

Coercive Growth: Forced Resettlement and Ethnicity-Based Agglomeration

Shanon Hsuan-Ming Hsu
University of Chicago

May 6, 2025

Abstract

While the clustering of people can raise productivity through social interactions, social divisions such as ethnic segregation and tension may limit these benefits. I study how such divisions shape the gains from agglomeration, leveraging an ethnic-based resettlement program in 1950s British Malaya that forcibly relocated 600,000 rural Chinese into villages across the country. I find that areas with higher resettlement remained more densely populated and had higher Chinese population shares decades later, driven by both the program and subsequent migration. Moreover, these areas were wealthier and more industrialized, with greater labor market specialization. However, the economic benefits were concentrated among the Chinese population. Other ethnic groups saw only marginal gains when employed outside agriculture. Evidence suggests that segregation and deeper cultural and linguistic barriers hindered cross-ethnic spillovers. To evaluate the aggregate effects of resettlement, I estimate a quantitative spatial model allowing agglomeration externalities to vary by sector and ethnic composition. While resettlement raised total output, the economic gains did not outweigh the welfare losses from forced relocation. *JEL codes:* J15, N15, O15, R11, R23.

I am grateful to my advisors Richard Hornbeck, Oeindrila Dube, Michael Dinerstein, and Jonathan Dingel for their guidance and support throughout the project. I thank Sarah Ling Francis for her excellent research assistance, Hugh Alexander and Alexander Poole for their help with archival research, and David Baillargeon for sharing 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 Kunievisky, Michael Pollmann, James Robinson, Jordan Rosenthal-Kay, Martin Rotemberg, and Jeanne Sorin, as well as seminar participants at the University of Chicago, the Mountain West Economic History Conference, NEUDC 2024, and UEA 2025. 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 raise productivity by reducing transaction and search costs and fostering knowledge spillovers (Marshall, 1890; Duranton and Puga, 2004). Yet, these gains from proximity hinge on the density and nature of local social interactions. Social divisions such as ethnic segregation and tension, common in many urban settings, can hinder these productive interactions. This raises a fundamental question: How do ethnic divisions shape agglomeration economies and, more broadly, patterns of economic development?

To answer this question, I study an ethnic-based resettlement program implemented during the 1950s Malayan Emergency in British Malaya, which forcibly relocated roughly 600,000 rural Chinese into villages across the country. Leveraging this variation, I examine how the concentration of one ethnic group (Chinese) affected local economic outcomes across regions and groups over the following 50 years. To interpret these effects, 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 program’s aggregate impacts and evaluate counterfactual policies for economic development in the presence of heterogeneous agglomeration forces.

This setting provides a useful context for studying how ethnic divisions shape agglomeration economies for three reasons. First, the resettlement program shifted ethnic composition alongside population size while limiting self-selection. Second, the Chinese community in British Malaya specialized in industrial and urban sectors, where agglomeration externalities typically emerge, suggesting that Chinese concentration could generate productivity gains. Third, deep cultural, religious, and linguistic divisions existed across ethnic groups in Malaysia—especially between Chinese and the majority Malay population—and ethnic segregation and tensions were pervasive. These contrasts suggest that different groups may have benefited differently from the influx of Chinese, even holding other factors constant.

To estimate the local effects of resettlement, I leverage the program’s wartime objectives to construct counterfactual resettlement and isolate quasi-random variation. During the Malayan Emergency (1948–1960), a guerrilla conflict between the British and Malayan communists, the British forcibly relocated nearly one-tenth of the population—mostly ethnic Chinese—into roughly 500 “New Villages.” The goal was to sever ties to communist insurgents, so the colonial government moved people from remote, jungle-adjacent areas, where insurgents were based, to more accessible locations. The program was implemented in two stages. First, suitable sites were selected based on security access, with most located along main roads or rivers (Figure 1). Second, rural Chinese populations were relocated to these

sites in a way that minimized dislocation from their original settlements.

Following this procedure, I construct counterfactual resettlement in two steps. First, I randomly permute New Village locations, conditional on distance to transportation networks and other key covariates the British considered, such as land use and the local Chinese population. Second, I use a gravity model to predict the number of people resettled to each site, assuming that resettlement costs increased with distance. I repeat this procedure 1,000 times and take the average predicted resettlement density across permutations to obtain each county’s *expected* resettlement density (Borusyak and Hull, 2023).¹

The estimation compares areas with similar expected resettlement density and prewar characteristics but different realized resettlement. The identifying variation comes from the precise placement of New Villages 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 resettlement based on unobserved economic fundamentals. Consistent with this, I show that geographic features and prewar economic activity—such as agricultural suitability and proximity to major cities or industrial facilities—are balanced across areas with varying residual resettlement densities.

I document that the program persistently reshaped the population distribution in Peninsular Malaysia. Counties with higher resettlement saw a sharp increase in Chinese population between 1947 and 1957, while the non-Chinese population remained stable. After mobility restrictions were lifted in 1960, these areas attracted internal migrants from all ethnic groups. By 2000, they continued to have higher population densities and a larger Chinese share. On average, resettling 1% of the 1947 population led to a 1.3% increase in population by 2000.

The influx of Chinese into resettled areas substantially transformed local economic structure and raised productivity. Census data from 1980 and 1991 show that counties with higher resettlement had greater employment and a larger share in non-agricultural sectors such as manufacturing, trade, and services. This shift reflected both a higher concentration of Chinese—who were more likely to work outside agriculture—and a smaller agricultural share among non-Chinese. Chinese households earned higher incomes in these counties, especially those employed outside agriculture. In contrast, non-Chinese saw only marginal gains in the non-agricultural sector, with effects less than half as large as those for Chinese.

These comparisons are made within ethnic groups and sectors across counties with varying resettlement densities. Thus, the results are not driven by comparative advantage or broader cultural differences across ethnic groups. The ethnic gap in income gains also appears within agriculture, suggesting that the productivity benefits of co-ethnic concentration extended

¹Dell and Olken (2020) use a similar approach to identify the effects of proximity to sugar plants established in colonial Java by specifying counterfactual plant locations.

beyond specific sectors.

I find that labor pooling contributed to ethnicity-based agglomeration economies. By the 1980s, Chinese individuals in more resettled counties had higher labor force participation, and their occupations and industries were more concentrated. Educational attainment was also higher, especially among younger cohorts who had not yet completed or begun schooling at the time of resettlement. These patterns are consistent with higher co-ethnic density improving worker-firm matching, raising the returns to formal employment, specialization, and education (Kim, 1989; Dauth et al., 2022). In contrast, non-Chinese individuals showed no significant differences in these labor market outcomes, and their educational gains were much smaller. While education accounts for part of the income gains among Chinese, the ethnic gap persists even after controlling for education.

The evidence suggests that segregation and social divisions limited gains for non-Chinese individuals. Non-Chinese households benefited more when living near Chinese communities, with no measurable effects beyond five kilometers. Gains were also larger in counties with higher initial Chinese population shares or where some non-Chinese residents spoke Chinese—settings likely to involve lower cross-ethnic frictions. These patterns point to ethnic tensions or cultural and linguistic barriers as likely obstacles to cross-ethnic spillovers.

The differential income gains from resettlement suggest that agglomeration forces vary by sector and local ethnic composition. This heterogeneity implies that place-based or industrial policies that shift population across regions or sectors can have aggregate consequences. However, the cross-sectional analysis captures only relative differences across counties and does not show the program’s overall impact.

To evaluate aggregate effects, I develop a quantitative spatial model with two ethnic groups and two sectors. The model incorporates a Roy (1951)-type framework for migration and occupation choices—the key margins of adjustment following resettlement. Extending the work of Allen and Donaldson (2022) and Peters (2022), I allow productivity spillovers to vary by sector and the ethnic composition of local employment, so that within-ethnic spillovers may differ from cross-ethnic spillovers. Local amenities also depend on ethnic composition, as in Diamond (2016). In this two-period model, individuals begin with an initial location, choose where to migrate based on heterogeneous preferences and movement costs, and then select a sector in which they are more productive.

To bring the model to the data, I use the 1957 population distribution—observed after most resettlement was completed—as the initial population and treat 1980 as the equilibrium outcome. The forced resettlement provides an exogenous shift in population for identifying agglomeration parameters. Due to migration frictions, individuals tended to move near their resettled locations, so the 1957 distribution continued to shape the 1980 equilibrium.

The relationship between the local Chinese share and the Chinese wage premium (relative to non-Chinese) within a location-sector, controlling for occupational composition, identifies the strength of within- versus cross-ethnic productivity spillovers. In addition, the relationship between local employment size and wages—adjusting for composition effects and general equilibrium forces—identifies sector-specific scale economies.

The estimates reveal sizable variation in agglomeration externalities across sectors and regions, shaped by ethnic composition. In a neoclassical model with downward-sloping labor demand, an increase in non-agricultural labor supply would lower relative wages in the sector, pushing workers into agriculture. Instead, I find higher non-agricultural wages and lower agricultural employment shares in more densely resettled counties, consistent with strong external economies of scale in non-agriculture. I estimate that a 1% increase in non-agricultural employment raises labor productivity by 0.22%. By contrast, agriculture exhibits local diminishing returns to scale due to the fixed land supply: a 1% increase in agricultural employment reduces labor productivity by 0.12%. I also estimate stronger within-ethnic productivity spillovers than cross-ethnic spillovers, with the latter being positive only in non-agriculture. Finally, the estimated amenity spillovers suggest homophily—a preference for living near others of the same ethnic background.

I use the estimated model to evaluate the aggregate economic and welfare impacts of the resettlement program. To do so, I simulate a “no resettlement” 1980 equilibrium using the 1947 population distribution as the initial condition, instead of the resettled 1957 distribution, holding all parameters and location fundamentals fixed. Assuming the 1947 distribution would have persisted to 1957 absent resettlement, I measure the program’s impact by comparing this counterfactual to the observed 1980 equilibrium. I find that resettlement increases aggregate output by 2%, with two-thirds of the gain driven by labor reallocation to more productive regions and sectors. Moving 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.

While output increases, the welfare effects are more nuanced. The average welfare gain is 4.8%, but this calculation abstracts from the costs of forced movement by treating the shift from the 1947 to the 1957 distribution as costless. To benchmark these costs, I invert the model to calculate the minimum place- and ethnicity-specific wage subsidies required to induce voluntary movement to the observed 1957 distribution. I find that the total required subsidy for the resettled population exceeds the program’s aggregate economic gains, suggesting a net welfare loss from forced relocation. These subsidies serve as a lower bound on utility loss (in monetary terms), as resettled individuals likely valued their assigned destinations less than the average resident, biasing the required compensation downward. The

results highlight an intergenerational tradeoff: while younger cohorts benefit from improved economic opportunities, the original resettled generation bears the welfare loss.

A large literature documents the productivity benefits of density and geographic concentration. Studies find that agglomeration economies arise from firm establishments (Greenstone, Hornbeck and Moretti, 2010), infrastructure development (Kline and Moretti, 2014; Heblich, Redding and Sturm, 2020), improved market access (Ahlfeldt et al., 2015), political designations (Smith and Kulka, 2024), and refugee resettlement (Peters, 2022; Ciccone and Nimczik, 2024).² While this literature emphasizes the role of density, less is known about how social composition shapes agglomeration gains. I study an ethnic-based resettlement program that exogenously shifted local ethnic composition to examine its effects on group-specific outcomes, complementing research on the distributional consequences of agglomeration across skill groups (Baum-Snow and Pavan, 2013; Baum-Snow, Freedman and Pavan, 2018).

The relationship between diversity and economic performance has long been debated. In the urban context, Marshall (1890) emphasizes gains from industrial specialization, while Jacobs (1969) highlights the role of urban variety in fostering innovation.³ A broader literature links ethnic divisions to economic outcomes through firm productivity (Hjort, 2014), social cohesion (Esteban, Mayoral and Ray, 2012; Guarnieri, 2025), trust (Alesina and La Ferrara, 2002), and public goods provision (Easterly and Levine, 1997; Alesina, Baqir and Easterly, 1999; Miguel and Gugerty, 2005).⁴ I show that in mid-20th century Malaysia, deep ethnic divisions limited cross-ethnic spillovers and muted the productivity gains from agglomeration.⁵ This highlights a distinct channel through which diversity may constrain development.

By treating ethnic composition as an endogenous amenity, I build on work showing how local demographics influence residential sorting (Diamond, 2016; Weiwu, 2024; Almagro and Domínguez-Iino, 2024). I extend this line of research by showing that ethnic composition also directly affects local productivity, contributing to a broader literature on how neighborhood exposures shape long-run economic outcomes (Chetty, Hendren and Katz, 2016; Ioannides, 2012; Chyn and Katz, 2021; Chetty et al., 2025; Chyn, Collinson and Sandler, 2025).⁶

²See Duranton and Puga (2004) and Rosenthal and Strange (2004) for reviews.

