robin Hood, and Atkinson indices. Table G.1 reports the means and standard deviations of these measures across 282 villages in 1921 and 1932. For every measure, the mean in 1932 is higher than that in 1921. Table G.2 reports the correlation matrix for seven measures, including our baseline Gini coefficient. All measures are highly correlated.

Table G.1: Alternative Measures of Land Inequality: Descriptive Statistics

	1921		1932	
Measurement	Mean	S.D.	Mean	S.D.
Gini (raw)	0.6094	0.0047	0.6213	0.0044
Gini	0.6074	0.0052	0.6160	0.0048
Theil	0.8106	0.0133	0.8387	0.0137
Robin Hood	0.4686	0.0038	0.4753	0.0039
Atkinson ($\epsilon = 0.25$)	0.1800	0.0027	0.1864	0.0028
Atkinson ($\epsilon = 0.50$)	0.3199	0.0043	0.3313	0.0044
Atkinson ($\epsilon = 0.75$)	0.4274	0.0052	0.4426	0.0052

Table G.2: Correlation Matrix of 1932 Land-Inequality Measures

Measurement	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Gini (raw)	1.000	0.958	0.957	0.984	0.979	0.992	0.994
(2) Gini	0.958	1.000	0.914	0.957	0.937	0.949	0.949
(3) Theil	0.957	0.914	1.000	0.955	0.994	0.978	0.953
(4) Robin Hood	0.984	0.957	0.955	1.000	0.975	0.983	0.979
(5) Atkinson 0.25	0.979	0.937	0.994	0.975	1.000	0.994	0.979
(6) Atkinson 0.50	0.992	0.949	0.978	0.983	0.994	1.000	0.995
(7) Atkinson 0.75	0.994	0.949	0.953	0.979	0.979	0.995	1.000

We next replace our baseline Gini coefficient with these alternative measures and repeat the baseline OLS and IV estimations. The results are reported in Table G.3. Our findings remain qualitatively unchanged across all specifications. Additionally, the estimates using the Theil index should be interpreted with caution because we lack an upper bound and further information for the largest landholding group, Group 12. Within Groups 1-11, the results remain sensitive to upper-tail holdings; they do not identify concentration in the omitted open-ended group. Furthermore, the estimated effect based on the Atkinson index becomes monotonically stronger as ϵ increases. This pattern suggests that the effect of market access on land inequality is stronger when more weight is placed on small landholders. Together, these two findings point to a pattern of "upward reallocation."

Building on the findings in Table G.1 and our baseline regression specification, we provide further quantitative evidence on the trend toward greater inequality. We replace the dependent variable with two outcomes: the difference between the 1932 and 1921 Gini coefficients, and an indicator equal to one if this difference is positive. The results are reported in Table G.4. Whether or not we control for initial

Table G.3: Robustness Estimates with Alternative Land-Inequality Measures

Dependent Variable	Baseline OLS	Baseline IV
Gini (raw), 1932	-0.0093***	-0.0248***
	(0.0034)	(0.0077)
Gini, 1932	-0.0078**	-0.0257***
	(0.0030)	(0.0067)
Theil, 1932	-0.0252**	-0.0813**
	(0.0117)	(0.0314)
Robin Hood, 1932	-0.0092***	-0.0233***
	(0.0028)	(0.0073)
Atkinson ($\epsilon = 0.25$), 1932	-0.0050**	-0.0158***
	(0.0022)	(0.0057)
Atkinson ($\epsilon = 0.50$), 1932	-0.0081**	-0.0243***
	(0.0033)	(0.0082)
Atkinson ($\epsilon = 0.75$), 1932	-0.0099**	-0.0280***
	(0.0038)	(0.0090)

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

inequality in 1921, the coefficient on market distance remains negative and statistically significant.

