Coercive Growth:
Forced Resettlement and Ethnicity-Based Agglomeration

Shanon Hsuan-Ming Hsu
University of Chicago

November 1, 2024

Click here for the latest version

Abstract

How do social divisions affect the benefits of agglomeration? While the clustering of people can enhance productivity through social interactions, social divisions such as ethnic segregation and tension may limit these gains. To answer this question, I leverage an ethnic-based resettlement program that forcibly relocated 600,000 rural Chinese into compact villages in 1950s British Malaya. I find that, decades later, areas with higher resettlement had persistently higher population densities and concentrations of Chinese, driven by both the program's direct impact and internal migration. Moreover, these areas were wealthier, more industrialized, and exhibited greater labor market specialization. However, the economic benefits primarily accrued to the Chinese population, while other ethnic groups saw only marginal gains when geographically integrated with the Chinese and working in non-agricultural sectors. To assess the overall impact of the program, I estimate a quantitative spatial model that allows agglomeration externalities to vary by sector and ethnic composition. While the resettlement increased aggregate output, the gains were insufficient to offset the welfare losses from the program's coercive nature.

I am deeply grateful to my advisors Richard Hornbeck, Oeindrila Dube, Michael Dinerstein, and Jonathan Dingel for their guidance and support throughout the project. I also thank Sarah Ling Francis for her excellent research assistance, and Hugh Alexander and Alexander Poole for their help with archival research. Special thanks to David Baillargeon for sharing key data. For helpful comments and suggestions, I thank Rodrigo Adão, Scott Behmer, Fiona Burlig, Joshua Dean, Tomás Domínguez-Iino, Rachel Glennerster, Esteban Rossi-Hansberg, Estéfano Rubio, Thomas Hierons, Anders Humlum, Peter Hull, Ed Jee, Furkan Kilic, Chanwool Kim, Michael Kremer, Nadav Kunievsky, Michael Pollmann, James Robinson, Jordan Rosenthal-Kay, Martin Rotemberg, and Jeanne Sorin, as well as seminar participants at the University of Chicago and the Mountain West Economic History Conference. I also thank the Department of Statistics Malaysia (DOSM) for providing data, and the data entry teams at Digital Divide Data. This project was funded by the Pearson Institute at the University of Chicago. All errors are my own. Contact: shmhsu@uchicago.edu.

I Introduction

The clustering of people and economic activity can enhance productivity by reducing transaction and search costs and fostering knowledge spillovers (Marshall, 1890; Duranton and Puga, 2004). Yet, these benefits of proximity hinge on the density and nature of local, social interactions. Divisions such as ethnic segregation and tension—common in urban settings across many countries—can hinder these productive interactions. This raises a critical question: how do social divisions shape the benefits of agglomeration and broader patterns of economic development?

To answer this question, I leverage an ethnic-based resettlement program during the 1950s Malayan Emergency in British Malaya, which forcibly relocated roughly 600,000 rural Chinese into compact villages across the country. This setting provides a natural experiment to study how social divisions affect agglomeration benefits, as the resettlement shifted both population size and ethnic composition while limiting self-selection. Using this program as a source of variation, I examine how the concentration of one ethnic group (Chinese) impacted local economic outcomes across regions and different groups over the next 50 years. To interpret the findings, I develop a quantitative spatial model that allows agglomeration externalities to vary by sector and ethnic composition, using the resettlement as a population shifter to identify the model's agglomeration parameters. I then use the model to quantify the aggregate impacts of the resettlement and explore counterfactual policies for promoting economic development in the presence of these heterogeneous agglomeration forces.

Chinese communities in British Malaya specialized in industrial and urban sectors, where agglomeration externalities typically emerge. In contrast, most Malays engaged in subsistence agriculture (Purcell, 1947). These economic divides, along with cultural, religious, and linguistic differences, contributed to ethnic segregation and tensions between the groups (Hirschman, 1986).

During the Japanese occupation (1942–1945), many Chinese were displaced to rural areas, where their presence raised security concerns during the Malayan Emergency (1948–1960)—a guerrilla conflict between the British and the Malayan communists. To prevent these rural Chinese from supporting the insurgents, the British implemented a large-scale resettlement program, forcibly relocating nearly one-tenth of the population—primarily Chinese—into about 500 "New Villages" in more accessible areas. Two-thirds of these villages were built around or integrated into existing settlements, leading to a significant rise in urban populations (Sandhu, 1964).

I identify the local impacts of the resettlement by leveraging the program's wartime objectives and hastiness to specify counterfactual resettlement and isolate exogenous variation. The policy was implemented in two stages: first, suitable sites were selected based on security access, with most scattered along main roads or rivers; second, the rural Chinese were relocated to these sites in a way that minimized dislocation from their original settlements. Following this procedure, I specify counterfactual resettlement in two steps. First, I randomly permute counterfactual locations for the New Villages, conditional on their distance to the transportation network and other key covariates considered by the British, such as land use and the initial rural Chinese population in the county. Next, I use a gravity model to predict the number of people resettled to each potential site, assuming resettlement costs increased with the distance they were moved. I repeat this permutation procedure a thousand times and average the resettlement density in each county to obtain the *expected* resettlement density (Borusyak and Hull, 2023).

The estimation compares areas with varying actual resettlement densities while controlling for expected density and key covariates related to the initial transportation network and population distribution. The identifying variation comes from the exact locations of the New Villages relative to the average location along major transportation routes and deviations from the dislocation-minimizing plan. Historical accounts suggest that the British lacked the capacity or intent to fine-tune site selection and relocation based on unobserved economic fundamentals. I show that geographic and prewar economic characteristics—such as ruggedness, agricultural suitability, and proximity to major industrial facilities and cities—are balanced across locations with varying residual resettlement densities.

The resettlement program had a significant and lasting impact on population distribution. In the short term, counties with higher resettlement saw a sharp increase in their Chinese population between 1947 and 1957, while the non-Chinese population remained unchanged. For every 1% of the population resettled, the local population grew by 0.69%, accounting for 77% of these counties' population growth during this period. After mobility restrictions were lifted in 1960, more densely resettled counties began attracting internal migrants from all ethnic groups. By 2000, these counties still had higher population densities and a larger share of Chinese residents, with each 1% of population resettled in 1947 leading to a 1.3% higher population in 2000.

The influx of Chinese into resettled areas reshaped the economic structure due to their historical specialization in the non-agricultural sector. I find that by 1970, counties with higher resettlement had more manufacturing firms, especially in industries which the Chinese had a strong pre-Emergency presence. Firms in these county-industries were also larger, suggesting productivity gains from agglomeration economies. By the 1980s, employment in these counties was significantly higher, particularly in non-agricultural sectors such as manufacturing, trade, and services. This shift was driven by a higher concentration of Chinese workers, who were more likely to work outside agriculture, and by a smaller agricultural share among non-Chinese workers in more resettled areas.

The higher density of workers and firms facilitated a greater division of labor, especially among the Chinese population. I find that by the 1980s, Chinese individuals in counties with higher resettlement had higher labor force participation and greater employment specialization. In contrast, non-Chinese workers in these counties showed no significant differences in these labor market outcomes. Additionally, Chinese individuals in more resettled counties achieved greater educational attainment, especially among younger cohorts who had not completed their education by the time of resettlement, consistent with a greater division of labor increasing the returns to higher and more specialized education.

Resettlement-induced agglomeration economies led to higher worker productivity, particularly among the Chinese population. Chinese households in more densely resettled counties earned higher incomes, while non-Chinese households saw only marginal and statistically insignificant income gains in the non-agricultural sector. Notably, even Chinese agricultural workers saw higher incomes in these counties, suggesting the productivity benefits were not confined to specific industries but were more pronounced among ethnic Chinese.

Geographic segregation between ethnic groups explains half of the smaller agglomeration benefits for the non-Chinese population. I find that in counties where Chinese and non-Chinese communities were more geographically integrated, non-Chinese households saw greater productivity gains from resettlement. For every additional kilometer of separation between the communities, non-Chinese income gains dropped by one-fifth, with no positive effect beyond 5 kilometers. However, even in fully integrated areas, the non-Chinese population saw only half the income gains from resettlement compared to the Chinese, suggesting that deeper factors—such as ethnic tensions and isolated social networks—may have hindered productive interactions between groups.

Cross-ethnic frictions in agglomeration spillovers imply that spatial externalities vary with local ethnic composition, suggesting that population resettlements can impact aggregate economic outcome. For example, moving Chinese individuals from areas with relatively few Chinese—where the combined within-ethnic and cross-ethic spillovers are weaker—to areas with a higher concentration of Chinese—where those spillovers are stronger—could enhance overall productivity. A similar logic applies to industries: reallocating labor from sectors with smaller external economies of scale to those with larger economies can increase aggregate productivity—a common rationale for industrial policies.

To interpret the empirical findings from the resettlement and evaluate its aggregate impacts, I develop a quantitative spatial model with two ethnic groups and two sectors, incorporating a Roy (1951)-type framework for migration and occupation choices—the key adjustment margins following the resettlement shock. Extending the work of Allen and

Donaldson (2020) and Peters (2022), I allow agglomeration forces to vary by sector and ethnic composition. Local amenities are endogenously shaped by ethnic composition, as in Diamond (2016), and cross-group productivity spillovers can differ from within-group spillovers depending on local ethnic composition. In this two-period model, individuals start with an initial location and have heterogeneous preferences across locations and productivities across location-sectors. Agents decide where to migrate, subject to movement costs, and, after moving, which sector to work in to maximize their utility given their preferences and productivities.

To bring the model to the data, I use the 1957 population distribution—observed after most of the resettlement was completed—as the initial population, and treat the 1980 data as the equilibrium outcomes. The forced resettlement serves as an exogenous population shifter for identifying the model's agglomeration parameters. Due to migration costs, individuals tended to move to areas near their resettled locations, so the population shifts by 1957 continued to affect the 1980 equilibrium distribution. The relationship between the local Chinese share and the wage premium earned by Chinese workers within a location-sector—controlling for differences in occupational structures across locations—identifies the strength of within-ethnic versus cross-ethnic productivity spillovers. Additionally, the relationship between wages and local employment size—after adjusting for composition effects and general equilibrium forces—identifies the scale economies in each sector.

The estimates reveal that agglomeration externalities vary across sectors and regions, shaped by ethnic composition. A neoclassical model with downward-sloping demand would predict that an increased supply of industrial labor lowers relative non-agricultural wages, pushing workers into agriculture. The higher non-agricultural wages and smaller share of agricultural employment observed in more densely resettled counties suggests that labor demand in the non-agricultural sector may have been flat or even upward-sloping. The model attributes this pattern to external economies of scale in the non-agricultural sector: a 1% increase in non-agricultural employment raises productivity by 0.25%. By contrast, the agricultural sector exhibits local diminishing returns to scale due to the fixed land supply, where a 1% increase in agricultural employment reduces productivity by 0.1%. Additionally, I estimate that within-ethnic spillovers are stronger than cross-ethnic spillovers, with the latter being positive only in the non-agricultural sector. Lastly, the estimated positive amenity spillovers with respect to local ethnic share indicate homophily—a preference for living near others of the same ethnic background.

To assess the aggregate impacts of the Emergency resettlement, I simulate a "no resettlement" equilibrium using the 1947 population distribution as the initial condition, as opposed to the resettled 1957 distribution, holding all parameters and location fundamentals constant. Assuming the 1947 distribution was in a steady state that would have continued to 1957 without resettlement, I calculate the resettlement's aggregate impact in 1980 by comparing this no-resettlement equilibrium to the observed 1980 equilibrium. I find that the resettlement program increases aggregate output by 2%, with two-thirds of the gains driven by labor reallocation to more productive regions and sectors. Relocating rural Chinese from remote areas to locations with better market access and higher industrial productivity raises their output, while freeing up rural lands for Malays, who benefit from improved agricultural productivity.

Despite the net output gain, the coercive nature of the program raises welfare concerns. To benchmark the economic gains in welfare terms, I invert the model to calculate the place-based wage subsidies required to voluntarily relocate people from the 1947 population distribution to the 1957 resettled distribution. I find that forced resettlement reduces welfare, as the costs of implementing this voluntary migration far outweigh the modest productivity benefits from resettlement.

Finally, given stronger external economies in the non-agricultural sector and limited cross-ethnic spillovers that hinder Malay participation, subsidizing Malay industrialization could improve economic outcomes while reducing inequality. I find that an 18% wage subsidy for Malays in the non-agricultural sector—designed to equalize non-agricultural employment shares across ethnic groups—increases aggregate output by 1%. However, this policy also has significant distributional consequences, crowding out Chinese workers from the non-agricultural sector and reducing their incomes.

Since Marshall (1890), a vast body of literature has documented the productivity benefits of density and geographic concentration (Glaeser et al., 1992; Duranton and Puga, 2004; Moretti, 2004b; Rosenthal and Strange, 2004; Ahlfeldt et al., 2015; Davis and Dingel, 2019; Heblich et al., 2020). This paper contributes to the relatively few studies that use natural experiments to provide direct evidence of agglomeration spillovers (see, e.g., Greenstone et al., 2010; Kline and Moretti, 2014; Peters, 2022; Smith and Kulka, 2023). Among these, my work is closely related to Peters (2022), which examines how refugee settlement in postwar Germany spurred industrialization and growth. By studying an ethnic-based resettlement that not only altered population sizes but also changed the social composition across regions—unlike the resettled refugees in that study, who shared similar social backgrounds with the incumbent population—I highlight how local population composition shapes agglomeration benefits across different groups. In doing so, I contribute to the literature on the distributional consequences of agglomeration (Baum-Snow and Pavan, 2013; Baum-Snow et al., 2018; Ahlfeldt and Pietrostefani, 2019) and complement the work of Ananat et al. (2013, 2018), who argue that segregation and lower levels of cross-race social interactions reduce the returns to agglomeration, contributing to a larger black-white wage gap in bigger cities.

I also contribute to the literature concerning the aggregate implications of placed-based policies (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014; Neumark and Simpson, 2015; Fajgelbaum and Gaubert, 2020; Rossi-Hansberg et al., 2019). As emphasized by Glaeser and Gottlieb (2008), when agglomeration elasticities are constant across regions, spatial reallocation of economic activity leads only to distributional effects, with no aggregate consequence. However, this result assumes no spatial transfers, and Fajgelbaum and Gaubert (2020) have shown that introducing spatial transfers can improve aggregate outcomes in a decentralized equilibrium. On the other hand, Rossi-Hansberg et al. (2019) finds stronger productivity spillovers within "cognitive non-routine" occupations as a source of non-constant agglomeration elasticities, motivating spatial redistribution of specific types of workers. My work adds to this discussion by highlighting cross-ethnic frictions as another source of non-constant agglomeration elasticities.

This paper also adds to the body of work on neighborhood effects (Glaeser et al., 1996; Kling et al., 2007; Chetty et al., 2016, 2018; Ioannides, 2012; Fogli and Guerrieri, 2019; Ambrus et al., 2020; Chyn and Katz, 2021; Redding and Sturm, 2024). While much of this research focuses on high-income countries, I provide new evidence on how neighborhood composition matters economically in a developing country. By treating ethnic composition as part of a location's amenities, my work connects to studies that examine how endogenous amenities derived from demographic characteristics shape the sorting of heterogeneous agents (Bayer et al., 2004, 2007; Diamond, 2016; Gechter and Tsivanidis, 2023; Tsivanidis, 2023; Weiwu, 2023; Almagro and Domínguez-Iino, 2024). This paper extends this line of research by introducing the productivity effects of demographic composition on local residents, linking to the ethnic-enclave literature, which highlights the role of social capital in labor market outcomes among migrants in enclave neighborhoods (Borjas, 1992; Bertrand et al., 2000; Edin et al., 2003; Munshi, 2003; Aizer and Currie, 2004; Damm, 2009; Beaman, 2011; Abramitzky et al., 2024; Eriksson, 2019).

This paper also builds on prior research into the economic impacts of forced migration and villagization (Hilhorst and Leeuwen, 2000; Whittaker, 2012; Bazzi et al., 2016; Abel, 2019; Bazzi et al., 2019; Becker et al., 2020; Carlitz et al., 2022; Peters, 2022; Sarvimäki et al., 2022; Carrillo et al., 2023). Forced villagization has been a common policy for development or nation-building in many countries, with previous studies often finding neutral or negative economic effects for villagers, likely due to increased inter-group tensions when diverse populations are forced to live together (Dippel, 2014). My main contribution is using a spatial model to quantify the aggregate effects of forced resettlement.

Finally, this paper ties into the literature on the rapid postwar economic development of several East and Southeast Asian countries, particularly the role of state intervention in driving this growth (Haggard, 1990; Amsden, 1992; Mundial, 1993; Wade, 2004; Lane, 2022). The higher returns to scale I find in the non-agricultural sector, compared to agriculture, suggest that industrial policies promoting structural transformation—common among the East Asian Tigers and later "Look East" followers like Malaysia—could spur economic take-off through self-reinforcing productivity spillovers. Additionally, the stronger within-group agglomeration spillovers identified from resettled Chinese populations may partly reflect the greater collective action in Chinese communities, a factor that Dell et al. (2018) link to historical exposure to a centralized state institution.

The rest of the paper is organized as follows. Section II provides the historical context. Section III describes the data. Section IV discusses the empirical strategy and Section V examines the local impacts of resettlement. Section VI lays out the model, which I estimate in Section VII.B. Section VIII explores counterfactual policies, and Section IX concludes.

II Historical Background

The agglomeration effects of Chinese clustering depend on the types of economic activities the Chinese population specialized in before resettlement, while local spillovers to other ethnic groups hinge on the degree of cross-ethnic interaction. This section first describes the distinct economic roles of the ethnic Chinese in British Malaya and their relationship with the majority Malay population. I then discuss the objectives of the British colonial administration in resettling rural Chinese during the Malayan Emergency.

II.A Ethnic Chinese and Social Divisions in British Malaya

By the end of World War II, British Malaya's population was 49% Malay, 39% Chinese, and 12% Indian and other ethnic groups (Appendix Table A.1). Historically, Chinese immigrants had specialized in the colony's industrial and urban sectors, initially through their work in tin mining and rubber plantations—the key export industries during the colonial period. Over time, many Chinese transitioned into manufacturing and commerce. In contrast, the majority Malay population primarily engaged in subsistence agriculture, particularly paddy rice and coconut cultivation (Ginsburg, 1958; Lee and Tan, 2000). By 1947, 60% of Chinese worked in agriculture compared to 80% of Malays (Appendix Figure A.1). Industrial jobs were concentrated in towns, leading to higher urbanization among the Chinese, with 40% living in urban areas by 1947, compared to only 10% of Malays (Del Tufo, 1947). These distinct economic roles contributed to ethnic tensions between the two groups.

Cultural, religious, and linguistic differences further deepened these social divisions. The majority of Malays were Muslim, while the Chinese were predominantly non-Muslim and followed cultural practices, such as pork consumption, that conflicted with Muslim values.

These differences limited intermarriage between the two groups. Language barriers further reinforced these divisions: in 1947, fewer than 1% of Chinese spoke Malay, and by 1980, only 25% were fluent (Del Tufo, 1947; Khoo and Perangkaan, 1983). As a result, ethnic groups remained segregated both spatially and socially, even in urban areas, with few opportunities for cross-ethnic interaction in schools or workplaces (Hirschman, 1986).

By the late 1940s, about one-third of the Chinese population lived in rural areas near the jungle fringe (Sandhu, 1964, p. 150). Many of these rural Chinese—often referred to as "squatters" because they occupied land without formal legal titles—had previously lived in urban areas but were displaced to the countryside during the Japanese occupation (1942—1945), which severely disrupted the industrial sectors.[1]

II.B The Briggs Plan: Emergency Resettlement

The large population of Chinese squatters in remote areas became a security concern for the British during the Malayan Emergency (1948—1960), a guerrilla war between British forces and communist insurgents. Many squatters aided the communists by providing information and supplies, with some becoming part of their non-military support network.[2] The squatters' isolated locations near the jungle, where the communists operated, made it difficult for the British to prevent this aid (Humphrey, 1971, p. 49; Loh, 1988, pp. 106–107).

To address this, the British launched the "Briggs Plan," a large-scale resettlement program that forcibly relocated squatters to secure areas when their original settlements were deemed unsafe.[3] The plan had two main objectives: to control populated areas for intelligence gathering and to cut off the communists' supply lines, forcing them into unfavorable confrontations with British forces (Briggs, 1951, p. 7). Speed was critical to prevent the communists from adapting to the new conditions (Sunderland, 1964, p. 161; Humphrey, 1971, p. 106). Additionally, the plan aimed to minimize dislocation and economic disruption to avoid fueling resentment among the Chinese population, while still prioritizing security objectives (Humphrey, 1971, pp. 181–182).

The state government implemented the program rapidly: beginning in mid-1950, most resettlements were completed by the end of 1952 (Sandhu, 1964, pp. 159–161). The process involved selecting sites, clearing land, marking house plots and roads, and issuing removal

[1] The Squatter Committee Report, The National Archives in the UK (hereafter, "TNA"), CO 717/178. Many urban Chinese fled to the countryside to escape conflict or forced labor under Japanese rule (Humphrey, 1971, pp. 39, 47; Loh, 1988, pp. 57–60). Reduced demand and price volatility for tin and rubber during World War I and the Great Depression may have also contributed to the rural shift among the Chinese (Humphrey, 1971, p. 43; Loh, 1988, pp. 23, 27–29).

[2] The Malayan communist Party (MCP) had strong ties to the Malayan People's Anti-Japanese Army (MPAJA), which had previously fought the Japanese. The communists' non-military network was also called the "Min Yuen."

[3] TNA: CO 717/178.

notices to the squatters.[4] Squatters typically received fewer than 14 days' notice before relocation, with even shorter notice in cases where resistance or escape was expected.[5] Military transport was used to move the squatters and their belongings to the new sites, after which the original settlements were burned down (Sandhu, 1964, p. 160).

Although all resettlement areas were referred to as "New Villages," only one-third were completely new, while the rest were built around or integrated into existing villages or towns (Sandhu, 1964, p. 163; Humphrey, 1971, p. 98). Most villages shared similar layouts and amenities, typically including a school, police station, and community center. Strict mobility regulations, such as dusk-to-dawn curfews and police checkpoints, were enforced at village entrances (Humphrey, 1971, pp. 118, 358).

By the end of the Emergency, approximately 573,000 people had been resettled into 480 New Villages. The population of these villages was predominantly Chinese (86%), with smaller numbers of Malays (9%) and other groups (5%) (Sandhu, 1964, p. 159).[6] These villages continued to grow after the Emergency, and many still exist today.[7]

II.C Determinants of Resettlement Density

Understanding the factors that shaped resettlement density is crucial for evaluating its impacts. The resettlement program was implemented in two stages: first, selecting village sites; and second, relocating squatters with an emphasis on minimizing dislocation. The British outlined several key criteria for choosing these sites (Humphrey, 1971, pp. 95–97), and this section discusses how these criteria, along with other idiosyncratic factors, influenced the variation in resettlement density.

Security and defensibility. The primary goal of the resettlement was security. Village sites were ideally located near major roads to ensure easy access for police (Sandhu, 1964, p. 164; Dhu Renick, 1965, p. 9). In more remote areas lacking major roads, sites were placed near navigable waterways (Humphrey, 1971, p. 96). Figure 1 shows that most New Villages were located along main transportation routes. For defensibility, villages were also ideally sited on elevated hills away from observation points, although the data does not show a meaningful correlation with elevation.

Land availability. Land acquisition costs were another important consideration. Financial constraints led the British to prioritize state-owned land or land with low commercial

[4] TNA: CO 1022/29.

[5] In cases of suspected communist ties or resistance, relocations were conducted at dawn without prior notice (Humphrey, 1971, p. 102).

[6] Roughly half of the resettled population were squatters, while the remainder were legitimate landholders.

[7] The Ministry of Housing and Local Government of Malaysia estimated that 465 New Villages remained in 1972, with a population of roughly one million. By 2005, about 450 villages were still in existence (Lee and Tan, 2000, p. 262).

value (Humphrey, 1971, p. 367). Many New Villages were established on state-owned rubber estates. Since different land types likely varied in latent productivity, I control for land use patterns in the analysis.

Economic sustainability. Villages were ideally located on well-drained land, suitable for agriculture, and with access to water supplies (Dhu Renick, 1965, p. 9; Humphrey, 1971, p. 96). However, due to a shortage of qualified staff for surveying and planning—driven by financial, time, and manpower constraints—village siting decisions were often made by individuals with limited field knowledge. As a result, many villages were poorly sited, prone to flooding, or unsuitable for agriculture (Humphrey, 1971, p. 107).[8] A 1954 survey found that 31% of the sites were unlikely to remain sustainable after the Emergency (Corry, 1954). This suggests that economic considerations were not the primary focus in site selection. I will show that geographic characteristics related to productivity—such as ruggedness and agricultural suitability—do not correlate with resettlement density once I control for proximity to transportation.

Proximity to squatter populations. To minimize disruption, village sites were ideally located near the original squatter settlements (Sandhu, 1964, p. 160; Humphrey, 1971, pp. 96–97). Most relocations occurred within 20 miles of the original settlements (Sandhu, 1964, p. 160). As such, the initial distribution of squatters played a crucial role in determining how many people were resettled to a region. Since Chinese squatters self-selected into locations before resettlement, I condition on the per-resettlement population in the analysis.

Other idiosyncratic factors. Several additional factors influenced resettlement distribution. First, due to the lack of field surveys, the British had limited knowledge of squatter population distribution and continued discovering new squatter areas throughout the program. A 1952 newspaper article noted, "The Government had only the haziest idea of the numbers [of squatters]: it was first believed that there were 318,500...."[9] Newly discovered squatters often had to be relocated to more distant sites because no suitable locations had been selected near their original settlements. Moreover, as the communists adapted and shifted their supply bases, some areas initially considered secure later became unsafe, leading to further relocations to distant sites. This resulted in some locations housing more people than initially planned.

[8] Examples include Batu Rakit/Pulai (Trengganu), sited on sandy wasteland; Jemaluang (Johore), located on tin tailings; and Kampung Abdullah (Johore), which regularly flooded (Sandhu, 1964, p. 161). See also Notes on Planning and Housing Aspects of Resettlement and the Development of New Villages (Arkib Negara Malaysia, hereafter, "ANM", 1953).

[9] TNA: CO 1022/29, p. 63. Between 1952 and 1953, the estimated number of people needing resettlement remained at 80,000–90,000, even though over 150,000 had already been resettled during that period (Humphrey, 1971, p. 123).

The British focus on speed and security, combined with the absence of comprehensive field surveys, led to plausibly exogenous variations in population resettlement. Sites were chosen from many similarly suitable locations along roads, and limited knowledge of squatter distribution or shifting communist risks often prevented the British from resettling people in their original locations. I will leverage these variations to construct a population shifter, which I use to examine the local effects of resettlement.

III Data

III.A Emergency Resettlement

I measure the resettled population from a 1954 official report (henceforth, the "Corry report") to the High Commissioner (Corry, 1954), conducted shortly after most of the resettlement had been completed.[10] The report provides the names, populations, and forms of local government for 439 villages, along with qualitative details villages conditions. Village locations are taken from Baillargeon (2021), who geolocated 430 of the New Villages listed in the report. These villages account for approximately 540,000 people, covering 94% of the total resettled population by the end of the Emergency.[11] The geolocation was based on village names and the states, and I cross-validated the resettled population figures using a 1958 study by the Malayan Christian Council (Council, 1958).[12]

To measure key covariates related to the resettlement, I digitized several historical maps, including a 1942 road and railway map, a 1943 land utilization map, 1945 topographical maps, a 1945 military map showing prewar industrial facilities, a 1947 population census map, and a 1957 "Black Areas" map showing regions with higher communist activity and Emergency regulations.[13] I also obtained elevation data from the Shuttle Radar Topography Mission (SRTM, 2000), crop suitability data from the Food and Agriculture Organization's Global Agro-Ecological Zones (GAEZ) database, and terrain ruggedness from Riley et al. (1999).

[10] The report had four main objectives: (i) to assess the sufficiency of agricultural land and overall economic conditions; (ii) to evaluate the long-term sustainability of the villages and estimate potential out-migration after the Emergency; (iii) to examine land ownership among villagers; and (iv) to estimate the number of Chinese still in rural areas requiring resettlement.

[11] An estimated 573,000 people had been resettled by the end of the Emergency (Sandhu, 1964, p. 159).

[12] The 1958 survey included around 100 additional, smaller villages built after 1954. For villages documented in both sources, the population figures are consistent.

[13] The road/railway map and industrial facilities map are from the U.S. Office of Strategic Services (U.S. Office of Strategic Services, 1942, 1944). The topographical maps are from the HIND 1076 map series Survey of India Offices (P.Z.O.), and the land use maps are from the GSGS 4474 series War Office (1943).

III.B Outcome Data

Data on Population. I digitized Malaysia's Census of Population at the county level for the years 1931, 1947, 1957, 1970, 1980, 1991, and 2000.[14] The data provides population counts by ethnic group for each county. To account for changes in county boundaries over time, I created time-consistent borders based on the 1947 boundaries, grouping counties with overlapping geographies across different years. I excluded nine counties with populations in 1947 but no reported populations in 1957 or 1970, as these are likely enumeration errors rather than true zeros. My baseline sample includes 777 counties. For regressions involving the 1931 population, I generated a separate set of 614 grouped counties based on the 1931 boundaries. The median county width is 8–9 kilometers.

Data on Economic Structure. I measure county employment by ethnic group, industry, and occupation from the 1980 and 1991 Population Census. Additionally, I obtain prewar aggregate employment data by industry and ethnic group from the 1947 census.

For manufacturing data, I digitized the 1970 Directory of Manufacturing, which lists approximately 12,000 registered manufacturing establishments in Peninsular Malaysia.[15] The directory provides each establishment's name, address, main products, industry, and employment size, which I georeferenced to their respective counties based on the provided addresses.