³Empirically, Henderson, Kuncoro and Turner (1995) shows that mature manufacturing industries in postwar U.S. cities benefited from specialization, with spillovers among firms within the same industry. In contrast, Glaeser et al. (1992) finds that employment growth was driven by cross-industry spillovers.

⁴Ashraf and Galor (2013) document a hump-shaped relationship between genetic diversity and development. See also Alesina and La Ferrara (2005) for a review.

⁵This finding aligns with Ananat, Fu and Ross (2013); Ananat, Shihe and Ross (2018), who show that segregation and limited interracial interactions reduce agglomeration benefits and widen racial wage gaps in U.S. cities.

⁶Related work on ethnic enclaves highlights how social capital in migrant communities affects labor market

The literature on forced displacement documents mixed economic consequences. Some studies find improved outcomes for the resettled (Becker et al., 2020; Nakamura, Sigurdsson and Steinsson, 2022; Sarvimäki, Uusitalo and Jäntti, 2022), while others report negative effects when destinations have lower income levels (Carrillo, Charris and Iglesias, 2023), are poorly matched to migrants’ skills (Bazzi et al., 2016), or group together populations without a shared governance history (Dippel, 2014).⁷ I find that much of the aggregate output gain from Chinese resettlement reflects relocation to urban areas better suited for industrial activity and aligned with Chinese comparative advantage. But because pre-resettlement location choices reflected not only economic incentives but also preferences, the program imposed welfare losses despite economic gains. I quantify these effects using a spatial general equilibrium model that accounts for spillovers in both destinations and origins. This analysis contributes to the literature on the aggregate effects of place-based policies (Glaeser and Gottlieb, 2008; Fajgelbaum and Gaubert, 2020), showing how cross-ethnic frictions introduce a new source of spatial heterogeneity in agglomeration elasticities.⁸

Finally, this paper contributes to the literature on postwar development in East and Southeast Asia (Haggard, 1990; Amsden, 1992; Mundial, 1993; Wade, 2004; Lane, 2025). After World War II, the East Asian Tigers and “Look East” followers like Malaysia adopted industrial policies aimed at structural transformation. The strong non-agricultural externalities I estimate suggest that such policies can generate self-reinforcing productivity gains. The co-ethnic agglomeration effects arising from Chinese resettlement may also reflect collective action within Chinese communities, which Dell, Lane and Querubin (2018) link to historical exposure to centralized state institutions.

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

II Historical Background

The agglomeration effects of Chinese clustering depend on the economic activities in which the ethnic Chinese specialized before resettlement, while local spillovers to other groups

outcomes (Edin, Fredriksson and Åslund, 2003; Munshi, 2003; Damm, 2009; Beaman, 2011), human capital accumulation (Borjas, 1992), and welfare participation (Bertrand, Luttmer and Mullainathan, 2000; Aizer and Currie, 2004).

⁷Other studies examine political outcomes, including effects on social capital (Abel, 2019) state- and nation-building (Bazzi et al., 2019; Carlitz et al., 2024), and voting behavior (Kok et al., 2025). In the same Malaysian setting, Kok et al. (2025) find that proximity to New Villages reduced support for the pro-Malay nationalist party. For a broader review, see Becker (2022).

⁸Relatedly, Rossi-Hansberg, Sarte and Schwartzman (2019) find that productivity spillovers are stronger within “cognitive non-routine” occupations, motivating spatial redistribution of specific worker types.

depend on the degree of cross-ethnic interaction. In this section, I first describe the economic roles of the Chinese population in British Malaya and their relationship with the majority Malay population. I then discuss the colonial government’s objectives 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 others (Appendix Table A.2). Chinese immigrants had long concentrated in the colony’s industrial and urban sectors, initially through employment in tin mining and rubber plantations—the key export industries of the colonial economy. Over time, many moved into manufacturing and commerce. In contrast, most Malays engaged in subsistence agriculture, particularly paddy rice and coconut cultivation (Ginsburg, 1958; Lee and Tan, 2000). By 1947, 80% of Malays worked in agriculture, compared to 60% of Chinese (Appendix Figure A.1). Industrial jobs were concentrated in towns, contributing to higher urbanization among the Chinese: 40% lived in urban areas by 1947, compared to just 10% of Malays (Del Tufo, 1947).

Besides these distinct economic roles, cultural, religious, and linguistic differences further deepened social divisions. The majority of Malays were Muslim, while the Chinese were largely non-Muslim and followed cultural practices that conflicted with Islamic norms, such as pork consumption. Inter-marriage was rare, and language barriers persisted: in 1947, fewer than 1% of Chinese spoke Malay, and by 1980, only 25% were fluent (Del Tufo, 1947; Khoo, 1983). As a result, ethnic groups remained segregated even in cities, with limited interaction in schools or workplaces (Hirschman, 1986).

II.B The Briggs Plan: Emergency Resettlement

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” due to their lack of legal land titles—had been pushed from towns to the countryside during the Japanese occupation (1942–1945), which disrupted industrial employment.⁹

These Chinese squatters became a security concern during the Malayan Emergency (1948–1960), a guerrilla conflict between British forces and communist insurgents. Many squatters supported the insurgents by supplying food and information, and some participated in the communists’ non-military support network.¹⁰ Their proximity to jungle areas,

⁹The Squatter Committee Report, The National Archives in the UK (hereafter, “TNA”), CO 717/178. Many urban Chinese fled to avoid conflict or forced labor (Humphrey, 1971, pp. 39, 47; Loh, 1988, pp. 57–60). Early downturns in tin and rubber production during World War I and the Great Depression may have also contributed to this rural shift (Humphrey, 1971, p. 43; Loh, 1988, pp. 23, 27–29).

¹⁰The Malayan communist Party (MCP) had strong ties to the Malayan People’s Anti-Japanese Army

where the insurgents were based, made British surveillance and control especially difficult (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.¹¹ The plan aimed to deny insurgents access to supplies and intelligence, while forcing them into direct confrontation with British forces (Briggs, 1951, p. 7). Although the plan prioritized military objectives, it also sought to minimize dislocation and economic disruption (Humphrey, 1971, pp. 181–182).

The state government implemented the program rapidly, as speed was critical to prevent the insurgents from adapting (Sunderland, 1964, p. 161; Humphrey, 1971, p. 106). 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 notices.¹² Squatters were typically given fewer than 14 days’ notice; in areas where resistance or escape was likely, relocations occurred without warning.¹³ The military provided transport 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 built from scratch; the rest expanded upon existing settlements (Sandhu, 1964, p. 163; Humphrey, 1971, p. 98). Most villages followed a standard layout and included basic amenities such as a school, police station, and community center. Access and mobility were tightly controlled, with dusk-to-dawn curfews and police checkpoints 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 was overwhelmingly Chinese (86%), with smaller shares of Malays (9%) and others (5%) (Sandhu, 1964, p. 159).¹⁴ Households were typically allocated a 1/6-acre house lot and, if previously farmers, an additional 2 acres of agricultural land (Sandhu, 1964). Many of these villages continued to grow after the Emergency and still exist today.¹⁵

(MPAJA), which had previously resisted Japanese occupation. The communists’ non-military network was also referred to as *Min Yuen*.

¹¹TNA: CO 717/178.

¹²TNA: CO 1022/29.

¹³In cases of suspected communist ties, relocations were conducted at dawn without prior notice (Humphrey, 1971, p. 102).

¹⁴About half of those resettled were squatters; the rest were legitimate landholders.

¹⁵In 1972, the Ministry of Housing and Local Government of Malaysia reported 465 remaining New Villages with a combined population of one million. By 2005, about 450 remained (Lee and Tan, 2000, p. 262).

II.C Determinants of Resettlement Density

Understanding the determinants of resettlement density is crucial for evaluating its impacts. The British outlined several criteria for the program (Humphrey, 1971, pp. 95–97), which was implemented in two stages: first, selecting village sites; second, relocating squatters while minimizing dislocation. This section discusses how these criteria, along with ad hoc factors, shaped variation in resettlement density.

Security and defensibility. Security was the primary objective. Sites were ideally located near major roads or navigable waterways to facilitate police access (Sandhu, 1964, p. 164; Dhu Renick, 1965, p. 9; Humphrey, 1971, p. 96). Most New Villages were located along main transportation routes (Figure 1). For defensibility, sites were preferably on elevated terrain and away from observation points, though elevation does not correlate with resettlement density in the data.

Land availability. Land acquisition costs also shaped siting decisions. Financial constraints led the British to prioritize state-owned land or low-value land (Humphrey, 1971, p. 367). Many villages were established on public rubber estates. Since land types likely varied in latent productivity, I control for land use in the analysis.

Economic sustainability. Villages were ideally located on well-drained land with water access and agricultural potential (Dhu Renick, 1965, p. 9; Humphrey, 1971, p. 96). In practice, however, a shortage of trained personnel led to poor siting decisions: many villages were prone to flooding or unsuitable for farming (Humphrey, 1971, p. 107).¹⁶ An official report in 1954 found that 31% of sites were unlikely to remain viable after the Emergency (Corry, 1954), suggesting that economic considerations were not central to siting decisions. In line with this, I show that location characteristics related to productivity or amenities—such as ruggedness, rice suitability, and access to public goods—do not correlate with resettlement density once I control for proximity to transportation.

Proximity to squatter settlements. To reduce disruption, village sites were ideally located near original squatter communities (Humphrey, 1971, pp. 96–97). Most relocations occurred within 20 miles of the original settlements (Sandhu, 1964, p. 160). The spatial distribution of squatters thus played a key role in shaping local resettlement density. Because these locations were self-selected before resettlement, I control for pre-resettlement population distribution in the analysis.

¹⁶Examples 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).

Deviations from minimizing dislocation. Despite the stated goal of minimizing dislocation, several factors led to longer relocations. Without reliable field surveys, the British had limited knowledge of squatter distributions and continued to discover new settlements 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....”¹⁷ Given limited site capacity near some settlements, these newly discovered squatters were likely relocated to more distant areas. As insurgents shifted operations, some areas initially deemed secure later became unsafe. These dynamics led certain sites to house more people than expected based on the surrounding Chinese population.

Summary. The British emphasis on speed and security, combined with limited information and planning capacity, generated plausibly exogenous variation in population resettlement. Sites were selected from many similarly suitable locations along major roads, while poor planning amid shifting insurgency risks led to longer relocations. I later use this variation to construct a population shifter and examine the local effects of resettlement.

III Data

This section describes the data on population resettlement and key economic outcomes. Additional details are provided in Appendix A.

III.A Emergency Resettlement

I measure the resettled population using a 1954 official report to the High Commissioner (Corry, 1954)—henceforth, the “Corry report”—compiled shortly after most resettlement had been completed.¹⁸ It documents 439 resettlement sites (“New Villages”) with their names, populations, forms of local government, and qualitative descriptions. I use village locations from Baillargeon (2021), who geolocated 430 of these villages based on village names and states, accounting for roughly 540,000 people—94% of the estimated total resettled population by the end of the Emergency. I cross-validate population figures using a 1958 Malayan Christian Council survey (Council, 1958), which shows consistent population counts for villages documented in both sources.¹⁹

To specify counterfactual resettlement, I collected and digitized several historical maps: a 1942 road and railway map, a 1943 land utilization map, 1945 topographical maps, a 1947

¹⁷TNA: 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).

¹⁸The report had four aims: (i) assess agricultural land sufficiency and economic conditions; (ii) evaluate village sustainability and potential out-migration after the Emergency; (iii) examine land ownership among villagers; and (iv) estimate the number of remaining Chinese requiring resettlement.

¹⁹The 1958 survey included around 100 additional, smaller villages built after 1954.

population census map, and a 1957 “Black Areas” map showing regions under Emergency regulations due to communist activity.²⁰ I measure initial squatter settlements by overlaying the land-use, population census, and Black Areas maps. I define a cluster of Chinese population in 1947 as a squatter settlement if it lies within the Black Areas and within 5 kilometers of forest (Appendix Figure A.2).²¹

III.B Outcome Data

Population. I digitized Malaysia’s Census of Population at the county level for 1931, 1947, 1957, 1970, 1980, 1991, and 2000.²² The census provides population counts by ethnic group for each county. To account for changes in county boundaries over time, I construct time-consistent borders based on the 1947 boundaries, grouping counties with overlapping geographies across years. I exclude nine counties with populations in 1947 but no reported populations in 1957 or 1970, as these are likely enumeration errors. The resulting baseline sample includes 777 counties, with a median width of 8–9 kilometers. For regressions using 1931 data, I create a separate set of 614 grouped counties based on the 1931 boundaries.