Table G.4: Market and the Rise of Land Inequality

	(1)	(2)	(3)	(4)
Dependent Variable	Δ Land Gini		$\mathbf{1}\{\Delta\text{Land Gini} > 0\}$	
Log Market Distance	-0.0066*	-0.0093***	-0.0736**	-0.0838**
	(0.0033)	(0.0034)	(0.0346)	(0.0348)
Gini, 1921		-0.2184***		-0.8229*
		(0.0540)		(0.4675)
Log Population, 1931	-0.0100	-0.0052	0.0156	0.0336
	(0.0075)	(0.0056)	(0.0658)	(0.0661)
% Female, 1931	-0.3154*	-0.2802*	-0.4292	-0.2966
	(0.1888)	(0.1430)	(0.9724)	(0.9522)
% Mainland Origin, 1926	0.0125	0.0022	0.2937	0.2546
	(0.0367)	(0.0355)	(0.3023)	(0.3022)
% Net Migration, 1931	-0.0939*	-0.0632	0.3979	0.5136
	(0.0520)	(0.0468)	(0.4179)	(0.4075)
Constant	0.2547*	0.3380***	0.5127	0.8263
	(0.1512)	(0.1253)	(0.9042)	(0.8415)
Prefecture FE	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y
Initial Gini		Y		Y
Adjusted R-squared	0.1359	0.2049	0.0427	0.0474
Observations	282	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

G.3 Alternative Measures, Sensitivity, and Placebos for Market Proximity

We further consider the specification and measurement of our core independent variable. In our baseline specification, we take the logarithm of market distance to reduce its right skewness. Column (1) of Table G.5 removes this logarithmic transformation, and the result remains robust. We also consider two alternative transformations: the inverse hyperbolic sine (IHS) transformation and the Box–Cox transformation. The IHS transformation handles values close to zero while retaining properties similar to the logarithmic transformation. The Box–Cox transformation allows the data to determine the functional form, reducing concerns that our results depend on the choice of the log transformation. The results, reported in columns (2) and (3), are qualitatively similar to our main results. We also use the numbers of markets within 5 and 10 kilometers of each village centroid as positive measures of market access. Columns (4) and (5) report the results. The positive estimates indicate that villages with more nearby markets, and thus higher market density, also have greater land inequality.

Table G.5: Alternative Market Access Measurement

	Non-logged Transformed	IHS Transformed	Box-Cox Transformed	# Markets <5km	# Markets <10km
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Land Gini, 1932				
{*Market Access*}	-0.0018***	-0.0096**	-0.0038***	0.0056***	0.0036***
	(0.0006)	(0.0037)	(0.0013)	(0.0021)	(0.0008)
Prefecture FE	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y
Baseline controls	Y	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y	Y
Observations	282	282	282	282	282
Adjusted R^2	0.7808	0.7797	0.7816	0.7783	0.7863

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Our baseline estimation uses the log distance from each village to its nearest market as the main explanatory variable. One concern is that overlapping market areas may weaken the use of the nearest market as a proxy for market access. To address this concern, we remove the logarithmic transformation and calculate the distance from each village to every market. We then replace the distance to the nearest market with the distance to the second-nearest and other nearby markets. The results are reported in Figure G.2. The point estimate is largest in absolute value when we use the nearest market. Once we use the second-nearest or other markets, the estimate falls sharply and becomes much less statistically significant. For more distant markets, the estimates are close to zero. This provides a natural placebo test, showing that village-level land inequality is not sensitive to markets located far away.

We also conduct 2,000 random reassignments of market distance across villages. This nonparametric permutation test rejects the null hypothesis that the observed relationship is due to random assignment. The results are reported in Figure G.3. The estimates from the random assignments follow an approximately normal distribution centered around zero, and almost none are as large in absolute value as our