Income Data. Since the census lacks direct income or wage data, I use the 1988–1989 Second Malaysian Family Life Survey (MFLS-2) to estimate household incomes. I train a statistical model using the survey data to predict household earnings based on district fixed effects, household size, and asset ownership (e.g., automobile, motorcycle, bicycle, phone, refrigerator, and TV), including interactions of these indicators. Appendix Table A.2 presents the statistical model. I then apply this model to the 1980 census 2% microdata to generate income measures for the broader population.[16]

Education and Segregation Data. I measure educational attainment from the 1980 census microdata, which includes indicators for primary, secondary, and higher education completion, as well as years of schooling.

I obtain data on Chinese vernacular and other national schools from Malaysia's Ministry of Education. This dataset provides a comprehensive list of all primary and secondary schools in 2022, including their names, number of teachers and students, and exact geographic

[14] I use "county" to refer to the administrative unit "mukim" in Malaysia. The 1931 Census was the first to document population by county.

[15] All establishments were required to register under the Registration of Business Ordinance 1957.

[16] Conducted by RAND and Malaysia's National Population and Family Development Board, the MFLS-2 survey provides demographic and socioeconomic information on nearly 3,000 households. Although the survey sample is designed to be representative of Peninsular Malaysia, its geographic coverage is limited, with only 174 counties included.

coordinates. I identify Chinese vernacular schools based on their names and cross-reference them with a historical list published in the 1959 Directory of Singapore and Malaya (Ju, 1959), which I georeferenced using the provided address information. To measure geographic segregation between Chinese and other ethnic communities, I use the locations of Chinese vernacular schools as a proxy for Chinese settlements and other national schools as a proxy for non-Chinese settlements.

Migration Data. I measure migration flows from the 1980 census, which tabulates population by place of last residence at the district level (66 districts in total).[17] From this, I construct a matrix of bilateral migration flows between district pairs to estimate migration costs. The 1980 census microdata also includes indicators of internal and external migration, along with the number of years individuals have resided in their current locations.

Buildings Data. I measure built-up volumes from 1975 onwards using data from the Global Human Settlement Layer (GHSL) project. The volumes are calculated using surface and height data at a 100-meter resolution, sourced from Sentinel-2 and Landsat satellite imagery.

IV Empirical Strategy

The resettlement of rural Chinese was not entirely random. In this section, I explain how I isolate the random component of the program to construct a population shifter. I then assess the balance of geographic and pre-resettlement characteristics to provide evidence supporting the identifying assumptions.

IV.A Empirical Specification

My goal is to examine how the increased Chinese population density resulting from the Emergency resettlement impacted local economic outcomes and how these effects differed across ethnic groups. I focus on county-level outcomes because counties are small enough to capture the fine variation in resettlement, yet large enough to account for local spillovers from agglomeration. The key challenge is converting the site-level variation in resettlement into an exogenous population shifter at the county level.

Consider the following reduced-form model:

$$Y_c = \beta ResettleDensity_c + \alpha \mathbf{1}\{ResettleDensity_c > 0\} + \lambda \widehat{ResettleDensity}_c + \gamma X_c + \varepsilon_c. \quad (1)$$

I define county resettlement density as the standardized inverse hyperbolic sine of the total resettled population per county area. This log-like transformation shifts population density

[17] A district is the administrative unit above a county, and each district contains several counties.

in percent terms while accommodating zeros.[18] The coefficient β represents the effect of a one-standard-deviation increase in resettlement density (or "Higher Resettlement").

To ensure comparisons are made only among counties that received resettlement—since those without resettlement were distinct in several ways—I include an indicator for whether a county received any resettlement. Non-resettled counties, such as densely populated areas along the west coast or remote regions deep in the jungle, were typically unsuitable for resettlement and had distinct economic potential.[19]

I control for a set of pre-period characteristics in X_c that were related to the resettlement and could directly affect post-period outcomes. First, I include state fixed effects, as the program was implemented by state governments, each with its own economic and land policies. Second, since areas closer to the transportation network received more resettlement and had better market access, I control for the county's pre-period road density and average distances to roads, rail stations, and the coast. Third, because the program targeted areas with larger initial Chinese populations, which were often more urbanized, I control for 1947 county population density and the Chinese population share. Fourth, since rubber and tin estates were often located on state-owned lands prioritized by the program, I control for pre-period land use shares for rubber plantations and mining. Lastly, I control for county area as county sizes vary across the sample.[20]

Despite the extensive set of covariates in Equation (1), a concern remains that they may not fully account for non-random exposure to resettlement, which depends on a county's position within the broader transportation network and population distribution. For example, two counties with identical road densities could receive different resettled populations if their neighboring counties had distinct road networks, which could affect a county's market access and, in turn, economic outcomes (Donaldson and Hornbeck, 2016). Similarly, counties near more urban areas with larger pre-period, self-selected Chinese populations may differ in location productivity or amenities compared to those near less urbanized areas. If these exposures are not fully captured by X_c, they could introduce omitted variable bias, as highlighted by Borusyak and Hull (2023).

To address this issue, I leverage knowledge of the program to specify counterfactual resettlement.[21] By controlling for the expected resettlement density ($\widehat{ResettleDensity}_c$)—the

[18]This functional form is motivated by efficiency, as it will be used as an instrument for the logarithm of population density when estimating the model's key agglomeration parameters. I explore robustness to using a logarithm transformation with an imputed value for zeros in Appendix Section A.2.

[19]Non-resettled counties are included to improve efficiency, and I will show in Appendix Section A.2 that the results are robust when restricting the analysis to only resettled counties.

[20]Appendix Section A.2 examines robustness to excluding the largest and smallest counties.

[21]Dell and Olken (2020) use a similar approach to identify the impact of proximity to sugar plants established in colonial Java by specifying counterfactual locations for these plants.

average of all possible counterfactual resettlements—I isolate plausibly exogenous variation at the site level (Borusyak and Hull, 2023).[22] Figure 2 illustrates this process using the state of Johor and a single covariate, distance to roads. I approximate $\widehat{ResettleDensity}_c$ by averaging the counterfactual resettlement density across a thousand permutations.

Each permutation is conducted independently for each state as follows.

(i). Randomly (and uniformly) permute counterfactual New Village sites (denoted by i), conditional on (i) distance to roads or rivers;[23] (ii) land use; and (iii) the county's squatter population decile.[24]

(ii). Calculate the counterfactual number of people resettled to each site using the gravity equation:

$$\sum_{j=1}^{J} n_{j\to i} = \sum_{j=1}^{J} n_j \times \frac{d_{ji}^{-\psi}}{\sum_{s=1}^{I} d_{js}^{-\psi}}, \tag{2}$$

where n_j is the initial population of Chinese squatters at origin j, d_{ji} is the distance between origin j and site i, and ψ is the resettlement cost elasticity with respect to distance, which governs how costly it was to relocate people to farther sites.[25]

(iii). Calculate the counterfactual county resettlement density by summing the counterfactual resettled population across sites in each county and dividing by the county area.

The identification assumptions are twofold. First, I assume that the British were equally likely to select resettlement sites with similar proximity to the transportation network and observable characteristics, without targeting locations based on unobserved productivity or amenities. Second, I assume that the British aimed to minimize dislocation; however, as discussed in Section II, poor planning and idiosyncratic communist risks led to longer relocations that were unrelated to unobserved location fundamentals.[26] The identifying variation comes from the specific location of New Villages relative to other comparable sites along the transportation network, as well as from unpredicted distant relocations beyond nearby squatter populations.

[22] Borusyak and Hull (2023) note that controlling for or re-centering by the expected resettlement purge the omitted bias.

[23] If no roads were accessible within a 5-kilometer buffer but a river was, the permutation is conditional on the distance to the nearest river.

[24] I measure the distribution of squatters by overlaying the land-use maps, population census map; and the map of "Black Areas." I define a cluster of Chinese population as a squatter settlement if it was located within the Black Areas and within a 5 kilometer radius of a forest (Appendix Figure A.2).

[25] I calibrate ψ as 0.65 by minimizing the sum of squared differences between the actual and predicted number of people resettled.

[26] For a formal discussion of the identification assumptions, see Appendix A.1.

Figure 3, Panel A maps the New Villages against the expected county resettlement density, highlighting the spatial clustering and overlap between actual resettlement areas and expected density. This pattern aligns with the British strategy of targeting areas with denser road networks and larger pre-existing squatter populations. Panel B shows the identifying variation after residualizing the expected resettlement density and the covariates specified in Equation (1).

For individual or household-level outcomes, I estimate the following regression:

$$Y_{igc} = \beta_g ResettleDensity_c + \alpha_g \mathbf{1}\left\{ResettleDensity_c > 0\right\} + \lambda_g \widehat{ResettleDensity}_c + \gamma_g X_{ic} + \varepsilon_{igc}, \tag{3}$$

where i denotes the individual or household, and g denotes the ethnic group. The control variables, X_{ic}, include the same controls as X_c in Equation (1) and, in some specifications, also include individual or household characteristics.

To estimate the effects of county resettlement density and pre-period Chinese employment share on manufacturing firm outcomes at the county-industry level, I estimate:

$$\begin{aligned} Y_{cj} = & \beta_1 ResettleDensity_c + \beta_2 ResettleDensity_c \times ChineseIndustries_j \\ & + \alpha \mathbf{1}\left\{ResettleDensity_c > 0\right\} + \lambda \widehat{ResettleDensity}_c + \gamma X_c + \delta_j + \varepsilon_{cj}, \end{aligned} \tag{4}$$

where $ChineseIndustries_j$ indicates high Chinese exposure, defined as industries where the 1947 Chinese employment share exceeded 80%, and δ_j denotes industry fixed effects to absorb nationwide industry-specific shocks. The coefficient β_1 represents the impact of a one-standard-deviation increase in resettlement density on manufacturing outcomes in industries with low Chinese exposure.[27] The coefficient β_2 captures the additional effect in industries with high Chinese exposure.

To estimate percentage changes for variables like the number of establishments or employment, which may be zeros for some counties or industries, I use the Poisson Pseudo Maximum Likelihood (PPML) estimator (Silva and Tenreyro, 2006). This estimator captures both the extensive and intensive margins, remains unaffected by the unit of the outcome variables, and provides effects relative to the baseline mean (Chen and Roth, 2023).

I report Conley standard errors that account for spatial correlation within a 30-kilometer radius (Conley, 1999). The 30-kilometer cutoff is based on the localized nature of resettlement, typically within 20 miles, beyond which the treatment can be considered independent. The baseline Conley standard errors are comparable to those clustered by district (across 66

[27] Low Chinese exposure industries include food products, woods products, textiles, and other miscellaneous products; see Appendix Figure A.6.

districts), and are 10–15% higher with distance cutoffs up to 50 kilometers.

IV.B Pre-Characteristic Balance

This section examines the balance of location fundamentals and county characteristics measured before resettlement. If the procedure effectively isolates the as-good-as-random component of the resettlement program, we would expect the residualized county resettlement density to be orthogonal to pre-period characteristics.

Table 1 reports the relationship between county resettlement density and various location characteristics. Columns 1–4 focus on geography, such as elevation, ruggedness, and suitability for rice and coconut cultivation—the primary food crops in Malaysia. Columns 5–8 examine access to amenities and public goods, including the distance to the nearest police station, post/telegraph office, hospital, and Chinese temple. Columns 9–12 examine pre-period economic activities, including land use shares for rubber plantations and mining—the two major export industries in British Malaya—as well as proximity to industrial facilities and major commercial or administrative centers such as Singapore, George Town, Malacca, Ipoh, and Kuala Lumpur.[28]

Panel A shows that the raw correlations within each state align with the resettlement strategy, which prioritized areas along transportation networks. Counties with higher resettlement were located closer to public goods (Columns 5–7), industrial factories, and major cities (Columns 11–12). The prevalence of rubber plantations (Column 9) is also consistent with historical records that many resettlement areas were on state-owned rubber estates.

Panel B shows that after controlling for key resettlement covariates, such as road networks and initial population distributions, location characteristics are generally balanced. Panel C shows that adding expected resettlement density to the regression does not significantly change the estimates, suggesting that the potential omitted variable bias from broader network effects is largely accounted for by the county covariates. An exception is the land share for rubber, which becomes more balanced after conditioning on expected resettlement. In the main specification, I control for land-use shares for both rubber and mining. Overall, the magnitudes of the estimates are small—for instance, a one-standard-deviation increase in resettlement density corresponds to just a 16-meter increase in elevation. These results support the plausibility of the identification assumptions.

Notably, counties with higher resettlement do not show higher agricultural suitability—in fact, they appear to be less suitable—despite agricultural potential being a factor considered in selecting resettlement sites (Columns 3–4). This finding aligns with historical accounts

[28] Industrial facilities include key strategic industries such as airplane and automotive repair facilities, engineering shops, shipyards, chemical plants, power plants, rubber and tin plants, food and clothing manufacturers, among others.

that economic considerations were secondary to security and speed in the program.

V Results

This section examines the local effects of resettlement in receiving areas over the following five decades. I begin by showing how the resettlement led to persistent shifts in population distribution. I then present the economic impacts and discuss their implications.

V.A Population Growth and Changes in Ethnic Composition

Figure 4 shows the estimates from Equation (1) on county population growth (Panel A) and changes in the Chinese population share (Panel B). By 1957, shortly after resettlement was largely completed, counties with higher resettlement saw a significant increase in the Chinese population, with no notable changes among other ethnic groups. This led to a lasting rise in the Chinese population share that continues to the present. Importantly, there were no significant pre-period population changes in counties that later experienced higher resettlement, supporting the assumption that these areas did not initially have faster growth or higher labor demand.

Table 2 shows that from 1947 to 1957, counties with one standard deviation higher resettlement density experienced a 9.4% increase in overall population density. Since a one standard deviation increase in resettlement density corresponds to 13.6% of the 1947 population, this implies that for every 1% of the population resettled, local population grew by 0.69% during this period. This growth accounted for 77% of the total population increase in these counties, with the rise driven by an influx of Chinese, resulting in a 4.8 percentage point increase in the Chinese population share (Column 4).

After 1960, when mobility restrictions were lifted, counties with higher resettlement continued to experience population growth while maintaining a larger share of Chinese residents. By 1980, counties with one standard deviation higher resettlement density had 11% greater overall population density and a 5 percentage point higher Chinese share (Table 2, Column 2). Between 1980 and 2000, these counties saw additional population growth across all ethnic groups, while the higher Chinese composition persisted (Columns 3 and 6). Post-Emergency population growth in the more resettled counties was driven by internal migration rather than higher fertility. Chinese residents in these counties were 5 percentage points (12%) more likely to be internal migrants who voluntarily moved in after 1960, with no increase in fertility rates (Appendix Table A.4).[29]

The denser population in the more resettled areas was accompanied by a greater increase in build-up volumes (Appendix Table A.3). By 1975, counties with higher resettlement had

[29]Interestingly, the fertility rate among non-Chinese women was slightly lower in more resettled areas.

33% more buildings.[30] The larger percentage increase in buildings compared to population growth suggests a relatively elastic housing supply.

V.B Economic Structure

The influx of Chinese constituted a skill-biased labor supply shock, given their historical specialization in industrial and urban sectors. This section examines how resettlement shaped the local economic structure in the decades following the Emergency.

Table 3 shows differences in employment structure in the 1980s for counties with one standard deviation higher resettlement density. Employment in the primary sector (agriculture and mining) was 11% higher (Column 1), while non-agricultural employment was 29% higher (Column 2).[31] The nearly three times larger effect on non-agricultural employment reflects both the Chinese population's tendency to work in the non-agricultural sector (Panel B) and a shift of Malays out of agriculture (Panel C). Within non-agricultural industries, employment increases in the secondary and tertiary sectors were similar (Appendix Table A.5).[32]

On the firm side, Table 4 shows that counties with higher resettlement had more and larger manufacturing establishments in industries that historically employed a higher share of Chinese workers. Column 1 reports that counties with one standard deviation higher resettlement density had 21% more manufacturing establishments in industries where the majority of pre-Emergency employment was Chinese.[33] Column 2 shows that these manufacturing firms in Chinese-dominated industries were also larger, with 2 percentage points (5%) more likely to have at least one full-time employee.[34]

V.C Specialization, Human Capital Accumulation, and Household Income

The resettlement of the Chinese population shifted local labor markets toward the non-agricultural sector, potentially generating productivity gains through agglomeration externalities. However, the extent to which non-Chinese populations benefited from the increased Chinese density may depend on their level of interaction with the Chinese community. This section explores these mechanisms in more detail.

Table 5 shows that, in 1980, Chinese individuals in counties with higher resettlement

---

[30] Satellite images reveal more clustering of buildings around New Villages, despite relatively uniform settlement patterns in surrounding areas before the Emergency (Appendix Figure A.5).

[31] Throughout the paper, I use "agriculture" and "primary sector" (or "non-agriculture" and "non-primary sector") interchangeably.

[32] The secondary sector includes manufacturing, utilities, and construction. The tertiary sector includes wholesale/retail trade, transport, communication, finance, and services.

[33] The only four manufacturing industries with less than 80% Chinese employment shares were food products, wood products, textiles, and miscellaneous manufacturing. See Appendix Figure A.6 for the Chinese employment shares of each manufacturing industry.

[34] Less than half of the manufacturing establishments in 1970 employed full-time workers.

were more likely to participate in the labor market (Panel A). Additionally, those who participated exhibited greater occupational and industrial specialization, as measured by the Herfindahl-Hirschman Index (HHI) (Panels B and C).[35] In contrast, non-Chinese workers in more resettled counties showed no significant differences in these labor market outcomes (Column 2). These patterns align with labor market pooling as a channel of agglomeration, where a denser labor market improves the quality of matches between workers and firms, thereby increasing returns to formal labor market participation and specialization (Marshall, 1890).

Table 6 shows that counties with higher resettlement density had greater educational attainment. In counties with one standard deviation higher resettlement density, Chinese individuals in 1980 had, on average, 0.4 additional years of schooling (7.7%), were 3.6 percentage points (6%) more likely to complete primary education, and 3.9 percentage points (14%) more likely to complete secondary education (Column 1). Non-Chinese individuals also experienced improvements in educational attainment, but the effects were smaller (Column 2). The increase in education among the Chinese was primarily driven by younger cohorts under age 50, who had not completed their education by the time of resettlement (Appendix Table A.6). In contrast, resettlement density had no significant effect on educational outcomes for individuals aged 50 or older, suggesting that more resettled counties were not initially more industrialized or populated by a better-educated workforce during the colonial period. These results are consistent with the idea that a greater division of labor increases the returns to acquiring higher and more specialized education (Kim, 1989).

The benefits of agglomeration translated into higher household incomes. Table 7 shows that, in 1980, Chinese households in counties with one standard deviation higher resettlement density had income 11% higher than that of Chinese households in less resettled areas (Panel A, Column 1).[36] For non-Chinese households, the effect was smaller, with a statistically insignificant 3.7% increase compared to non-Chinese households in less resettled areas (Column 2). As a result, the income difference between Chinese and other ethnic groups increased by 7.3% in these counties (Column 3). Panels B and C break down the results by households employed in the primary and non-primary sectors, based on the industry of the household head. In counties with higher resettlement, Chinese households earned 7.3% more in the primary sector and 12.1% more in the non-primary sector (Column 1). In con-

[35] Formally, the HHI index of occupation for ethnic group e in county n is defined as $HHI^{e}_{n,occ} \equiv \sum_k (L^e_{nk}/L^e_n)^2$, where L^e_{nk} denotes total employment of group e in occupation k, county n; and $L^e_n = \sum_k L^e_{nk}$. This index captures the extent to which local employment of an ethnic group was concentrated in a few occupations or spread across many.

[36] For the impact on household asset ownership—the main predictors for household income—see Appendix Table A.7.

trast, non-Chinese households saw income gains only in the non-primary sector, with a 4.4% increase that was not statistically significant (Column 2).

Geographic segregation between ethnic groups accounts for half of the smaller agglomeration benefits observed for the non-Chinese population. Table 8 shows that for every additional kilometer of separation between these communities, the income gains for the non-Chinese population dropped by one-fifth, with no positive effects beyond 5 kilometers. Even in geographically integrated areas, however, the non-Chinese population saw only half the income gains from resettlement compared to their Chinese counterparts. This suggests that deeper factors, such as ethnic tensions and isolated social networks, may have impeded productive interactions between groups and contributed to ongoing segregation.

V.D Alternative Mechanisms and Robustness

The Chinese-specific patterns—higher labor force participation, greater specialization, higher education, and higher income—are consistent with agglomeration economies from labor market pooling and barriers to cross-group interactions. However, the Emergency resettlement may have influenced factors beyond the distribution of Chinese population that affected economic outcomes and potentially benefited the Chinese community more than others. This section explores alternative mechanisms and evaluates the robustness of the results.

One alternative explanation for the Chinese-specific outcomes is that the resettled squatters were initially more industrial and productive than the existing Chinese population, driving the income premium without agglomeration benefits. However, this is unlikely, as the squatters were more likely to have been agricultural, having self-selected into rural areas before the resettlement. Historical accounts suggest that 60% of resettled squatters were agriculturalists (Sandhu, 1964, p. 169). Data from the Second Malaysian Family Life Survey (1988) further supports this, showing that resettled Chinese were more likely to have started in agriculture and were less educated than other Chinese of similar age in the same state (Appendix Table A.8).[37]

Land ownership might also explain the better outcomes for Chinese in resettled areas, as resettled families were allocated house lots, with those who were previously farmers additionally receiving agricultural land (Sandhu, 1964).[38] However, data shows that resettled Chinese households actually owned *less* land than other Chinese and non-Chinese households in the same state, suggesting that land ownership was not the primary driver of better outcomes (Appendix Table A.8).[39]

[37]The Second Malaysian Family Life Survey (1988–1989) includes migration history. I identified 64 resettled Chinese as those who were not born in a New Village but had "migrated" there before 1960.

[38]Resettled families were typically allotted 1/6 of an acre for house lots and an additional 2 acres of agricultural land if they had been farmers, with lease terms of 20–30 years depending on the State.

[39]Many resettled Chinese received land titles much later, and some were unaware of their land rights

Another possibility is that access to Chinese schools in New Villages reduced education costs, potentially improving educational attainment and income. However, the data shows that counties with higher resettlement did not have better access to Chinese schools; in fact, access to non-Chinese national schools was slightly better in these counties (Appendix Table A.9). Therefore, differences in school availability are unlikely to explain the disparities in education and income between groups in more resettled areas.

Furthermore, the Chinese income premium in more resettled counties remains, though smaller and less precisely estimated, even after controlling for the household head's years of schooling—a factor likely endogenous to Chinese agglomeration (Appendix Table A.10). While part of the income premium in the industrial sector can be attributed to higher education levels, its persistence in the agricultural sector—potentially due to linkages with downstream Chinese-owned firms that purchased and processed agricultural products—suggests that ethnicity played a significant role beyond employment in high-paying sectors.

Appendix Section A.2 examines the robustness of the results across various alternative specifications and sample restrictions. First, the results are robust to using a logarithmic transformation of the population shifter with imputed values for zeros. Second, alternative specifications of counterfactual resettlements generates similar estimates. Third, controlling for neighboring road and population characteristics, as well as pre-period proximity to industrial and urban areas, yields similar results. The estimates are also robust to excluding counties with the largest or smallest areas, high-density prewar towns, and counties with extreme resettlement densities.

V.E Discussion

In summary, the resettlement of Chinese during the Malayan Emergency had a lasting impact on population distribution and economic activity across regions, sectors, and ethnic groups. Industries that demanded skills specialized by Chinese labor flourished in areas with higher Chinese density, generating positive externalities for local workers and firms. The increased productivity and wages, particularly in the non-agricultural sector, spurred internal migration and a structural change of the local economy.

While neoclassical theory suggests that the marginal workers entering the non-agricultural sector after resettlement would be less productive than those already in the sector, thereby lowering wages, the opposite occurred: non-agricultural wages increased, and the share of agricultural employment declined. This shift points to external economies of scale in the non-agriculture sector. In contrast, despite the exit of less productive workers from agriculture, overall agricultural productivity did not improve in counties with higher resettlement,

(Strauch, 1981, pp. 63–72).

suggesting local diminishing returns to scale in agriculture due to the fixed land supply.

The distinct outcomes for Chinese in resettled counties align with the benefits of labor market pooling, likely lowering search costs and improving job matching, thereby raising the returns to specialization. Knowledge sharing and spillovers may have also been greater among the local Chinese. However, these benefits were largely confined to the Chinese population, suggesting cross-ethnic frictions. The New Villages were ethnically segregated, limiting opportunities for intergroup interactions. Moreover, cultural and religious differences, long-standing tensions, and language barriers likely hindered economic and social interactions between Chinese and Malays (Greif, 1993; Fehr et al., 1997; Hjort, 2014).

These cross-ethnic barriers suggest that ethnic composition affects local productivity externalities. A key implication of the spatial variation in agglomeration externalities, as noted by Glaeser and Gottlieb (2008), is that placed-based policies like population resettlement can have broader aggregate effects. For example, relocating Chinese populations from areas with lower concentrations of Chinese—where spillovers to Chinese are weaker—to areas with higher concentrations—where spillovers are stronger—could increase overall productivity if the gains at the destination exceed the losses at the origin. The same logic applies to industries: reallocating labor from sectors with smaller external economies of scale to those with larger external economies would similarly raise aggregate productivity—a common justification for industrial policies.

However, the cross-sectional analysis shows only relative impacts across counties and does not capture the program's aggregate effects. Moreover, the comparison holds exogenous location fundamentals constant without considering that the program typically relocated people from remote areas to better-connected locations along transportation networks. A general equilibrium model is needed to interpret the empirical results and assess the overall impact of resettlement. Such a model would also provide a framework for evaluating counterfactual policies aimed at promoting economic development, particularly in the context of heterogeneous agglomeration forces across different groups and sectors.

VI A Quantitative Model of Migration, Occupation, and Agglomeration

I develop a spatial general equilibrium model that extends the work of Allen and Donaldson (2020) and Peters (2022), incorporating agglomeration forces that vary by industry and local ethnic composition. To capture the key adjustments following the resettlement shock, the model includes migration and occupation choices in a Roy (1951)-type framework. Individuals from different ethnic groups have heterogeneous regional preferences and sectoral comparative advantages. Regions are linked through migration and trade, with the resettlement shock shifting the initial population distribution and shaping the longer-term distribution due to movement costs.

VI.A Environment

The model features N regions and two sectors $k \in \{A, M\}$: Agriculture (A) and Manufacturing (M). Individuals are characterized by two ethnic groups $e \in \{c, m\}$: Chinese (c) and Malays (m), and are initially endowed with a location. They decide where to migrate after drawing a regional taste shock. After moving, they draw idiosyncratic productivity for each sector and choose between working in agriculture or manufacturing. Finally, consumption and production take place.

Production. Each region n produces a unique good in both sectors A and M, following Armington (1969). In each sector-region, a continuum of perfectly competitive firms produces a homogeneous regional variety. The production technology exhibits constant returns to scale, with labor as the only input. The regional production function is $Q_{nk} = H_{nk}$, where H_{nk} denotes the total labor, measured in efficiency units (defined later), employed in sector k of region n and summed across ethnic groups. Labor from Chinese and Malay workers is assumed to be perfectly substitutable in the production function.

Firms in sector k and region n choose labor H_{nk} to maximize profit, taking the local sectoral wage (per efficiency unit) and prices as given. In equilibrium, the no-arbitrage condition implies that $p_{nrk} = (\tau_{nr}/\tau_{nn})p_{nnk}$, where p_{nrk} is the price of goods produced in sector k in region n and sold in region r; $\tau_{nr} \geq 1$ is the iceberg trade cost between regions n and r; p_{nnk} is the price of sector-k goods sold locally; and τ_{nn} is the within-region trade cost.[40] Under perfect competition, firms earn zero profit in equilibrium, leading to the condition $w_{nk} = p_{nnk}/\tau_{nn}$, where w_{nk} is the wage per efficiency unit in sector k of region n.

Consumption. Individuals of ethnicity e living in region n derive utility from consuming both agricultural and manufacturing goods, as well as from enjoying the amenity specific to location n. Their utility function is:

$$U_n^e(C_A, C_M) = a_n^e \left(\frac{C_A}{\alpha}\right)^{\alpha} \left(\frac{C_M}{1-\alpha}\right)^{1-\alpha}$$
$$C_k = \left(\sum_{r=1}^{N} c_{rk}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}},$$

where $\sigma > 1$ denotes the constant elasticity of substitution across regional varieties, assumed to be the same for both sectors, and a_n^e captures the (endogenous) amenity value for

[40] Without loss of generality, within-region trade can be costly ($\tau_{nn} > 1$), meaning local wages decrease as τ_{nn} increases, which is isomorphic to worse location fundamentals.

individuals of ethnicity e in location n.[41]

Utility maximization implies that the indirect utility of an individual from group-e in region n with income y_n^e is given by $a_n^e y_n^e / P_n$, where $P_n \equiv P_{nA}^{\alpha} P_{nM}^{1-\alpha}$ is the ideal price index in region n. $P_{nk} \equiv (\sum_{l=1}^{N} \tau_{ln}^{1-\sigma} w_{lk}^{1-\sigma})^{1/(1-\sigma)}$ is the price index for sector-k goods in region n.

Migration. Each individual starts with an initial location and decides where to migrate, subject to a moving cost, after drawing a regional taste shock. At the time of migration, individuals know their ethnicity but have not yet observed their skill realizations. An individual i of ethnicity e draws an idiosyncratic taste for region n, denoted u_{in}^e, from a Fréchet distribution:

$$F_n^e(u) = \exp\left(-\bar{a}_n^e u^{-\nu}\right),$$

where $\bar{a}_n^e$ is the scale parameter capturing any exogenous, ethnicity-specific amenity in region n, and ν is the shape parameter, with a higher value indicating less variation in tastes across regions.

Individual i from group e values the amenities in region n, denoted a_{in}^e, based on both an idiosyncratic taste u_{in}^e and the local population distribution:

$$a_{in}^e = u_{in}^e f_a(L_n^e, L_n^{e'}),$$

where $f_a(\cdot)$ is parameterized as a function of total population size and ethnic composition:

$$f_a(L_n^e, L_n^{e'}) = (L_n)^{\beta} \left(\frac{L_n^e}{L_n}\right)^{\beta^e}.$$

The parameters β and β^e capture congestion effects and amenity spillovers. Specifically, β^e allows for homophily, a preference for living near others of the same ethnicity, which may also reflect ethnic tensions between groups.