Employment. I measure employment by ethnic group across occupations and industries at the county level, using the 1980 and 1991 Population Censuses. Because the 1980 tabulations do not disaggregate employment by ethnicity, I impute ethnic breakdowns for each occupation and industry using corresponding ethnic shares calculated from the 2% census microdata.²³ For prewar aggregate employment data by industry and ethnic group, I use the 1947 tabulated census.

Income and education. I measure household incomes using both the 1980 census microdata and the Second Malaysian Family Life Survey (MFLS-2) from 1988–1989. The census provides data on household asset ownership but lacks direct measures of income or wages, while the MFLS-2 contains household earnings but has limited geographic coverage.²⁴ To generate income measures for the broader population, I train a model using the MFLS-2 to predict earnings based on observable household characteristics and apply it to the census microdata. The model includes household assets (e.g., automobile, motorcycle, refrigerator,

²⁰See Appendix Table A.1 for sources.

²¹The 1947 census map does not report the ethnic composition of each settlement. I use county-level composition from the tabulation, assuming uniform ethnic shares across settlements within a county.

²²I use “county” to refer to the administrative unit *mukim*. The 1931 Census was the first to document population by county.

²³The 1980 tables report the employed population aged 10 years and above by 1-digit occupation and industry; the 1991 tables report the employed population aged 15–64 years by similar categories.

²⁴The MFLS-2 was conducted by RAND and Malaysia’s National Population and Family Development Board. It provides demographic and socioeconomic data on nearly 3,000 households and is representative of Peninsular Malaysia. However, it covers only 174 counties.

TV), pairwise interactions of these assets, household size, and district fixed effects.²⁵

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.

Migration. I measure migration flows from the 1980 tabulated census, which reports population by place of last residence at administrative the district level (66 in total).²⁶ I construct a matrix of bilateral migration flows between district pairs to estimate migration costs. The 1980 census microdata also provides individual-level indicators of internal and external migration, along with the number of years individuals have resided in their current locations.

Firms. For historical manufacturing outcomes, I digitized the 1970 Directory of Manufacturing, which lists approximately 12,000 registered establishments in Peninsular Malaysia.²⁷ The directory provides each establishment’s name, address, main products, industry classification, and employment size, which I geocode to counties based on the provided addresses.

For more recent firm outcomes by ethnic ownership, I extracted and cleaned data from the Orbis Historical Disk (accessed in 2024), following the procedure in Kalemli-Özcan et al. (2024). I identify each firm’s ultimate owner(s) as those holding more than 50% ownership (or at least 25% if no single owner holds a majority), using the provided global ownership indicators.²⁸ The final sample includes firms with positive revenue, a Malaysian ultimate person owner, a valid NAICS industry code, and a location in Peninsular Malaysia. I restrict the sample to 2011–2015, the years with the most complete coverage. The resulting sample contains roughly 110,000 firms per year.²⁹

IV Empirical Strategy

The resettlement of rural Chinese was not entirely random. This section explains how I isolate the quasi-random component of the program to construct a population shifter. I then assess balance in geographic and pre-resettlement characteristics to evaluate the plausibility of the identifying assumptions.

IV.A Empirical Specification

I examine how the increase in Chinese population density from Emergency resettlement affected local economic outcomes, and how these effects varied 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

²⁵See Appendix Table A.3 for the model.

²⁶A district is the administrative unit above a county.

²⁷All establishments were required to register under the Registration of Business Ordinance 1957.

²⁸See Appendix A for details.

²⁹Appendix Figure A.8 shows the number of firms meeting the sample criteria by year, from 2003 to 2022.

challenge is converting site-level resettlement into an exogenous population shifter at the county level.

Consider the following reduced-form model:

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

I define county resettlement density as the standardized inverse hyperbolic sine of resettled population per county area.³⁰ I refer to a one-standard-deviation increase in resettlement density as “Higher Resettlement”; the coefficient β captures its effect.

I control for a set of pre-period characteristics in X_c that shaped resettlement decisions and could directly affect post-period outcomes. First, I include state fixed effects, as the program was implemented by state governments with varying economic and land policies. Second, because areas closer to transportation networks received more resettlement and had better market access, I control for pre-period road density and average distances to roads, rail stations, and the coast. Third, because the program aimed to minimize dislocation of resettled Chinese and areas with larger initial Chinese populations were often more urbanized, I control for 1947 county population and the Chinese population share. Fourth, since resettlement sites were often located on state-owned rubber or tin estates, I control for pre-period land use shares for rubber plantations and mining. Finally, I control for county area, which varies across the sample.³¹

To ensure comparisons are made only among similar areas, I include an indicator for whether a county received any resettlement. Counties without resettlement, such as dense west coast cities or remote jungle regions, were generally unsuitable for the program and had distinct economic potential.³²

Despite these controls, a concern remains that they may not fully account for non-random exposure to resettlement. Exposure depends not only on a county’s own characteristics but also on its position within the broader transportation network and population distribution. For instance, two counties with similar road densities could receive different resettlement flows if their neighbors varied in connectivity. Likewise, counties near more urbanized areas with larger initial Chinese populations may also receive more resettlement. These neighboring characteristics were correlated with market access (Donaldson and Hornbeck, 2016) and

³⁰I later use county resettlement density as an instrument for log population density in the structural estimation. The log-like transformation improves first-stage power while accommodating zeros. Appendix B.3 shows that results are robust to using a log transformation while restricting the sample to resettled counties.

³¹Appendix B.3 examines robustness to excluding the largest and smallest counties.

³²I include counties without any resettlement to improve the precision of covariate estimates. Appendix B.3 shows that results are similar when restricting to only resettled counties.

reflected selection of Chinese populations, which could correlate with unobserved productivity or amenities. Such exposures could introduce omitted variable bias if they are not fully captured by X_c (Borusyak and Hull, 2023).

To address this concern, I leverage institutional knowledge to specify counterfactual resettlement. Specifically, I isolate plausibly exogenous variation by controlling for the expected resettlement density, $\overline{ResettleDensity}_c$, defined as the average density across all counterfactual resettlements.³³

I construct $\overline{ResettleDensity}_c$ using a permutation procedure conducted independently within each state:

- (i). Randomly (and uniformly) permute counterfactual New Village sites (denoted by i), conditional on (i) distance to roads or rivers;³⁴ (ii) land use; and (iii) the county's squatter population decile.
- (ii). Calculate the counterfactual number of people resettled to each site using the gravity equation:

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

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.³⁵

- (iii). Aggregate the counterfactual resettled population across sites in each county and divide by county area to obtain the counterfactual county resettlement density.

Figure 2 illustrates the procedure for Johor using a single covariate: distance to roads. I approximate $\overline{ResettleDensity}_c$ by averaging counterfactual resettlement densities across 1,000 permutations.

The identification assumptions are twofold. First, I assume the British were equally likely to select sites with similar proximity to transportation networks and observable characteristics, without targeting locations based on unobserved productivity or amenities. Second,

³³Controlling for or re-centering by the expected resettlement purge potential omitted variable bias (Borusyak and Hull, 2023). In practice, they deliver qualitatively similar results.

³⁴If no roads are accessible within a 5-kilometer buffer but a river is, the permutation is conditional on distance to the nearest river.

³⁵The parameter ψ governs how costly it was to relocate people to more distant sites. I calibrate $\psi = 0.65$ by minimizing the sum of squared differences between actual and predicted resettlement across villages.

I assume they aimed to minimize dislocation, but poor planning and unpredictable communist activity led to idiosyncratic relocations unrelated to unobserved location fundamentals.³⁶ The identifying variation comes from the precise siting of New Villages relative to comparable locations along the transportation network, as well as from distant relocations beyond nearby squatter populations.

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

For individual and household outcomes, I estimate:

$$Y_{iec} = \beta_e \text{ResettleDensity}_c + \lambda_e \overline{\text{ResettleDensity}_c} + \gamma_e X_{ic} + \alpha_e \mathbf{1}\{\text{ResettleDensity}_c > 0\} + \varepsilon_{iec}, \quad (3)$$

where i denotes individuals or households and e denotes ethnic groups. The controls X_{ic} include the county covariates from Equation (1), and, in some specifications, additional individual or household characteristics to improve precision.

I use the Poisson Pseudo Maximum Likelihood (PPML) estimator to estimate elasticities for outcomes such as the number of establishments or employment, which may be zero in some counties or industries (Silva and Tenreyro, 2006). PPML captures both extensive and intensive margins and is invariant to the units of the dependent variable (Chen and Roth, 2023).

I report Conley standard errors that account for spatial correlation within a 30-kilometer radius (Conley, 1999). The cutoff is based on the localized nature of resettlement, typically within 20 miles, beyond which the treatment is plausibly independent. These standard errors are similar to those clustered at the district level (66 districts) and increase modestly (by 10–15%) when the cutoff is extended to 50 kilometers.

IV.B Pre-Characteristic Balance

This section assesses whether county resettlement density is orthogonal to pre-resettlement characteristics, as expected if the empirical strategy successfully isolates quasi-random variation. I examine balance on geography, access to amenities, and pre-existing economic activity.

Table 1 reports the relationship between resettlement density and various pre-period characteristics. Columns 1–4 examine geographic features such as elevation, terrain ruggedness,

³⁶For a formal discussion, see Appendix B.1.

and suitability for rice and coconut cultivation—the main food crops in Malaysia. Columns 5–8 examine proximity to public amenities: the nearest police station, post/telegraph office, hospital, and Chinese temple. Columns 9–12 assess pre-existing economic activity: land use for rubber and mining—the two major export industries—and proximity to industrial facilities and major cities such as Singapore, George Town, Malacca, Ipoh, and Kuala Lumpur.³⁷

Panel A shows that raw within-state correlations align with the program’s strategy, which prioritized areas along major transportation networks and in more urbanized regions. Counties with higher resettlement density were closer to public services (Columns 5–7), industrial facilities (Column 11), and major cities (Column 12). These areas also had more rubber cultivation (Column 9), consistent with historical accounts that many resettlement sites were located on state-owned rubber estates. Notably, these counties were not more suitable for agriculture. If anything, they were slightly less suitable, despite agricultural potential being a stated criterion in site selection (Columns 3–4). This finding supports historical accounts that emphasized security and speed over economic factors in resettlement.

Panel B adds controls for key resettlement determinants, including road networks and initial population distributions. With these controls, location characteristics are largely balanced, suggesting that much of the potential bias from broader network effects is accounted for. One exception is rubber land share, which becomes balanced only after additionally controlling for expected resettlement density (Panel C). Throughout the paper, the main specification includes controls for rubber and mining land use.

The magnitudes of estimated correlations are small. For example, a one-standard-deviation increase in resettlement density corresponds to just a 16-meter increase in elevation. Overall, the results suggest that the identification assumptions are plausible.

V Local Impacts of Resettlement

This section examines the local impacts of resettlement in receiving areas over the following decades. I begin by showing how the resettlement persistently shaped the population distribution. I then show its economic impacts and discuss their implications.

V.A Population Growth and Ethnic Composition

Figure 4 shows changes in total population by ethnic group (Panel A) and Chinese population share (Panel B) from a one standard deviation increase in county resettlement density. Prior to resettlement, population trends were similar across counties with varying resettlement density, supporting the identifying assumption that these areas did not initially differ in growth or labor demand. After resettlement, counties with higher resettlement density saw a

³⁷Industrial facilities include engineering shops, shipyards, chemical plants, power plants, and rubber and tin processing plants.

sharp and persistent rise in both the Chinese population and its share of the total population. In the short run, only the Chinese population increased, consistent with the program’s targeting of ethnic Chinese. Over time, non-Chinese populations also began moving into these counties.

Table 2 documents these changes. From 1947 to 1957, a one standard deviation increase in county resettlement density raised total population by 9.4% and the Chinese share by 4.8 percentage points (Column 1). Since one standard deviation corresponds to 13.6% of the 1947 population, this implies that each 1% resettled raised local population by 0.69%. This direct effect explains over three-quarters of population growth in resettled counties during the decade.

After mobility restrictions were lifted in 1960, population growth in higher-resettlement counties continued, and the higher Chinese share persisted. By 2000, these counties had 18% greater population density and a 4.1 percentage point higher Chinese share (Column 3). These long-run changes were driven by internal migration, not fertility: Chinese residents in counties with higher resettlement were more likely to have voluntarily moved in after 1960, with no significant difference in fertility rates (Appendix Table A.5).³⁸

Since employment and income—the main economic outcomes—are measured in 1980, it is useful to note that by that year, a one standard deviation increase in resettlement density raised total population by 11% and the Chinese share by 5 percentage points (Column 2). Given a baseline share of 0.42, this corresponds to a 12% increase in Chinese share.