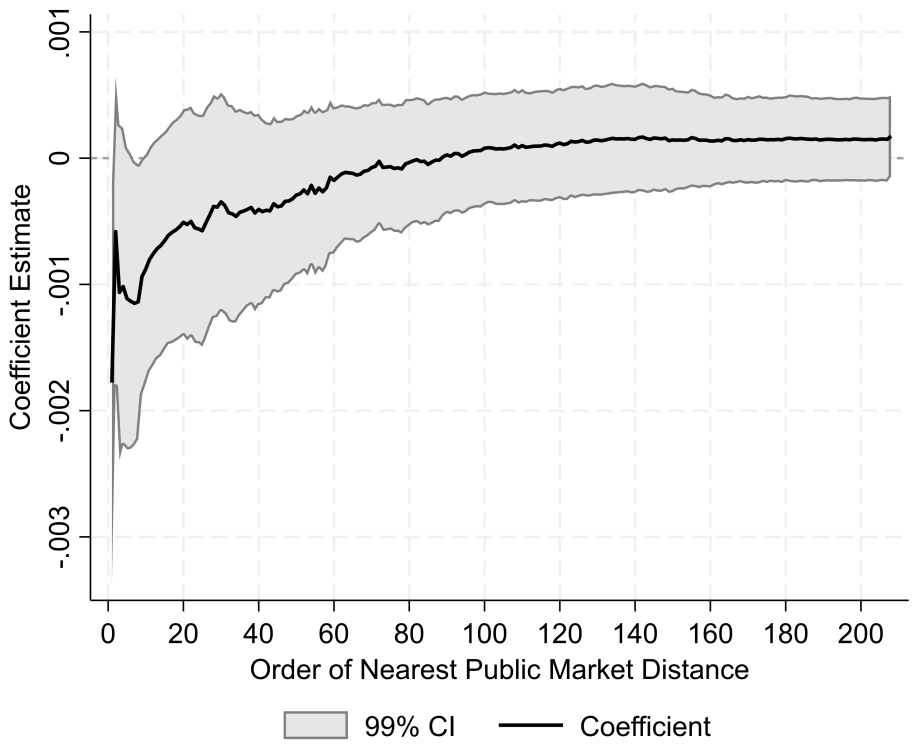


Figure G.2: Placebo Distances and Sensitivity

observed estimate.

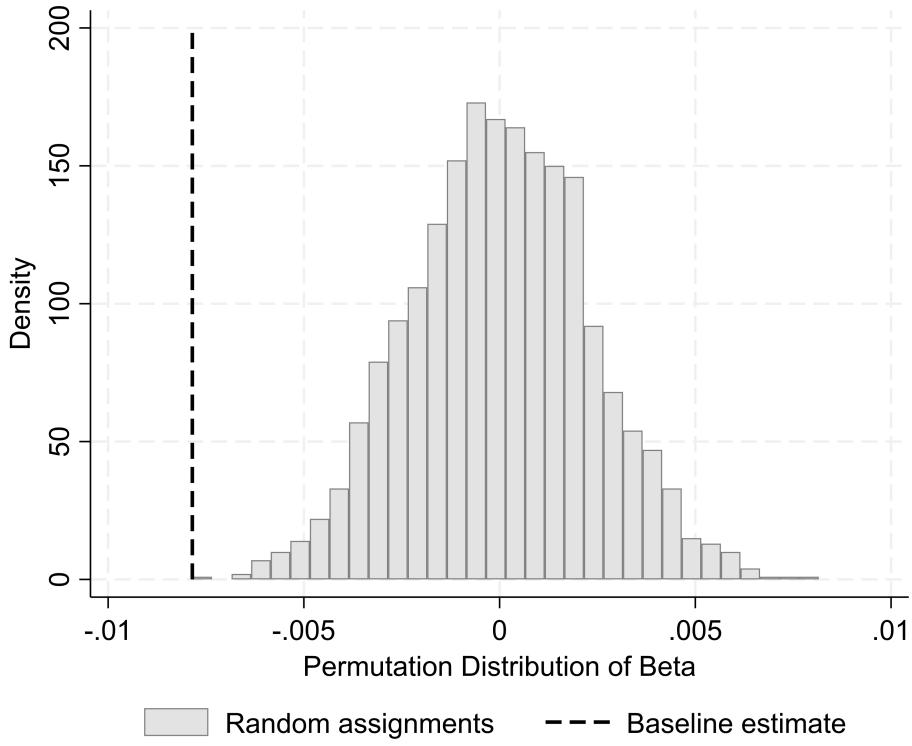


Figure G.3: Random Assignments

G.4 Model Specification

We next discuss several specification issues related to our baseline estimation model. The first concerns the functional form of market distance. Our baseline specification assumes that log market distance, which serves as a proxy for market access, linearly predicts land inequality in 1932. To reduce concerns about this linearity assumption, we allow market distance to enter the estimating equation in quadratic or cubic form. The results are reported in columns (1) and (2) of Table G.6. We find no evidence that these nonlinear terms significantly predict land inequality. When we further include the linear term in columns (3) and (4), only the linear term remains statistically significant. This pattern suggests that, after the logarithmic transformation, distance to the nearest market has an approximately linear and monotonic relationship with observed land inequality.