The indirect utility for individual i of ethnicity e from origin r living in destination n, with an idiosyncratic preference a_{in}^e, is given by:

$$V_{irn}^e = \eta_{rn}^{-1} a_{in}^e \Gamma_\theta \bar{w}_n^e P_n^{-1},$$

where η_{rn} denotes the iceberg migration cost from r to n, and $\Gamma_\theta \bar{w} n^e P n^{-1}$ is the real wage in region n, with $\Gamma_\theta \equiv \Gamma(1 - 1/\theta)$ and $\Gamma(\cdot)$ denoting the Gamma function.

Since the indirect utility is a Fréchet random variable u_{in}^e multiplied by a constant

[41] Technically, a_n^e should include an individual subscript, such as a_{in}^e, where i denotes the individual, since it varies idiosyncratically across individuals. For simplicity, individual subscripts are omitted here.

$\eta_{rn}^{-1} L_n^{\beta} (L_n^e/L_n)^{\beta^e} \Gamma_\theta \bar{w}_n^e P_n^{-1}$, it is itself Fréchet distributed. This implies that the share of ethnicity e initially residing in region r who choose to migrate to n is given by:

$$m_{rn}^e = \frac{\left(\eta_{rn}^{-1} V_n^e\right)^\nu}{\sum_{l=1}^N \left(\eta_{rl}^{-1} V_l^e\right)^\nu},$$

where the mean utility of residing in region n for ethnicity e is

$$V_n^e = \left(\bar{a}_n^e\right)^{1/\nu} \left(L_n\right)^\beta \left(\frac{L_n^e}{L_n}\right)^{\beta^e} \bar{w}_n^e P_n^{-1}. \tag{5}$$

Thus, the bilateral migration flow of ethnic group e from r to region n is

$$L_{rn}^e = \eta_{rn}^{-\nu} \times \frac{\check{L}_r^e}{(\Pi_r^e)^\nu} \times \frac{L_n^e/\bar{L}}{(\mathcal{V}_n^e)^{-\nu}}, \tag{6}$$

where I define two migration market access terms as:

$$\Pi_r^e \equiv \left(\sum_{l=1}^N (\eta_{rl}^{-1} V_l^e)^\nu\right)^{1/\nu}, \tag{7}$$

$$\mathcal{V}_n^e \equiv V_n^e \left(L_n^e/\bar{L}\right)^{-1/\nu}. \tag{8}$$

Here, $\bar{L}$ is the total population in the country, normalized to 1.

The term Π_r^e represents the overall value for group e to move *out of* region r, while $\mathcal{V}_n^e$ captures the value of moving *into* region n. These terms are referred to as the outward or inward migration market access in the trade literature (Anderson and van Wincoop, 2003).

Sectoral Labor Supply. Individuals with heterogeneous productivity earn income by inelastically supplying one unit of labor. After migration, each individual draws a vector of efficiency units in sectors A and M, denoted as $\Lambda_i = (\Lambda_{iA}, \Lambda_{iM})$, where Λ_{ik} represents the effective labor individual i provides if she works in sector k. These efficiency units are drawn independently from a Fréchet distribution:

$$F_{nk}^e(\Lambda) = \exp\left(-\phi_{nk}^e \Lambda^{-\theta}\right),$$

where the scale parameter ϕ_{nk}^e captures the average productivity of ethnicity e in sector k and region n. This parameter reflects both the absolute and comparative advantage of different ethnic groups, as well as fixed location fundamentals that make a region more productive in specific sectors. The shape parameter θ governs the dispersion of efficiency units, with a higher θ indicating less dispersion in productivity.

Due to human capital externalities, individuals from ethnic group e in region n and sector k has net efficiency units λ^e_{ink} that depend on both their own skill Λ^e_{ink} and the local population distribution:

$$\lambda^e_{ink} = \Lambda^e_{ink} f_\lambda(L^e_{nk}, L^{e'}_{nk}),$$

where I parameterize $f_\lambda(\cdot)$ as a function of sectoral population size and ethnic composition:

$$f_\lambda(L^e_{nk}, L^{e'}_{nk}) \equiv (L_{nk})^{\gamma_k} \left(\frac{L^e_{nk}}{L_{nk}}\right)^{\gamma^e}.$$

The parameters γ_k and γ^e govern the strength of productivity spillovers from local interactions, which depend on the number of workers in the sector via γ_k and on ethnic composition of local workforce via γ^e. Specifically, the elasticity of productivity with respect to the concentration of ethnic group e in region n is given by:

$$\frac{\partial \ln \lambda^e_{nk}}{\partial \ln L^e_n} = \underbrace{\left(\gamma_k \frac{L^e_{nk}}{L_{nk}} + \gamma^e \left(1 - \frac{L^e_{nk}}{L_{nk}}\right)\right)}_{\text{Direct effect}} \underbrace{\left(1 + \frac{\partial \ln \pi^e_{nk}}{\partial \ln L^e_n}\right)}_{\text{Indirect/GE effect}}, \tag{9}$$

where π^e_{nk} is the share of ethnic group e working in sector k in region n.

A larger population of ethnic group e working in sector k directly affects that group's productivity, assuming their occupational structure remains fixed, and indirectly through changes in the occupational share. The firm term in Equation (9) is a weighted average of γ_k and γ^e, where the weights correspond to the share of workers in sector k from ethnicity e. If group e dominates sector k in region n (i.e., L^e_{nk}/L_{nk} is large), the agglomeration elasticity for group e approaches γ_k. Conversely, if group e' dominates the local sector, the elasticity for group e approaches γ^e. The indirect effect scales the agglomeration elasticity based on how the larger size of group e influences its local occupational share, driven by changes in relative wages in the general equilibrium.

There are also cross-ethnic productivity spillovers. For ethnic group $e' \neq e$, the elasticity of group e''s efficiency with respect to the concentration of group e is given by:

$$\frac{\partial \ln \lambda^{e'}_{nk}}{\partial \ln L^e_n} = \underbrace{(\gamma_k - \gamma^e)\frac{L^e_{nk}}{L_{nk}}}_{\text{Direct effect}} \underbrace{\left(1 + \frac{\partial \ln \pi^{e'}_{nk}}{\partial \ln L^e_n}\right)}_{\text{Indirect/GE effect}}. \tag{10}$$

Similar to within-ethnic spillovers, cross-ethnic spillovers from e to e' have both direct and indirect components. When $\gamma_k > \gamma^e$, the direct effect in sector k tends to be positive, and vice versa. The magnitude of the spillover is also proportional to the share of group e in sector k.

Given the Fréchet distributed efficiency units Λ_k^e, the share of individuals of ethnicity e in region n who choose to work in sector k is given by:

$$\pi_{nk}^e = \phi_{nk}^e \left(\frac{w_{nk}^e}{\bar{w}_n^e} \right)^\theta , \tag{11}$$

where

$$w_{nk}^e = w_{nk} \left(L_{nk} \right)^{\gamma_k} \left(\frac{L_{nk}^e}{L_{nk}} \right)^{\gamma^e} , \tag{12}$$

and the average wage (up to a scale) for ethnicity e in region n is:

$$\bar{w}_n^e = \left(\phi_{nA}^e \left(w_{nA}^e \right)^\theta + \phi_{nM}^e \left(w_{nM}^e \right)^\theta \right)^{1/\theta} .$$

Trade. Bilateral trade flows from region n to region r incur an exogenous iceberg trade cost, $\tau_{nr} \geq 1$, where $\tau_{nr} = 1$ represents frictionless trade. Given consumer preferences, the trade flow expenditure on sector-k goods from r to n (with goods flowing from n to r), denoted by X_{nrk}, follows a standard gravity form:

$$X_{nrk} = X_{rk} \frac{\tau_{nr}^{1-\sigma} (w_{nk})^{1-\sigma}}{\sum_{l=1}^N \tau_{lr}^{1-\sigma} (w_{lk})^{1-\sigma}},$$

where $X_{rk} = \alpha_k Y_r$ is the total expenditure of region r on sector-k goods, with $\alpha_A = \alpha$ and $\alpha_M = 1 - \alpha$. The total income of region r is given by $Y_n = w_{rA} H_{rA} + w_{rM} H_{rM}$.

This equation can be rewritten as:

$$X_{nrk} = \alpha_k \tau_{nr}^{1-\sigma} \times \frac{Y_n / \bar{Y}}{\mathcal{P}_{nk}^{1-\sigma}} \times \frac{Y_r}{P_{rk}^{1-\sigma}}, \tag{13}$$

where I define two trade market access terms, similar to migration flows:

$$P_{rk} \equiv \left(\sum_{l=1}^N \tau_{lr}^{1-\sigma} w_{lk}^{1-\sigma} \right)^{1/(1-\sigma)} , \tag{14}$$

$$\mathcal{P}_{nk} \equiv w_{nk}^{-1} \left(Y_n / \bar{Y} \right)^{1/(1-\sigma)} . \tag{15}$$

Here, $\bar{Y} \equiv \sum_r Y_r$ is the total income of the economy, which is normalized to one as the numeraire.

As with migration flows, (the inverse of) P_{rk} represents the inward trade market access for sector-k goods in region r, while (the inverse of) $\mathcal{P}_{nr}$ represents the outward trade market access for sector-k goods from region n.

VI.B Static Equilibrium

Given any strictly positive initial population vector $\{\check{L}_r^e\}$ and a set of location fundamentals $\{\phi_{nk}^e, \bar{a}_n^e, \tau_{nr}, \eta_{nr}\}$, an equilibrium is a vector of prices $\{w_{nk}, p_{nk}\}$ and quantities $\{L_{nk}^e, H_{nk}\}$, such that (i) firms and consumers act optimally, and (ii) both goods and labor markets clear in all regions.

The goods market clearing condition is:

$$w_{nk}H_{nk} = \sum_{r=1}^{N} \alpha_k (w_{rA}H_{rA} + w_{rM}H_{rM}) \frac{\tau_{nr}^{1-\sigma} w_{nk}^{1-\sigma}}{\sum_{l=1}^{N} \tau_{lr}^{1-\sigma} w_{lk}^{1-\sigma}}. \tag{16}$$

The labor market clearing condition is:

$$H_{nk} = \sum_e H_{nk}^e = \sum_e L_n^e \pi_{nk}^e \left(\Gamma_\theta \bar{w}_n^e w_{nk}^{-1}\right) \tag{17}$$

$$L_n^e = \sum_r \check{L}_r^e \frac{\left(\eta_{rn}^{-1} V_n^e\right)^\nu}{\sum_{l=1}^{N} \left(\eta_{rl}^{-1} V_l^e\right)^\nu}. \tag{18}$$

Equation (16) embeds two underlying conditions: (i) total sectoral sales in each region equal labor payments, and (ii) a region's total income is fully spent on goods from all locations. Equation (17) indicates that a region's total efficiency units in sector k are the sum of the contributions from both ethnic groups. The contribution from group e is the product of their sectoral employment ($L_n^e \pi_{nk}^e$) and their average efficiency units ($\Gamma_\theta \bar{w}_n^e w_{nk}^{-1}$). Equation (18) comes from the migration flow identity, which states that the equilibrium population of ethnicity e in a region is the sum of migration flows of ethnicity-e individuals from all regions.

Using equations (5), (11), and (12), we can substitute out V_n^e, π_{nk}^e, and $\bar{w}_n^e$, replacing them with exogenous parameters and endogenous outcomes $\{w_{nk}, L_n^e\}$. The equilibrium is then characterized by a system of $6 \times N$ equations (16–18) in $6 \times N$ unknowns $\{w_{nk}, H_{nk}, L_n^e\}$.

Existence and Uniqueness. I prove the existence of equilibrium by construction, using an iterative procedure described in Appendix B.3. The process involves three nested loops: the outer loop solves for population distribution, the second loop solves for sectoral wages and prices given the population distribution, and the inner loop solves for occupational shares given population and wages. Convergence is ensured by congestion forces like idiosyncratic migration preferences, sectoral productivity, and imperfect substitution across regional varieties.

When agglomeration forces (γ_k and γ^e) outweigh congestion forces, the model may exhibit multiple equilibria, where economic activity concentrates in one set of location-sector in one equilibrium and in a different set in another. I verify the uniqueness of equilibrium at baseline parameters by starting the algorithm from different initial values and confirming convergence to the same outcomes.[42] Note that equilibrium is conditional on a given initial population, and uniqueness here refers to having at most one equilibrium given an initial population.

VII Identification and Estimation

To estimate the model, I use the 1957 population distribution—observed after most of the resettlement had been completed—as the initial population, and treat the 1980 data as the equilibrium outcomes.

I make a set of parametric assumptions for migration and trade costs. I assume symmetric bilateral migration and trade costs that increase with distance. Migration costs are modeled as $\eta_{rn} = (d_{rn}/d_{min})^{\kappa}$, where d_{min} is the minimum within-county distance, and $\kappa \geq 0$ is the the distance elasticity of migration costs. Similarly, trade costs are modeled as $\tau_{nr} = (d_{rn}/d_{min})^{\xi}$, where $\xi \geq 0$ is the distance elasticity of trade costs.[43]

The model is characterized by a tuple of location fundamentals $\{\phi^e_{nk}, \bar{a}^e_n\}$ and 11 structural parameters:

$$\Theta \equiv \{ \underbrace{\alpha, \sigma}_{\text{Preference}} , \underbrace{\xi, \kappa}_{\text{Trade/Migration}} , \underbrace{\theta, \gamma_A, \gamma_M, \gamma^e}_{\text{Productivity}}, \underbrace{\nu, \beta, \beta^e}_{\text{Amenity}} \}.$$

I externally set or calibrate three parameters: the elasticity of substitution across regional varieties (σ), the distance elasticity of trade cost (ξ), and the migration elasticity (ν). Estimates of σ in the literature typically range from 4 to 9, and I set $\sigma = 8$ based on recent estimates from Vietnam (Balboni, forthcoming).[44] Since I do not have direct data on trade flows within Malaysia, I follow Monte et al. (2018) and set $\xi(1-\sigma) = -1.29$, which implies, given $\sigma = 8$, a distance elasticity of trade cost of $\xi = 0.18$. While estimates of ν are rare, particularly for developing countries, existing studies suggest values between 2 and 4.[45] I set

[42] I also explore deriving sufficient conditions for uniqueness following Allen et al. (forthcoming), but this approach did not yield informative conditions (see Appendix B.4).

[43] Cross-county distances d_{rn} (for any $r \neq n$) are measured by the Euclidean distance between centroids, while within-county distances d_{rr} are calculated from the centroid to the boundary. I allow within-county costs to exceed 1 (except for the smallest county, which is normalized to 1) to account for varying county sizes. This normalization is without loss of generality, because higher migration costs reduce utility as if amenities $\bar{a}n^e$ were worse, and costly trade lowers productivity similarly to a reduction in ϕnk^e.

[44] Donaldson and Hornbeck (2016) estimate a value of 9.22 in 19th century U.S., Peters (2022) estimates 5.02 in post-war Germany, and Balboni (forthcoming) estimates 7.92 in Vietnam in 2009.

[45] See, for example, Monte et al. (2018); Morten and Oliveira (2024); Bryan and Morten (2019); Tombe

$\nu = 3$ as my baseline.

The remainder of the section is structured as follows. I first discusses how I identify and estimate the remaining eight parameters. I then discuss the estimation results and compare them to the existing literature.

VII.A Identification

I begin by introducing a proposition that establishes the identification of the market access terms and the agricultural expenditure share α. I then discuss the identification of the remaining model parameters and the recovery of location fundamentals.

Market access terms. From the equilibrium conditions, I derive four key relationships involving trade and migration market access terms: (i) total sales equals labor payments; (ii) total income equals total expenditure; (iii) final population equals total in-migration; and (iv) initial population equals total out-migration.[46] These conditions yield the following system of equations:

$$\mathcal{P}_{nk}^{1-\sigma} = \frac{\alpha_k}{\Omega_{nk}} \sum_r \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1}, \tag{19}$$

$$P_{rk}^{1-\sigma} = \sum_n \tau_{nr}^{1-\sigma} Y_n \mathcal{P}_{nk}^{\sigma-1}, \tag{20}$$

$$(\mathcal{V}_n^e)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e \left(\Pi_r^e\right)^{-v}, \tag{21}$$

$$(\Pi_r^e)^{v} = \sum_n \eta_{rn}^{-v} L_n^e \left(\mathcal{V}_n^e\right)^{\nu}, \tag{22}$$

where $\Omega_{nk} \equiv w_{nk} H_{nk} / Y_n$ is the share of income in region n generated by sector k.

Proposition 1. *Given observed data on* $\{Y_n, \Omega_{nk}, \check{L}_n^e, L_n^e\}$ *and parameter values* $\{\tau_{nr}^{1-\sigma}, \eta_{nr}^{-\nu}\}$, *there exists a unique scalar* α *and a set of values (up to scale) for* $\{\mathcal{P}_{nk}^{\sigma-1}, P_{rk}^{\sigma-1}, (\mathcal{V}_n^e)^{\nu}, (\Pi_r^e)^{v}\}$ *that satisfy equations* (19)–(22).

Proof. See Appendix C.1. □

This proposition shows that the market access terms can be identified (up to scale) without needing to know the agglomeration parameters $\gamma_k, \gamma^e, \beta, \beta^e$, even in the presence of multiple equilibria.

Migration cost elasticity. Since migration cost elasticity κ enters multiplicatively with taste dispersion ν in the migration cost function, I estimate their product, $\tilde{\kappa} \equiv \kappa\nu$. Using

and Zhu (2019).

[46] See Appendix C.1 for details.

non-linear least squares, I minimize the difference between the model-predicted district-to-district migration flows and the observed flows (see Appendix C.2 for further details).

The identification assumption is that the difference between observed and predicted migration flows is due to classical measurement errors, which are uncorrelated with geography or other unobservable factors influencing the migration market access terms. As the sample size increases, these errors vanish, and the observed flows converge to the model predictions under the true $\tilde{\kappa}$. Although proving uniqueness is difficult, Appendix Figure A.7 suggests that the loss function is convex, suggesting the existence of a unique $\tilde{\kappa}$ that minimizes it.

A complication is that the observed migration flows may not align with the 24-year frequency used in the model. The 1980 census reports migration flows based on the "place of last residence," with an average residency of 12 years in the microdata. Assuming stable migration shares, I convert the observed 12-year shares to 24-year shares, following the method of Artuç et al. (2010) and Caliendo et al. (2019).[47]

Skill dispersion. The shape parameter θ governs the dispersion of Fréchet-distributed productivity across individuals, with higher values of θ indicating less dispersion. Given that individual potential earnings are also Fréchet distributed, let y^e_{ink} denote the earnings of individual i of ethnicity e, working in sector k and residing in region n. The distribution assumption implies:

$$\frac{\mathrm{Var}[y^e_{ink}]}{\mathbb{E}[y^e_{ink}]^2} = \frac{\Gamma(1-\frac{2}{\theta}) - \Gamma(1-\frac{1}{\theta})^2}{\Gamma(1-\frac{1}{\theta})^2}. \tag{23}$$

The variance of y^e_{ink}, normalized by the squared expectation, is a function of θ. This variance approaches infinity as θ approaches 2 from above and decreases monotonically toward 0 as θ increases.[48] This implies that there exists a unique value of θ for any given normalized variance, meaning θ is identified by this moment.[49]

Productivity spillovers. The parameters $\gamma_A, \gamma_M, \gamma^e$, which govern productivity spillovers, affect expected earnings in each sector and, therefore, occupational choices. By rewriting

[47] I first calculate the 12-year migration shares matrix, with each row summing to 1. Assuming constant migration shares over two 12-year periods, I square the matrix to obtain the 24-year migration shares matrix, $\hat{m}_{jh}$. This approach also eliminates zeros in the shares, allowing me to take logs as in Equation (A-5).

[48] For the variance of Fréchet-distributed y^e_{ink} to exist, θ must be greater than 2.

[49] Since θ is assumed to be constant across locations, sectors, and ethnic groups, it is over-identified in data where variance may differ across these dimensions.

the occupation choice equation (11) using the trade market access terms, we get:

$$\ln \bar{w}^e_n = \gamma_k \ln L_{nk} + \gamma^e \ln \left(\frac{L^e_{nk}}{L_{nk}} \right) - \frac{1}{\theta} \ln \pi^e_{nk}$$
$$- \left(\frac{1}{\sigma - 1} \right) \ln \left(\mathcal{P}_{nk} \right)^{\sigma-1} - \left(\frac{1}{\sigma - 1} \right) \ln Y_n + \underbrace{\frac{1}{\theta} \ln \phi^e_{nk}}_{\text{error term}}, \quad \forall k \in \{A, M\}. \quad (24)$$

The left-hand side is the average wage of ethnic group e in region n, which is observed in the data.[50] Local employment in sector k shifts the average wage through γ_k, while the ethnic composition affects wages via γ^e.

To identify γ^e, we subtract Equation (24) for one group from the other, eliminating region-industry-specific terms:

$$\ln \left(\frac{\bar{w}^c_n}{\bar{w}^m_n} \right) = \gamma^e \ln \left(\frac{L^c_{nk}}{L^m_{nk}} \right) - \frac{1}{\theta} \ln \left(\frac{\pi^c_{nk}}{\pi^m_{nk}} \right) + \underbrace{\frac{1}{\theta} \ln \left(\frac{\phi^c_{nk}}{\phi^m_{nk}} \right)}_{\text{error term}}. \quad (25)$$

This equation represents a relative (inverse) demand curve for sector k, where the negative $1/\theta$ term reflects the neoclassical force driving downward-sloping demand when the within-ethnic agglomeration effect (governed by γ^e) is not too strong.[51]

Unobserved productivity for ethnicity e in sector k and region n enters as the error term, which is typically positively correlated with local population due to individuals sorting into more productive areas. This selection bias tends to inflate the OLS estimate of γ_k. Similarly, individuals from ethnic group e who are more productive may sort into specific location-sector pairs to exploit better fundamentals, leading to an upward bias in the OLS estimate of γ^e. Conversely, classical measurement errors in the population distribution may attenuate the estimates of both γ_k and γ^e, biasing them downward.

To address these biases, I employ an instrumental variable strategy, using exogenous resettlement variation that shifted the equilibrium population in 1980. I denote the residualized resettlement density from Section V as $Z_n^{(own)}$. I also construct a neighboring resettlement shifter, denoted $Z_n^{(neighbor)}$, which is defined as the average county resettlement density of neighboring counties, after controlling for baseline characteristics and the expected resettlement density of those counties. Since relocations were not driven by location productivity

[50]The left-hand side of (24) does not vary with k due to the Fréchet property and the assumption of a constant shape parameter across industries. Consistent with this, Appendix Figure A.8 shows that average household log earnings within a county and ethnic group are similar across agriculture and non-agriculture.

[51]Here, only the share matters, not the quantity, since Chinese and Malays are assumed to be perfect substitutes. If both groups have equal probabilities of working in sector k, differences in ethnic population would affect relative wages only through γ^e.

(conditional on covariates), these instruments are plausibly orthogonal to location fundamentals ϕ^e_{nk}, leading to the following identifying moment conditions:[52]

$$\mathbb{E}[Z_n \ln \phi^e_{nk}] = 0,\ \forall k, e; \quad Z_n \in \{Z_n^{(own)}, Z_n^{(neighbor)}\}. \tag{26}$$

Amenity spillovers. The amenity spillover parameters, β and β^e, affect migration choices. The value of residing in region n, as expressed in Equation (5), can be rewritten using migration market access as:

$$\begin{aligned} \ln \bar{w}^e_n = {} & (-\beta + \beta^e) \ln L_n + \left(\frac{1}{\nu} - \beta^e\right) \ln L^e_n + \frac{1}{\nu} \ln (\mathcal{V}^e_n)^\nu \\ & + \left(\frac{\alpha}{\sigma - 1}\right) \ln P^{\sigma-1}_{nA} + \left(\frac{1-\alpha}{\sigma - 1}\right) \ln P^{\sigma-1}_{nM} - \underbrace{\frac{1}{\nu} \ln \bar{a}^e_n}_{\text{error term}}. \end{aligned} \tag{27}$$

To identify β^e, we can express the Chinese wage premium as:

$$\ln\left(\frac{\bar{w}^c_n}{\bar{w}^m_n}\right) = \left(\frac{1}{\nu} - \beta^e\right) \ln\left(\frac{L^c_n}{L^m_n}\right) + \frac{1}{\nu} \ln\left(\frac{(\mathcal{V}^c_n)^\nu}{(\mathcal{V}^m_n)^\nu}\right) - \underbrace{\frac{1}{\nu} \ln\left(\frac{\bar{a}^c_n}{\bar{a}^m_n}\right)}_{\text{error term}}. \tag{28}$$

This equation represents a relative (inverse) labor supply curve across space. The first term indicates that, when the amenity spillover β^e is not too strong, the neoclassical force $1/\nu$ predicts an upward-sloping supply curve. Intuitively, if β^e is large, a higher Chinese share becomes an attractive amenity for Chinese individuals, making them willing to accept lower wages. The inward migration market access term, $\mathcal{V}^e_n$, reflects the potential migrants of group e from other counties, which is a labor supply shifter. The error term captures unobserved factors that make a county more appealing to a particular group, which implies that the OLS estimate of β^e tends to be biased upward.

To address endogeneity, I again use population shifters from the resettlement program, assuming the program did not target areas based on amenity fundamentals. This assumption yields the following identifying moment conditions:

$$\mathbb{E}[Z_n \ln \bar{a}^e_n] = 0,\ \forall e, \quad Z_n \in \{Z_n^{(own)}, Z_n^{(neighbor)}\}. \tag{29}$$

Recovery of location fundamentals. The exogenous location fundamentals are recovered (up to scale) as residuals from Equations (24) and (27). Specifically, I recover ϕ^e_{nk} as

[52] These moment conditions identify the three productivity spillover parameters: γ_A, γ_M, and γ^e. Specifically, γ^e is identified from Equation (25) using an instrument for $\ln(L^e_{nk}/L_{nk})$, given θ. Then, after moving all terms in Equation (24) to the left except for $\ln L_{nk}$, the same instruments are used to identify γ_k.

the residuals from (24) after estimating γ_k and γ^e. Similarly, I recover $\bar{a}_n^e$ as the residuals from (27).

VII.B Estimation

The estimation proceeds as follows. First, I estimate the migration cost elasticity with respect to distance, $\tilde{\kappa}$, and use it to compute the migration cost matrix $\eta_{nr}^{-\nu}$. Next, I iteratively solve for the market access terms and the agricultural expenditure share α, based on Proposition 1. Then, I estimate the shape parameter of Fréchet skills, θ, by targeting the population-weighted average of the normalized wage variance within an (n, k, e) cell. Finally, I estimate the parameters $\{\gamma_A, \gamma_M, \gamma^e, \beta, \beta^e\}$ using a generalized method of moments (GMM) estimator, based on the moment conditions in Equations (26) and (29).

To mitigate small-sample biases, I exclude counties with fewer than five households from the 1980 census microdata and weight the estimations by the number of households, resulting in a sample of 698 counties. I bootstrap the entire procedure to obtain the standard errors of the parameter estimates.[53]

Table 9 documents the parameter estimates, which I discuss and compare with the existing literature in turn.

Migration cost elasticity. The estimated elasticity of migration costs with respect to distance, κ, is 0.47. This value aligns with the range of existing estimates in the literature. For example, Bryan and Morten (2019) find an elasticity of 0.37 in Indonesia between 1995 and 2012, while Peters (2022) reports an elasticity of 1.09 in post-war Germany in 1955.[54]

Skill dispersion. The shape parameter θ, which governs the dispersion of productivity draws, affects how individuals select into sectors based on their comparative advantage. A higher θ indicates less dispersion in skills. My estimate of 3.35 falls within the range found in the literature. For instance, Lagakos and Waugh (2013) estimate a θ of 5.3 for agriculture and 2.7 for the non-agricultural sector in the U.S. between 1996 and 2010. Similarly, Hsieh et al. (2019) report values between 1.5 and 2.6 for the U.S. from 1960 to 2012.[55]

Productivity spillovers. I estimate that local employment in the non-agricultural sector increases labor productivity with an elasticity of $\gamma_M = 0.25$, while the agricultural sector

[53]In each bootstrap iteration, I sample individuals from the census microdata with replacement at the district level and aggregate the outcomes to the county level. Administrative districts are larger than counties, with 66 districts in total.

[54]Bryan and Morten (2019) estimate migration costs non-parametrically, rather than assuming proportionality to distance. I translate their Figure 3 into my setting, where $1 - \eta_{nr}^{-1} \approx -0.5 + 0.147 \ln d_{nr}$. This implies that their distance elasticity varies with distance, unlike the constant elasticity assumed in my model. For comparison, I use the average log distance of 7.5 in their setting, resulting in $\partial \ln \eta_{nr} / \partial \ln d_{nr} \approx 0.37$.

[55]One reason their estimates may be lower is that wage variance in their model reflects differences in (endogenous) educational attainment in addition to idiosyncratic productivity draws.

shows a smaller, negative elasticity of $\gamma_A = -0.1$. My estimate for non-agricultural sectors is higher than the 0.2 estimated by Kline and Moretti (2014) but lower than the 1.25–3.1 range reported by Greenstone et al. (2010).[56] Although estimates for agriculture are scarce, my finding of a smaller elasticity aligns with the general understanding that agglomeration effects in agriculture are weaker than in industrial sectors.[57] Moreover, since my model does not account for land input in agricultural production, the negative elasticity also reflects diminishing returns to labor with a fixed amount of land.

I estimate a notable productivity spillover elasticity with respect to ethnic composition, $\gamma^e = 0.15$. This suggests that, holding county population constant, an increase in the Chinese employment share enhances the productivity of local Chinese workers. The effect on Malay workers is more nuanced and depends on the sector. Since $\gamma^e < \gamma_M$, Equation (10) indicates that Malays in non-agricultural sectors benefit from an increase in the Chinese population. However, because $\gamma^e > \gamma_A$, an increase in the Chinese population reduces Malays' agricultural productivity. These predictions are consistent with empirical evidence showing that Malays in non-agricultural sectors in more resettled areas experienced marginal income gains, while those in agriculture did not.