V.B Economic Structure and Household Income

The influx of Chinese constituted a skill-biased labor supply shock, given their historical concentration in industrial and urban sectors. This section examines how resettlement reshaped the local economic structure and household income in subsequent decades.

Table 3 shows that by the 1980s, counties with higher resettlement density had 11% more agricultural employment (Column 1) and 29% more non-agricultural employment (Column 2).³⁹ The much larger increase in non-agricultural employment reflects both the Chinese tendency to work in non-agriculture (Panel B) and a shift of Malays out of agriculture in these counties (Panel C).⁴⁰

Table 4 shows that average household incomes were higher in more resettled counties, particularly for Chinese households. In 1980, Chinese households in counties with higher

³⁸Interestingly, non-Chinese women had slightly lower fertility rates in counties with higher resettlement.

³⁹Throughout the paper, I use “agricultural sector” or simply “agriculture” to refer to agriculture, hunting, forestry, fishing, and mining, though it primarily reflects agriculture.

⁴⁰Within the non-agricultural sector, employment effects were similar in secondary and tertiary industries (Appendix Table A.6). The effect was slightly larger in finance and business activities; see Appendix Figure A.6, Panel A, for a more detailed breakdown by industry.

resettlement density earned 11% more than their counterparts in less resettled counties (Panel A, Column 1).⁴¹ For non-Chinese households, the income gain was smaller and statistically insignificant at 4% (Column 2).

Panels B and C examine income effects by sector of employment, based on the industry of the household head.⁴² Chinese households in more resettled counties earned 7% more in agriculture and 13% more in non-agriculture, compared to their counterparts in less resettled counties (Column 1). Non-Chinese households saw modest income gains only in non-agriculture, with a 5% increase that was not statistically significant (Column 2). The ethnic gap in income gains is statistically significant (Column 3).⁴³

V.C Labor Pooling, Specialization, and Ethnicity-Based Agglomeration

The clustering of resettled Chinese may have generated productivity gains through agglomeration externalities, with greater benefits for the Chinese community. One potential channel is labor pooling, where a denser co-ethnic workforce improves matching between workers and firms (Marshall, 1890). This section investigates this mechanism by examining outcomes for both workers and firms.

Table 5, Column 1, shows that by 1980, Chinese individuals in counties with higher resettlement were more likely to participate in the labor force (Panel A). Among those employed, their industries and occupations were more concentrated (Panels B and C), consistent with greater specialization.⁴⁴ They were also more likely to work in managerial occupations, including legislators, senior officials, and managers (Panel D).⁴⁵ In contrast, non-Chinese individuals in more resettled counties showed no significant differences in these labor market outcomes (Column 2).

Table 6 shows that educational attainment was also higher in counties with higher resettlement density. By 1980, Chinese individuals in these counties had 0.4 additional years of schooling (8%), were 3.7 percentage points (7%) more likely to complete primary education,

⁴¹Appendix Table A.8 shows similar patterns across multiple household assets, which are strong predictors of income.

⁴²For households where the head lacked valid employment information or was unemployed, I use the eldest household member with valid employment.

⁴³Within non-agriculture, income gains are similar across the secondary and tertiary sectors. Effects may vary across more detailed industries, but sample sizes at the county-industry level are too small for precise estimation.

⁴⁴Concentration is measured by the Herfindahl-Hirschman Index (HHI). Formally, the occupational HHI for ethnic group e in county n is defined as $HHI_{n,occ}^e \equiv \sum_k (L_{nk}^e / L_n^e)^2$, where L_{nk}^e is total employment of group e in occupation k ; and $L_n^e = \sum_k L_{nk}^e$ is total employment of group e in county n . Higher HHI values indicate greater concentration. Occupations and industries are measured at the 1-digit level based on the 1991 census; see Appendix Figure A.6.

⁴⁵The share of Chinese employed as legislators, senior officers, or managers was 1.2 percentage points higher in counties with higher resettlement density—an increase of about 21% relative to the 5.6% national average in these managerial roles.

and 4.1 percentage points (14%) more likely to complete secondary education (Column 1). Non-Chinese individuals also had higher education, but to a lesser extent (Column 2). These improvements were driven entirely by younger cohorts under age 50, who had not completed schooling by the time of resettlement (Appendix Table A.7).

Table 7 shows that manufacturing firms relying on Chinese labor entered and scaled more in counties with higher resettlement density.⁴⁶ By 1970, these counties had 24% more manufacturing establishments in industries where ethnic Chinese made up over 80% of pre-resettlement employment (Column 1).⁴⁷ Establishments in these county-industries were also larger: they were 2 percentage points (5%) more likely to employ at least one full-time worker (Column 2), consistent with higher productivity.⁴⁸

Finally, Table 8 shows that Chinese-owned firms benefited more from higher Chinese density than non-Chinese-owned firms.⁴⁹ Using firm-level data from Orbis (2011–2015), I find that counties with higher resettlement had 52% more Chinese-owned firms and 31% more non-Chinese-owned firms (Panel A).⁵⁰ Chinese-owned firms in these counties were also larger, earning 19% more revenue than their counterparts in the same NAICS 2-digit industries in less resettled counties (Panel B, Column 1). In contrast, non-Chinese-owned firms saw smaller and statistically insignificant revenue gains of 8% (Column 2).

V.D Discussion

The resettlement of Chinese had persistent effects on population distribution and economic activity. Industries that relied on Chinese-specialized skills expanded in areas with higher Chinese density, generating positive externalities for both workers and firms. These productivity gains, particularly in non-agriculture, spurred internal migration and structural transformation of local economies.

While neoclassical theory predicts that marginal entrants to non-agriculture would be less productive than incumbents, thereby lowering wages, the opposite occurred: non-agricultural wages increased, and agricultural employment share declined. This shift suggests external economies of scale in the non-agricultural sector. By contrast, agricultural productivity saw little improvement in more resettled counties, even as less productive workers exited. This

⁴⁶I estimate regressions at the county-by-industry level, including county resettlement density, its interaction with an indicator for industries where over 80% of 1947 employment was Chinese, baseline county covariates, and industry fixed effects.

⁴⁷Ethnic Chinese historically dominated manufacturing in British Malaya. In 1947, only food products, wood products, textiles, and miscellaneous manufacturing had less than 80% Chinese employment. See Appendix Figure A.7.

⁴⁸Fewer than half of manufacturing establishments employed full-time workers in 1970.

⁴⁹I estimate regressions at the county-industry-year level, including county resettlement density, industry-by-year fixed effects, and their interactions with baseline county covariates. See Appendix A.2 for details on firm ownership classification.

⁵⁰Appendix A.2 describes how I define ultimate ownership and classify firms by owner ethnicity.

is consistent with local diminishing returns to labor in agriculture, given fixed land supply.

Labor pooling likely contributed to the external economies in non-agriculture. A denser Chinese workforce may have reduced search costs and improved worker-firm matching (Dauth et al., 2022), raising the returns to specialization and education (Kim, 1989). In line with this mechanism, Chinese workers in more resettled counties took on more specialized roles and were more likely to hold managerial positions (Appendix Figure A.6). These patterns suggest greater division of labor and organizational complexity typical of urban economies. Income gains were also larger in initially more populated counties (Appendix Table A.9, Panel A), consistent with stronger pooling effects in urban settings.

These Chinese-specific gains suggest barriers to cross-ethnic spillovers. Geographic segregation reduced gains: income effects for non-Chinese declined with distance from Chinese communities and disappeared beyond five kilometers. But even in the most integrated counties, non-Chinese households saw only half the income gains of Chinese households (Appendix Table A.9, Panel B). This points to deeper barriers—such as cultural and religious differences, language, and ethnic tensions—that limited cross-ethnic interactions. Consistent with this interpretation, non-Chinese households saw larger gains in counties with higher initial Chinese population shares or where some non-Chinese residents spoke Chinese—patterns that, while endogenous, suggest lower cross-ethnic frictions.⁵¹

Most New Villages were predominantly Chinese and offered limited opportunities for intergroup interaction. Villagers typically spent leisure time with same-village neighbors, many of whom attended the same Chinese primary schools. Even those working outside the village were often employed alongside Chinese coworkers (Lee and Tan, 2000).

V.E Alternative Mechanisms and Robustness

Beyond ethnicity-based agglomeration, resettlement may have affected outcomes through other channels that disproportionately benefited the Chinese. This section examines alternative mechanisms and assesses the robustness of the main results.

Characteristics of resettled Chinese. One possibility is that resettled squatters were more industrial or productive than the existing Chinese population, driving up average income without agglomeration benefits. This is unlikely, as most squatters had self-selected into rural areas before resettlement, and historical accounts suggest that 60% were agriculturalists (Sandhu, 1964, p. 169). Survey data from the late 1980s show that resettled Chinese were more likely to begin in agriculture and were less educated than other Chinese of similar

⁵¹Appendix Table A.9, Panels C–D, report the estimates. In the 1980 census, about 0.9% of non-Chinese individuals in Malaysia reported speaking any Chinese. Their income gains from resettlement were similar in magnitude and statistically indistinguishable from those of Chinese. While Chinese language acquisition is endogenous and may reflect unobserved traits, it is suggestive of lower frictions with the Chinese community.

age in the same state (Appendix Table A.10).⁵²

Selective migration. Another potential explanation is selective migration of wealthier Chinese into more resettled counties after 1960. The data do not support this as a main driver: income gains among Chinese households remain, though slightly attenuated, when restricting to non-migrants (Appendix Table A.11). Instead, income gains among non-Chinese disappear in the non-migrant sample (Column 2), suggesting that their gains were largely driven by selective in-migration of those better positioned—economically or socially—to benefit from Chinese agglomeration.

Land ownership. Better outcomes for Chinese in resettled areas might also reflect land ownership, as resettled families were allocated house lots and, if previously farmers, agricultural land (Sandhu, 1964). However, survey data from 1990 suggests that resettled Chinese households actually owned less land than other households in the same state (Appendix Table A.10). Land ownership is thus unlikely to explain their improved outcomes.⁵³

Education. Educational improvements may partly reflect better access to schooling or shifting preferences toward education. On the supply side, Chinese schools in New Villages may have lowered the cost of education. However, school data show that counties with higher resettlement density did not have better access to Chinese schools and had only slightly better access to national (non-Chinese) schools (Appendix Table A.12).⁵⁴

On the demand side, forced migration may have increased the value placed on education (Becker et al., 2020). Since the data do not distinguish resettled from other Chinese, I cannot separate this mechanism from broader agglomeration effects. Still, the results largely hold when controlling for education—an outcome that may itself reflect agglomeration economies (Appendix Table A.13). Education accounts for part of the income gains in non-agriculture, but the persistence of higher agricultural income among Chinese suggests that the results are not solely driven by sorting of the educated into higher-paying industrial jobs.⁵⁵

Roads. During the Emergency, the British colonial government built roads to connect remote villages to the main transport network when no road or river access was available. These early infrastructure investments may have had persistent effects through path dependence. While road access would not necessarily favor one ethnic group over another,

⁵²Resettled Chinese are identified from the Second Malaysian Family Life Survey as individuals not born in a New Village but who had “migrated” there before 1960.

⁵³Lease terms ranged from 20 to 30 years depending on the state. Many received land titles only years later, and some were unaware of their land rights (Strauch, 1981, pp. 63–72).

⁵⁴One caveat is that the school data are from 2022 and may not reflect school access in the 1980s. To assess this concern, I digitize a directory of Chinese schools around 1960, which shows similar school coverage.

⁵⁵Higher agricultural income among Chinese households may reflect linkages with downstream Chinese-owned firms that purchased and processed agricultural products.

I examine its relationship with county resettlement density in the two decades following resettlement (Appendix Table A.14). In 1961, shortly after the Emergency, counties with higher resettlement density had slightly higher density of local roads but no better access to main roads (Panel D, Column 1), consistent with historical accounts. Over the next two decades, road access improved modestly in these counties, possibly reflecting input sharing as a channel of agglomeration economies, though the differences are not statistically significant (Column 3). Overall, the differences are small in magnitude—about 0.4 kilometers closer to minor roads and 12 meters per square kilometer higher road density—and the main results are robustness to directly controlling for road access (Appendix Table A.16).