Table G.6: Flexible Functional-Form Assumptions for Market Distance

	(1)	(2)	(3)	(4)
Dependent Variable		Land Gini, 1932		
Log Market Distance			-0.0035**	-0.0030***
			(0.0014)	(0.0010)
Log Market Distance^2	-0.0000		0.0001	
	(0.0000)		(0.0000)	
Log Market Distance^3		-0.0000		0.0000
		(0.0000)		(0.0000)
Prefecture FE	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y
Adjusted R-squared	0.7756	0.7735	0.7823	0.7830
Observations	282	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Another important feature of our model is that both the dependent and explanatory variables are measured using 1932 as the cutoff year. Column (2) of Table G.7 instead constructs market distance using markets established by 1921. The estimated coefficient is -0.0084 and is significant at the 1 percent level. Its absolute magnitude is even larger than the estimate of -0.0078 based on the 1932 cutoff in column (1). We also include the initial Gini coefficient in 1921 on the right-hand side of the estimating equation. Although this control is important for our IV estimation, it may raise a concern in the OLS estimation: we may overlook the continued establishment of new markets between 1921 and 1932 (see Figure G.4), which changed the distance to the nearest market for some villages. We therefore add the change in log distance to the nearest market between 1921 and 1932 and focus on villages whose market distance changed. In columns (3) and (4), we find that this change does not significantly predict land inequality in 1932. One possible explanation is that most of the relevant variation is captured by markets established in earlier years. Path dependence from these early markets may therefore be important for explaining later inequality.

The third feature of our model is the assumption that villages within the same prefecture share similar characteristics along different dimensions. Prefecture fixed effects therefore absorb many variables

Table G.7: Sources of Variation in Market Proximity

	(1)	(2)	(3)	(4)
Dependent Variable		Land Gini, 1932		
Log Market Distance	-0.0078**		-0.0108***	
	(0.0030)		(0.0031)	
Log Market Distance (1921)		-0.0084***		-0.0110***
		(0.0030)		(0.0032)
Δ Log Market Distance			-0.0036	-0.0050
			(0.0083)	(0.0082)
Prefecture FE	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y
Adjusted R-squared	0.7790	0.7800	0.7650	0.7651
Observations	282	282	222	222