Although there are no direct comparisons for ethnicity-based spillovers in the literature, similar externalities have been examined using other demographic characteristics, such as education and occupation. For instance, Moretti (2004a) estimates wage elasticities of 0.14 for college graduates and 0.21 for high school graduates with respect to college share in a city.[58] Rossi-Hansberg et al. (2019) estimate wage elasticities with respect to the share of workers in "cognitive non-routine" occupations, finding substantial elasticities of 1.3 for workers in these occupations and 0.84 for those in non-cognitive roles.[59]

Amenity spillovers. I estimate the amenity spillover elasticity with respect to local population size at $\beta = -0.03$. This small value suggests that congestion forces—such as increased traffic or higher housing prices—are relatively weak. As discussed in Bryan and Morten (2019), extending the model to include housing as a non-traded good implies that

[56] See a discussion in Kline and Moretti (2014).

[57] See Combes and Gobillon (2015) and Melo et al. (2009) for a review of density-productivity elasticity, typically between 0.02 and 0.09 in developed countries. Estimates for developing countries are less common but tend to be above 0.1.

[58] Moretti (2004a) finds that a 1 percentage point increase in the share of college-educated workers leads to a 1.3% wage increase. I convert this to an elasticity, assuming an average college share of 0.25 in 1990. Diamond (2016) finds higher elasticities—0.31 for college graduates and 0.93 for non-college workers—though these estimates include substitution effects between high- and low-skilled workers.

[59] I assume perfect substitutability between Chinese and non-Chinese workers. If they are imperfect substitutes, which is plausible, the true γ^e could be even higher. Neoclassical forces suggest that an influx of Chinese workers would lead to larger wage gains for non-Chinese workers due to complementarity between groups. Therefore, a stronger within-group spillover would be required to explain the limited wage gains among non-Chinese workers in the data.

the amenity spillover can be decomposed as $\beta = \beta_a - \delta\beta_r$, where β_a represents the pure amenity spillover, β_r is the inverse of housing supply elasticity, and δ is the share of income spent on housing. Using the resettlement shocks as a demand shifter and housing prices from the MFLS-2 survey, I estimate $\beta_r \approx 0.3$, corresponding to a housing supply elasticity of 3.3 (Appendix Table A.14). This estimate is higher than U.S. estimates, which range from 1 to 3 (Gyourko et al., 2008; Saiz, 2010).[60] In 1980, housing expenditure accounted for 17.6% of total spending, implying a pure amenity spillover of $\beta_a = \beta + \delta\beta_r = 0.02$.[61]

There are few estimates of the β in low-income countries. Bryan and Morten (2019) report a value of 0.04, though with limited precision. Allen and Donaldson (2020) estimate both contemporaneous and historical amenity spillovers using U.S. data from 1800 to 2000, finding a -0.26 contemporaneous spillover and a 0.31 historical spillover (based on population 50 years prior). Since my model does not differentiate between contemporaneous and historical effects, my estimate reasonably falls between these two values.

My baseline estimate of the amenity spillover elasticity with respect to ethnic composition is $\beta^e = 0.13$. The positive β^e suggests that an increase in the population of an ethnic group raises the utility of people from that same group more than those from the other group. The stronger within-ethnic amenity spillover is consistent with the economies of scale in the provision of urban amenities, such as restaurants or entertainment, as discussed in Duranton and Puga (2004). It also aligns with the presence of social frictions, as reflected in consumption segregation documented in Davis et al. (2019).[62]

VIII Counterfactual Experiments

Estimates of agglomeration parameters show that agglomeration elasticities vary across space, suggesting potential aggregate effects of resettling populations.[63] This section uses the estimated model to evaluate the aggregate impact of the resettlement program and to explore policy counterfactuals. In Section VIII.A, I simulate a "no resettlement" equilibrium by using the 1947 population distribution instead of the resettled 1957 distribution, while holding all parameters and location fundamentals fixed. This equilibrium serves as the baseline for all counterfactual comparisons. Section VIII.B examines how barriers to productive spillovers across ethnic groups affect aggregate productivity. Finally, Section VIII.C assesses

[60] In Indonesia, Bryan and Morten (2019) estimate a value of 4, though with limited statistical power.

[61] The expenditure category is "gross rent, fuel, and power." In 1973, the same category accounted for 14.9% of expenditures. See Department of Statistics Malaysia (1980).

[62] There are no direct comparisons for ethnicity-based amenity spillovers. The closest comparison comes from Fajgelbaum and Gaubert (2020), who use Diamond's (2016) estimates of amenity spillovers by college share. Similar to productivity spillovers, the authors calibrate four constant amenity spillover elasticities: $(\gamma^A_{UU}, \gamma^A_{SU}, \gamma^A_{US}, \gamma^A_{SS}) = (-0.43, 0.18, -1.24, 0.77)$, where γ^A_{SU} denotes the marginal amenity spillover of a college graduate (S) on the utility of a non-college graduate (U)), and so on.

[63] For the distribution, see Appendix Figures A.9 and A.10.

the impact of an industrial policy that subsidizes Malays in manufacturing.

VIII.A Aggregate Impact of Forced Resettlement

Section V showed that the resettlement program had significant distributional impacts across locations and ethnic groups. While regions that received Chinese resettlement appear to benefit from the increased Chinese density, the areas from which they were removed likely experience negative effects. This section examines the overall impact of the resettlement program.

Table 10 reports the changes in aggregate economic outcomes from the "no resettlement" equilibrium to the observed 1980 equilibrium. This comparison provides the aggregate impact of resettlement as of 1980, assuming the 1947 distribution was in a steady state that would have continued to 1957 without the intervention. I calculate that the relocation of Chinese squatters from remote areas to regions with better market access and higher manufacturing productivity increases the Chinese manufacturing employment share by 1.2 percentage points (1.6%) and raises manufacturing productivity by 1.7%. However, Chinese agricultural productivity declines, as the resettled areas are less suited for agriculture.

The relocation also allows Malays to move into vacated agricultural land, slightly increasing their agricultural employment share in the resettled baseline (Column 2). This process boosts Malay productivity in both sectors: reduced Chinese participation in agriculture increases Malay agricultural productivity due to local diminishing returns to scale ($\gamma_A < 0$), and Malay manufacturing workers benefit from external economies of scale in manufacturing as the Chinese shift to that sector. Although the program primarily affects the Chinese population, the Malays also experience notable economic changes, highlighting the broader general equilibrium effects of place-based policies at this scale.[64]

The aggregate output changes from the no-resettlement to the resettled equilibrium can be decomposed as follows:

$$\sum_{n,k,e}\left(\tilde{y}^e_{nk}\tilde{L}^e_{nk}-y^e_{nk}L^e_{nk}\right)=\sum_{n,k,e}\left(\tilde{L}^e_{nk}-L^e_{nk}\right)y^e_{nk}+\sum_{n,k,e}\left(\tilde{y}^e_{nk}-y^e_{nk}\right)L^e_{nk}$$
$$+\sum_{n,k,e}\left(\tilde{y}^e_{nk}-y^e_{nk}\right)\left(\tilde{L}^e_{nk}-L^e_{nk}\right),$$

where y^e_{nk} denotes the baseline average per-capita output of group e in county n and sector

[64]The higher manufacturing employment share in the no-resettlement counterfactual may appear inconsistent with the cross-sectional comparisons in Section V, which showed relatively higher manufacturing shares of Malays in counties with more resettlement, all else equal. This discrepancy is due to the missing-intercept problem: the higher manufacturing share is relative to less resettled areas, not average changes. Moreover, the cross-sectional comparisons focused on destinations receiving resettlement, without considering the original locations from which Chinese squatters were relocated.

k, and $\tilde{y}^{e}_{nk}$ denotes its value in the resettled equilibrium. The first term captures output changes due to labor reallocation, holding per-capita output fixed. The second term reflects changes in per-capita output, holding labor allocation fixed. The last term captures joint effects from the interaction between labor reallocation and changes in per-capita output.

Overall, the resettlement program increases aggregate output by 2% compared to the no-resettlement baseline (Table 10, Column 3). This gain is primarily driven by labor reallocation toward more productive sectors and regions—with productivity fixed at baseline, labor reallocation alone accounts for two-thirds of the output increase. The total expenditure of the resettlement program up to 1954 was approximately 133 million Malayan Dollars or $43 million (Dhu Renick, 1965), roughly 0.5% of Malaysia's 1980 GDP (adjusted for inflation). Therefore, the net output gain from resettlement amounts to 1.5%.

Despite the output gain, the forced nature of the resettlement raises welfare concerns. To compare the economic gains with the utility losses from coercion, I invert the model to solve for the place-based wage subsidies required to voluntarily relocate the Chinese and Malays from the 1947 population distribution to the 1957 resettled distribution. Since migration decisions depend only on relative prices across regions, I calculate the minimum, weakly positive ad-valorem subsidies that would achieve this voluntary migration.[65] The results indicate that forced resettlement reduces welfare, as implementing such a voluntary resettlement program would cost the British colonial government nearly 1.4 times the total output, far exceeding the program's economic gains.

VIII.B Reducing Cross-Ethnic Frictions

Section V showed that Chinese workers benefited more from Chinese density than Malays did, indicating frictions in agglomeration spillovers between ethnic groups. However, these cross-ethnic frictions also negatively affect the Chinese population, especially since they make up only one third of the total population. This section explores the economic impact of reducing those frictions.

Table 11 shows the changes in economic outcomes when cross-ethnic frictions in productivity spillovers are reduced to half of their baseline level (i.e., $\tilde{\gamma}^{e} = 0.5\gamma^{e}$). Reducing these frictions increases Malay participation in manufacturing by 1.1 percentage points (2%) (Column 2). As more Malays shift into manufacturing, Chinese productivity in agriculture improves, and their concentration in manufacturing decreases. This shift alleviates diminishing returns in agriculture (Column 1) while fostering greater interaction between Chinese and Malays in that sector.

Halving cross-ethnic frictions leads to 4.8% increase in aggregate output, with 84% of the

[65] See Appendix D.1 for more details.

gains driven by productivity improvements. Notably, despite the Chinese comprising only a third of the population, their contribution to the output gains is nearly equal to that of the Malays. This finding underscores the significant economic benefits of enabling minority groups to interact productively with the majority.

VIII.C Wage Subsidies for Malays in Manufacturing

Given the strong external economies in manufacturing, shifting labor from agriculture to manufacturing has the potential to boost aggregate output. A key postwar challenge for the Malaysian government was integrating Malays and other indigenous populations into the industrial sector, as Malays were predominantly engaged in low-productivity agriculture. The previous section highlighted how cross-ethnic frictions hindered Malay industrialization by limiting productivity spillovers between groups. In this section, I evaluate an industrial policy that provides wage subsidies to Malays in manufacturing.

This policy is inspired by the New Economic Policy (NEP) from the Second Malaysia Plan (1971–1975), which aimed to restructure society and *eliminate the identification of race with economic function*.[66] The government implemented various measures to promote Malay entrepreneurship, such as providing access to credit, training, and quotas in higher education. At the same time, regulations like the Industrial Coordination Act imposed costs on non-Malay industrial participation, requiring firms to meet certain Malay ownership or employment quotas.

I simulate an 18% wage subsidy for Malays in manufacturing, funded by an 7.5% income tax on all citizens to balance the government budget.[67] The subsidy rate is chosen to align the Malay manufacturing employment share with that of the Chinese, as targeted by the NEP. Table 12 reports the results of the policy. The subsidy raises Malay participation in manufacturing by 10.3 percentage points (19%), crowding out Chinese participation by 10 percentage points (13%). Total manufacturing employment increases by 3.4 percentage points (5%).

The policy boosts Malay productivity by 4.1%, with a significant 9% increase in agricultural productivity due to the labor shift. However, this structural change comes at a cost to Chinese workers: their overall productivity declines as more are pushed out of manufacturing into the less productive agricultural sector, despite a slight increase in Chinese agricultural productivity due to the alleviated diminishing returns from Malays exiting.

Despite these distortions, aggregate output increases by 1.1%. Although manufacturing

[66]The racial riots of 1969 were pivotal in shaping affirmative action within the NEP. See Koon (1997); Jomo (2017) for further discussion.

[67]Income tax is non-distortionary as it applies uniformly across sectors, with no leisure-work trade-off in the model.

productivity declines slightly, the loss is mitigated by the sector's strong local increasing returns. Moreover, agricultural productivity rises significantly as labor shifts out of agriculture, leading to an overall productivity improvement and a net gain in aggregate output.

However, the tax-funded subsidies reduce consumption and lower overall welfare. The Chinese experience an 11.1% drop in utility, while Malays see a 3.3% loss, resulting in an overall welfare decline of 6%. This underscores the trade-offs of affirmative industrial policies: while they can promote industrialization and growth, they may come at the expense of welfare and exacerbate ethnic tensions.

IX Conclusion

This paper examines how social divisions affect agglomeration benefits through a large-scale, ethnic-based resettlement program that forcibly relocated rural Chinese into villages in British Malaya during the 1950s, reshaping both economic and social landscapes. Despite the coercive nature of the program, which likely yielded fewer benefits than voluntary migration, the increased density of local Chinese and the concentration of industrial human capital led to greater divisions of labor and significant productivity gains. However, these benefits were not equally shared across ethnic groups.

The local economic impacts of resettlement were shaped by heterogeneous agglomeration externalities that varied by sector and ethnic composition. The influx of industrial labor prompted a shift away from agriculture, driven by higher returns to scale in non-agricultural sectors compared to the land-constrained agricultural sector. Segregation and barriers to social interactions between the Chinese and Malays limited cross-ethnic productivity spillovers, resulting in most gains from Chinese agglomeration accruing to the Chinese community.

To assess the aggregate impact of the resettlement program, I estimate a quantitative spatial model that incorporates these heterogeneous agglomeration forces. The program generated output gains by reallocating labor from remote, less productive areas to regions with better market access and higher industrial productivity. However, this forced relocation ultimately reduced welfare by disregarding individuals' psychological costs of moving and their preferences about where to live.

While this paper highlights ethnic background as a barrier to social interactions and agglomeration benefits, similar frictions can emerge from other demographic factors, such as caste, culture, religion, and gender. For instance, caste norms in India have historically restricted inter-caste interactions, potentially limiting the benefits of urbanization and slowing economic development. These findings also carry policy implications for the settlement of refugees and immigrants, as their differing social backgrounds may hinder economic integration with incumbent populations.

References

Abel, Martin (2019) "Long-Run Effects of Forced Resettlement: Evidence from Apartheid South Africa," *The Journal of Economic History*, 79 (4), 915–953.

Abramitzky, Ran, Leah Boustan, and Dylan Shane Connor (2024) "Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office," *The Journal of Economic History*, 84 (2), 352–394.

Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani (2019) "The Economic Effects of Density: A Synthesis," *Journal of Urban Economics*, 111, 93–107.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf (2015) "The Economics of Density: Evidence From the Berlin Wall," *Econometrica*, 83 (6), 2127–2189.

Aizer, Anna, and Janet Currie (2004) "Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California," *Journal of Public Economics*, 88 (12), 2573–2585.

Allen, Treb, Costas Arkolakis, and Xiangliang Li (forthcoming) "On the Equilibrium Properties of Spatial Models," *American Economic Review: Insights*.

Allen, Treb, and Dave Donaldson (2020) "Persistence and Path Dependence in the Spatial Economy," Working Paper 28059, National Bureau of Economic Research.

Almagro, Milena, and Tomás Domínguez-Iino (2024) "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," Working Paper 32304, National Bureau of Economic Research.

Ambrus, Attila, Erica Field, and Robert Gonzalez (2020) "Loss in the Time of Cholera: Long-Run Impact of a Disease Epidemic on the Urban Landscape," *American Economic Review*, 110 (2), 475–525.

Amsden, Alice H. (1992) *Asia's Next Giant: South Korea and Late Industrialization*: Oxford University Press.

Ananat, Elizabeth, Shihe Fu, and Stephen L Ross (2013) "Race-Specific Agglomeration Economies: Social Distance and the Black-White Wage Gap," Working Paper 18933, National Bureau of Economic Research.

Ananat, Elizabeth, Fu Shihe, and Stephen L. Ross (2018) "Race-specific Urban Wage Premia and the Black-white Wage Gap," *Journal of Urban Economics*, 108, 141–153.

Anderson, James E., and Eric van Wincoop (2003) "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (1), 170–192.

Armington, Paul S. (1969) "A Theory of Demand for Products Distinguished by Place of Production (Une théorie de la demande de produits différenciés d'après leur origine) (Una teoría de la demanda de productos distinguiéndolos según el lugar de producción)," *Staff Papers (International Monetary Fund)*, 16 (1), 159–178.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren (2010) "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 100 (3), 1008–45.

Baillargeon, David (2021) "Spaces of Occupation: Colonial Enclosure and Confinement in British Malaya," *Journal of Historical Geography*, 73, 24–35.

Balboni, Clare (forthcoming) "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities," *American Economic Review*.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan (2018) "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics*, 10 (4), 1–42.

Baum-Snow, Nathaniel, and Ronni Pavan (2013) "Inequality and City Size," *Review of Economics and Statistics*, 95 (5), 1535–1548.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007) "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115 (4), 588–638.

Bayer, Patrick, Robert McMillan, and Kim Rueben (2004) "An Equilibrium Model of Sorting in an Urban Housing Market," Working Paper 10865, National Bureau of Economic Research.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong (2016) "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106 (9), 2658–2698.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong (2019) "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 109 (11), 3978–4025.

Beaman, Lori A. (2011) "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.," *The Review of Economic Studies*, 79 (1), 128–161.

Becker, Sascha O, Irena Grosfeld, Pauline Grosjean, Nico Voigtländer, and Ekaterina Zhuravskaya (2020) "Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers," *American Economic Review*, 110 (5), 1430–1463.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan (2000) "Network Effects and Welfare Cultures," *The Quarterly Journal of Economics*, 115 (3), 1019–1055.

Borjas, George J. (1992) "Ethnic Capital and Intergenerational Mobility," *The Quarterly Journal of Economics*, 107 (1), 123–150.

Borusyak, Kirill, and Peter Hull (2023) "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 91 (6), 2155–2185.

Briggs, Harold (1951) *Report on the Emergency in Malaya from April, 1950 to November,* *1951*, Kuala Lumpur: Government Press, Lieut.-General, Director of Operations.

Bryan, Gharad, and Melanie Morten (2019) "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 127 (5), 2229–2268.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019) "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87 (3), 741–835.

Carlitz, Ruth, Ameet Morjaria, Joris M Mueller, and Philip Osafo-Kwaako (2022) "State-Building in a Diverse Society," Working Paper 30731, National Bureau of Economic Research.

Carrillo, Bladimir, Carlos Charris, and Wilman Iglesias (2023) "Moved to Poverty? A Legacy of the Apartheid Experiment in South Africa," *American Economic Journal: Economic Policy*, 15 (4), 183–221.

Chen, Jiafeng, and Jonathan Roth (2023) "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 139 (2), 891–936.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2018) "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," Working Paper 25147, National Bureau of Economic Research.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz (2016) "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106 (4), 855–902.

Chyn, Eric, and Lawrence F Katz (2021) "Neighborhoods Matter: Assessing the Evidence for Place Effects," *Journal of Economic Perspectives*, 35 (4), 197–222.

Combes, Pierre-Philippe, and Laurent Gobillon (2015) "Chapter 5 - The Empirics of Agglomeration Economies," in Duranton, Gilles, J. Vernon Henderson, and William C. Strange eds. *Handbook of Regional and Urban Economics*, 5 of Handbook of Regional and Urban Economics, 247–348: Elsevier.

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

Corry, WCS (1954) "A General Survey of New Villages," *Kuala Lumpur: Government Printers (being a report to His Excellency Sir Donald MacGullivray)*, 24.

Council, Malayan Christian (1958) *A Survey of the New Villages in Malaya: Malayan Christian Council*: Malayan Christian Council.

Damm, Anna Piil (2009) "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence," *Journal of Labor Economics*, 27 (2), 281–314.

Davis, Donald R, and Jonathan I Dingel (2019) "A Spatial Knowledge Economy," *American Economic Review*, 109 (1), 153–170.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales (2019) "How Segregated is Urban Consumption?" *Journal of Political Economy*, 127 (4), 1684–1738.

Del Tufo, Moroboë Vincenzo (1947) *Malaya Comprising the Federation of Malaya and the Colony of Singapore: A Report on the 1947 Census of Population*: Crown Agents for the Colonies.

Dell, Melissa, Nathan Lane, and Pablo Querubin (2018) "The Historical State, Local Collective Action, and Economic Development in Vietnam," *Econometrica*, 86 (6), 2083–2121.

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

Department of Statistics Malaysia (1980) "Report of the Household Expenditure Survey: Peninsular Malaysia (1980), Sabah & Sarawak (1982)."

Dhu Renick, Rhoderick (1965) "The Emergency Regulations of Malaya Causes and Effect," *Journal of Southeast Asian History*, 6 (2), 1–39.

Diamond, Rebecca (2016) "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000," *American Economic Review*, 106 (3), 479–524.

Dippel, Christian (2014) "Forced Coexistence and Economic Development: Evidence from Native American Reservations," *Econometrica*, 82 (6), 2131–2165.

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

Duranton, Gilles, and Diego Puga (2004) "Chapter 48 - Micro-Foundations of Urban Agglomeration Economies," in Henderson, J. Vernon, and Jacques-François Thisse eds. *Cities and Geography*, 4 of Handbook of Regional and Urban Economics, 2063–2117: Elsevier.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund (2003) "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment*," *The Quarterly Journal of Economics*, 118 (1), 329–357.

Eriksson, Katherine (2019) "Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants During the Age of Mass Migration," *European Review of Economic History*, 24 (3), 427–446.

Fajgelbaum, Pablo D, and Cecile Gaubert (2020) "Optimal Spatial Policies, Geography, and Sorting," *The Quarterly Journal of Economics*, 135 (2), 959–1036.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997) "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65 (4), 833–860.

Fell, H (1960) "Population Census of the Federation of Malaya," *Kuala Lumpur: Department of Statistics.*

Fogli, Alessandra, and Veronica Guerrieri (2019) "The End of the American Dream? Inequality and Segregation in US Cities," Working Paper 26143, National Bureau of Economic Research.

Gechter, Michael, and Nick Tsivanidis (2023) "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills," *Unpublished manuscript*.

Ginsburg, Norton Sydney (1958) *Malaya*, Publications of the American Ethnological Society, Seattle,: University of Washington Press, 533.

Glaeser, Edward L., and Joshua D. Gottlieb (2008) "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer (1992) "Growth in Cities," *Journal of political economy*, 100 (6), 1126–1152.

Glaeser, Edward L, Bruce Sacerdote, and Jose A Scheinkman (1996) "Crime and Social Interactions," *The Quarterly Journal of Economics*, 111 (2), 507–548.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010) "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118 (3), 536–598.

Greif, Avner (1993) "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review*, 83 (3), 525–548.

Gyourko, Joseph, Albert Saiz, and Anita Summers (2008) "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban studies*, 45 (3), 693–729.

Haggard, Stephan (1990) *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries*, Cornell studies in political economy: Cornell University Press.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm (2020) "The Making of the Modern Metropolis: Evidence from London," *The Quarterly Journal of Economics*, 135 (4), 2059–2133.

Hilhorst, Dorothea, and Mathijs van Leeuwen (2000) "Emergency and Development: The Case of Imidugudu, Villagization in Rwanda," *Journal of Refugee Studies*, 13 (3), 264–280.

Hirschman, Charles (1986) "The Making of Race in Colonial Malaya: Political Economy and Racial Ideology," in *Sociological forum*, 1, 330–361, Springer.

Hjort, Jonas (2014) "Ethnic Divisions and Production in Firms," *The Quarterly Journal of Economics*, 129 (4), 1899–1946.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow (2019) "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 87 (5), 1439–1474.

Humphrey, John Weldon (1971) *Population Resettlement in Malaya*: Northwestern University.

Ioannides, Yannis M (2012) *From Neighborhoods to Nations: The Economics of Social Interactions*: Princeton University Press.

Jomo, Kwame Sundaram (2017) "The New Economic Policy and Interethnic Relations in Malaysia," in *Global minority rights*, 239–266: Routledge.

Ju, Shi Jie Shu (1959) *Directory of Singapore and Malaya*: Shi Jie Shu Ju, 924.

Khoo, Teik Huat, and Malaysia Jabatan Perangkaan (1983) *Laporan Am Banci Penduduk = General Report of the Population Census*, Kuala Lumpur: Jabatan Perangkaan Malaysia.

Kim, Sunwoong (1989) "Labor Specialization and the Extent of the Market," *Journal of Political Economy*, 97 (3), 692–705.

Kline, Patrick, and Enrico Moretti (2014) "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *The Quarterly Journal of Economics*, 129 (1), 275–331.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007) "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75 (1), 83–119.

Koon, Heng Pek (1997) "The New Economic Policy and the Chinese Community in Peninsular Malaysia," *The Developing Economies*, 35 (3), 262–292.

Lagakos, David, and Michael E Waugh (2013) "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103 (2), 948–980.

Lane, Nathan (2022) "Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea," *Available at SSRN 3890311*.

Lee, Kam Hing, and Chee-Beng Tan (2000) *The Chinese in Malaysia*: Oxford University Press.

Loh, Francis Kok-Wah (1988) *Beyond the Tin Mines: Coolies, Squatters and New Villagers in the Kinta Valley, Malaysia, c. 1880-1980*: Singapore: Oxford University Press.

Marshall, Alfred (1890) *Principles of Economics*: London: Macmillan.

Melo, Patricia C, Daniel J Graham, and Robert B Noland (2009) "A Meta-Analysis of Estimates of Urban Agglomeration Economies," *Regional Science and Urban Economics*, 39 (3), 332–342.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg (2018) "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108 (12), 3855–3890.

Moretti, Enrico (2004a) "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data," *Journal of Econometrics*, 121 (1-2), 175–212.

——— (2004b) "Workers' Education, Spillovers, and Productivity: Evidence from Plant-

Level Production Functions," *American Economic Review*, 94 (3), 656–690.

Morten, Melanie, and Jaqueline Oliveira (2024) "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *American Economic Journal: Applied Economics*, 16 (2), 389–421.

Mundial, Banco (1993) *The East Asian Miracle: Economic Growth and Public Policy*: World Bank.

Munshi, Kaivan (2003) "Networks in the Modern Economy: Mexican Migrants in the US Labor Market," *The Quarterly Journal of Economics*, 118 (2), 549–599.

Neumark, David, and Helen Simpson (2015) "Place-Based Policies," in *Handbook of Regional and Urban Economics*, 5, 1197–1287: Elsevier.

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

Peters, Michael (2022) "Market Size and Spatial Growth-Evidence From Germany's Post-War Population Expulsions," *Econometrica*, 90 (5), 2357–2396.

Purcell, Victor (1947) "Chinese Settlement in Malacca," *Journal of the Malayan Branch of the Royal Asiatic Society*, 20 (1 (141), 115–125.

Redding, Stephen J, and Daniel M Sturm (2024) "Neighborhood Effects: Evidence from Wartime Destruction in London,"Technical report, National Bureau of Economic Research.

Riley, Shawn J, Stephen D DeGloria, and Robert Elliot (1999) "Index that Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 5 (1-4), 23–27.

Rosenthal, Stuart S, and William C Strange (2004) "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, 4, 2119–2171: Elsevier.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman (2019) "Cognitive Hubs and Spatial Redistribution," Working Paper 26267, National Bureau of Economic Research.

Roy, Andrew Donald (1951) "Some Thoughts on the Distribution of Earnings," *Oxford economic papers*, 3 (2), 135–146.

Saiz, Albert (2010) "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 125 (3), 1253–1296.

Sandhu, Kernial Singh (1964) "Emergency Resettlement in Malaya," *The Journal of Tropical Geography*, 18, 180.

Sarvimäki, Matti, Roope Uusitalo, and Markus Jäntti (2022) "Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations," *Journal of the European Economic Association*, 20 (6), 2497–2539.

Silva, JMC Santos, and Silvana Tenreyro (2006) "The Log of Gravity," *The Review of Eco-

nomics and statistics, 88 (4), 641–658.

Smith, Cory, and Amrita Kulka (2023) "Agglomeration Over the Long Run: Evidence from County Seat Wars."

Strauch, Judith (1981) *Chinese Village Politics in the Malaysian State*: Harvard University Press.

Sunderland, Riley (1964) "Resettlement and Food Control in Malaya."

Survey of India Offices (P.Z.O.) (1944) "Malaysia, Malaya, Yala, Series: HIND 1076, 1944, 1:253 440."

Tombe, Trevor, and Xiaodong Zhu (2019) "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 109 (5), 1843–1872.

Tsivanidis, Nick (2023) "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's Transmilenio," *University of California, Berkeley.*

U.S. Office of Strategic Services (1942) *Roads and Railroads of Malaya*, Washington: OSS Repro. Section, provisional edition.

——— (1944) *Malaya Industrial Facilities*, Washington: OSS Repro. Section, provisional edition, `https://catalog.archives.gov/id/176201037`.

Vlieland, CA (1931) *British Malaya, A Report on the 1931 Census*: Crown Agents for the Colonies.