Robustness. Appendix B.3 shows that the main results are robust to a range of alternative specifications and sample restrictions. First, results are similar when using a logarithmic transformation of the population shifter. Second, alternative specifications of counterfactual resettlement generate similar estimates. Third, controlling for neighboring road and population characteristics, as well as pre-period proximity to industrial and urban areas, yields similar results. Results are also robust to excluding counties with the largest or smallest areas, high-density prewar towns, and counties with extreme resettlement densities.

VI A Quantitative Model of Migration, Occupation, and Agglomeration

The empirical results suggest that agglomeration elasticities vary by sector and local ethnic composition. This heterogeneity implies that place-based or industrial policies that reallocate population across regions or sectors can have aggregate consequences (Glaeser and Gottlieb, 2008).⁵⁶

However, the cross-sectional analysis shows only relative differences across counties and does not capture the aggregate impact of the program. The comparison across places along the road network with similar location fundamentals also disregard the fact that resettlement generally moved people from remote areas to places with better market access.

To assess the program’s aggregate impact and evaluate counterfactual policies, I develop a spatial general equilibrium model that extends Allen and Donaldson (2022) and Peters (2022). The model features agglomeration forces that vary by sector and ethnic composition. It incorporates migration and occupation choices in a Roy (1951)-type framework, allowing individuals from different ethnic groups to have heterogeneous location preferences and sectoral comparative advantage. Regions are linked through trade and migration. The

⁵⁶Relocating Chinese populations from areas with lower Chinese density—where spillovers to Chinese are weaker—to areas with higher Chinese density—where spillovers are stronger—could increase aggregate output if the gains at the destination exceed the losses at the origin. Similarly, reallocating labor from sectors with weaker agglomeration to those with stronger agglomeration can boost productivity—a common rationale for industrial policy.

resettlement shifts the initial population distribution, which continues to shape the equilibrium due to mobility costs.

VI.A Environment

The model features N regions and two sectors $k \in \{A, M\}$: Agriculture (A) and Non-Agriculture (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 non-agriculture. Finally, consumption and production take place.

Production. Each region n produces a unique good in each sector, following Armington (1969). In each region-sector, a continuum of perfectly competitive firms produces a homogeneous regional variety. Production exhibits constant returns to scale with labor as the only input. Output is given by $Q_{nk} = H_{nk}$, where H_{nk} denotes 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 production.

Firms in sector k and region n choose labor H_{nk} to maximize profit, taking the local sectoral wage (per efficiency unit) and output prices as given. In equilibrium, no-arbitrage implies that the price of sector- k goods produced in region n and sold in region r is given by $p_{nrk} = (\tau_{nr}/\tau_{nn})p_{nnk}$, where p_{nnk} is the price of sector- k goods sold locally; $\tau_{nr} \geq 1$ is the iceberg trade cost between regions n and r ; and τ_{nn} is the within-region trade cost.⁵⁷

Under perfect competition, firms earn zero profit in equilibrium, implying $w_{nk} = p_{nnk}/\tau_{nn}$, 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 agricultural and non-agricultural goods, and from the local amenity in 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 $\sigma > 1$ is the constant elasticity of substitution across regional varieties, assumed identical across sectors. The term a_n^e captures the (endogenous) amenity value for individuals

⁵⁷Without loss of generality, within-region trade can be costly ($\tau_{nn} > 1$). An increase in τ_{nn} lowers local wages and is isomorphic to worse production fundamentals.

of ethnicity e in region n .⁵⁸

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

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

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

where the scale parameter $\bar{a}_n^e$ captures exogenous, ethnicity-specific amenities in region n , and ν is the shape parameter, with a higher ν implying less dispersion in location preferences across individuals.

The realized amenity value for individual i of ethnicity e in region n is:

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

where $f_a(\cdot)$ depends on the local population size and ethnic composition:

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

Here, β captures congestion effects, while β^e governs the strength of homophily—a preference for living near co-ethnics—which may also reflect ethnic tensions.

The indirect utility of individual i from group e , moving from origin r to destination n , 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 .⁵⁹

Since V_{irn}^e is a Fréchet random variable scaled by a constant, it is itself Fréchet distributed.

⁵⁸Technically, a_n^e varies at the individual level (e.g., a_{in}^e), but individual subscripts are omitted for simplicity.

⁵⁹The term $\Gamma_\theta \equiv \Gamma(1 - 1/\theta)$ is a constant derived from Fréchet distribution, where $\Gamma(\cdot)$ denotes the Gamma function.

This implies that the share of individuals from ethnicity e initially in r who choose to move to n is:

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

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

$$V_n^e = (\bar{a}_n^e)^{1/\nu} (L_n)^\beta \left(\frac{L_n^e}{L_n} \right)^{\beta^e} \bar{w}_n^e P_n^{-1}. \quad (4)$$

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

$$L_{rn}^e = \eta_{rn}^{-\nu} \times \frac{\check{L}_r^e}{(\Pi_r^e)^\nu} \times \frac{L_n^e / \bar{L}}{(\mathcal{V}_n^e)^{-\nu}}, \quad (5)$$

where the two migration market access terms are defined as:

$$\Pi_r^e \equiv \left(\sum_{l=1}^N (\eta_{rl}^{-1} V_l^e)^\nu \right)^{1/\nu}, \quad (6)$$

$$\mathcal{V}_n^e \equiv V_n^e (L_n^e / \bar{L})^{-1/\nu}, \quad (7)$$

and $\bar{L}$ denotes total population in the economy, normalized to 1.

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

Sectoral labor supply. Individuals supply labor inelastically and earn income based on their heterogeneous productivity. After migrating, each individual draws a vector of efficiency units in agriculture (A) and non-agriculture (M), denoted $\Lambda_i = (\Lambda_{iA}, \Lambda_{iM})$, where Λ_{ik} is the effective labor individual i provides if employed in sector k . These draws follow a Fréchet distribution:

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

where the scale parameter ϕ_{nk}^e captures the average productivity of ethnicity e in sector k and region n . This parameter reflects both ethnic comparative advantage and 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 across individuals.

Productivity also depends on local agglomeration externalities. Let λ_{ink}^e denote the net efficiency of individual i of ethnicity e working in sector k in region n . It depends on their skill draw Λ_{ink}^e and the local population distribution:

$$\lambda_{ink}^e = \Lambda_{ink}^e f_\lambda(L_{nk}^e, L_{nk}^{e'}),$$

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

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

The parameters γ_k and γ^e govern the strength of productivity spillovers with respect to sector size and ethnic composition, respectively.

Specifically, the elasticity of group- e productivity in sector k and region n with respect to group- e population in n is:

$$\frac{\partial \ln \lambda_{nk}^e}{\partial \ln L_n^e} = \underbrace{\left(\gamma_k \frac{L_{nk}^e}{L_{nk}} + \gamma^e \left(1 - \frac{L_{nk}^e}{L_{nk}} \right) \right)}_{\text{Direct effect}} \underbrace{\left(1 + \frac{\partial \ln \pi_{nk}^e}{\partial \ln L_n^e} \right)}_{\text{Indirect/GE effect}}, \quad (8)$$

where π_{nk}^e is the share of group- e individuals working in sector k in region n . The first term reflects the direct impact of a larger co-ethnic population on group- e 's productivity. It is a weighted average of γ_k and γ^e , where the weights correspond to the share of group e in sector k . When ethnic group e dominates the sector in region n , the elasticity approaches γ_k ; when group e' dominates, it approaches γ^e . The second term captures indirect effects through general equilibrium adjustments, scaling the direct effect based on how a larger group- e population affects sectoral choice via changes in relative wages.

Cross-ethnic spillovers are also allowed. The elasticity of group- e' ($e' \neq e$) productivity with respect to the size of group e is:

$$\frac{\partial \ln \lambda_{nk}^{e'}}{\partial \ln L_n^e} = \underbrace{(\gamma_k - \gamma^e) \frac{L_{nk}^e}{L_{nk}}}_{\text{Direct effect}} \underbrace{\left(1 + \frac{\partial \ln \pi_{nk}^{e'}}{\partial \ln L_n^e} \right)}_{\text{Indirect/GE effect}}. \quad (9)$$

The sign of the direct effect depends on whether own-group spillovers (γ^e) are stronger or weaker than sectoral spillovers (γ_k). It is positive when $\gamma_k > \gamma^e$, and its magnitude scales with group e 's local employment share in sector k .

Given the Fréchet distributed efficiency draws, the share of group- e individuals in region

n who work in sector k is given by:

$$\pi_{nk}^e = \phi_{nk}^e \left(\frac{w_{nk}^e}{\bar{w}_n^e} \right)^\theta, \quad (10)$$

where w_{nk}^e is the effective wage per efficiency unit (adjusted for externalities):

$$w_{nk}^e = w_{nk} (L_{nk})^{\gamma_k} \left(\frac{L_{nk}^e}{L_{nk}} \right)^{\gamma^e}, \quad (11)$$

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

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

Trade. Trade from region n to r faces exogenous iceberg costs, denoted $\tau_{nr} \geq 1$, where $\tau_{nr} = 1$ represents frictionless trade. Given CES 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 the standard gravity equation:

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

where $X_{rk} = \alpha_k Y_r$ is 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 $Y_r = w_{rA} H_{rA} + w_{rM} H_{rM}$.

The gravity equation can be rewritten as:

$$X_{nrk} = \alpha_k \tau_{nr}^{1-\sigma} \times \frac{Y_n / \bar{Y}}{\mathcal{P}_{nk}^{1-\sigma}} \times \frac{Y_r}{P_{rk}^{1-\sigma}}, \quad (12)$$

where I define two trade market access terms:

$$P_{rk} \equiv \left(\sum_{l=1}^N \tau_{lr}^{1-\sigma} w_{lk}^{1-\sigma} \right)^{1/(1-\sigma)}, \quad (13)$$

$$\mathcal{P}_{nk} \equiv w_{nk}^{-1} (Y_n / \bar{Y})^{1/(1-\sigma)}. \quad (14)$$

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

As with migration, (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 $\{\tilde{L}_r^e\}$ and a set of location fundamentals $\{\phi_{nk}^e, \bar{a}_n^e, \tau_{nr}, \eta_{nr}\}$, an equilibrium consists of a set of prices $\{w_{nk}, p_{nk}\}$ and quantities $\{L_{nk}^e, H_{nk}\}$, such that (i) firms and consumers act optimally, and (ii) goods and labor markets clear in every region.

The goods market clearing condition is:

$$w_{nk}H_{nk} = \sum_{r=1}^N \alpha_k (w_{rA}H_{rA} + w_{rM}H_{rM}) \frac{\tau_{nr}^{1-\sigma} w_{nk}^{1-\sigma}}{\sum_{l=1}^N \tau_{lr}^{1-\sigma} w_{lk}^{1-\sigma}}. \quad (15)$$

This condition states that total income earned in sector k of region n equals total sales of sector- k goods to all destinations. It embeds two underlying conditions: (i) all sectoral revenue is paid out as wages, and (ii) all regional income is fully spent on goods from all regions.

The labor market clearing condition is:

$$H_{nk} = \sum_e H_{nk}^e = \sum_e L_n^e \pi_{nk}^e (\Gamma_\theta \bar{w}_n^e w_{nk}^{-1}) \quad (16)$$

$$L_n^e = \sum_r \tilde{L}_r^e \frac{(\eta_{rn}^{-1} V_n^e)^\nu}{\sum_{l=1}^N (\eta_{rl}^{-1} V_l^e)^\nu}. \quad (17)$$

Equation (16) states that total labor in sector k of region n equals the sum of efficiency units contributed by both ethnic groups. Each group's contribution is the product of its sectoral employment ($L_n^e \pi_{nk}^e$) and average efficiency units ($\Gamma_\theta \bar{w}_n^e w_{nk}^{-1}$). Equation (17) follows from the migration flow identity: the equilibrium population of group e in region n equals the sum of migration flows from all origins.

Using Equations (4), (10), and (11), 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 (15–17) in $6 \times N$ unknowns $\{w_{nk}, H_{nk}, L_n^e\}$.

Existence and uniqueness. I prove the existence of equilibrium by construction, using the iterative procedure described in Appendix C.3. The process features three nested loops: the outer loop updates population distribution; the middle loop solves for sectoral wages and prices given the population; and the inner loop solves for occupational shares given population and wages. Convergence is ensured by congestion forces, such as idiosyncratic migration preferences, sectoral heterogeneity in productivity, imperfect substitution across regional varieties, and external congestion externalities.