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

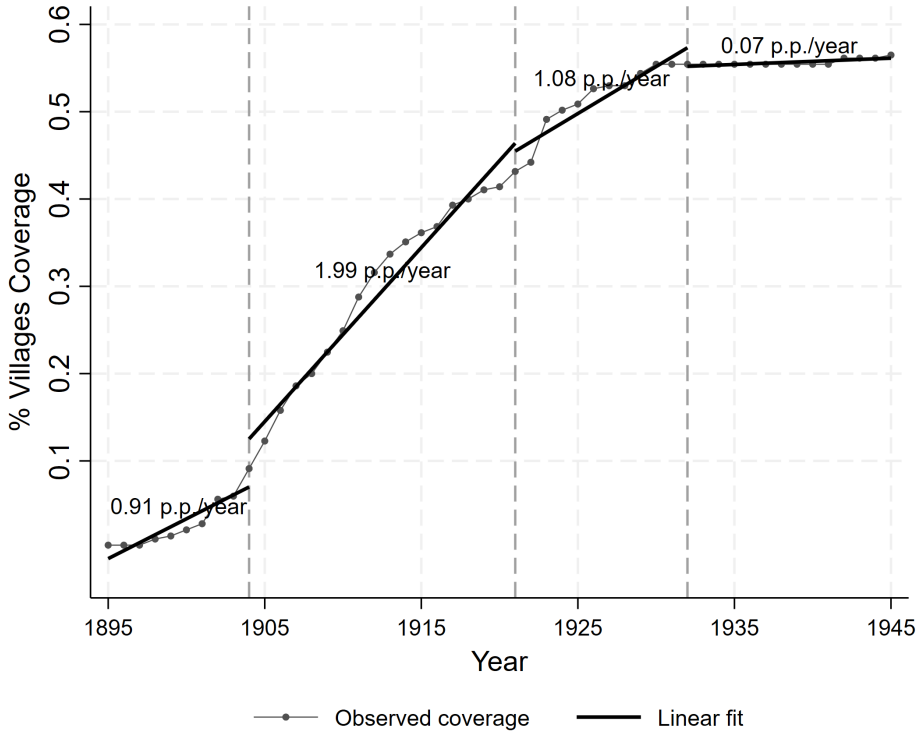


Figure G.4: Expansion of Public Markets, 1895–1945

reported only at the prefecture level. Given Taiwan's spatial characteristics, we further introduce additional spatial fixed effects and indicators. Column (1) of Table G.8 mechanically divides Taiwan into northern, central, and southern regions based on latitude and controls for fixed effects across these three regions. Column (2) adds an indicator for coastal villages, while column (3) adds an indicator for island villages. Column (4) includes both coastal and island indicators to account for spatial features related to the sea. Column (5) adds an indicator for villages near Aboriginal territories. Column (6) includes all these spatial controls. Accounting for these systematic spatial differences does not substantially change our estimated coefficient.

Table G.8: Additional Spatial Controls and Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Land Gini, 1932					
Log Market Distance	-0.0078**	-0.0078**	-0.0077**	-0.0077**	-0.0072**	-0.0069**
	(0.0031)	(0.0030)	(0.0031)	(0.0031)	(0.0032)	(0.0033)
Latitude-tercile FE	Y					Y
Coastal Dummy		Y		Y		Y
Island Dummy			Y	Y		Y
Aboriginal Territory Dummy					Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y	Y	Y
Adjusted R-squared	0.7775	0.7782	0.7787	0.7778	0.7793	0.7767
Observations	282	282	282	282	282	282

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we examine the standard-error assumptions in our baseline model. We first use conventional and heteroskedasticity-robust standard errors, with the results reported in Table G.9. The statistical significance of both the OLS and IV estimates remains consistent. We then account for spatial correlation using Conley (1999) standard errors, allowing error terms to be spatially correlated within cutoff distances of 10, 25, and 50 kilometers. The use of Conley (1999) standard errors does not affect our statistical inference. Finally, we adopt a more flexible spatial-clustering assumption by dividing villages into 10×10 and 15×15 spatial grids based on longitude and latitude. These alternative clustering methods also do not substantially change our statistical inference.

Table G.9: Robustness to Different Standard-Error Assumptions

Standard error assumption	Baseline OLS	Baseline IV	# Clusters
Baseline coefficient	-0.0078	-0.0257	–
Baseline standard error	(0.0030)**	(0.0067)***	61
Conventional standard error	(0.0029)***	(0.0080)***	–
Robust HC1 standard error	(0.0025)***	(0.0077)***	–
Conley spatial correction, 10 km cutoff	(0.0025)***	(0.0078)***	–
Conley spatial correction, 25 km cutoff	(0.0030)***	(0.0074)***	–
Conley spatial correction, 50 km cutoff	(0.0034)**	(0.0063)***	–
Cluster by 10-by-10 equal-width coordinate grid	(0.0033)**	(0.0063)***	44
Cluster by 15-by-15 equal-width coordinate grid	(0.0027)***	(0.0080)***	74

Notes: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

G.5 Weak-Instrument Inference

In the baseline results from our instrumental variables estimation, the first-stage F statistic is 45.74, exceeding the conventional threshold for weak-instrument concerns ($F > 10$). We illustrate this strong relationship visually in Figure 4a.