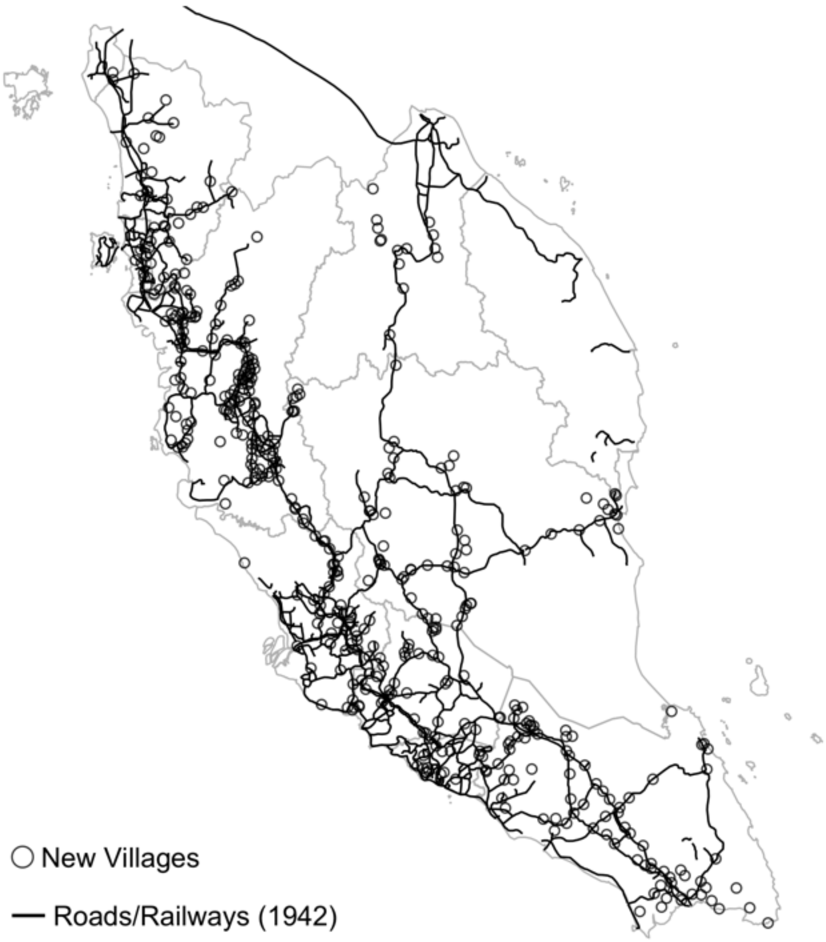
Wade, Robert (2004) *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*: Princeton University Press.

War Office (1943) *Malaya: Land Utilization Map*, London: GSGS.

Weiwu, Laura (2023) "Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities," MIT, mimeograph.

Whittaker, Hannah (2012) "Forced Villagization During the Shifta Conflict in Kenya, ca. 1963–1968," *The International journal of African historical studies*, 45 (3), 343–364.

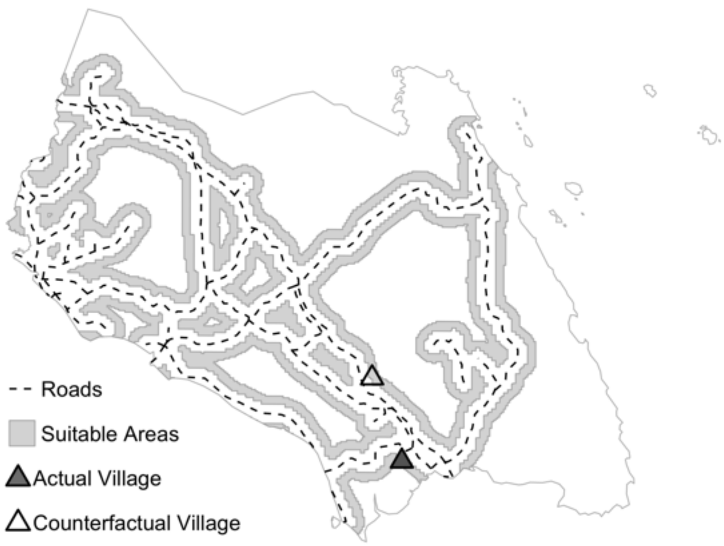
Figure 1. The New Villages and Transportation Network



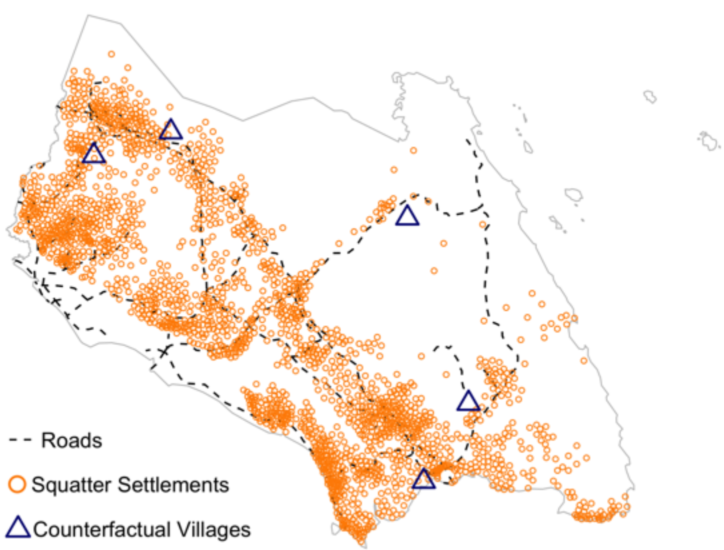
Notes: This figure shows the location of the New Villages (round circles) and the roads and railways in 1942 (line). The gray polygons indicate state boundaries. Data on the New Villages are from the Corry report. Data on roads and railways from U.S. Office of Strategic Services (1942).

Figure 2. Counterfactual Site Selection and Relocation

Panel A. Counterfactual Site Selection



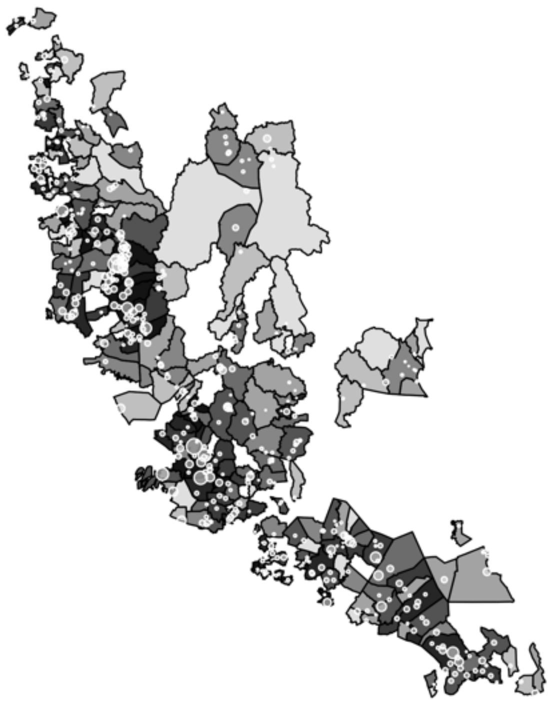
Panel B. Counterfactual Relocation



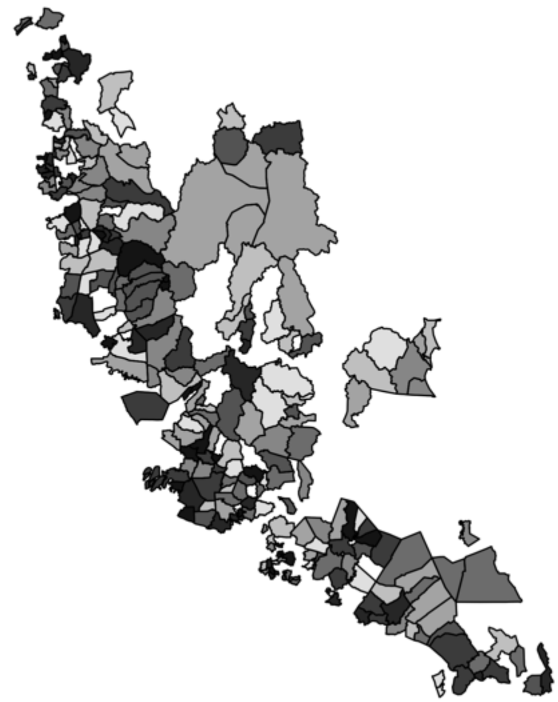
Notes: This figure illustrates the counterfactual site selection and relocation for the state of Johor. Panel A shows the selection of counterfactual sites. The solid triangle represents an actual New Village, and the dashed lines denote the road and rail network. Gray shaded areas indicate regions that are equidistant from the actual village and equally suitable for resettlement. The hollow triangle represents a counterfactual village location, randomly drawn from these suitable areas. Panel B illustrates the relocation of squatters to the counterfactual sites. Orange circles denote the initial squatter settlements.

Figure 3. County Resettlement Density, Expected and Residualized

Panel A. County Resettlement Density, Expected



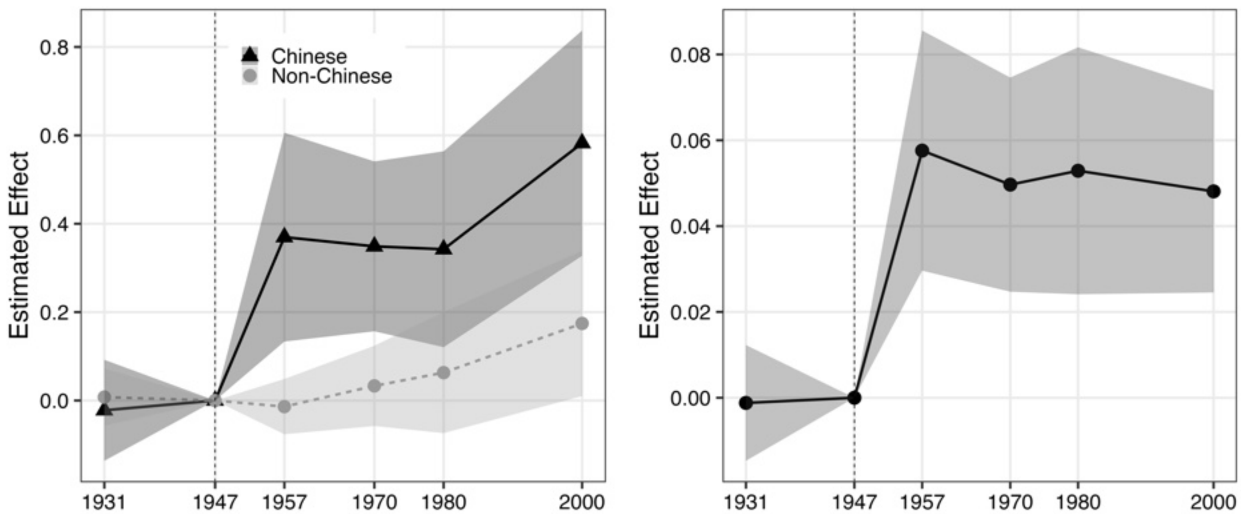
Panel B. County Resettlement Density, Residualized



Notes: This figure shows the expected and residualized county resettlement density, with darker shades indicating higher resettlement density deciles. The white bubbles denote the New Villages, with bubble sizes proportional to the resettled population. The sample is restricted to the 249 counties that contain at least one New Village. Panel A shows the expected resettlement density, calculated using Equation (A.1). Panel B shows the residualized resettlement density after controlling for state fixed effects, expected resettlement density, and baseline covariates: (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. Data on resettlement from the Corry report.

Figure 4. Changes in Population Distribution from 1931 to 2000, by County Resettlement Density

Panel A. Population Growth　　　　Panel B. Changes in Chinese Share



Notes: Regressions control for the expected resettlement density, whether a county has any resettlement, (log) county area, (log) distance to nearest road, road density, distance to nearest rail station, distance to coastline, 1947 Chinese population share and population density, and the land shares of rubber and mining. The shaded region reflects the 95% confidence interval under Conley standard errors, with a distance cutoff of 30 kilometers.

Table 1. Balance of Location Fundamentals and Pre-Period Characteristics

	Geography				Amenities				Economic Activities			
	Elev.	Rugged.	Rice Suitab.	Coconut Suitab.	Dist. Police	Dist. Post	Dist. Hosp.	Dist. Temple	Land Use Rubber	Land Use Mining	Dist. Factory	Dist. Cities
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Within State												
Higher Resettlement	0.19	3.86	−0.04	−0.01	−0.46	−0.35	−1.60	0.00	0.07	0.01	−0.97	−5.79
	(0.12)	(3.19)	(0.01)	(0.01)	(0.42)	(0.34)	(0.69)	(1.37)	(0.02)	(0.01)	(0.85)	(1.67)
Panel B. Baseline Controls												
Higher Resettlement	0.07	−2.22	−0.05	−0.01	0.44	0.51	0.59	3.22	0.05	0.02	−0.64	−1.03
	(0.16)	(4.93)	(0.02)	(0.02)	(0.49)	(0.42)	(0.93)	(1.96)	(0.01)	(0.01)	(1.01)	(1.90)
Panel C: Expected Resettlement												
Higher Resettlement	0.16	−7.19	−0.05	−0.02	0.47	0.79	0.35	2.14	0.02	0.02	−0.93	0.54
	(0.27)	(5.84)	(0.03)	(0.03)	(0.61)	(0.55)	(0.96)	(2.21)	(0.03)	(0.01)	(1.09)	(2.44)
Mean	0.94	62.77	1.21	1.12	9.35	11.33	23.50	66.13	0.24	0.01	26.23	87.69
Standard Deviation	1.51	74.19	0.23	0.21	8.19	8.68	19.45	47.40	0.30	0.07	18.08	69.82
# Counties	777	777	777	777	777	777	777	777	777	777	777	777

Notes: This table shows the relationship between county characteristics and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Columns 1–4 show the effects of higher resettlement on geography: elevation (column 1), ruggedness (column 2), suitability for padi rice cultivation (column 3), and suitability for coconut cultivation (column 4). Columns 5–8 show the effects on amenities as measured in 1945: distance to the nearest police station (column 5), post or telegraph office (column 6), hospital (column 7), and Chinese temple (column 8). Columns 9–12 report the effects on per-period economic activities: land use share for rubber in 1943 (column 9), land use share for mining in 1943 (column 10), distance to industrial facilities in 1945 (column 11), and distance to major cities: Singapore, George Town, Malacca, Ipoh, and Kuala Lumpur (column 12). Panel A reports regression estimates with state fixed effects. Panel B additionally includes the baseline controls (except for land use shares for rubber and mining): an indicator for any resettlement, (log) county area, road density, distance to roads, distance to rail stations, distance to the coastline, log population in 1947, and Chinese population share in 1947. Panel C additionally controls for the expected resettlement density. The unit of observation is the county. Elevation data from SRTM; ruggedness from Nunn and Puga (2012); rice and coconut suitability from FAO GAEZ v4; locations of police stations, post/telegraph offices, hospitals, and temples from HIND 1076 topographical maps (Survey of India Offices , P.Z.O.); population density from the 1947 census; land use shares from GSGS 4474 land utilization maps; and prewar industrial facilities from the US National Archive, RG226 (U.S. Office of Strategic Services, 1944). Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 2. Post-Resettlement Population Distribution, by County Resettlement Density

	Log Population, by Year:			Chinese Share of Population, by Year:		
	1957	1980	2000	1957	1980	2000
	(1)	(2)	(3)	(4)	(5)	(6)
Higher Resettlement	0.094	0.108	0.177	0.048	0.050	0.041
	(0.034)	(0.062)	(0.075)	(0.012)	(0.011)	(0.011)
# Counties	777	777	777	777	777	777

Notes: This table shows the relationship between measures of population distribution from 1957 to 2000 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Columns 1–3 report the effect of resettlement density on log county population in 1957 (column 1), 1980 (column 2), and 2000 (column 3). Columns 4–6 report the effect of resettlement density on the Chinese share of county population in 1957 (column 4), 1980 (column 5), and 2000 (column 6). Columns 1–3 are estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator. Columns 4–6 are estimated using OLS. All regressions include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county. Data from the tabulated Census of Population. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 3. Sectoral Employment in 1980–1991, by County Resettlement Density

	Primary Sector (1)	Non-Primary Sector (2)	Difference (2) − (1) (3)
Panel A. Total Employment			
Higher Resettlement	0.108	0.286	0.178
	(0.037)	(0.129)	(0.140)
# County-Years	1,554	1,554	
Panel B. Chinese Employment			
Higher Resettlement	0.267	0.342	0.074
	(0.058)	(0.174)	(0.172)
# County-Years	1,516	1,502	
Panel C. Non-Chinese Employment			
Higher Resettlement	−0.008	0.241	0.249
	(0.045)	(0.108)	(0.126)
# County-Years	1,516	1,502	

Notes: This table shows the relationship between sectoral employment in 1980–1991 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A shows the effect of resettlement on total employment in the primary sector (column 1), the non-primary sector (column 2), and the difference between the two (column 3). Panels B and C show the effects on Chinese employment and non-Chinese employment, respectively. The primary sector is comprised of agriculture, hunting, forestry, fishing, mining, and quarrying. The non-primary sector is comprised of manufacturing; utility; construction; wholesale and retail trade; transport and communication; and finance, business and other services. All regressions are estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator and include state-year fixed effects and the main controls interacted with year: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county-year. Data from the Census of Population in 1980 and 1991. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 4. Manufacturing Activity in 1970, by County Resettlement Density and Pre-Period Industry Share of Chinese Employment

	Total Number of Manuf. Establishments (1)	Share of Employer Establishments (2)
Higher Resettlement	0.009	0.001
	(0.126)	(0.020)
Higher Resettlement × Chinese Industries	0.217	0.020
	(0.082)	(0.009)
# County-Industries	15,540	2,142

Notes: This table shows the relationship between measures of manufacturing activity in 1970 and county resettlement density by industries of pre-period Chinese employment share. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. "Chinese Industries" is an indicator for industries with more than 80% Chinese employment in 1947, comprised of all manufacturing industries except for food products, wood products, textiles, and other miscellaneous manufacturing (see Appendix Figure A.6). Column 1 reports the effect on the total number of manufacturing establishments, estimated with the Poisson pseudo-maximum-likelihood (PPML) estimator. Column 2 reports the OLS estimates on the share of establishments with at least one full-time employee. All regressions include state fixed effects, 2-digit industry fixed effects, and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county-industry. Data on manufacturing establishments are from the Directory of Manufacturing in 1970. Data on Chinese employment share are from the tabulated Population Census in 1947. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 5. Participation and Specialization in the Labor Market in 1980–1991, by County Resettlement Density

	Chinese Individuals (1)	Non-Chinese Individuals (2)	Difference (2) − (1) (3)
Panel A. Labor Force Participation			
Higher Resettlement	0.015	0.002	0.014
	(0.008)	(0.007)	(0.010)
Mean of Outcome	0.578	0.569	
# Counties	524	745	
Panel B. Occupation Specialization Index			
Higher Resettlement	0.016	−0.006	0.022
	(0.008)	(0.008)	(0.009)
Mean of Outcome	0.255	0.257	
# Counties	752	776	
Panel C. Industry Specialization Index			
Higher Resettlement	0.018	−0.006	0.023
	(0.008)	(0.009)	(0.009)
Mean of Outcome	0.306	0.266	
# Counties	752	776	

Notes: This table shows the relationship between county resettlement density and labor force participation in 1980, as well as measures of specialization in the labor market in 1991. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel presents the effect of resettlement density on different labor market outcomes: labor force participation rate in 1980 (Panel A); employment concentration across occupations in 1991, measured by the Herfindahl-Hirschman Index (HHI) (Panel B); and employment concentration across industries in 1991, also measured by the HHI (Panel C). Column 1 shows estimates for Chinese individuals, column 2 shows estimates for non-Chinese individuals, and column 3 shows the differences between columns 1 and 2. All regressions are estimated by OLS and include state fixed effects and the main controls: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to the coastline; Chinese population share of the county in 1947; log population density of the county in 1947; share of lands used for rubber cultivation in 1944; and share of lands used for mining in 1944. The unit of observation for Panel A is the individual, and for Panels B and C, it is the county. Data is sourced from the 2% individual-level Census of Population microdata in 1980 and county-level tabulation in 1991. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 6. Educational Attainment in 1980, by County Resettlement Density

	Chinese Individuals (1)	Non-Chinese Individuals (2)	Difference (1) − (2) (3)
Panel A. Years of Schooling			
Higher Resettlement	0.414	0.098	0.315
	(0.223)	(0.123)	(0.150)
Panel B. Primary Education			
Higher Resettlement	0.036	0.018	0.018
	(0.017)	(0.012)	(0.010)
Panel C. Secondary Education			
Higher Resettlement	0.039	0.012	0.027
	(0.022)	(0.012)	(0.017)
# Individuals	31,507	57,345	

Notes: This table shows the relationship between educational attainment in 1980 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel shows the effect of resettlement density on a different outcome of education: years of schooling (Panel A); completion of primary education (Panel B); and completion of secondary education (Panel C). Column 1 reports results for Chinese households, column 2 reports results for non-Chinese households, and column 3 reports the difference in these estimates. All regressions are estimated by OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the individual. The sample is restricted to individuals aged 20 or above from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 7. Household Income in 1980, by County Resettlement Density

	Chinese Households (1)	Non-Chinese Households (2)	Difference (1) − (2) (3)
Panel A. Log Earnings			
Higher Resettlement	0.111	0.037	0.073
	(0.052)	(0.031)	(0.037)
# Households	10,622	22,706	
Panel B. Log Earnings, Primary Sector			
Higher Resettlement	0.073	−0.008	0.082
	(0.036)	(0.040)	(0.044)
# Households	1,660	8,066	
Panel C. Log Earnings, Non-Primary Sector			
Higher Resettlement	0.121	0.044	0.077
	(0.052)	(0.030)	(0.033)
# Households	8,962	14,640	

Notes: This table shows the relationship between household income and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A, Columns 1 and 2 show the effect of resettlement density on log household earnings for Chinese households (column 1) and non-Chinese households (column 2), respectively. Column 3 reports the difference between the estimates in columns 1 and 2. Panel B restricts the sample to households whose head is employed in the primary sector, comprised of agriculture and mining. Panel C restricts the sample to households whose head is employed outside the primary sector. All regressions are estimated by OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the household. Data from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 8. Household Income in 1980, by County Resettlement Density and Distance Between Ethnic Communities

	Log Household Earnings, by Ethnic Group:		
	Chinese Households (1)	Non-Chinese Households (2)	Difference (1) − (2) (3)
Higher Resettlement	0.118	0.060	0.058
	(0.047)	(0.032)	(0.030)
Higher Resettlement × Community Distance	−0.006	−0.013	0.007
	(0.004)	(0.004)	(0.003)
# Households	10,622	22,706	

Notes: This table shows the relationship between household income and county resettlement density by the distance between Chinese and non-Chinese communities. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. "Community Distance" is the average distance between Chinese (primary/secondary) schools and non-Chinese schools in 2022 in a county. Column 1 shows the estimates on log household earnings for Chinese households, and column 2 shows the estimates for non-Chinese households. Column 3 reports the difference between the estimates in columns 1 and 2. All regressions are estimated by OLS and include state fixed effects, the "community distance", and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the household. Data from the 2% individual-level Census of Population microdata in 1980 and the Ministry of Education. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table 9. Parameter Estimates

Parameter (1)	Description (2)	Value (3)	SE (4)
Panel A. Estimated Parameters			
κ	Distance elasticity of migration costs	0.472	(0.003)
α	Expenditure share on agriculture	0.310	(0.001)
θ	Skill dispersion	3.352	(0.031)
γ_A	Productivity spillover w.r.t. size, agric.	-0.103	(0.124)
γ_M	Productivity spillover w.r.t. size, manuf.	0.247	(0.073)
γ^e	Productivity spillover w.r.t. ethnic share	0.146	(0.053)
β	Amenity spillover w.r.t. size	-0.031	(0.046)
β^e	Amenity spillover w.r.t. ethnic share	0.133	(0.092)
Panel B. External Parameters			
σ	Elasticity of substitution	8.00	
ν	Migration elasticity	3.00	
ξ	Distance elasticity of trade costs	0.18	

Notes: This tables shows the estimates of model parameters. Panel A reports parameters estimated from the data, and Panel B reports the remaining three parameters that are assumed or calibrated using external moments provided by other papers. Column 1 shows the Greek symbol of the parameters, column 2 shows the parameter description, and column 3 lists the estimates. Column 4 reports the bootstrap standard errors in parentheses.

Table 10. The Impact of the Emergency Resettlement

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Manufacturing Employment Share	1.19	−1.14	−0.35
Output per Capita	1.60	2.26	1.97
For Agriculture	−0.51	1.24	0.16
For Manufacturing	1.73	3.53	2.92
Aggregate Output (% Baseline Output)	0.69	1.28	1.97
From Reallocation of Labor	0.39	0.95	1.34
From Changes in Productivity	−0.20	0.05	−0.16
From Joint Changes	0.50	0.29	0.79

Notes: This table documents changes in various outcomes from the baseline equilibrium to a "no-resettlement" counterfactual, using the 1947 population distribution as the initial condition (as opposed to the resettled 1957 distribution), relative to their baseline values. The results are divided into three sections. The first section shows the percentage point changes in the share of manufacturing employment for Chinese (column 1), non-Chinese (column 2), and for the economy as a whole (column 3). The second section shows the percent changes in overall output per capita (column 3) and by ethnic group (columns 1–2), as well as per capital output for agriculture and manufacturing separately. The third section, first row, shows the change in aggregate output for Chinese (column 1), non-Chinese (column 2), and in total (column 3), as a percentage of the total baseline output. The subsequent rows break down the total change into contributions from the reallocation of labor, differences in productivity (or output per capita), and joint changes due to cross-term interactions of the first two components.

Table 11. The Impact of Halving Cross-Ethnic Frictions

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Manufacturing Employment Share	−2.14	1.07	−0.02
Output per Capita	4.84	4.79	4.81
For Agriculture	5.23	3.19	4.74
For Manufacturing	5.46	5.02	4.85
Aggregate Output (% Baseline Output)	2.09	2.72	4.81
From Reallocation of Labor	−0.72	0.70	−0.02
From Changes in Productivity	2.29	1.74	4.03
From Joint Changes	0.53	0.27	0.80

Notes: This table documents changes in various outcomes from the baseline economy to a counterfactual equilibrium where the friction in productivity spillover across ethnic groups is halved (i.e., $hatgamma^{e} = 0.5\gamma^{e}$), relative to the baseline values. The results are divided into three sections. The first section shows the percentage point changes in the share of manufacturing employment for Chinese (column 1), non-Chinese (column 2), and for the economy as a whole (column 3). The second section shows the percent changes in overall output per capita (column 3) and by ethnic group (columns 1–2), as well as per capital output for agriculture and manufacturing separately. The third section, first row, shows the change in aggregate output for Chinese (column 1), non-Chinese (column 2), and in total (column 3), as a percentage of the total baseline output. The subsequent rows break down the total change into contributions from the reallocation of labor, differences in productivity (or output per capita), and joint changes due to cross-term interactions of the first two components.

Table 12. The Impact of Subsidizing Malays in Manufacturing

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Manufacturing Employment Share	−9.97	10.34	3.44
Output per Capita	−2.83	4.05	1.07
For Agriculture	1.26	8.95	11.29
For Manufacturing	−0.96	−2.89	−4.35
Aggregate Output (% Baseline Output)	−1.23	2.30	1.07
From Reallocation of Labor	−1.22	1.51	0.29
From Changes in Productivity	−0.27	1.37	1.10
From Joint Changes	0.26	−0.58	−0.32
Welfare	−11.12	−3.34	−5.99

Notes: This table documents changes in various outcomes from the baseline economy to a counterfactual equilibrium with an 18% subsidy for Malays in manufacturing. The results are divided into four sections. The first section shows the percentage point changes in the share of manufacturing employment for Chinese (column 1), non-Chinese (column 2), and for the economy as a whole (column 3). The second section shows the percent changes in overall output per capita (column 3) and by ethnic group (columns 1-2), as well as per capital output for agriculture and manufacturing separately. The third section, first row, shows the change in aggregate output for Chinese (column 1), non-Chinese (column 2), and in total (column 3), as a percentage of the total baseline output. The subsequent rows break down the total change into contributions from the reallocation of labor, differences in productivity (or output per capita), and joint changes due to cross-term interactions of the first two components. The last section shows the percent changes in welfare by ethnic group (columns 1–2) and for the economy as a whole (column 3).

Appendices

A Empirical Results

A.1 Identification Assumptions

This section formalizes the identification assumptions outlined in Section IV. The reduced-form model for the effect of county resettlement density on outcomes is:

$$Y_c = \beta ResettleDensity_c + \alpha \mathbf{1}\left\{ResettleDensity_c > 0\right\} + \lambda \widehat{ResettleDensity}_c + \gamma X_c + \varepsilon_c.$$

To illustrate the method and assumptions, I express county resettlement density in terms of the smaller geographic units at which resettlement occurred:

$$ResettleDensity_c = f_c(g_1, g_2) \equiv \text{asinh}\left(\frac{\sum_{i \in c} g_{1i} \times g_{2i}}{area_c}\right),$$

where $i = 1, \ldots, I$ denotes "sites" smaller than the county, with a total of I sites in the state. The variable g_{1i} indicates whether site i was selected as a resettlement area, with $g_1 \equiv \{g_{1i}\}_{i=1}^{I}$. The number of people resettled to site i is denoted by $g2i$, with $g_2 \equiv \{g_{2i}\}_{i=1}^{I}$. $area_c$ denotes the area of county c.[68]

Leveraging knowledge of the program, I make two assumptions regarding site selection (g_1) and the number of people resettled to the sites (g_2). First, I assume that, conditional on a vector of site characteristics $w_1 \equiv \{w_{1i}\}_{i=1}^{I}$—where w_{1i} includes the distance of site i to the transportation network, land-use type, and the decile of the county's squatter population—the selection of a site is orthogonal to location fundamentals $\varepsilon \equiv \{\varepsilon_c\}_{c=1}^{C}$, where C is the number of counties in the state. Second, I assume that, conditional on the selected sites g_1 and a vector of characteristics $w_2 \equiv \{w_{2i}\}_{i=1}^{I}$—which includes the distance of site i to to each initial squatter settlement and the population of that settlement—the number of people resettled to a site is orthogonal to ε. These assumptions are formalized as follows.

Assumption 1. (Resettlement Exogeneity)

(i) (Site selection) $g_1 \perp\!\!\!\perp \varepsilon \mid w_1$: conditional on distance to transportation, land-use patterns, and the decile of county squatter population, site selection was exogenous.[69]

(ii) (Number resettled) $g_2 \perp\!\!\!\perp \varepsilon \mid (g_1, w_2)$: conditional on the selected sites and the initial distribution of Chinese squatters, the number resettled to a site was exogenous.

Under Assumption 1, the potential omitted variable in Equation (1) is the conditional expectation of resettlement density given $w \equiv (w_1, w_2)$, denoted as $\mathbb{E}[f_c(g_1, g_2)|w]$. As shown

[68] Without loss, each site contains at most one New Village. The variables $g1i$ and g_{2i} are interdependent, as no resettlement occurs at a site not selected for resettlement.