When agglomeration forces (γ_k and γ^e) outweigh congestion forces, the model may admit

multiple equilibria, with economic activity concentrating in different sets of location-sectors across equilibria. I verify uniqueness at baseline parameters by initializing the algorithm from different starting values and confirming convergence to the same outcomes.⁶⁰ Note that equilibrium is conditional on the initial population distribution; uniqueness here refers to the existence of 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. Bilateral migration and trade costs are both symmetric and increasing in distance. Migration costs take the form $\eta_{rn} = (d_{rn}/d_{min})^\kappa$, where d_{min} is the minimum within-county distance, and $\kappa \geq 0$ is the distance elasticity of migration costs. Similarly, trade costs are given by $\tau_{nr} = (d_{rn}/d_{min})^\xi$, where $\xi \geq 0$ is the distance elasticity of trade costs.⁶¹

The model is characterized by a tuple of location fundamentals $\{\phi_{nk}^e, \bar{a}_n^e\}$ and 11 structural parameters:

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

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). I set $\sigma = 8$ based on a recent estimate from Vietnam (Balboni, 2025).⁶² Since internal trade flow data are unavailable for Malaysia, I follow Monte, Redding and Rossi-Hansberg (2018) and calibrate ξ such that $\xi(1 - \sigma) = -1.29$, which implies $\xi = 0.18$, given $\sigma = 8$. Finally, I set $\nu = 3$, consistent with the range of 2 to 4 reported by existing studies in developing countries.⁶³

The remainder of the section is structured as follows. I first discusses how I identify

⁶⁰I also explore deriving sufficient conditions for uniqueness following Allen, Arkolakis and Li (2024), but the conditions are not informative in my setting (see Appendix C.4).

⁶¹Cross-county distances d_{rn} are measured by the Euclidean distance between centroids; within-county distances d_{rr} are calculated from the centroid to the nearest boundary. I allow within-county costs to exceed 1—except for the smallest county, which is normalized to 1—to account for differences in county size. This normalization is without loss of generality: higher migration costs reduce utility similarly to a decline in amenity $\bar{a}_n^e$, and higher trade costs lowers productivity as if ϕ_{nk}^e were lower.

⁶²Estimates of σ typically range from 4 to 9. Donaldson and Hornbeck (2016) estimate 9.22 in 19th century U.S., Peters (2022) estimates 5.02 in post-war Germany, and Balboni (2025) estimates 7.92 in Vietnam in 2009.

⁶³Estimates of ν are rare, especially for developing countries. See, for example, Bryan and Morten (2019); Tombe and Zhu (2019); Morten and Oliveira (2024).

and estimate the remaining eight parameters. I then discuss the estimation procedure and results.

VII.A Identification

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

Market access terms. From the equilibrium conditions, I derive four relationships involving trade and migration market access: (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.⁶⁴ These conditions yield the following system of equations:

$$\mathcal{P}_{nk}^{1-\sigma} = \frac{\alpha_k}{\Omega_{nk}} \sum_r \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1}, \quad (18)$$

$$P_{rk}^{1-\sigma} = \sum_n \tau_{nr}^{1-\sigma} Y_n \mathcal{P}_{nk}^{\sigma-1}, \quad (19)$$

$$(\mathcal{V}_n^e)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e (\Pi_r^e)^{-v}, \quad (20)$$

$$(\Pi_r^e)^v = \sum_n \eta_{rn}^{-v} L_n^e (\mathcal{V}_n^e)^\nu, \quad (21)$$

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 (18)–(21).*

Proof. See Appendix D.1. □

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

Migration cost elasticity. Since the migration cost elasticity κ enters the migration cost function multiplicatively with taste dispersion ν , I estimate their product, $\tilde{\kappa} \equiv \kappa\nu$. Using non-linear least squares, I minimize the difference between the model-predicted and observed district-to-district migration flows (see Appendix D.2 for details). I assume a common κ across ethnic groups, as separate estimates by group yield similar values (see Section VII.B).

⁶⁴See Appendix D.1 for derivations.

Identification relies on the assumption that deviations between observed and predicted flows are due to classical measurement errors, uncorrelated with geography or other unobserved determinants of migration market access. As the sample size grows, these errors average out, and observed flows converge to the model predictions under the true $\tilde{\kappa}$. Appendix Figure A.9 shows that the loss function is convex, indicating a unique minimizing value of $\tilde{\kappa}$.

Migration flows in the model are defined over a 24-year period, whereas the census data report flows with an average of 12 years.⁶⁵ To match the model's time horizon, I convert the observed 12-year migration shares into 24-year shares, assuming stable migration patterns over time (Artuç, Chaudhuri and McLaren, 2010; Caliendo, Dvorkin and Parro, 2019).⁶⁶

Skill dispersion. The shape parameter θ governs the dispersion of Fréchet-distributed productivity across individuals, with larger θ indicating less dispersion. Let y_{ink}^e denote the earnings of individual i of ethnicity e working in sector k and residing in region n . Since earnings also follow a Fréchet distribution, it follows that:

$$\frac{\text{Var}[y_{ink}^e]}{\mathbb{E}[y_{ink}^e]^2} = \frac{\Gamma(1 - \frac{2}{\theta}) - \Gamma(1 - \frac{1}{\theta})^2}{\Gamma(1 - \frac{1}{\theta})^2}.$$

This normalized variance approaches infinity as θ approaches 2 from above and decreases monotonically toward 0 as θ increases. Thus, any observed normalized variance maps to a unique value of θ ; that is, θ is identified.

Productivity spillovers. The parameters $\gamma_A, \gamma_M, \gamma^e$ govern productivity spillovers, affecting expected earnings in each sector and hence occupational choices. Rewriting the occupation choice equation (10) using trade market access terms yields:

$$\begin{aligned} \ln \bar{w}_n^e = & \gamma_k \ln L_{nk} + \gamma^e \ln \left(\frac{L_{nk}^e}{L_{nk}} \right) - \frac{1}{\theta} \ln \pi_{nk}^e \\ & - \left(\frac{1}{\sigma - 1} \right) \ln (\mathcal{P}_{nk})^{\sigma-1} - \left(\frac{1}{\sigma - 1} \right) \ln Y_n + \underbrace{\frac{1}{\theta} \ln \phi_{nk}^e}_{\text{error term}}, \quad \forall k \in \{A, M\}. \end{aligned} \quad (22)$$

The left-hand side is the average wage of ethnic group e in region n , which is observed in the data.⁶⁷ The size of sectoral employment shifts wages through γ_k , while its ethnic composition

⁶⁵The 1980 tabulated census reports migration flows based on the “place of last residence,” and the micro-data indicate an average residency of 12 years.

⁶⁶I first compute a 12-year migration share matrix (with each row summing to 1) and square it to obtain the 24-year matrix $\hat{m}_{jh}$. This transformation also smooths out zeros in the data, allowing the use of logs in estimation (see Equation (A-6)).

⁶⁷The Fréchet assumption, combined with a common shape parameter across sectors, implies that average wage for an ethnic group within a region are equal across sectors. Appendix Figure A.11 supports this

shifts wages through γ^e .

To identify γ^e , I subtract Equation (22) for one group from the other, removing region-sector-specific terms:

$$\ln \left(\frac{\bar{w}_n^c}{\bar{w}_n^m} \right) = \gamma^e \ln \left(\frac{L_{nk}^c}{L_{nk}^m} \right) - \frac{1}{\theta} \ln \left(\frac{\pi_{nk}^c}{\pi_{nk}^m} \right) + \underbrace{\frac{1}{\theta} \ln \left(\frac{\phi_{nk}^c}{\phi_{nk}^m} \right)}_{\text{error term}}. \quad (23)$$

This equation represents a relative (inverse) labor demand curve for sector k . The $-1/\theta$ term reflects the neoclassical force driving downward-sloping demand when γ^e is not too large.⁶⁸

Unobserved productivity ϕ_{nk}^e enters as the error term and likely correlates positively with local population, as individuals sort into more productive locations. This selection bias tends to inflate the OLS estimates of γ_k and γ^e . In contrast, classical measurement errors in the population distribution attenuate both estimates, biasing them downward.

To address these biases, I use an instrumental variable strategy. Specifically, I instrument for local ethnic composition and sector size using exogenous variation from the resettlement program. I denote the residualized resettlement density from Section V as $Z_n^{(own)}$. I also construct another instrument, $Z_n^{(neighbor)}$, based on neighboring resettlement: the average resettlement density of neighboring counties after controlling for baseline characteristics and their expected resettlement density. Since resettlement was not targeted based on location productivity (conditional on covariates), these instruments are plausibly orthogonal to ϕ_{nk}^e , leading to the following moment conditions:

$$\mathbb{E}[Z_n \ln \phi_{nk}^e] = 0, \quad \forall k, e; \quad Z_n \in \{Z_n^{(own)}, Z_n^{(neighbor)}\}. \quad (24)$$

These moment conditions identify the three productivity spillover parameters: γ_A , γ_M , and γ^e . I identify γ^e using Equation (23), instrumenting for $\ln(L_{nk}^e/L_{nk})$ given θ . I then identify γ_k using Equation (22) by instrumenting for $\ln L_{nk}$ after moving all other terms to the left-hand side.

Amenity spillovers. The amenity spillover parameters, β and β^e , shape migration patterns. The value of residing in region n , from Equation (4), can be rewritten using migration

assumption by showing that average log household earnings within a county-ethnicity pair are similar across sectors.

⁶⁸Only relative shares matter here because Chinese and Malays are perfect substitutes in production. If both groups are equally likely to work in sector k , differences in ethnic composition would affect relative wages only through γ^e .

market access as:

$$\begin{aligned} \ln \bar{w}_n^e = & (-\beta + \beta^e) \ln L_n + \left(\frac{1}{\nu} - \beta^e \right) \ln L_n^e + \frac{1}{\nu} \ln (\mathcal{V}_n^e)^\nu \\ & + \left(\frac{\alpha}{\sigma - 1} \right) \ln P_{nA}^{\sigma-1} + \left(\frac{1 - \alpha}{\sigma - 1} \right) \ln P_{nM}^{\sigma-1} - \underbrace{\frac{1}{\nu} \ln \bar{a}_n^e}_{\text{error term}}. \end{aligned} \quad (25)$$

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

$$\ln \left(\frac{\bar{w}_n^c}{\bar{w}_n^m} \right) = \left(\frac{1}{\nu} - \beta^e \right) \ln \left(\frac{L_n^c}{L_n^m} \right) + \frac{1}{\nu} \ln \left(\frac{(\mathcal{V}_n^c)^\nu}{(\mathcal{V}_n^m)^\nu} \right) - \underbrace{\frac{1}{\nu} \ln \left(\frac{\bar{a}_n^c}{\bar{a}_n^m} \right)}_{\text{error term}}. \quad (26)$$

This equation represents a relative (inverse) labor supply curve. When β^e is not too large, the neoclassical force $1/\nu$ predicts an upward-sloping relationship between relative wages and population share. If β^e is large, however, a higher Chinese share becomes an amenity for Chinese individuals, making them more willing to accept lower wages. The term $\mathcal{V}_n^e$ captures potential group- e migrants from other counties and shifts local labor supply. The error term captures unobserved amenities that make a county more attractive to a group, likely biasing the OLS estimate of β^e upward.

To address this endogeneity, I use the same resettlement-based instruments. Assuming the program did not target areas based on amenity fundamentals, the following moment conditions identify β and β^e :

$$\mathbb{E}[Z_n \ln \bar{a}_n^e] = 0, \quad \forall e, \quad Z_n \in \{Z_n^{(own)}, Z_n^{(neighbor)}\}. \quad (27)$$

Location fundamentals. I recover the exogenous location fundamentals (up to scale) as structural residuals from Equations (22) and (25). Specifically, I recover ϕ_{nk}^e from Equation (22) after estimating γ_k and γ^e , and $\bar{a}_n^e$ from Equation (25) after estimating β and β^e .

VII.B Estimation

The estimation proceeds in four steps. First, I estimate the migration cost elasticity with respect to distance, $\tilde{\kappa}$, and use it to compute the migration cost matrix $\eta_{nr}^{-\nu}$. Second, I iteratively solve for the market access terms and the agricultural expenditure share α , following Proposition 1. Third, I estimate the shape parameter of Fréchet skills, θ , by targeting the population-weighted average of normalized wage variance within (n, k, e) cells. Finally, I estimate the agglomeration parameters $\{\gamma_A, \gamma_M, \gamma^e, \beta, \beta^e\}$ using the generalized method of moments (GMM), based on the moment conditions in Equations (24) and (27).

To mitigate small-sample concerns, I exclude counties with fewer than five employed

households in the 1980 census microdata and weight the estimation by household count. The final sample includes 685 counties. I bootstrap the entire procedure to obtain standard errors.⁶⁹

Table 9 reports the parameter estimates. I find strong productivity spillovers in non-agriculture: the elasticity of productivity with respect to local employment is $\gamma_M = 0.22$, in line with estimates for manufacturing in the literature.⁷⁰ In contrast, agricultural productivity declines with local employment ($\gamma_A = -0.12$), consistent with diminishing returns to labor when land is fixed.