One potential concern is that, as additional control variables enter the estimating equation, the correlation between the instrument and the endogenous explanatory variable may be partially absorbed. For example, when column (3) additionally controls for distance to clean water sources and hospitals, the first-stage F statistic declines to 26.22. Although this value remains well above the commonly used threshold of 10, to avoid potential weak-identification problems affecting conventional 2SLS inference, we further implement four weak-instrument-robust procedures: the Fuller(1)-LIML estimator, the Anderson–Rubin test, the conditional likelihood-ratio test (CLR), and the tF adjustment (Lee et al., 2022), and report the corresponding 95% confidence intervals in Table G.10. The results show that the confidence intervals obtained from all the weak-instrument-robust inference procedures exclude zero, suggesting that inferential distortions arising from potential weak-instrument bias are unlikely to affect our findings.

Table G.10: Weak-Instrument Robustness

Method	95% Confidence Interval
Fuller–LIML ($\alpha = 1$)	$[-0.03837, -0.01213]$
Anderson–Rubin	$[-0.04234, -0.01322]$
CLR	$[-0.04190, -0.01346]$
tF adjustment	$[-0.04134, -0.01005]$

G.6 Plausible Exogeneity of the Instrument

Our instrument is orthogonal to a range of potential confounding factors, and the instrumental-variable estimates remain robust even after the inclusion of additional controls. Furthermore, we consider direct effects of pesthouses outside the public-market channel, which would violate the instrument's assumption.

Conley, Hansen and Rossi (2012) develop methods for inference when the exclusion restriction is relaxed. Specifically, we allow the instrument to enter the land-inequality equation directly with coefficient γ, and calculate 95% confidence intervals for the coefficient on log market distance at alternative values of γ. In our preferred specification, the direct coefficient on the instrument would have to be approximately -0.00555, or more negative, before the 95% confidence interval for the coefficient on log public-market distance includes zero. At this threshold, the adjusted IV estimate is -0.0123 and its 95% confidence interval is $[-0.0248, 0.0001]$. The IV point estimate reaches zero only when $\gamma = -0.01067$, in which case the direct effect of the instrument accounts for the entire reduced-form relationship.

As an empirical calibration, the estimated coefficient on log hospital distance in the corresponding land-inequality specification is -0.0032. Because hospital distance and pesthouse distance are different variables, we do not interpret -0.0032 as an identified estimate of γ. Rather, we use it as a benchmark for the magnitude of a possible direct effect operating through public-health accessibility or related non-market channels. If we set $\gamma = -0.0032$, a one-standard-deviation increase in the instrument directly changes land Gini by approximately 0.062 standard deviations. This calibration represents about 30% of the reduced form and approximately 58% of the direct effect required to eliminate statistical significance. The adjusted IV estimate remains -0.0180, with a 95% confidence interval of $[-0.0307, -0.0053]$. Hence, the direct effect would have to be approximately 1.7 times the hospital-distance benchmark for the IV estimate to lose statistical significance, and approximately 3.3 times the benchmark for the IV point estimate to be reduced to zero. We believe that the main result is robust to a modest relaxation of the exclusion restriction calibrated to the observed hospital-distance coefficient, even when we allow the instrument to be imperfectly exogenous.

Table G.11: Bayesian Reconciliation of OLS and IV Estimates

Specification	(κ, ρ_{T^*u})	OLS	IV	$P(\varnothing)$	$P(\text{Valid})$	ρ_{Zu}	β
Design-anchored	$(0.8, 1.0] \times [0.0, 0.9]$	-0.10	-0.31	0.000	0.996	0.06	-0.41
		(0.03)	(0.09)			[-0.15, 0.29]	[-0.96, -0.08]
High-error stress	$(0.5, 0.8] \times [0.0, 0.9]$	-0.10	-0.31	0.000	0.981	0.12	-0.51
		(0.03)	(0.09)			[-0.12, 0.38]	[-1.16, -0.10]
Broad envelope	$(0.5, 1.0] \times [0.0, 0.9]$	-0.10	-0.31	0.000	0.996	0.10	-0.47
		(0.03)	(0.09)			[-0.14, 0.34]	[-1.08, -0.08]