[69] A weaker assumption of mean independence between g_1 and ε, conditional on w, suffices for identification.

by Borusyak and Hull (2023), β is identified when this expected resettlement density is used to re-center the county resettlement density or is controlled for.

To estimate the expected resettlement density, I use the design of the Briggs Plan and make two further assumptions about the distributions of g_1 and g_2, denoted by $G_1(\cdot)$ and $G_2(\cdot)$, respectively.

Assumption 2. (Resettlement Design)

(i) (Equally suitable sites) $G_1(g_1|w_1)$ is uniform: sites were equally likely to be selected conditional on their distance to transportation, land-use patterns, and the decile of county squatter population.

(ii) (Minimizing dislocation) $\mathbb{E}[f_c(g_1, g_2)|g_1, w] = f_c(g_1, \bar{g}_2(g_1, w))$: conditional on the selected village sites and the initial distribution of Chinese squatters, the counterfactual resettlement density followed a gravity-based resettlement plan:

$$\bar{g}_2(g_1, w) = \sum_{j=1}^{J} n_{j\to i} = \sum_{j=1}^{J} n_j \times \frac{d_{ji}^{-\psi}}{\sum_{s=1}^{I} d_{js}^{-\psi}},$$

where n_j is the initial population of Chinese squatters at origin j, d_{ji} is the distance between origin j and site i, and ψ is the resettlement cost elasticity with respect to distance.

Assumption 2(i) implies that the British considered observationally similar sites as equally suitable; and Assumption 2(ii) imposes that they aimed to minimize dislocation but faced idiosyncratic shocks that occasionally resulted in longer relocations. Under these assumptions, the expected resettlement density can be expressed as

$$\begin{aligned}\mathbb{E}\left[f_c(g_1, g_2) \mid w\right] &= \int_{G_1}\int_{G_2} f_c\left(g_1, g_2\right) dG_2\left(g_2|g_1, w\right) dG_1\left(g_1|w\right) \\ &= \int_{G_1} f_c\left(g_1, \bar{g}_2\right) dG_1\left(g_1|w\right),\end{aligned}$$

where the first equality follows from the law of iterated expectation and the second equality follows from Assumption 2(ii).

I estimate the conditional expected resettlement density using a permutation procedure. Each permutation $s = 1, 2, \ldots, S$ is performed independently for each state as follows:

(i). Randomly (and uniformly) permute counterfactual New Village locations $g_1^{(s)}$, conditional on covariates w_1.

(ii). Calculate the gravity resettlement populations for all counterfactual sites $\bar{g}_2^{(s)}$.

(iii). Calculate the counterfactual county resettlement density as $f_c(g_1^{(s)}, \bar{g}_2^{(s)})$.

The expected resettlement density is then approximated by averaging the counterfactual county resettlement density across permutations:

$$\widehat{ResettleDensity}_c \equiv \frac{1}{S}\sum_{s=1}^{S} f_c(g_1^{(s)}, \bar{g}_2^{(s)}).$$

A comparison between the actual resettlement pattern and the expected resettlement density suggests that Assumption 2(ii) is reasonable (Appendix Figure A.3). The actual county resettlement density centers around the expected resettlement density. The village-level resettled population also matches well with the model prediction.

A.2 Robustness

The analysis requires a set of assumptions about what defines a plausible counterfactual resettlement scheme. This section shows that the results are robust to alternative specifications of the expected resettlement density and different covariate choices, as well as sample restrictions.

First, the baseline population shifter is defined as the inverse hyperbolic transformation of the number of resettled persons per unit area. This log-like transformation is motivated by efficiency, as it aims to shift population density in percent terms. I show in Appendix Table A.11 that the results are similar when using a logarithm transformation with an imputed value for zeros.

I also examine alternative specifications for the counterfactual resettlement sites. The baseline analysis assumes that the British prioritized moving people to nearby rivers when no roads were within 5 kilometers. Appendix Table A.11 shows that the estimates are similar when assuming a preference for roads over rivers up to 10 kilometers. Additionally, the baseline approach randomly permutes counterfactual villages across space, allowing them to be sited arbitrarily close to each another. Although the minimum distance between villages in the data is only 200 meters—suggesting that closely spaced villages are feasible—most sites are at least 1 kilometer apart. Appendix Table A.11 shows that the results are robust to enforcing a minimum spacing of 1 kilometer between counterfactual villages.

I show that the expected number of squatters resettled to each counterfactual village site is also robust to different definitions of squatters and variations in resettlement cost elasticity. In the baseline, squatters are defined as Chinese communities originally residing within 5 kilometers of the forest. Appendix Table A.11 shows similar estimates when using a

10-kilometer cutoff. In addition, the main analysis models the expected number of resettled populations for each counterfactual site using a gravity equation, with cost elasticity based on the distance relocated. A higher cost elasticity suggests that counterfactual resettlement density is closer to the original squatter density. While the baseline analysis calibrates elasticity based on observed populations in the New Villages, I show in Appendix Table A.11 that the results are robust to different values of this elasticity.

I include the expected resettlement density in the regression to capture potential omitted variables related to nearby roads and population. Appendix Table A.12 shows that, as expected, the estimates remain stable when I additionally control for the transportation and population covariates of neighboring counties. The results are also robust to controlling for features of productivity fundamentals, including agricultural productivity (ruggedness, paddy rice suitability, and coconut suitability) and industrial productivity (distance to prewar industrial facilities and major cities). The robustness of these results is not surprising, given the balance result established in Section IV.B.

The counties in the baseline sample vary in size, with some large, sparsely populated counties inland and smaller, more populated counties along the coast. This is why I control for county area in the main specification. In Appendix Table A.13, I show that the results are not sensitive to excluding large and small counties or the most densely populated prewar towns. The estimates are also robust to excluding counties with extreme resettlement density.

Although the identifying variation comes only from resettled counties, I include other counties in the baseline to help estimate the effects of covariates and improve efficiency. Appendix Table A.13 shows that the estimates using only the resettled counties are largely similar, though slightly larger and noisier. Lastly, the individual- or household-level outcomes are drawn from the 2% microdata in the 1980 Population Census, and counties with sampled Chinese only cover two-thirds of the baseline counties. The estimates are also similar when limited to these counties.

B Theoretical Results

B.1 Sectoral Labor Supply

I now derive the key equations pertaining to the sectoral labor supply. Individuals draw their efficiency units independently across sectors of agriculture and manufacturing $\Lambda^e = (\Lambda^e_A, \Lambda^e_M)$ from the joint distribution:

$$F^e_n(\Lambda_A, \Lambda_M) = \prod_{k=A,M} F^e_{nk}(\Lambda_k),$$

where the marginal probability distribution is Fréchet:

$$F^e_{nk}(\Lambda_k) = \exp\left(-\phi^e_{nk}\Lambda_k^{-\theta}\right).$$

After knowing their efficiency units, they choose the sector that pays higher earnings. Let w_{nk} be the wage per efficiency unit for industry k in region n. The earnings of individual i of ethnicity e in industry k, location n is thus

$$\begin{aligned} y^e_{ink} &= w_{nk}\lambda^e_{ink} \\ &= w_{nk}\Lambda^e_{ink} f(L^c_{nk}, L^m_{nk}) \\ &= w^e_{nk}\Lambda^e_{ink}, \end{aligned}$$

where

$$w^e_{nk} \equiv w_{nk} f(L^c_{nk}, L^m_{nk}).$$

Function $f(L^c_{nk}, L^m_{nk})$, which depends on local population distribution, captures human capital externalities.

Since y^e_{ink} equals a constant w^e_{nk} multiplied by a Fréchet random variable Λ^e_{ink}, it is also Fréchet distributed with shape θ and scale $\phi^e_{nk}(w^e_{nk})^\theta$. The expected earnings for ethnicity e in industry k and region n is thus $\Gamma_\theta\left(\phi^e_{nk}(w^e_{nk})^\theta\right)^{1/\theta}$.

For an individual of ethnicity e in region n, the probability of choosing to work in industry k is

$$\pi^e_{nk} \equiv \mathbb{P}(y^e_{ink} = \max_s y^e_{ins}) = \frac{\phi^e_{nk}(w^e_{nk})^\theta}{\sum_s \phi^e_{ns}(w^e_{ns})^\theta} = \phi^e_{nk}\left(\frac{w^e_{nk}}{\bar{w}^e_n}\right)^\theta,$$

where

$$\bar{w}^e_n \equiv \left(\phi^e_{nA}\left(w^e_{nA}\right)^\theta + \phi^e_{nM}\left(w^e_{nM}\right)^\theta\right)^{1/\theta}.$$

Since people of ethnicity e choose the sector that pays more and this process continues until the (e-specific) earning equalize across the two sectors, in equilibrium, the average wage for ethnic group e in region n is given by

$$\mathbb{E}[\max_k y^e_{ink}] = \Gamma_\theta\left(\sum_k \phi^e_k(w^e_{nk})^\theta\right)^{1/\theta} = \Gamma_\theta\bar{w}^e_n.$$

Moreover, due to the Fréchet property, ethnic group e in region n attain, on average, the same earning across the two sectors.

It follows that the average skill of group-e in region n, sector k, is given by

$$\mathbb{E}[\underbrace{y^e_{ink}/w^e_{nk}}_{\Lambda^e_{ink}}|y^e_{ink} = \max_s y^e_{ins}]f(L^c_{nk}, L^m_{nk}) = \Gamma_\theta \bar{w}^e_n w^{-1}_{nk}.$$

Notice that it can also be written in terms of occupation share as

$$\Gamma_\theta \left(\phi^e_{nk}\right)^{1/\theta} \left(\pi^e_{nk}\right)^{-1/\theta} f(L^c_{nk}, L^m_{nk}),$$

where the neoclassical force $(\pi^e_{nk})^{-1/\theta}$ implies that a higher share of labor supply tends to lower the average skill in the sector due to selection. In contrast, the externality term $f(L^c_{nk}, L^m_{nk})$ tends to increase the average skills in the number of population.

The aggregate sectoral earnings from ethnicity e in industry k and region n is the local population of ethnicity e multiplied by the share working in industry k and by their average sectoral earning conditional on choosing k:

$$w_{nk} H^e_{nk} = L^e_n \pi^e_{nk} \Gamma_\theta \bar{w}^e_n.$$

This implies that the aggregate human capital supply in industry k, region n is

$$\begin{aligned} H_{nk} &= \Gamma_\theta \sum_e L^e_n \phi^e_{nk} (w^e_{nk})^\theta w^{-1}_{nk} (\bar{w}^e_n)^{1-\theta} \\ &= \Gamma_\theta \sum_e L^e_n \phi^e_{nk} w^{-1}_{nk} w^e_{nk} (w^e_{nk})^{1-\theta} (\bar{w}^e_n)^{1-\theta} \\ &= \Gamma_\theta \sum_e L^e_n \phi^e_{nk} \left(L_{nk}\right)^{\gamma_k} \left(\frac{L^e_{nk}}{L_{nk}}\right)^{\gamma^e} \left(\frac{w^e_{nk}}{\bar{w}^e_n}\right)^{\theta-1}. \end{aligned}$$

B.2 Migration

Individuals of group e draw an idiosyncratic taste shock for each location and decide where to migrate before knowing their efficiency units. The taste shock u^e_n is assumed to drawn from the following location-specific Fréchet distribution

$$F^e_n(a) = \exp\left(-\bar{a}^e_n a^{-\nu}\right),$$

where the scale $\bar{a}^e_n$ captures the average attractiveness of location n for group e and the shape ν captures the dispersion of taste (which is assumed to be the same for all groups and locations).

The value of relocating from r to n for ethnicity e is

$$V_{rn}^{e}=\eta_{rn}^{-1}a_{n}^{e}\Gamma_{\theta}\bar{w}_{n}^{e}P_{n}^{-1}$$

where η_{rn} is the migration cost and the amenity term a_n^e depends on the local population:

$$a_{n}^{e}=u_{n}^{e}\left(L_{n}\right)^{\beta}\left(\frac{L_{n}^{e}}{L_{n}}\right)^{\beta^{e}}.$$

As V_{rn}^e is a Fréchet random variable u_n^e multiplied by a constant $\eta_{rn}^{-1}L_n^{\beta}(L_n^e/L_n)^{\beta^e}\Gamma_{\theta}\bar{w}_n^e P_n^{-1}$, it is itself Fréchet distributed. The distribution of V_{rn}^e thus implies that the probability of relocating from r to n for ethnicity e is

$$m_{rn}^{e}\equiv\mathbb{P}\left(V_{rn}^{e}=\max_{l}V_{rl}^{e}\right)=\frac{\bar{a}_{n}^{e}\left(\eta_{rn}^{-1}\left(L_{n}\right)^{\beta}\left(L_{n}^{e}/L_{n}\right)^{\beta^{e}}\bar{w}_{n}^{e}P_{n}^{-1}\right)^{\nu}}{\sum_{l=1}^{N}\bar{a}_{l}^{e}\left(\eta_{rl}^{-1}\left(L_{l}\right)^{\beta}\left(L_{l}^{e}/L_{l}\right)^{\beta^{e}}\bar{w}_{l}^{e}P_{l}^{-1}\right)^{\nu}}.$$

B.3 Iterative Procedure for Solving the Equilibrium

Given the model parameters and the inferred location fundamentals, I solve for equilibrium quantities and prices using an iterative approach with three nested loops. The outer loop solves for population by ethnic group $\{L_n^e\}$; the second loop, given $\{L_n^e\}$, solves for sector-specific wages per efficiency unit $\{w_{nk}\}$; and the third loop, taking $\{L_n^e, w_{nk}\}$ as given, solves for occupation shares $\{\pi_{nk}^e\}$, prices, and incomes. The iterative algorithm proceeds as follows.

The process starts with an initial guess for the equilibrium population distribution $\{L_n^e\}$, followed by the steps below.

1. Solve for wages $\{w_{nk}\}$:
   (a) Set an initial guess for wages $\{w_{nk}\}$.
   (b) Solve for occupational choices $\{\pi_{nk}^e\}$:
      i. Set an initial guess for $\{\pi_{nA}^e\}$ and calculate $\pi_{nM}^e = 1-\pi_{nA}^e$.
      ii. Calculate sectoral employment $L_{nk}^e = L_n^e\pi_{nk}^e$.
      iii. Calculate the average wage by ethnic group:

$$\bar{w}_{n}^{e}=\left(\phi_{nA}^{e}(w_{nA}^{e})^{\theta}+\phi_{nM}^{e}(w_{nM}^{e})^{\theta}\right)^{1/\theta},$$

where

$$w_{nk}^{e}=w_{nk}(L_{nk})^{\gamma_{k}}\left(\frac{L_{nk}^{e}}{L_{nk}}\right)^{\gamma^{e}}.$$

iv. Calculate the implied occupational shares:

$$\tilde{\pi}^e_{nk} \equiv \phi^e_{nk} \left(\frac{w^e_{nk}}{\bar{w}^e_n} \right)^\theta .$$

v. Update the occupational choices iteratively until convergence, using:

$$\pi^e_{nk,new} \equiv \iota \pi^e_{nk} + (1-\iota)\tilde{\pi}^e_{nk},$$

where $\iota \in (0,1)$ is the relaxation parameter in the Gauss-Seidel update. A lower ι accelerates the process but is more prone to overshooting and instability. I set $\iota = 0.95$ in practice.

(c) Calculate prices $\{p_{nrk}\}$ with $p_{nnk} = w_{nk}\tau_{nn}$ and $p_{nrk} = p_{nnk}\left(\frac{\tau_{nr}}{\tau_{nn}}\right)$, where τ_{nn} is the within-county trade cost, which can be greater than 1.

(d) Calculate labor efficiency $\{H_{nk}\}$ with $H_{nk} = \sum_e H^e_{nk}$, where $H^e_{nk} = \Gamma_\theta L^e_{nk}\left(\frac{\bar{w}^e_n}{w_{nk}}\right)$.

(e) Solve for regional income $\{Y_n\}$, such that $\sum_n Y_n = 1$.

i. Calculate total income of n by summing its trade flow expenditures over k and r:

$$Y_n = \sum_r \sum_k Y_r \underbrace{\alpha_k \left(\frac{p^{1-\sigma}_{nrk}}{\sum_l p^{1-\sigma}_{lrk}} \right)}_{\equiv \tilde{p}_{nrk}} = \sum_r \sum_k \tilde{p}_{nrk} Y_r = \sum_r \tilde{p}_{nr} Y_r,$$

where $\tilde{p}_{nr} \equiv \tilde{p}_{nrA} + \tilde{p}_{nrM}$. In matrix form, this can be written as:

$$Y = \tilde{P}Y \iff (I - \tilde{P})Y = \mathbf{0},$$

where

$$\tilde{P} = \begin{bmatrix} \tilde{p}_{11} & \cdots & \tilde{p}_{1N} \\ \vdots & \ddots & \vdots \\ \tilde{p}_{N1} & \cdots & \tilde{p}_{NN} \end{bmatrix}.$$

ii. Since this system has rank $N-1$, I impose $\sum_n Y_n = 1$ as a numeraire to pin down the level of Y. By dropping the last equation from above and replacing it with $\sum_n Y_n = 1$, I obtain:

$$\begin{bmatrix} 1-\tilde{p}_{11} & \cdots & -\tilde{p}_{1N} \\ \vdots & \ddots & \vdots \\ 1-\tilde{p}_{N-1,1} & & -\tilde{p}_{N-1,N} \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_{N-1} \\ Y_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

(f) Calculate the implied wages: $\tilde{w}_{nk} = Y_{nk}/H_{nk}$, where $Y_{nk} = \sum_r \tilde{p}_{nrk} Y_r$.

(g) Update wages iteratively until convergence:

$$w_{nk,new} \equiv \iota w_{nk} + (1-\iota)\tilde{w}_{nk}.$$

2. Calculate migration shares:

$$m^e_{rn} = \frac{\left(\eta^{-1}_{rn} V^e_n\right)^{\nu}}{\sum_{l=1}^N \left(\eta^{-1}_{rl} V^e_l\right)^{\nu}},$$

where

$$V^e_n = \left(\bar{a}^e_n\right)^{1/\nu} L^{\beta}_n \left(\frac{L^e_n}{L_n}\right)^{\beta^e} \bar{w}^e_n P^{-1}_n.$$

3. Calculate the implied population distribution: $\tilde{L}^e_n = \sum_r \check{L}^e_r m^e_{rn}$.

4. Update population iteratively until convergence:

$$L^e_{n,new} \equiv \iota L^e_n + (1-\iota)\tilde{L}^e_n.$$

B.4 Sufficient Conditions for Uniqueness of Equilibrium

This section applies Theorem 1 from Allen et al. (forthcoming) to derive sufficient conditions for the uniqueness of equilibrium. I rewrite the system of equations that characterizes the equilibrium in terms of 15 unknowns

$$\{w_{nA}, w_{nM}, P_{nA}, P_{nM}, H_{nA}, H_{nM}, \bar{w}^c_n, \bar{w}^m_n, L_{nA}, L_{nM}, L^c_n, L^m_n, L_n, \Pi^c_n, \Pi^m_n\}$$

and 15 equations:

$$
\begin{aligned}
w_{nk}^{\sigma} H_{nk} &= \sum_{r} \alpha_k \tau_{nr}^{1-\sigma} \left(w_{rA} H_{rA} + w_{rM} H_{rM} \right) P_{rk}^{\sigma-1} \\
P_{nk}^{1-\sigma} &= \sum_{r} \tau_{rn}^{1-\sigma} w_{rk}^{1-\sigma} \\
w_{nk} H_{nk} &= \sum_{e} \Gamma_{\theta} \left(\phi_{nk}^{e} \right)^{\frac{1}{1-\gamma^e\theta}} \bar{w}_{n}^{e} \left(w_{nk} \right)^{\frac{\theta}{1-\gamma^e\theta}} \left(L_{n}^{e} \right)^{\frac{1}{1-\gamma^e\theta}} \left(L_{nk} \right)^{\frac{\theta(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\
\left(\bar{w}_{n}^{e} \right)^{\theta} &= \sum_{k} \left(\phi_{nk}^{e} \right)^{\frac{1+\gamma^e\theta}{1-\gamma^e\theta}} \left(w_{nk} \right)^{\frac{\theta}{1-\gamma^e\theta}} \left(L_{n}^{e} \right)^{\frac{\gamma^e\theta}{1-\gamma^e\theta}} \left(L_{nk} \right)^{\frac{\theta(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\
L_{nk}^{1-(\gamma_k-\gamma^e)\theta} \left(w_{nk} \right)^{-\theta} &= \sum_{e} \left(\phi_{nk}^{e} \right)^{\frac{1+\gamma^e\theta}{1-\gamma^e\theta}} \left(\bar{w}_{n}^{e} \right)^{-\theta} \left(w_{nk} \right)^{\frac{\gamma^e\theta^2}{1-\gamma^e\theta}} \left(L_{n}^{e} \right)^{\frac{1}{1-\gamma^e\theta}} \left(L_{nk} \right)^{\frac{\gamma^e\theta^2(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\
\left(L_n \right)^{(\beta^e-\beta)\nu} \left(L_{n}^{e} \right)^{1-\beta^e\nu} \left(\bar{w}_{n}^{e} \right)^{-\nu} P_{nA}^{\nu\alpha} P_{nM}^{\nu(1-\alpha)} &= \sum_{r} \bar{a}_{n}^{e} \check{L}_{r}^{e} \eta_{rn}^{-\nu} \left(\Pi_{r}^{e} \right)^{-\nu} \\
\left(\Pi_{n}^{e} \right)^{\nu} &= \sum_{r} \eta_{nr}^{-\nu} \bar{a}_{r}^{e} \left(L_r \right)^{(\beta-\beta^e)\nu} \left(L_{r}^{e} \right)^{\beta^e\nu} \left(\bar{w}_{r}^{e} \right)^{\nu} P_{rA}^{-\nu\alpha} P_{rM}^{-\nu(1-\alpha)} \\
L_n &= \sum_{e} L_{n}^{e}.
\end{aligned}
$$

The equilibrium contains a set of $\mathcal{N} = \{1, \ldots, N\}$ locations and a set of $\mathcal{H} = 1, \ldots, H$ economic interactions (or endogenous variables), where $H = 15$. The $H \times H$ matrices B and Γ, as in Allen et al. (forthcoming), are given by

$$
B = \begin{bmatrix}
1 & 1 & \sigma-1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & 1 & \cdot & \sigma-1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1-\sigma & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 1-\sigma & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\frac{\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \frac{\theta(\gamma_A-\gamma^e)}{1-\gamma^e\theta} & \cdot & \frac{1}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \cdot & \cdot & \cdot \\
\cdot & \frac{\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \frac{\theta(\gamma_M-\gamma^e)}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \cdot & \cdot & \cdot \\
\frac{\theta}{1-\gamma^e\theta} & \frac{\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\theta(\gamma_A-\gamma^e)}{1-\gamma^e\theta} & \frac{\theta(\gamma_M-\gamma^e)}{1-\gamma^e\theta} & \frac{\gamma^e\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot \\
\frac{\theta}{1-\gamma^e\theta} & \frac{\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\theta(\gamma_A-\gamma^e)}{1-\gamma^e\theta} & \frac{\theta(\gamma_M-\gamma^e)}{1-\gamma^e\theta} & \cdot & \frac{\gamma^e\theta}{1-\gamma^e\theta} & \cdot & \cdot & \cdot \\
\frac{\gamma^e\theta^2}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & \cdot & -\theta & -\theta & \frac{\gamma^e\theta(\gamma_A-\gamma^e)}{1-\gamma^e\theta} & \cdot & \frac{1}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \cdot & \cdot & \cdot \\
\cdot & \frac{\gamma^e\theta^2}{1-\gamma^e\theta} & \cdot & \cdot & \cdot & \cdot & -\theta & -\theta & \cdot & \frac{\gamma^e\theta(\gamma_M-\gamma^e)}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \frac{1}{1-\gamma^e\theta} & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -\nu & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -\nu \\
\cdot & \cdot & -\nu\alpha & \nu(\alpha-1) & \cdot & \cdot & \nu & \cdot & \cdot & \cdot & \beta^e\nu & \cdot & (\beta-\beta^e)\nu & \cdot & \cdot \\
\cdot & \cdot & -\nu\alpha & \nu(\alpha-1) & \cdot & \cdot & \cdot & \nu & \cdot & \cdot & \cdot & \beta^e\nu & (\beta-\beta^e)\nu & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot
\end{bmatrix}
$$

$$
\Gamma = \begin{bmatrix}
\sigma & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \sigma & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & 1-\sigma & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1-\sigma & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \theta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \theta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
-\theta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1-(\gamma_A-\gamma^e)\theta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & -\theta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1-(\gamma_M-\gamma^e)\theta & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \nu\alpha & \nu(1-\alpha) & \cdot & \cdot & -\nu & \cdot & \cdot & \cdot & 1-\beta^e\nu & \cdot & (\beta^e-\beta)\nu & \cdot & \cdot \\
\cdot & \cdot & \nu\alpha & \nu(1-\alpha) & \cdot & \cdot & \cdot & -\nu & \cdot & \cdot & \cdot & 1-\beta^e\nu & (\beta^e-\beta)\nu & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \nu & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \nu \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot
\end{bmatrix}
$$

Next, I calculate $A = B\Gamma^{-1}$ and its spectral radius, denoted by $\rho(A)$ (i.e. its largest eigenvalue in absolute value). According to Theorem 1 in Allen et al. (forthcoming), a sufficient condition for uniqueness is that $\rho(A) < 1$. Although my baseline parameter values (Table 9) does not imply a spectral radius of A that is smaller than one, this is only a sufficient condition, so the equilibrium may still be unique. As noted in Remark 5 of Allen et al. (forthcoming), changing the system of equations through a change of variables may reduce the spectral radius, leading to a different sufficient condition that is more likely to hold.

C Structural Estimation

C.1 Market Access Terms

I derive four underlying conditions involving the trade and migration market access terms from the equilibrium conditions (16)–(18).

(i). Total sales equals payments to labor: $w_{nk}H_{nk} = \sum_r X_{nrk}$. Using Equation (13), this can be written as

$$\mathcal{P}_{nk}^{1-\sigma} = \frac{\alpha_k}{\Omega_{nk}} \sum_r \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1},$$

where $\Omega_{nk} \equiv w_{nk}H_{nk}/Y_n$ denotes the share of income in region n generated from sector k.

(ii). Total income equals total expenditure: $Y_r\alpha_k = \sum_n X_{nrk}$. This can be written as

$$P_{rk}^{1-\sigma} = \sum_n \tau_{nr}^{1-\sigma} Y_n \mathcal{P}_{nk}^{\sigma-1}.$$

(iii). Final population equals total in-migrations: $L_n^e = \sum_{r=1}^N L_{rn}^e$. Using Equation (6), this can be written as

$$\left(\mathcal{V}_n^e\right)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e \left(\Pi_r^e\right)^{-v}$$

(iv). Initial population equals total out-migrations: $\check{L}_r^e = \sum_{n=1}^N L_{rn}^e$. This can be written as

$$\left(\Pi_r^e\right)^{v} = \sum_n \eta_{rn}^{-v} L_n^e \left(\mathcal{V}_n^e\right)^{\nu}.$$

Putting these together, the derivation above yields a system of four equations:

$$\mathcal{P}_{nk}^{1-\sigma} = \frac{\alpha_k}{\Omega_{nk}} \sum_r \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1}, \tag{A-1}$$

$$P_{rk}^{1-\sigma} = \sum_n \tau_{nr}^{1-\sigma} Y_n \mathcal{P}_{nk}^{\sigma-1}, \tag{A-2}$$

$$\left(\mathcal{V}_n^e\right)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e \left(\Pi_r^e\right)^{-v}, \tag{A-3}$$

$$\left(\Pi_r^e\right)^{v} = \sum_n \eta_{rn}^{-v} L_n^e \left(\mathcal{V}_n^e\right)^{\nu}, \tag{A-4}$$

Given data on total income $\{Y_n\}$ and sectoral income shares $\{\Omega_{nk}\}$, the agricultural expenditure share α is identified. Since each region spends the same proportion of income on agricultural goods, the economy as a whole must also spend that same share in aggregate:

$$\alpha = \frac{\sum_n w_{nA} H_{nA}}{\sum_n w_{nA} H_{nA} + w_{nM} H_{nM}} = \frac{\sum_n Y_n \Omega_{nA}}{\bar{Y}} = \sum_n Y_n \Omega_{nA}.$$

The four equations (A-1)–(A-4) can be separated into two sets: one for the trade market access and one for migration market access. The equations for trade market access are:

$$\mathcal{P}_{nk}^{1-\sigma} = \sum_r \frac{\alpha_k}{\Omega_{nk}} \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1},$$
$$P_{nk}^{1-\sigma} = \sum_r \tau_{rn}^{1-\sigma} Y_r \mathcal{P}_{rk}^{\sigma-1}.$$

The migration market access equations are:

$$\left(\mathcal{V}_n^e\right)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e \left(\Pi_r^e\right)^{-v},$$
$$\left(\Pi_n^e\right)^{v} = \sum_r \eta_{nr}^{-v} L_r^e \left(\mathcal{V}_r^e\right)^{\nu}.$$

I can rewrite the first set of equations as:

$$x_{nk}^{-1} = \sum_r K_{nrk}^A y_{rk},$$
$$y_{nk}^{-1} = \sum_r K_{nr}^B x_{rk},$$

where $x_{nk} \equiv \mathcal{P}_{nk}^{\sigma-1}$ and $y_{nk} \equiv P_{nk}^{\sigma-1}$. Using Allen et al. (forthcoming), I compute matrices B_P and Γ_P as:

$$B_P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad \Gamma_P = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Thus, we have

$$A_P \equiv |B_p \Gamma_P^{-1}| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The spectral radius of A_P, which is the largest absolute value of its eigenvalues, is 1. Based on Theorem 1, part ii.b of Allen et al. (forthcoming), this guarantees the existence of a unique solution for $\{\mathcal{P}_{nk}^{\sigma-1}, P_{nk}^{\sigma-1}\}$ up to a scale.