The within-ethnic productivity spillover is estimated at $\gamma^e = 0.13$, implying that a higher Chinese share raises Chinese workers' productivity. The effect on Malay workers depends on the sector. Since $\gamma_M > \gamma^e > \gamma_A$, Equation (9) implies that a larger Chinese population increases Malay productivity in non-agriculture but reduces it in agriculture.⁷¹

For amenity spillovers, I estimate a small congestion elasticity with respect to total population ($\beta = -0.005$). Accounting for housing as a non-traded good, this implies a pure amenity spillover of 0.05.⁷² The within-group amenity spillover is $\beta^e = 0.13$, indicating that individuals prefer living near co-ethnics. This is consistent with scale economies in ethnicity-specific amenities and social frictions in residential or consumption choices (Duranton and Puga, 2004; Davis et al., 2019). Appendix D.3 discusses the full set of estimates and compares them to existing studies.

Estimated location fundamentals correlate with observable measures of productivity and amenities. Agricultural production fundamentals are negatively correlated with terrain ruggedness and positively with agricultural suitability (Appendix Table A.18, Column 1). Non-agricultural fundamentals are positively correlated with the number of manufacturing firms in 1970 (Column 2). Amenity fundamentals are positively correlated with the density of public services—such as police stations, post/telegraph offices, and schools—and negatively correlated with distance to these amenities (Columns 3–4).⁷³

⁶⁹In each bootstrap iteration, I sample individuals with replacement from the census microdata at the district level and re-aggregate outcomes to the county level. Districts are larger than counties, with 66 districts in total.

⁷⁰Kline and Moretti (2014) estimates 0.2 for the U.S. in the 1930s. Greenstone, Hornbeck and Moretti (2010) report values between 1.25 to 3.1 in more recent settings.

⁷¹I assume perfect substitutability between Chinese and non-Chinese workers. If they are imperfect substitutes, the estimated γ^e understates the true within-ethnic spillover. In that case, a stronger within-ethnic spillover would be required to rationalize the limited wage gains among non-Chinese workers, who benefit from complementarity with resettled Chinese.

⁷²The pure amenity spillover refers to the utility gain from local population net of housing market effects. See Appendix D.3 and Bryan and Morten (2019) for further discussion.

⁷³For the distribution of estimated fundamentals, see Appendix Figures A.12 and A.13.

VIII Counterfactuals

In this section, I use the estimated model to evaluate the resettlement program’s aggregate impact and explore how heterogeneous agglomeration forces shape the economy through policy counterfactuals. Section VIII.A simulates a counterfactual 1980 equilibrium without the resettlement program and compares it to the observed 1980 economy to quantify the program’s aggregate impact. This no-resettlement equilibrium serves as the baseline for all subsequent comparisons. Section VIII.B examines the role of cross-ethnic barriers in limiting productivity spillovers and shaping population distribution and aggregate productivity. Section VIII.C evaluates an industrial policy that subsidizes Malay wages in non-agriculture.

VIII.A Aggregate Impact of Forced Resettlement

While counties that received resettled Chinese populations benefited economically from higher Chinese density, the areas from which they were removed likely suffered. This section evaluates the distributional and aggregate impact of the resettlement program.

I simulate a counterfactual “no resettlement” equilibrium in 1980 by initializing the model with the 1947 population distribution—before resettlement—instead of the post-resettlement 1957 distribution, holding all parameters and location fundamentals fixed. I then compare this equilibrium to the observed 1980 economy to quantify the program’s aggregate impact.⁷⁴

Figure 5 maps how resettlement changes county outcomes. Panel A shows that resettlement shifts Chinese populations from inland and remote areas to more urban, coastal areas. Panel B shows that this shift raises Chinese income per capita (equal to output per capita in the model) in receiving counties, while income declines in counties that lose Chinese populations. Panel C shows that Malay income per capita generally declines in counties where Chinese density increases, except in major urban centers such as Kuala Lumpur. In those areas, many Malays work in non-agriculture, where strong external economies of scale boost their productivity despite limited cross-ethnic spillovers. At the same time, Chinese departures from rural areas raise Malay agricultural productivity through local diminishing returns to labor.

Table 10 quantifies these effects. Among Chinese, the program raises the share working in non-agriculture by 1 percentage point (1.5%) and increases their non-agricultural productivity by 1.6%. However, agricultural productivity declines, as resettled destinations are generally less suitable for farming (Column 1). Among Malays, the share working in non-agriculture falls as they move into agricultural lands made more productive by the departure

⁷⁴This approach assumes (i) the 1947 distribution was a steady state that would have persisted until 1957 without resettlement, and (ii) location fundamentals evolved after 1957, bringing the economy to the observed 1980 equilibrium. The first assumption is reasonable, as the direct effect of resettlement accounts for nearly 80% of total population growth in receiving counties during that decade (Section V).

of resettled Chinese.⁷⁵ Malay non-agricultural productivity rises, partly due to spillovers from incoming Chinese workers and partly due to selection, as less productive Malays shift into agriculture (Column 2). Although the program directly targets the Chinese population, Malays also experience notable changes through general equilibrium effects.

I decompose the aggregate output change as follows:

$$\begin{aligned} \sum_{n,k,e} \left(\tilde{y}_{nk}^e \tilde{L}_{nk}^e - y_{nk}^e L_{nk}^e \right) &= \sum_{n,k,e} \left(\tilde{L}_{nk}^e - L_{nk}^e \right) y_{nk}^e + \sum_{n,k,e} \left(\tilde{y}_{nk}^e - y_{nk}^e \right) L_{nk}^e \\ &\quad + \sum_{n,k,e} \left(\tilde{y}_{nk}^e - y_{nk}^e \right) \left(\tilde{L}_{nk}^e - L_{nk}^e \right), \end{aligned}$$

where y_{nk}^e is output per capita for group e in county n and sector k in the baseline, and $\tilde{y}_{nk}^e$ is its value in the resettled equilibrium. The first term isolates the effect of labor reallocation, holding output per capita fixed. The second term captures changes in per-capita output, holding labor allocation fixed. The third term reflects the interaction between labor reallocation and productivity changes.

Overall, the resettlement program raises aggregate output by 2%, driven largely by labor reallocation to more productive sectors and regions (Column 3). This reflects the fact that resettled destinations were along the transportation network, with better market access, and the model assigns these areas higher baseline productivity (Appendix Figure A.14). With productivity held fixed, labor reallocation alone accounts for about two-thirds of the output gain. The program's total expenditure was approximately 133 million Malayan Dollars (\$43 million), or 0.5% of Malaysia's 1980 GDP (adjusted for inflation), implying a net output gain of around 1.5%.⁷⁶

The output gain corresponds to a 4.8% increase in aggregate utility (Column 3), but this figure does not capture the welfare loss from forced relocation.⁷⁷ To benchmark the economic gain against this coercion cost, I proceed in two steps. First, I invert the model to recover the amenity fundamentals that would have sustained the 1947 population distribution. Then, I solve for the least-cost wage subsidies by place and ethnicity that would induce individuals to voluntarily relocate from the 1947 to the 1957 distribution.⁷⁸

⁷⁵The decline in Malay non-agricultural share may seem inconsistent with the cross-sectional results in Section V, which showed higher Malay non-agricultural shares in counties receiving more resettlement. This discrepancy reflects the missing intercept: the cross-sectional comparisons are relative to less resettled areas, not absolute levels, and exclude origin counties from which Chinese squatters were removed.

⁷⁶The expenditure estimates are for spending up to 1954. See Dhu Renick (1965) for details.

⁷⁷Chinese average utility increases by less than their output per capita because they were forcibly relocated to areas with lower (model-implied) amenity, on average. This is unsurprising, given that they did not choose to live in these destinations. In contrast, Malay utility rises by more than their output per capita because they moved to these areas after the Chinese relocation.

⁷⁸Since migration depends only on relative prices across regions, I calculate the minimum (weakly positive)

I interpret the total subsidy received by the resettled population as a lower bound on the utility loss, expressed in dollars. This is a lower bound for two reasons. First, the forcibly resettled likely value their destinations less than the average resident in the county, biasing the required subsidy downward. Second, the calculation omits psychological costs of forced displacement, which are not reflected in the amenity fundamentals.

The required subsidy for the resettled amounts to 17% of baseline output—exceeding the program’s economic gains—suggesting a net welfare loss from the forced resettlement. The result also underscores an intergenerational tradeoff: while younger cohorts benefit, the resettled generation suffers significant welfare losses.

VIII.B Reducing Cross-Ethnic Barriers to Productivity Spillovers

Section V showed that Chinese workers benefited more from incoming resettled Chinese than Malays, likely due to limited cross-ethnic productivity spillovers. This section examines how reducing these barriers affects population distribution, productivity, and welfare.

Figure 6 maps the effects of halving the cross-ethnic barriers to productivity spillovers (γ^e). Panel A shows that Malays move from rural to urban, Chinese-dense areas, while Chinese move in the opposite direction. This reshuffling suggests that strong within-ethnic spillovers partly drive observed ethnic segregation. Panel B shows that Chinese income per capita rises almost everywhere, especially in rural areas where Malays are the majority. Malay income increases more in urban areas, where Chinese are concentrated.

Table 11 summarizes the aggregate effects. Chinese non-agricultural employment falls, while Malay non-agricultural employment rises. Both groups see productivity gains across sectors, as rural Chinese and urban Malays now benefit more from stronger cross-ethnic spillovers. Aggregate output increases by 4.1%, with 89% of the gain coming from productivity improvements. Despite making up only one-third of the population, Chinese contribute nearly as much to the output gain as Malays, underscoring the economic benefits of integration.

On welfare, average Chinese utility rises by 1.9%—less than their 4.4% income gain—as they tend to move to rural areas with lower amenities (Column 1). Malay utility increases by 4.6%, exceeding their 4% income gain, due to improved amenity access in urban destinations (Column 2). Overall, welfare increases by 3.5% (Column 3).

These results suggest that cross-ethnic barriers constrain spatial and sectoral mobility, limiting gains from agglomeration. Policies that reduce these frictions can facilitate urbanization and generate sizable economic and welfare gains.

ad-valorem subsidies needed to achieve this voluntary migration. See Appendix E.1 for details.

VIII.C Wage Subsidies for Malays in Non-Agriculture

Given the strong external economies in non-agriculture, shifting labor from agriculture to non-agriculture can raise aggregate output. A key postwar objective of Malaysia was to integrate Malays into the industrial sector, as they predominantly worked in low-productivity agriculture. This section evaluates a policy that subsidizes Malay wages in non-agriculture.

I simulate an 18% wage subsidy for Malays in non-agriculture, financed by a uniform 7.3% income tax to balance the government budget. The subsidy rate is chosen to equalize Malay and Chinese non-agricultural employment shares, one of the goals of the New Economic Policy (NEP) introduced in the 1970s. Under the NEP, the government promoted Malay participation in industry through credit access, training programs, and higher education quotas.⁷⁹

Figure 7 maps the effects of the subsidy. As non-agriculture becomes more profitable for Malays, they migrate to urban areas and crowd out Chinese workers in the sector, who switch to agriculture and relocate to rural areas (Panel A). Urban Chinese incomes decline due to increased competition from Malay workers, while rural Chinese incomes generally rise as agricultural land becomes more available (Panel B). Malay incomes increase most in major urban centers like Kuala Lumpur, where the Malay population expands (Panel C).

Table 12 summarizes the aggregate effects. The share of Malays working in non-agriculture increases by 9.2 percentage points (17%), while the share of Chinese in non-agriculture declines by 8.1 percentage points (11%). Despite this ethnic reallocation, total non-agricultural employment share rises by 3.3 percentage points (5%).

Malay productivity increases by 3.4%, with a 9% gain in agriculture and a 3.5% decline in non-agriculture (Column 2). As Malays exit agriculture, selection and diminishing returns raise productivity among those who remain. The subsidy draws lower-productivity Malay workers into non-agriculture, but strong external economies limit the decline in average productivity.⁸⁰ For Chinese, the opposite shift—from non-agriculture to agriculture—lowers overall productivity by 2.1% (Column 1). Despite these distortions, aggregate output rises by 1%, driven largely by productivity gains (Column 3).