Notes: This table reports fully Bayesian estimates using the procedure of DiTraglia and García-Jimeno (2021). The dependent variable is the 1932 land Gini, the endogenous regressor is log distance to the nearest public market, and the instrument is log distance to the nearest pesthouse. All focal variables are standardized using population weights; all specifications include prefecture fixed effects, baseline demographic controls, and the 1921 land Gini. The parameter κ is the signal share in observed market distance, while ρ_{T^*u} captures selection between latent market distance and the second-stage error. The design-anchored specification reflects accurately recorded coordinates but imperfect proxying of effective market access; the high-error specification allows substantially more proxy noise. $P(\varnothing)$ is the share of posterior draws producing an empty identified set. $P(\text{Valid})$ is the share of posterior draws whose identified set contains $\rho_{Zu} = 0$. The parameter ρ_{Zu} measures the correlation between the instrument and the second-stage error, and β is the corrected standardized causal effect.

G.7 Heterogeneity of Complementary Sectors

The results in Table 7 show that the effect of market access varies systematically with commercial intensity. We therefore examine the composition of other economic sectors to address an alternative explanation: commercial intensity may simply be an inverse measure of activity in other sectors, such as agriculture and industry. The 1915 and 1920 population censuses report the number of workers in different sectors, allowing us to calculate sectoral intensity outside commerce. We estimate interaction specifications consistent with those in Table 7 and report the results in Table G.12. Across the different measures from 1915 and 1920, none of the interaction terms provides credible statistical evidence that the heterogeneity by commercial intensity is driven by simple complementarity with other major sectors.

Table G.12: Heterogeneity of Complementary Sectors

Industry Category	Agriculture + Livestock + Forestry + Fisheries 1915	Industry 1915	Agriculture 1920	Fishery 1920	Industry 1920
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Land Gini, 1932				
Log Market Distance × Sectoral Intensity	-0.0136	-0.0437	0.0035	0.0051	-0.0171
	(0.0087)	(0.0276)	(0.0096)	(0.0358)	(0.0250)
Log Market Distance	0.0047	-0.0044	-0.0097	-0.0076**	-0.0063*
	(0.0083)	(0.0030)	(0.0082)	(0.0033)	(0.0037)
Sectoral Intensity	-0.0480**	0.0465	-0.0191	-0.0499	0.0436
	(0.0191)	(0.0474)	(0.0175)	(0.0493)	(0.0579)
Prefecture FE	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y	Y