Similarly, the second set of equations can be rewritten as:

$$x_{ne}^{-1} = \sum_r K_{nre}^C y_{re}^{-1},$$
$$y_{ne} = \sum_r K_{nre}^D x_{re},$$

where $x_{ne} \equiv (\mathcal{V}_n^e)^{\nu}$ and $y_{ne} \equiv (\Pi_n^e)^{\nu}$. The corresponding matrices B_V and Γ_V are

$$B_V = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}; \quad \Gamma_V = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, we have:

$$A_V \equiv \left|B_V \Gamma_V^{-1}\right| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Since the spectral radius of A_V is also 1, by the same argument, there exists a unique solution for $\{(\mathcal{V}_n^e)^\nu, (\Pi_n^e)^\nu\}$ up to a scale.

C.2 Migration Cost Elasticity

The non-linear least squares estimation for migration cost elasticity proceeds as follows.

(i). Guess an initial $\tilde{\kappa}$ and calculate the corresponding migration costs $\eta_{rn}^\nu = (d_{rn}/d_{min})^{\tilde{\kappa}}$.

(ii). Using the initial and final population data $\check{L}_r^e$, L_n^e, solve for the migration market access terms $(\mathcal{V}_n^e)^\nu$, $(\Pi_n^e)^\nu$ as per Proposition 1.

(iii). Calculate the implied bilateral migration flows:

$$L_{rn} = \sum_e L_{rn}^e = \sum_e d_{rn}^{-\tilde{\kappa}} \times \frac{\check{L}_r^e}{(\Pi_r^e)^\nu} \times \frac{L_n^e}{(\mathcal{V}_n^e)^{-\nu}}.$$

(iv). Aggregate the model-implied migration flows to the district level and compute bilateral migration shares:

$$m_{jh} = \frac{\sum_{r\in j(r)} \sum_{n\in h(n)} L_{rn}}{\sum_{r\in j(r)} \sum_n L_{rn}},$$

where $j(r)$ and $h(n)$ denote the districts that counties r and n belong.

(v). Calculate the loss function as the sum of squared differences between the model-predicted and observed (log) migration shares:

$$loss \equiv \frac{1}{N_d^2} \sum_{j,h} \left(\ln m_{jh} - \ln \hat{m}_{jh}\right)^2, \tag{A-5}$$

where N_d is the total number of districts, and $\hat{m}_{jh}$ denotes the observed migration shares.

(vi). Search over the space of $\tilde{\kappa}$ to minimize the loss function.

D Counterfactuals

D.1 Place-Based Subsidies

This section describes how I solve for place-based wage subsidies—financed by a uniform income tax—that incentivize voluntary migration of Chinese and Malays from their 1947 population distribution to the 1957 resettled distribution.

Let ϵ_n^e denote the place-based, ad-valorem subsidy for group e in region n. The migration shares for group e from region r are then given by:

$$m_{rn}^e = \frac{\bar{a}_n^e \left(\eta_{rn}^{-1} (L_n)^\beta (L_n^e/L_n)^{\beta^e} (1+\epsilon_n^e) \bar{w}_n^e P_n^{-1}\right)^\nu}{\sum_{l=1}^N \bar{a}_l^e \left(\eta_{rl}^{-1} (L_l)^\beta (L_l^e/L_l)^{\beta^e} (1+\epsilon_l^e) \bar{w}_l^e P_l^{-1}\right)^\nu}.$$

Using the balance of migration flows, $L_n^e = \sum_r \check{L}_r^e m_{rn}^e$, we obtain:

$$L_n^e = \sum_r \check{L}_r^e \frac{\bar{a}_n^e \left(\eta_{rn}^{-1} (L_n)^\beta (L_n^e/L_n)^{\beta^e} (1+\epsilon_n^e) \bar{w}_n^e P_n^{-1}\right)^\nu}{\sum_{l=1}^N \bar{a}_l^e \left(\eta_{rl}^{-1} (L_l)^\beta (L_l^e/L_l)^{\beta^e} (1+\epsilon_l^e) \bar{w}_l^e P_l^{-1}\right)^\nu}.$$

Rearranging this, we derive the expression for the place-based subsidy:

$$(1+\epsilon_n^e)^{-\nu} = \frac{1}{L_n^e} \sum_r \check{L}_r^e \frac{\bar{a}_n^e \eta_{rn}^{-\nu} \left((L_n)^\beta (L_n^e/L_n)^{\beta^e} \bar{w}_n^e P_n^{-1}\right)^\nu}{\sum_{l=1}^N \bar{a}_l^e \eta_{rl}^{-\nu} \left((L_l)^\beta (L_l^e/L_l)^{\beta^e} (1+\epsilon_l^e) \bar{w}_l^e P_l^{-1}\right)^\nu}. \tag{A-6}$$

Since migration depends only on relative wages across regions, this system provides only $N-1$ equations, implying that ϵ_n^e is determined only up to a scale. To calculate the least-cost, weakly positive ad-valorem subsidies, I scale the solution vector such that the minimum subsidy across locations is zero. Specifically, let $\{\epsilon_n^e\}$ be any solution to (A-6). Then, the least-cost, weakly positive ad-valorem subsidies for group e, denoted $\{\varepsilon_n^e\}$, are defined by:

$$\varepsilon_n^e \equiv \frac{1+\epsilon_n^e}{1+\min_n(\epsilon_n^e)} - 1.$$

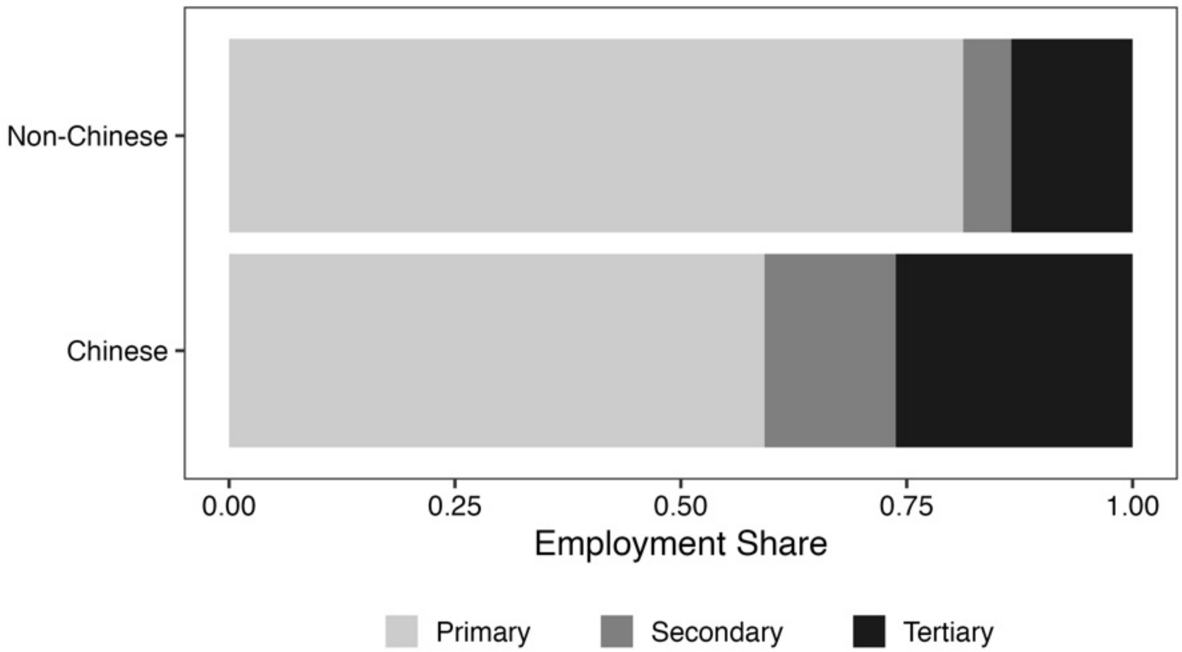
Finally, the required tax rate to balance the government's budget is given by:

$$t = \frac{\sum_e \sum_n \varepsilon_n^e L_n^e \bar{w}_n^e P_n^{-1}}{\sum_e \sum_n L_n^e \bar{w}_n^e P_n^{-1}},$$

where the numerator represents the total subsidy costs (in real terms) and the denominator is the tax base (total real income).

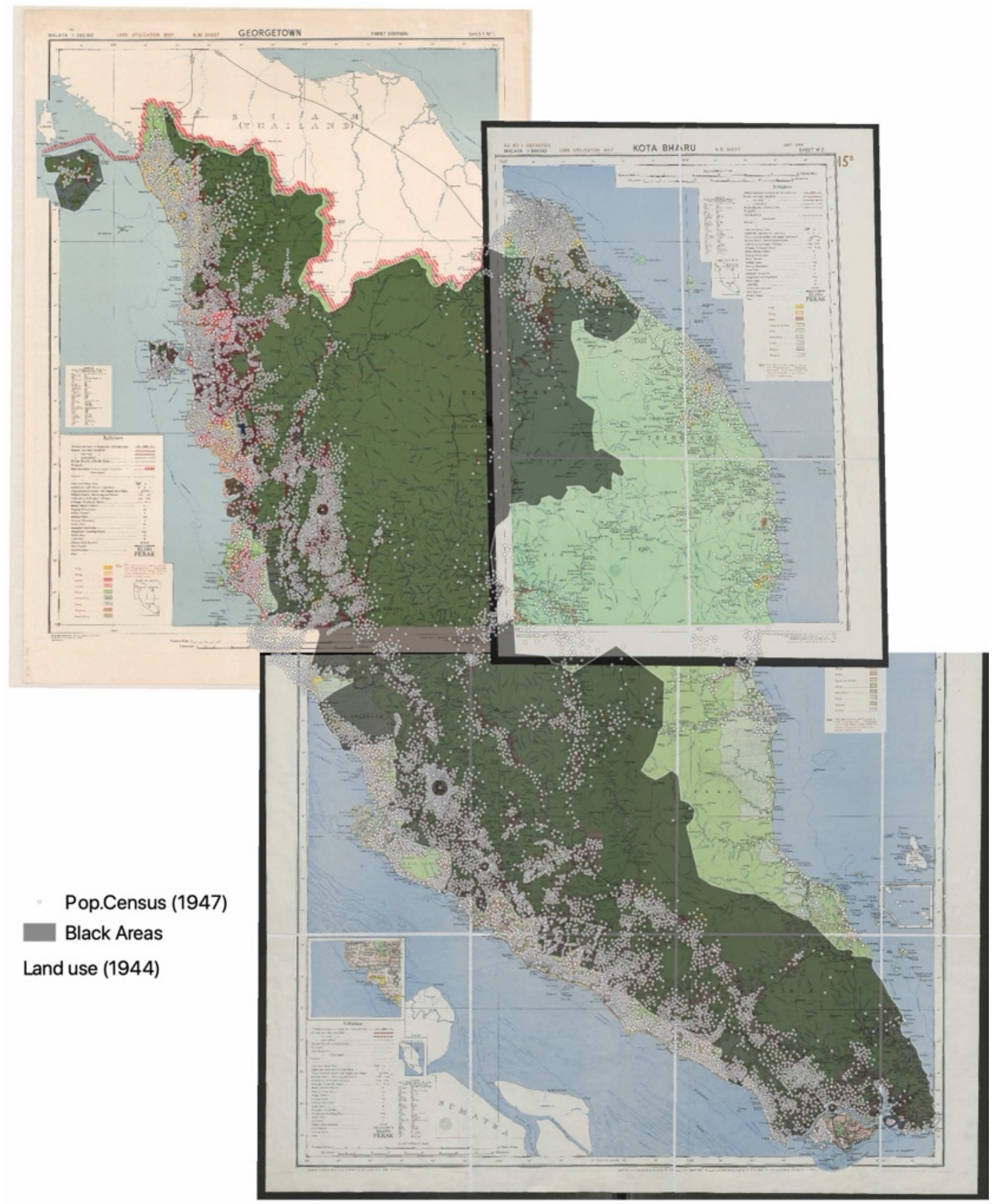
E Appendix Figures

Figure A.1. Employment Share in 1947, by Ethnic Group



Notes: This figure shows the employment share across the primary, secondary, and tertiary industries for Chinese and non-Chinese, respectively. The primary sector includes agriculture and mining. The secondary sector includes manufacturing, utility, and construction. The tertiary sector includes transportation, communication, commerce, finance, business, and other services. Data from the 1947 Census of Population (Del Tufo, 1947).

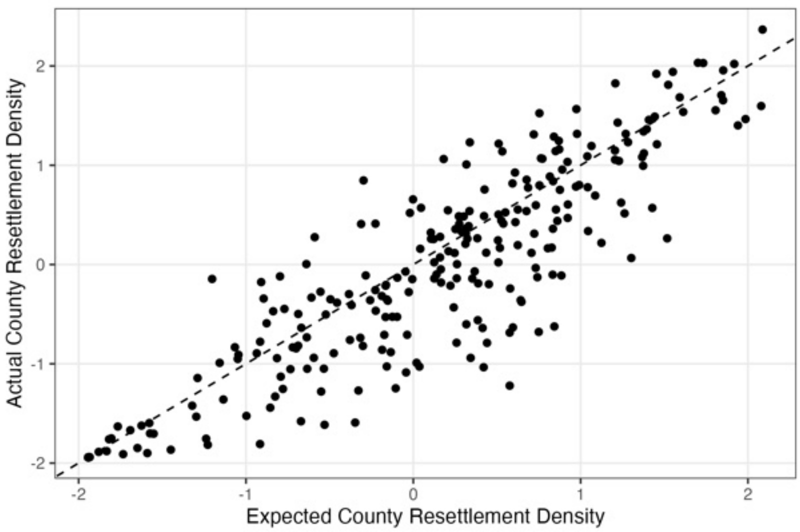
Figure A.2. Population Distribution of the Squatters



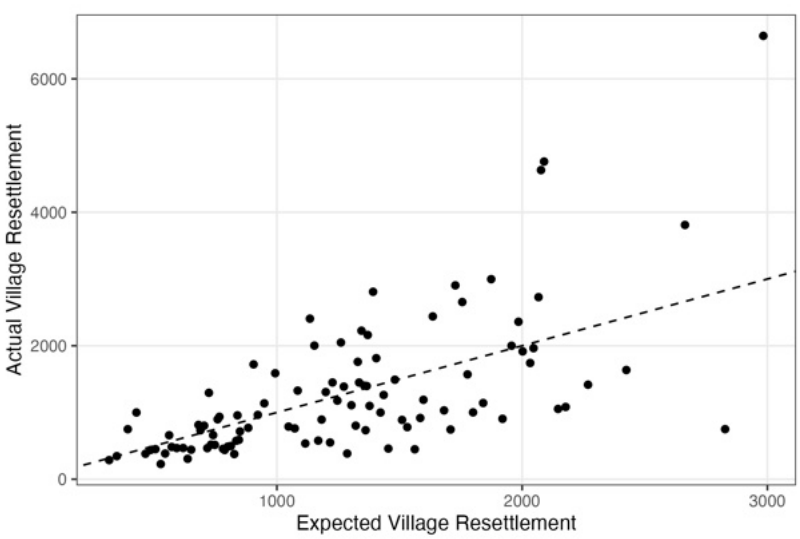
Notes: This figure shows the distribution of the squatters inferred from the intersection of three maps. The gray dots represent population clusters provided by the 1947 census. The areas shaded in dark are the "Black areas" with communist activities and were under various Emergency regulations. The areas shaded in green are areas classified as forest from land utilization maps in 1943 (War Office, 1943).

Figure A.3. Comparison of Actual and Predicted Resettlement, by County and Village

Panel A. County Resettlement Density, Compared to Expected Resettlement Density



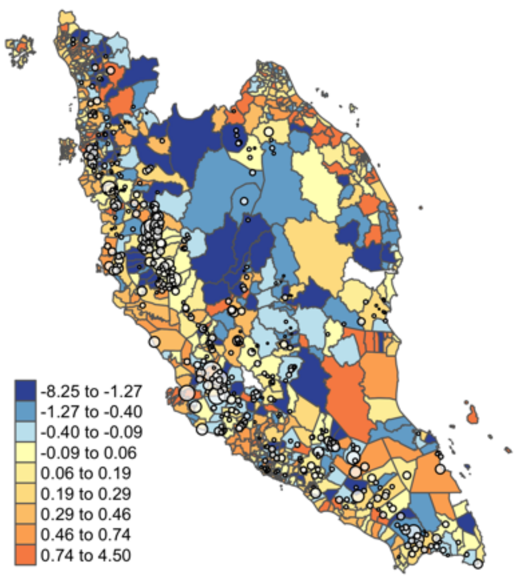
Panel B. Village Resettled Population, Compared to Expected Resettled Population



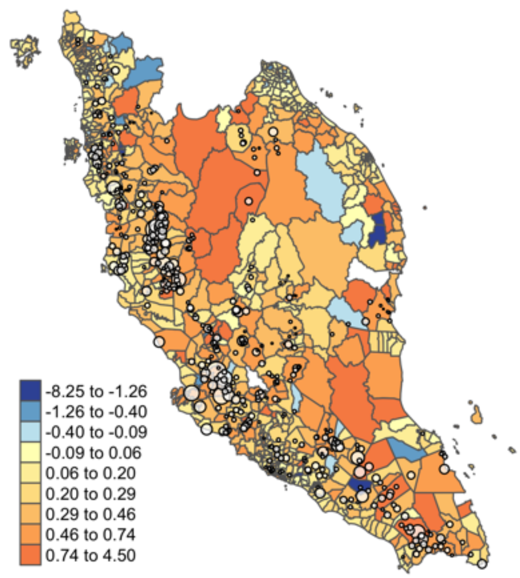
Notes: This figure compares predicted resettlement, based on the gravity model outlined in Appendix Section A.1, with the actual measured resettlement at the county and village levels. Panel A contrasts the measured county resettlement density with the expected resettlement density calculated from Equation (A.1), conditional on the actual locations of the New Villages. Panel B compares the measured resettled population of each village with the counterfactual population, also conditional on village locations. The counterfactual village resettlement is calculated using Equation (2), which models the dislocation-minimizing plan. Data from the Corry report.

Figure A.4. County Population Growth from 1947 to 1957, by Ethnic Group

Panel A. County Population Growth, Chinese

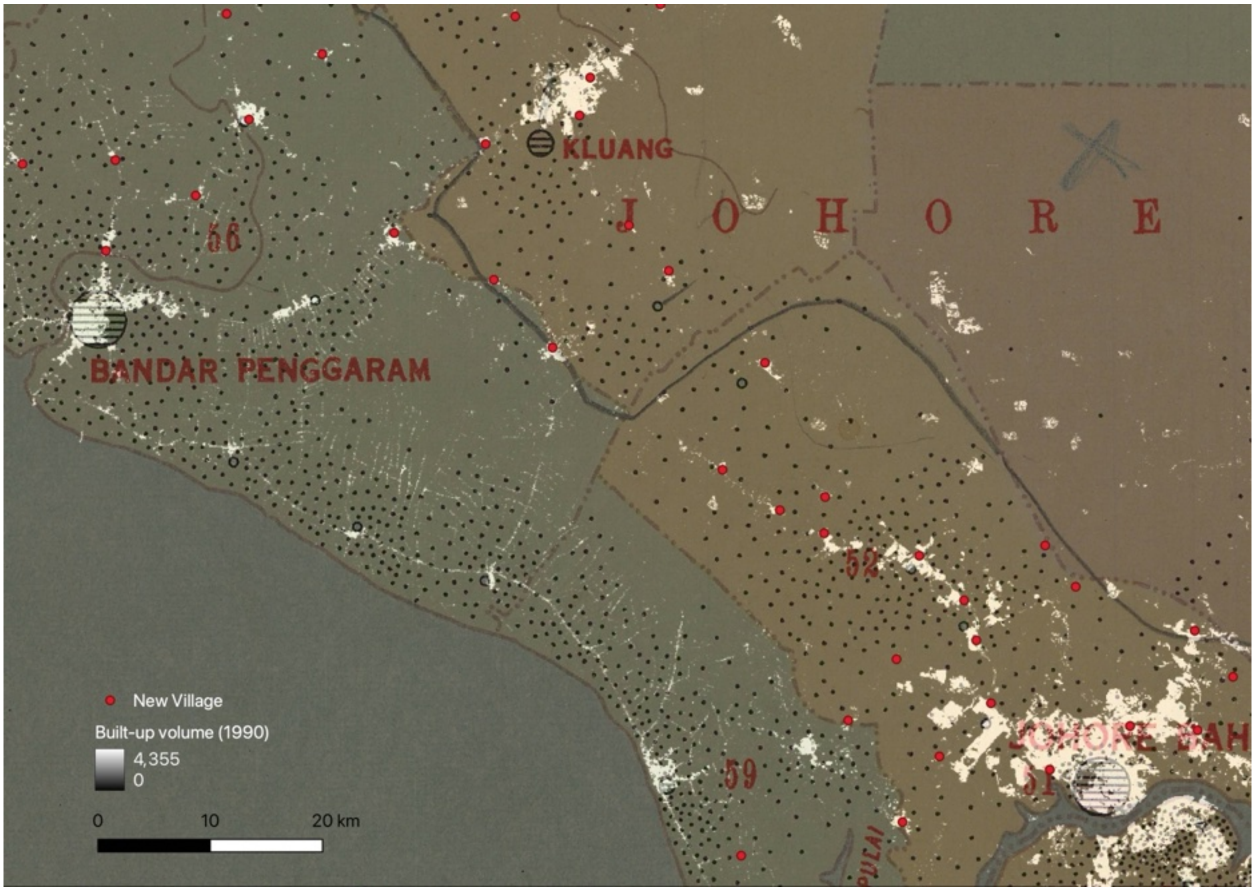


Panel B. County Population Growth, Non-Chinese



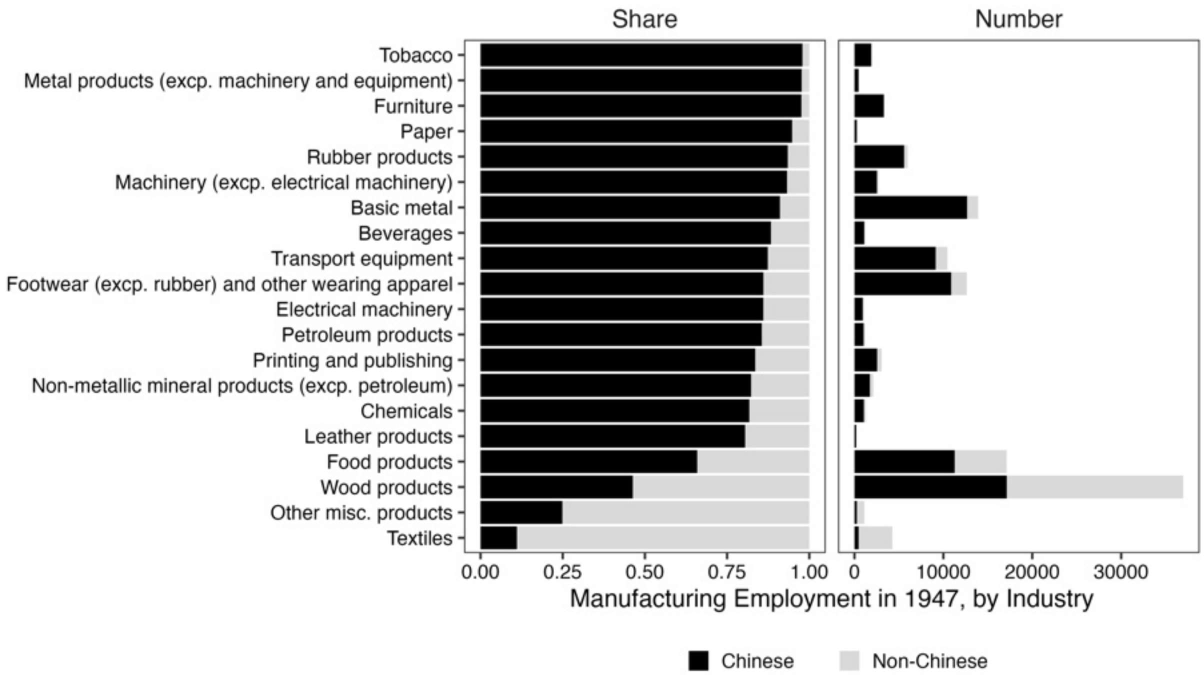
Notes: This figure shows county population growth from 1947 to 1957, by ethnic group. Panel A shows the log changes of Chinese population. Panel B shows the log changes of non-Chinese population. The white bubbles denote the New Villages, which are sized in proportion to the log resettled population in that village. Counties with missing population are shaded in white. Data from the tabulated Census of Population and the Corry report.

Figure A.5. Built-up Volume in 1990



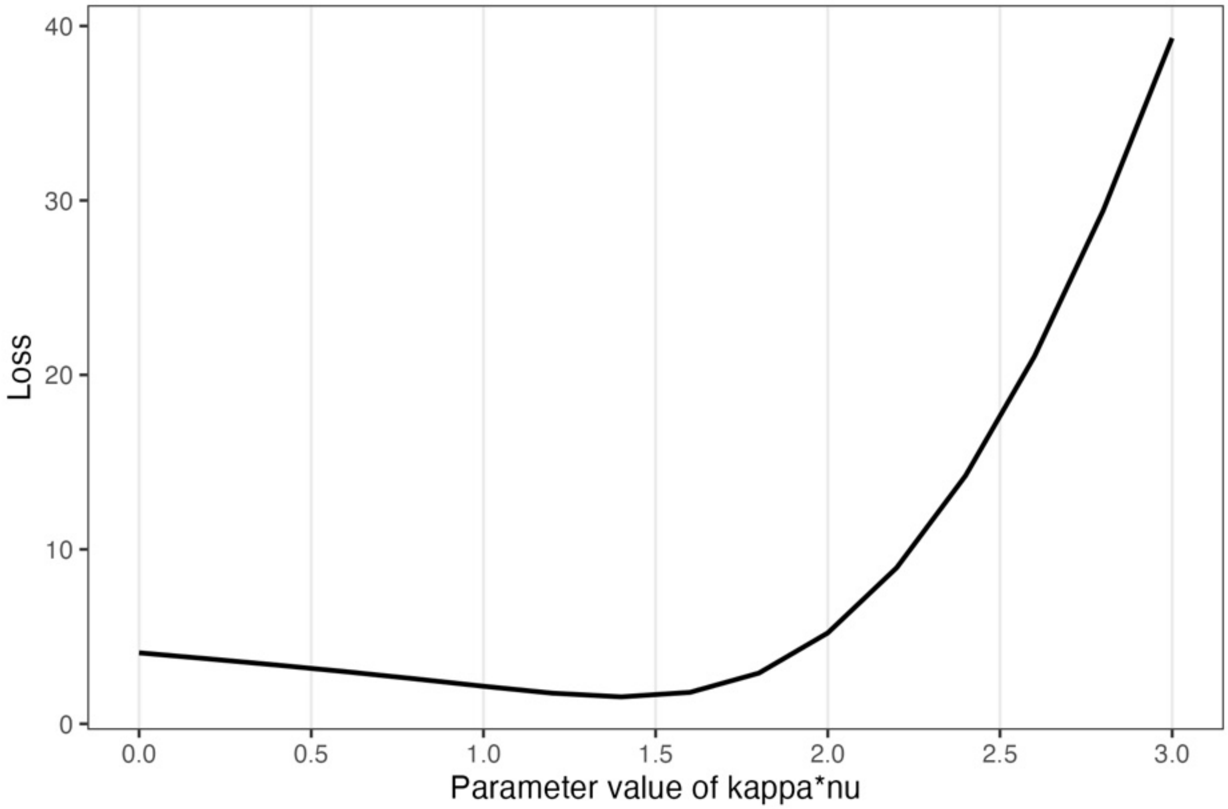
Notes: This figure shows the built-up volume in 1990 within a region of Johor, with New Villages marked by red dots. Built-up volumes (shaded in white) are calculated using 100-meter resolution surface and height data from Sentinel-2 and Landsat satellite imagery. Black dots represent population clusters from the 1947 Census. Built-up volume and New Village data are from the GHSL project and the Corry report.

Figure A.6. Chinese Manufacturing Employment Share and Number in 1947, by Industry



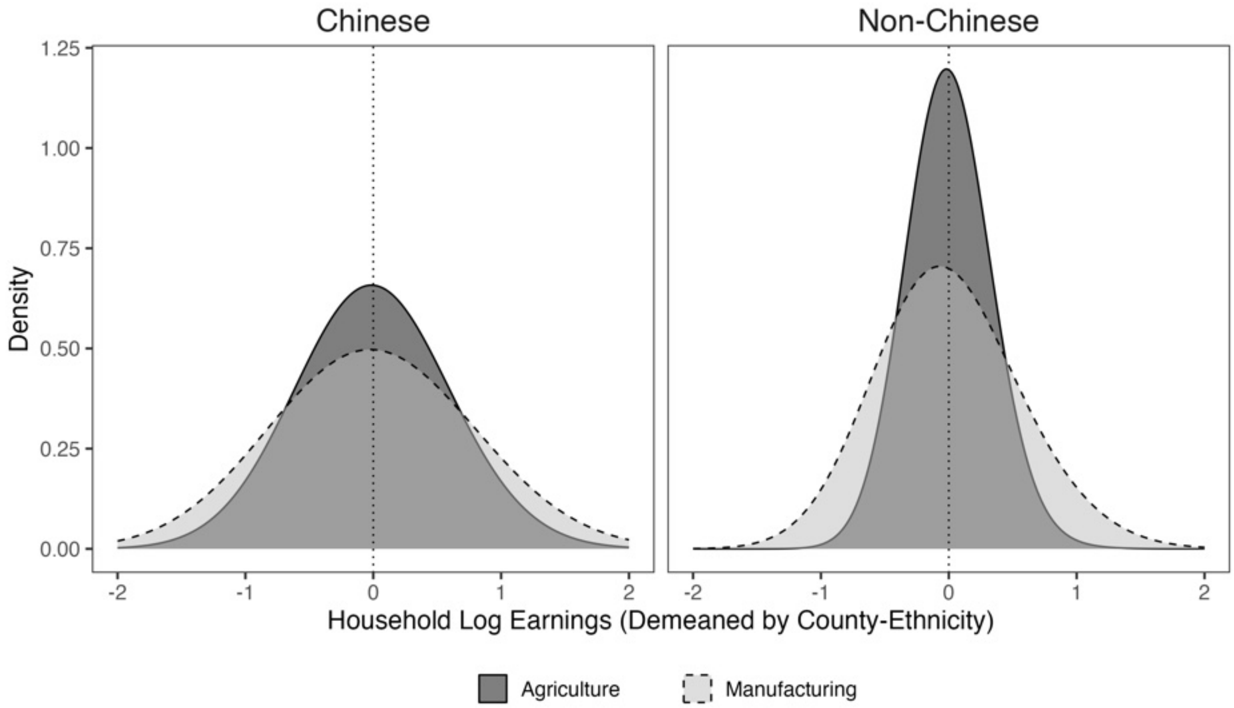
Notes: This figure shows the share (left panel) and number (right panel) of Chinese and non-Chinese employment across manufacturing industries in 1947. Dark bars denote Chinese employment, and grey bars denote non-Chinese employment. Data from the 1947 Census of Population (Del Tufo, 1947).