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

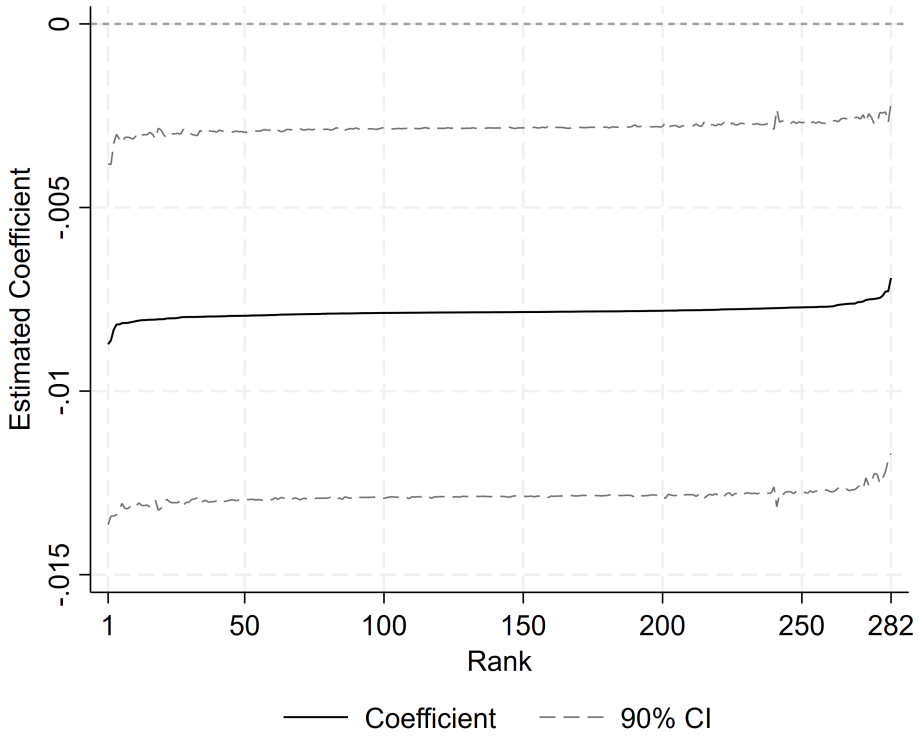
G.8 Falsification Tests

Table G.13: Falsification Tests

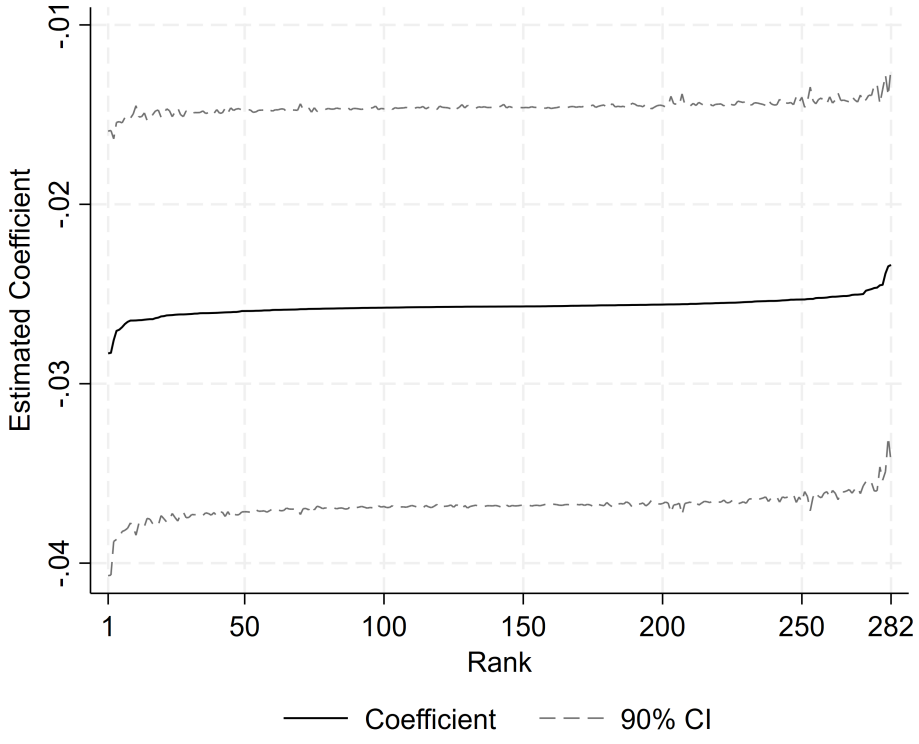
	Bottom Quartile of % Infectious Mortality	Post-1932 Market Access as Independent Variable	Non-infectious Mortality as Instrument
	(1)	(2)	(3)
Dependent Variable		Land Gini, 1932	
Log Market Distance	-0.0525	-0.8861	0.1010
	(0.0461)	(2.8444)	(0.1978)
Prefecture FE	Y	Y	Y
Population weighted	Y	Y	Y
Baseline controls	Y	Y	Y
Initial Gini	Y	Y	Y
First-stage F-statistic	2.82	0.08	0.21

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

G.9 Subsample Analysis



(a) OLS estimates



(b) IV estimates

Figure G.5: Leave-one-village-out estimates

Table G.14: Robustness to systematic sample restrictions

Excluded for	Political Center		Population $> p_{90}$		Market $< p_{10}$ or $> p_{90}$	
Estimator	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Land Gini, 1932					
Log Market Distance	-0.0080**	-0.0269***	-0.0079**	-0.0224***	-0.0094*	-0.0614***
	(0.0031)	(0.0068)	(0.0030)	(0.0069)	(0.0049)	(0.0224)
Prefecture FE	Y	Y	Y	Y	Y	Y
Population Weighted	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
Initial Gini	Y	Y	Y	Y	Y	Y
Mean Log Market Distance	1.2751	1.2751	1.3123	1.3123	1.2584	1.2584
Mean Land Gini, 1932	0.6130	0.6130	0.6105	0.6105	0.6227	0.6227
Observations	275	275	254	254	225	225

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.