Figure A.7. Convexity of the Loss Function in Estimating Migration Costs



Notes: This figure shows convexity of the loss function for estimating migration cost. The y-axis plots the loss from Equation (A-5), which is a function of observed bilateral migration flows and parameter value $\tilde{\kappa}$, shown in the x-axis. Data from the tabulated Census of Population in 1980.

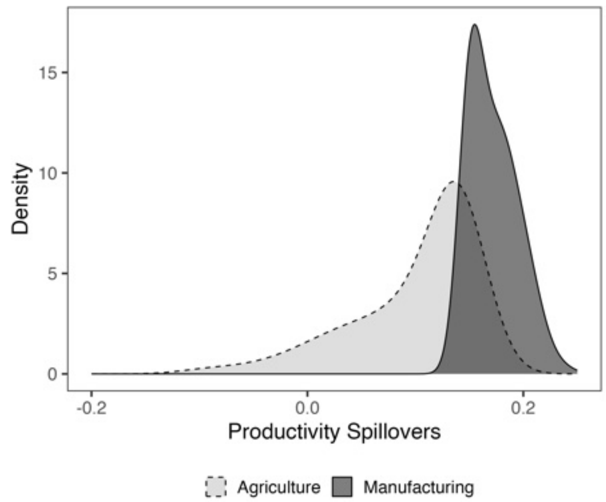
Figure A.8. Distribution of Demeaned Household Log Earnings, relative to County-Ethnicity Average



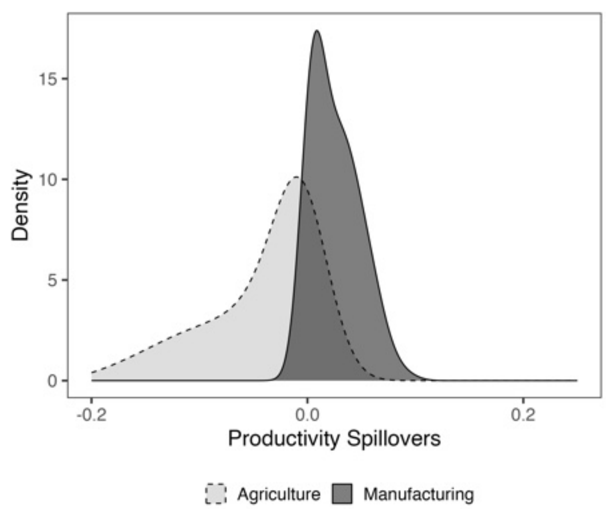
Notes: This figure shows the distribution of household log earnings demeaned by county-ethnicity averages for Chinese and non-Chinese households. The distribution are plotted separately for Agriculture (solid line) and Manufacturing (dashed line).

Figure A.9. Distribution of Marginal Productivity Spillovers, by Ethnic Group

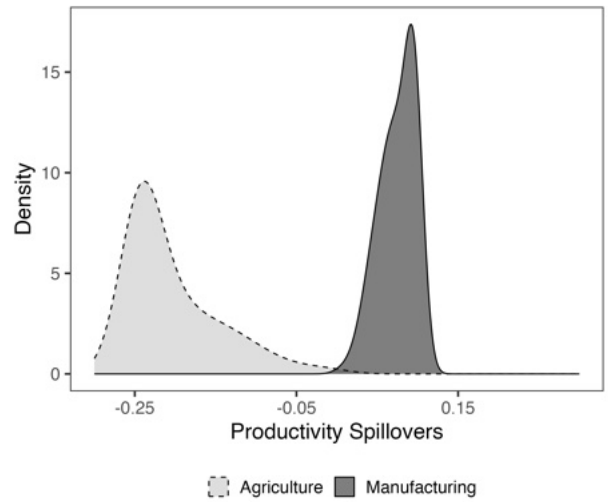
Panel A. Chinese-to-Chinese Spillover



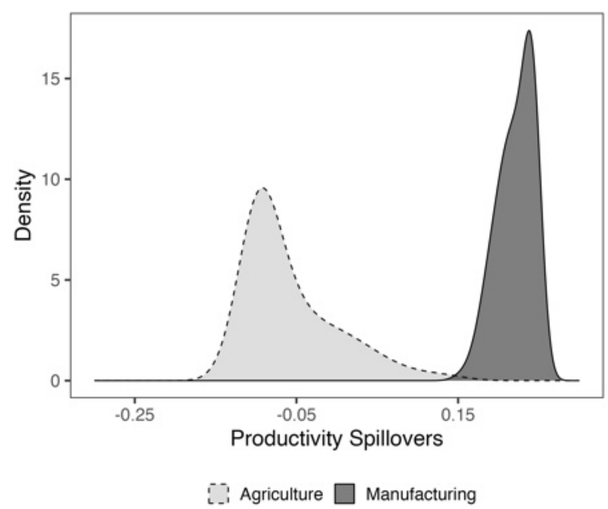
Panel B. Chinese-to-Malays Spillover



Panel C. Malays-to-Chinese Spillover



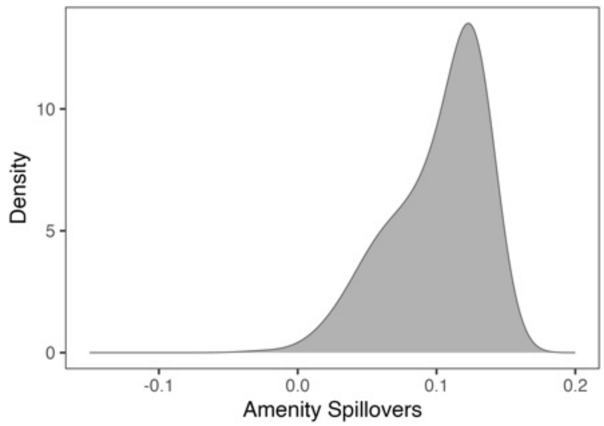
Panel D. Malays-to-Malays Spillover



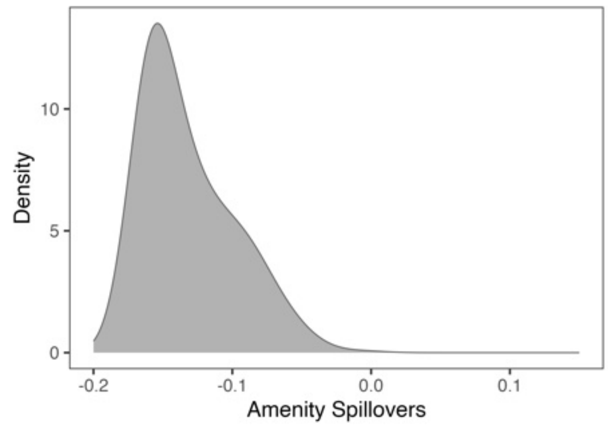
Notes: This figure shows the distribution of marginal productivity spillovers across counties, by pairwise combination of Chinese and Malays (non-Chinese). Panel A shows the elasticity of Chinese productivity with respect to local Chinese population; that is, the percent changes in Chinese productivity resulting from a 1 percent increase in the local Chinese population. Panel B shows the elasticity of Malays' productivity with respect to local Chinese population. Panel C shows the elasticity of Chinese productivity with respect to local Malays' population. Panel D shows the elasticity of Malays with respect to local Malays population. These elasticities are calculated from Equations (9) and (10), holding fixed occupational shares.

Figure A.10. Distribution of Marginal Amenity Spillovers, by Ethnic Group

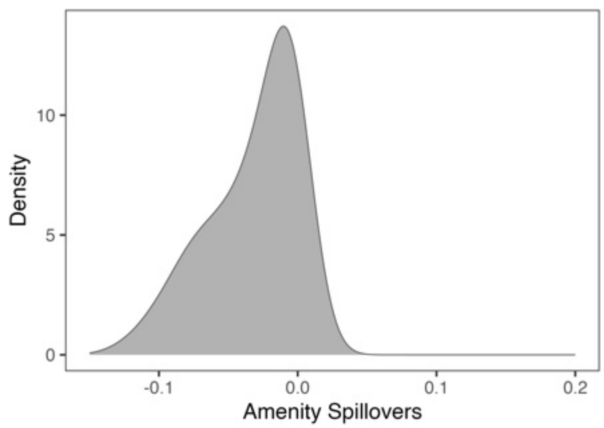
Panel A. Chinese-to-Chinese Spillover



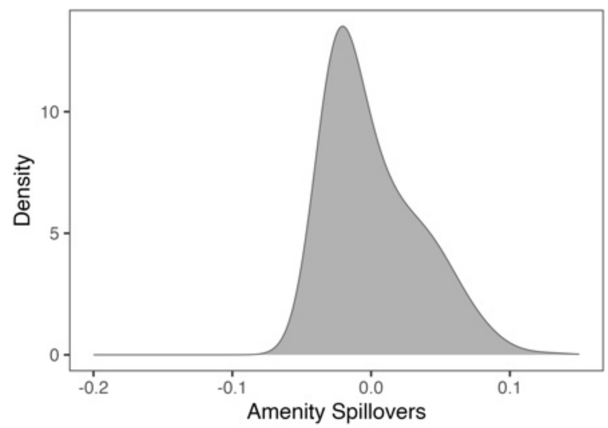
Panel C. Malays-to-Chinese Spillover



Panel B. Chinese-to-Malays Spillover



Panel D. Malays-to-Malays Spillover



Notes: This figure shows the distribution of marginal amenity spillovers across counties, by pairwise combination of Chinese and Malays (non-Chinese). Panel A shows the elasticity of Chinese utility with respect to local Chinese population; that is, the percent changes in Chinese utility resulting from a 1 percent increase in the local Chinese population. Panel B shows the elasticity of Malays' utility with respect to local Chinese population. Panel C shows the elasticity of Chinese utility with respect to local Malays' population. Panel D shows the elasticity of Malays with respect to local Malays' population. These elasticities are calculated similarly as in Equations (9) and (10) for the marginal productivity spillovers, holding fixed migration shares.

F Appendix Tables

Table A.1. Population in British Malaya from 1911 to 1957, by Ethnic Group

	Chinese		Malays		Indians and Others	
Year	Number (1)	Percent (2)	Number (3)	Percent (4)	Number (5)	Percent (6)
1911	692,228	30%	1,367,245	59%	239,169	12%
1921	855,863	29%	1,568,588	54%	439,172	17%
1931	1,284,094	34%	1,863,723	49%	572,205	17%
1947	1,882,700	39%	2,395,686	49%	529,594	12%
1957	2,328,480	37%	3,126,773	50%	695,923	13%

Notes: This table shows the population and share by ethnic group in British Malaya from 1911 to 1957. Columns 1 and 2 report the number of Chinese and its share in total population of a given year. Columns 3 and 4 report the same figures for Malays. Columns 5 and 6 report the same figures for Indians and other ethnic groups. Data from the Census of Population 1911–1957 (Vlieland, 1931; Del Tufo, 1947; Purcell, 1947; Fell, 1960).

Table A.2. Predicting Log Household Income in 1988

	Log Household Earning, 1988	
	(1)	(2)
Vehicle	0.567	0.608
	(0.288)	(0.319)
Motorcycle	0.002	0.098
	(0.157)	(0.159)
Bicycle	−0.113	−0.055
	(0.154)	(0.159)
Phone	1.006	0.907
	(0.400)	(0.398)
Refrigerator	0.424	0.253
	(0.176)	(0.179)
Television	0.182	0.117
	(0.142)	(0.144)
Household Size	0.050	0.054
	(0.010)	(0.010)
R^2	0.28	0.34
Fixed Effects	State	District
# Households	1,413	1,413

Notes: This table shows a linear model predicting (log) household income based on asset ownership and household size. The independent variables include household size and indicators for ownership of various household assets—vehicle, motorcycle, bicycle, phone, refrigerator, and television—as well as pairwise interactions between these asset indicators (not shown here due to space constraints). Column 1 includes state-district fixed effects. Column 2 includes state fixed effects. Data from the Second Malaysian Family Life Survey (1988–1989). Standard errors robust to heteroskedasticity reported in parentheses.

Table A.3. Build-up Volumes, by County Resettlement Density

	Log Build-up Volumes, by Year:		
	1975	1990	2005
	(1)	(2)	(3)
Higher Resettlement	0.333	0.257	0.197
	(0.110)	(0.087)	(0.080)
# Counties	776	776	776

Notes: This table shows the relationship between build-up volumes from 1975 to 2005 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Columns 1-3 report the effect of resettlement density on log county build-up volumes in 1975 (column 1), 1990 (column 2), and 2005 (column 3). All regressions are estimated using OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county. Data from the Global Human Settlement Layer (GHSL) project. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.4. Migration and Fertility in 1980, by County Resettlement Density

	Chinese Individuals (1)	Non-Chinese Individuals (2)	Difference (1) − (2) (3)
Panel A. Internal Migrant			
Higher Resettlement	0.046	0.004	0.042
	(0.029)	(0.018)	(0.022)
Mean of Outcome	0.39	0.47	
# Individuals	38,390	71,234	
Panel B. Internal Migrant After 1960			
Higher Resettlement	0.050	0.012	0.039
	(0.032)	(0.019)	(0.025)
Mean of Outcome	0.30	0.40	
# Individuals	38,258	70,976	
Panel C. Number of Children Born			
Higher Resettlement	−0.018	−0.144	0.126
	(0.112)	(0.061)	(0.105)
Mean of Outcome	4.01	4.16	
# Women	12,259	24,158	
Panel D. Log Household Size			
Higher Resettlement	0.012	−0.014	0.026
	(0.013)	(0.010)	(0.018)
Mean of Outcome	1.51	1.43	
# Households	10,831	23,489	

Notes: This table shows the relationship between migration and fertility outcomes in 1980 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel shows the effect of resettlement density on a different outcome: whether a person is an internal migrant (i.e., someone who moved to the current locality from another village or town within Malaysia) in Panel A; whether a person is an internal migrant who moved into the current locality within the last 20 years (or after 1960) in Panel B; the number of children born in Panel C; and log household size in Panel D. Column 1 reports the estimates for Chinese individuals, column 2 reports the estimates for non-Chinese individuals, and column 3 reports the difference between the estimates in columns 1 and 2. All regressions are estimated using OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the individual. The sample is restricted to individuals above age 20. Data from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.5. Secondary and Tertiary Employment in 1980–1991, by County Resettlement Density

	Secondary Industries (1)	Tertiary Industries (2)	Difference (2) − (1) (3)
Panel A. Total Employment			
Higher Resettlement	0.298	0.276	−0.021
	(0.124)	(0.137)	(0.063)
# County-Years	1,554	1,554	
Panel B. Chinese Employment			
Higher Resettlement	0.350	0.341	−0.009
	(0.160)	(0.196)	(0.061)
# County-Years	1,400	1,476	
Panel C. Non-Chinese Employment			
Higher Resettlement	0.233	0.234	0.001
	(0.106)	(0.116)	(0.073)
# County-Years	1,400	1,476	

Notes: This table shows the relationship between sectoral employment in 1980–1991 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A shows the effect of resettlement on total employment in the secondary sector (column 1), the tertiary sector (column 2), and the difference between the two (column 3). Panels B and C show the effects on Chinese employment and non-Chinese employment, respectively. The secondary sector is comprised of manufacturing; utility; and construction. The tertiary sector is comprised of wholesale and retail trade; transport and communication; and finance, business, and other services. All regressions are estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator and include state-year fixed effects and the main controls interacted with year: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county-year. Data from the Census of Population in 1980 and 1991. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.6. Educational Attainment in 1980, by Age Cohorts and County Resettlement Density

	Chinese Individuals, by Age Cohort:			Non-Chinese Individuals, by Age Cohort:		
	20–35	36–50	>50	20–35	36–50	>50
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Years of Schooling						
Higher Resettlement	0.415	0.314	0.072	0.117	0.096	−0.102
	(0.230)	(0.187)	(0.135)	(0.115)	(0.130)	(0.097)
Panel B. Primary Education						
Higher Resettlement	0.030	0.028	0.007	0.016	0.024	−0.004
	(0.014)	(0.017)	(0.016)	(0.011)	(0.014)	(0.012)
Panel C. Secondary Education						
Higher Resettlement	0.044	0.036	0.000	0.012	0.006	0.000
	(0.027)	(0.018)	(0.007)	(0.013)	(0.011)	(0.007)
# Individuals	15,597	8,843	7,067	30,087	15,056	12,202

Notes: This table shows the relationship between educational attainment and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel shows the effect of resettlement density on a different outcome of education: years of schooling (Panel A); completion of primary education (Panel B); and completion of secondary education (Panel C). Columns 1 to 3 report estimates for Chinese households for cohort aged 20–35 (column 1); 36–50 (column 2); and 36–50 (column 3). Columns 4 to 6 report the corresponding estimates for non-Chinese households. All regressions are estimated by OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the individual. Data from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.7. Household Asset Ownership, by County Resettlement Density

	Chinese Households (1)	Non-Chinese Households (2)	Difference (1) − (2) (3)
Panel A. Owned the House			
Higher Resettlement	0.045	0.022	0.022
	(0.015)	(0.017)	(0.018)
Panel B. Have Vehicle			
Higher Resettlement	0.058	0.020	0.037
	(0.028)	(0.012)	(0.022)
Panel C. Have Fridge			
Higher Resettlement	0.037	0.032	0.004
	(0.028)	(0.020)	(0.024)
Panel D. Have TV			
Higher Resettlement	0.035	0.003	0.032
	(0.013)	(0.018)	(0.015)
Panel E. Have Phone			
Higher Resettlement	0.034	0.013	0.021
	(0.022)	(0.008)	(0.016)
# Households	11,604	25,520	

Notes: This table shows the relationship between household asset ownership and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel shows the effect of resettlement density on a different indicator of asset ownership: the occupied house (Panel A); any motor car or van (Panel B); any refrigerator (Panel C); any black or color TV (Panel D); any phone (Panel E). Column 1 reports the estimates for Chinese households, column 2 reports the estimates for non-Chinese households, and column 3 reports the difference between the estimates in columns 1 and 2. All regressions are estimated by OLS and include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the household. Data from the 2% individual-level Census of Population microdata in 1980 and Second Malaysian Family Life Survey (1988–1989). Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.8. Characteristics of the Resettled Chinese

	Employed Agriculture, First Job (1)	Completed Primary Education (2)	Completed Secondary Education (3)	Acres of Land Owned (4)
Panel A. Comparison with Other Chinese				
Resettled Chinese	0.149	−0.125	−0.151	−105.390
	(0.069)	(0.063)	(0.023)	(78.932)
# Observations	896	989	989	395
Panel B. Comparison with Non-Chinese				
Resettled Chinese	−0.155	−0.049	−0.080	−46.747
	(0.062)	(0.053)	(0.057)	(16.517)
# Observations	3,153	3,627	3,627	1,299

Notes: This table shows the differences in characteristics between resettled Chinese and other residents living in the same state in 1988. The resettled Chinese were identified through migration history data from the Second Malaysian Family Life Survey, as detailed in the text. Panel A compares the resettled Chinese with other Chinese residents in the same state, and Panel B compares them with non-Chinese residents in the same state. Column 1 reports the estimated difference in the probability of being employed in agriculture for the first job; column 2 reports the estimated difference in primary education completion; column 3 reports the estimated difference in secondary education completion; and column 4 reports the estimated difference in the amount of land owned. All regressions include state-by-gender fixed effects and control for the individual's age and age squared. The unit of observation is the individual for columns 1–3 and the household for column 4. Data from the Second Malaysian Family Life Survey (1988–1989). Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.9. School Supply in 2022, by County Resettlement Density

	Chinese Schools (1)	Non-Chinese Schools (2)
Panel A. Elasticity of Schools with Respect to Population		
Higher Resettlement	−0.158	0.045
	(0.087)	(0.030)
# Counties	777	777
Panel B. Negative Log Distance to Schools		
Higher Resettlement	−0.036	0.051
	(0.032)	(0.025)
# Counties	777	777
Panel C. Teacher-to-Student Ratio		
Higher Resettlement	−0.032	−0.002
	(0.018)	(0.003)
# Counties	408	754

Notes: This table shows the relationship between measures of school supply in 2022 and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel shows the effect of resettlement density on a different measure of school access: elasticity of the number of schools with respect to ethnic population (Panel A); average negative log distance to schools (Panel B); and average teacher-to-student ratio (Panel C). Column 1 reports results for Chinese schools, and column 2 reports results for non-Chinese schools. Panel A is estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator, and Panels B and C are estimated using OLS. All regressions include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county. School data from the Ministry of Education. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.10. Household Income in 1980 by County Resettlement Density, Controlling for Household Head's Education

	Chinese Households (1)	Non-Chinese Households (2)	Difference (1) − (2) (3)
Panel A. Log Earnings			
Higher Resettlement	0.088	0.038	0.050
	(0.041)	(0.029)	(0.033)
# Households	10,622	22,706	
Panel B. Log Earnings, Primary Sector			
Higher Resettlement	0.066	−0.005	0.072
	(0.035)	(0.041)	(0.045)
# Households	1,660	8,066	
Panel C. Log Earnings, Non-Primary Sector			
Higher Resettlement	0.095	0.047	0.048
	(0.043)	(0.028)	(0.029)
# Households	8,962	14,640	

Notes: This table shows the relationship between household income and county resettlement density. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A, columns 1–2 show the effect of resettlement density on log household earnings predicted from asset ownership for Chinese households (column 1) and non-Chinese households (column 2), respectively. Column 3 reports the differences between the estimates in columns 1 and 2. Panel B restricts the sample to households whose head is employed in the primary sector, comprised of agriculture and mining. Panel C restricts the sample to households whose head is employed outside the primary sector. All regressions are estimated by OLS and include state fixed effects and the main controls—the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population density of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944—as well as the household head's years of schooling. The unit of observation is the household. Data from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.11. Robustness to Specifications of Counterfactual Resettlement Density

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total Population	Share of Chinese	Primary Sector	Non-Primary Sector	Chinese Households	Non-Chinese Households
	(1)	(2)	(3)	(4)	(5)	(6)
1. Baseline	0.108	0.050	0.108	0.286	0.111	0.037
	(0.062)	(0.011)	(0.037)	(0.129)	(0.052)	(0.031)
2. Log resettlement density	0.099	0.043	0.117	0.314	0.127	0.038
	(0.057)	(0.012)	(0.042)	(0.157)	(0.058)	(0.033)
3. Prioritize roads over river up to 10 km	0.091	0.053	0.123	0.292	0.108	0.031
	(0.066)	(0.011)	(0.039)	(0.144)	(0.053)	(0.030)
4. Minimum distance of 1 km between villages	0.091	0.054	0.123	0.292	0.109	0.031
	(0.066)	(0.011)	(0.039)	(0.141)	(0.053)	(0.030)
5. Squatters within 2.5 km of forest	0.105	0.053	0.121	0.284	0.107	0.033
	(0.062)	(0.011)	(0.039)	(0.132)	(0.051)	(0.032)
6. Squatters within 10 km of forest	0.098	0.054	0.125	0.292	0.104	0.028
	(0.064)	(0.012)	(0.039)	(0.142)	(0.054)	(0.030)
7. Lower resettlement cost elasticity ($\psi = 0.5$)	0.094	0.053	0.119	0.286	0.108	0.032
	(0.066)	(0.011)	(0.039)	(0.141)	(0.053)	(0.030)
8. Higher resettlement cost elasticity ($\psi = 0.8$)	0.093	0.054	0.123	0.286	0.106	0.029
	(0.065)	(0.011)	(0.039)	(0.144)	(0.053)	(0.030)

Notes: This table shows the robustness to different specifications of counterfactual resettlement density for the relationship between county resettlement density and the main outcome variables—log total population in 1980 (column 1), Chinese population share in 1980 (column 2), total employment in the primary sector in 1980–1991 (column 3), total employment in the non-primary sector in 1980–1991 (column 4), log earnings for Chinese households in 1980 (column 5), and log earnings for non-Chinese households in 1980 (column 6). The first row reports the baseline specification, including state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county for columns 1–4 and the households for columns 5–6. Rows 2–8 additionally control for a variant specification of the expected resettlement density. Data from the tabulated population Census in 1980 and 1991, as well as the 2% sample of microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.12. Robustness to Controls of Productive Location Fundamentals

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total Population (1)	Share of Chinese (2)	Primary Sector (3)	Non-Primary Sector (4)	Chinese Households (5)	Non-Chinese Households (6)
1. Baseline	0.108	0.050	0.108	0.286	0.111	0.037
	(0.062)	(0.011)	(0.037)	(0.129)	(0.052)	(0.031)
2. Neighboring roads	0.109	0.050	0.117	0.293	0.118	0.048
	(0.061)	(0.011)	(0.037)	(0.119)	(0.049)	(0.031)
3. Neighboring population	0.114	0.051	0.109	0.292	0.111	0.041
	(0.065)	(0.012)	(0.036)	(0.130)	(0.051)	(0.033)
4. Ruggedness	0.104	0.052	0.094	0.292	0.106	0.037
	(0.061)	(0.011)	(0.036)	(0.127)	(0.050)	(0.031)
5. Rice and coconut suitability	0.111	0.051	0.110	0.259	0.103	0.044
	(0.060)	(0.011)	(0.034)	(0.135)	(0.058)	(0.032)
6. Distance to prewar industrial sites	0.108	0.050	0.113	0.273	0.102	0.033
	(0.062)	(0.011)	(0.036)	(0.131)	(0.049)	(0.031)
7. Distance to major cities	0.108	0.050	0.108	0.273	0.094	0.041
	(0.062)	(0.011)	(0.037)	(0.128)	(0.051)	(0.032)
8. All above (rows 2–7)	0.119	0.053	0.105	0.275	0.109	0.057
	(0.061)	(0.011)	(0.029)	(0.121)	(0.051)	(0.036)

Notes: This table shows the robustness to including additional controls for location productivity for the relationship between county resettlement density and the main outcome variables—log total population in 1980 (column 1), Chinese population share in 1980 (column 2), total employment in the primary sector in 1980–1991 (column 3), total employment in the non-primary sector in 1980–1991 (column 4), log earnings for Chinese households in 1980 (column 5), and log earnings for non-Chinese households in 1980 (column 6). The first row reports the baseline specification, including state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. Rows 2–8 add additional controls to the baseline specification. The unit of observation is the county for columns 1–4 and the households for columns 5–6. Data from the tabulated population Census in 1980 and 1991, as well as the 2% sample of microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.13. Robustness to Sample

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total Population (1)	Share of Chinese (2)	Primary Sector (3)	Non-Primary Sector (4)	Chinese Households (5)	Non-Chinese Households (6)
1. Baseline	0.108	0.050	0.108	0.286	0.111	0.037
	(0.062)	(0.011)	(0.037)	(0.129)	(0.052)	(0.031)
2. Exclude 10 most densely populated towns	0.109	0.052	0.095	0.258	0.092	0.034
	(0.060)	(0.011)	(0.038)	(0.154)	(0.054)	(0.036)
3. Exclude top and bottom 1% counties by area	0.101	0.048	0.116	0.280	0.111	0.030
	(0.062)	(0.011)	(0.037)	(0.131)	(0.052)	(0.031)
4. Exclude top and bottom 1% resettled counties	0.100	0.059	0.109	0.248	0.113	0.026
	(0.059)	(0.014)	(0.035)	(0.128)	(0.058)	(0.035)
5. Only counties with sampled Chinese	0.077	0.046	0.102	0.265	0.111	0.047
	(0.059)	(0.013)	(0.037)	(0.134)	(0.052)	(0.032)
6. Only resettled counties	0.168	0.066	0.111	0.326	0.153	0.046
	(0.068)	(0.016)	(0.033)	(0.176)	(0.054)	(0.040)

Notes: This table shows the robustness to alternative choices of county sample for the relationship between county resettlement density and the main outcome variables—log total population in 1980 (column 1), Chinese population share in 1980 (column 2), total employment in the primary sector in 1980–1991 (column 3), total employment in the non-primary sector in 1980–1991 (column 4), log earnings for Chinese households in 1980 (column 5), and log earnings for non-Chinese households in 1980 (column 6). All regressions include state fixed effects and the main controls: the expected resettlement density; an indicator for any resettlement in the county; (log) county area; distance to the nearest road; road density of the county; distance to the nearest rail station; distance to coastline; Chinese population share of the county in 1947; (log) population of the county in 1947; the share of lands used for rubber cultivation in 1944; and the share of lands used for mining in 1944. The unit of observation is the county for columns 1–4 and the households for columns 5–6. Data from the tabulated population Census in 1980 and 1991, as well as the 2% sample of microdata in 1980. Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.

Table A.14. Housing Elasticity in 1989

	Log Rents (1989)	
	OLS	IV
	(1)	(2)
Panel A. Year 1980		
Log Population	0.266	0.326
	(0.054)	(0.155)
F-stat (1st Stage)		64.6
Panel B. Year 2000		
Log Population	0.267	0.270
	(0.059)	(0.122)
F-stat (1st Stage)		106.2
# Counties	103	103

Notes: This table shows the relationship between log housing rents in 1989 and log population in years 1980 (Panel A) and 2000 (Panel B). Column 1 reports the OLS estimates. Column 2 reports the IV estimates and the first-stage F statistics. The instrumental variable used is the residual resettlement density, as shown in Figure 3, Panel B. The unit of observation is the household. The sample is restricted to households reporting non-missing rent expenditure. Data from the Malaysian Family Life Survey (1988–1989). Conley standard errors with a distance cutoff of 30 kilometers are reported in parentheses.