Welfare effects are mixed. While aggregate welfare rises by 2% (Column 3), the gains accrue entirely to Malays, whose utility increases by 8.2% (Column 2); in contrast, Chinese utility falls by 10% (Column 1). The tax burden is shared, but only Malays benefit from the subsidy. As a result, Malay welfare increases by more than their productivity gain, while Chinese welfare declines by more than their productivity loss. Although controversial,

⁷⁹Following the 1969 racial riots, the NEP sought affirmative action to restructure society and *eliminate the identification of race with economic function*. See Koon (1997) and Jomo (2017) for further discussion.

⁸⁰Income per capita and productivity (output per capita) no longer coincide because income is affected by subsidies, which distort Malays' occupational choices.

industrial policies that target disadvantaged groups in sectors with strong agglomeration forces can raise overall output and reduce inequality.

IX Conclusion

This paper studies how social divisions shape the gains from agglomeration, leveraging a large-scale, ethnic-based resettlement program in 1950s British Malaya. The program forcibly relocated Chinese populations into villages, reshaping both economic and social structures. Despite its coercive nature—which likely limited the benefits relative to a voluntary scenario—the program generated productivity gains in receiving counties, spurring industrialization and greater division of labor. These gains, however, were unequally distributed across ethnic groups.

Local effects of resettlement were mediated by agglomeration externalities that varied by sector and ethnic composition. The influx of industrial labor shifted employment out of agriculture, driven by strong external economies in non-agriculture and diminishing returns in land-constrained agriculture. Denser labor markets promoted specialization and education, but segregation and cross-ethnic frictions limited spillovers to non-Chinese populations. As a result, income gains in more resettled counties accrued primarily to the Chinese community.

To assess aggregate effects, I develop and estimate a quantitative spatial general equilibrium model that incorporates heterogeneous agglomeration forces, migration, and occupational choices. Resettlement raised output by reallocating labor from remote, less productive areas to regions with better market access and industrial potential. However, the program ultimately reduced welfare by disregarding people’s location preferences.

While this paper focuses on ethnic divisions, similar frictions can arise along other social lines, such as caste, culture, religion, and gender. In India, for example, caste norms have long constrained intergroup interactions, potentially limiting agglomeration spillovers and slowing structural transformation. These findings also speak to refugee and migrant resettlement policies, where social integration plays a key role in realizing the economic gains from migration.

References

- Abel, Martin.** 2019. “Long-Run Effects of Forced Resettlement: Evidence from Apartheid South Africa.” *The Journal of Economic History*, 79(4): 915–953.
- 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.
- Alesina, Alberto, and Eliana La Ferrara.** 2002. “Who trusts others?” *Journal of Public Economics*, 85(2): 207–234.
- Alesina, Alberto, and Eliana La Ferrara.** 2005. “Ethnic Diversity and Economic Performance.” *Journal of Economic Literature*, 43(3): 762–800.
- Alesina, Alberto, Reza Baqir, and William Easterly.** 1999. “Public Goods and Ethnic Divisions.” *The Quarterly Journal of Economics*, 114(4): 1243–1284.
- Allen, Treb, and Dave Donaldson.** 2022. “Persistence and Path Dependence in the Spatial Economy.” National Bureau of Economic Research Working Paper 28059.
- Allen, Treb, Costas Arkolakis, and Xiangliang Li.** 2024. “On the Equilibrium Properties of Spatial Models.” *American Economic Review: Insights*, 6(4): 472—89.
- Almagro, Milena, and Tomás Domínguez-Iino.** 2024. “Location Sorting and Endogenous Amenities: Evidence from Amsterdam.” National Bureau of Economic Research Working Paper 32304.
- Amsden, Alice H.** 1992. *Asia’s Next Giant: South Korea and Late Industrialization*. Oxford University Press.
- 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.
- Ananat, Elizabeth, Shihe Fu, and Stephen L. Ross.** 2013. “Race-Specific Agglomeration Economies: Social Distance and the Black-White Wage Gap.” National Bureau of Economic Research Working Paper 18933.
- 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.” *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.
- Ashraf, Quamrul, and Oded Galor.** 2013. “The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development.” *American Economic Review*, 103(1): 1–46.
- Baillargeon, David.** 2021. “Spaces of Occupation: Colonial Enclosure and Confinement in British Malaya.” *Journal of Historical Geography*, 73: 24–35.
- Balboni, Clare.** 2025. “In Harm’s Way? Infrastructure Investments and the Persistence of Coastal Cities.” *American Economic Review*, 115(1): 77—116.
- Baum-Snow, Nathaniel, and Ronni Pavan.** 2013. “Inequality and City Size.” *Review of Economics and Statistics*, 95(5): 1535–1548.
- Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan.** 2018. “Why Has Urban Inequality Increased?” *American Economic Journal: Applied Economics*, 10(4): 1–42.
- 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

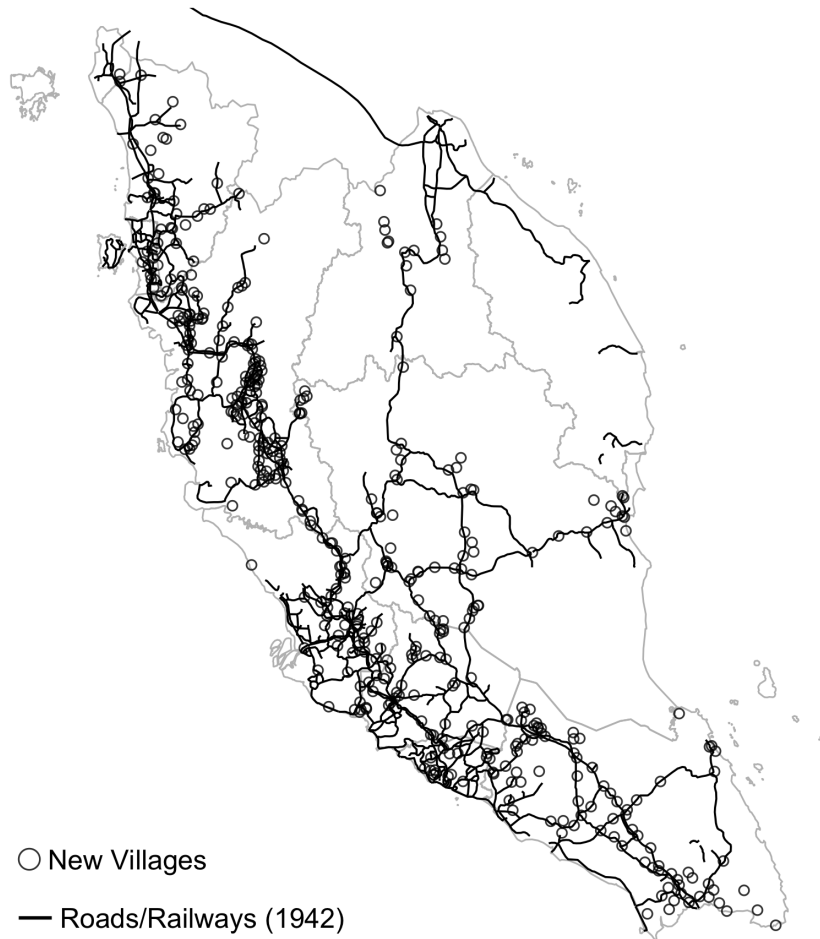
- in Diversity? How Intergroup Contact Can Foster Nation Building.” *American Economic Review*, 109(11): 3978–4025.
- Beaman, Lori A.** 2011. “Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.” *The Review of Economic Studies*, 79(1): 128–161.
- Becker, Sascha O.** 2022. “Forced Displacement in History: Some Recent Research.” *Australian Economic History Review*, 62(1): 2–25.
- 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 Mueller, and Philip Osafo-Kwaako.** 2024. “State Building in a Diverse Society.” *The Review of Economic Studies*. Forthcoming.
- 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.** 2025. “The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility.” *American Economic Review*. Forthcoming.
- 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.
- Chyn, Eric, Robert Collinson, and Danielle H Sandler.** 2025. “The Long-Run Effects of America’s Largest Residential Racial Desegregation Program: Gautreaux.” *The Quarterly Journal of Economics*. Forthcoming.
- Ciccone, Antonio, and Jan Nimczik.** 2024. “The Long-Run Effects of Immigration: Evidence Across a Barrier to Refugee Settlement.” Working Paper.
- 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*.
- 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.
- Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum.** 2022. "Matching in Cities." *Journal of the European Economic Association*, 20(4): 1478–1521.
- 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.
- 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.
- Dell, Melissa, Nathan Lane, and Pablo Querubin.** 2018. "The Historical State, Local Collective Action, and Economic Development in Vietnam." *Econometrica*, 86(6): 2083–2121.
- 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.
- 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 *Cities and Geography*. Vol. 4 of *Handbook of Regional and Urban Economics*, , ed. J. Vernon Henderson and Jacques-François Thisse, 2063–2117. Elsevier.
- Easterly, William, and Ross Levine.** 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics*, 112(4): 1203–1250.
- 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.
- Esteban, Joan, Laura Mayoral, and Debraj Ray.** 2012. "Ethnicity and Conflict: An Empirical Study." *American Economic Review*, 102(4): 1310—42.
- Fajgelbaum, Pablo D, and Cecile Gaubert.** 2020. "Optimal Spatial Policies, Geography, and Sorting." *The Quarterly Journal of Economics*, 135(2): 959–1036.
- Fell, H.** 1960. "Population Census of the Federation of Malaya." *Kuala Lumpur: Department of Statistics*.
- Galor, Oded, and Ömer Özak.** 2016. "The Agricultural Origins of Time Preference." *American Economic Review*, 106(10): 3064–3103.
- Ginsburg, Norton Sydney.** 1958. *Malaya. Publications of the American Ethnological Society*, Seattle,: University of Washington Press.
- 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.
- 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.
- Guarnieri, Eleonora.** 2025. "Cultural Distance and Ethnic Civil Conflict." *American Economic*

- Review*, 115(4): 1338–68.
- Haggard, Stephan.** 1990. *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries*. Ithaca, NY: 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.
- Henderson, Vernon, Ari Kuncoro, and Matt Turner.** 1995. “Industrial Development in Cities.” *Journal of Political Economy*, 103(5): 1067–1090.
- Hirschman, Charles.** 1986. “The Making of Race in Colonial Malaya: Political Economy and Racial Ideology.” Vol. 1, 330–361, Springer.
- Hjort, Jonas.** 2014. “Ethnic Divisions and Production in Firms.” *The Quarterly Journal of Economics*, 129(4): 1899–1946.
- 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.
- Jacobs, Jane.** 1969. *The Economy of Cities*. New York: Random House.
- 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.
- Khoo, Teik Huat,** ed. 1983. *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.
- Kok, Chun Chee, Gedeon J. Lim, Danial Shariat, Abu Siddique, and Shunsuke Tsuda.** 2025. “Interethnic Proximity and Political Development.” IZA Institute of Labor Economics IZA Discussion Paper 17776.
- Koon, Heng Pek.** 1997. “The New Economic Policy and the Chinese Community in Peninsular Malaysia.” *The Developing Economies*, 35(3): 262–292.
- Lane, Nathan.** 2025. “Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea.” *The Quarterly Journal of Economics*. Forthcoming.
- 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.
- Miguel, Edward, and Mary Kay Gugerty.** 2005. “Ethnic Diversity, Social Sanctions, and Public Goods in Kenya.” *Journal of Public Economics*, 89(11): 2325–2368.
- Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg.** 2018. “Commuting, Migration, and Local Employment Elasticities.” *American Economic Review*, 108(12): 3855–90.
- 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.
- Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson.** 2022. "The Gift of Moving: Intergenerational Consequences of a Mobility Shock." *The Review of Economic Studies*, 89(3): 1557–1592.
- 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.
- Rosenthal, Stuart S, and William C Strange.** 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*. Vol. 4, 2119–2171. Elsevier.
- Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman.** 2019. "Cognitive Hubs and Spatial Redistribution." National Bureau of Economic Research Working Paper 26267.
- Roy, Andrew Donald.** 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146.
- 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, J. M. C. Santos, and Silvana Tenreyro.** 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.
- Smith, Cory, and Amrita Kulka.** 2024. "Agglomeration Over the Long Run: Evidence from County Seat Wars." Working Paper.
- Strauch, Judith.** 1981. *Chinese Village Politics in the Malaysian State*. Harvard University Press.
- Sunderland, Riley.** 1964. *Resettlement and Food Control in Malaya*. Santa Monica, CA: RAND Corporation.
- Tombe, Trevor, and Xiaodong Zhu.** 2019. "Trade, Migration, and Productivity: A Quantitative Analysis of China." *American Economic Review*, 109(5): 1843–1872.
- U.S. Office of Strategic Services.** 1942. *Roads and Railroads of Malaya*. . Provisional ed., Washington: OSS Repro. Section.
- 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.** 2024. "Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities." Working Paper.

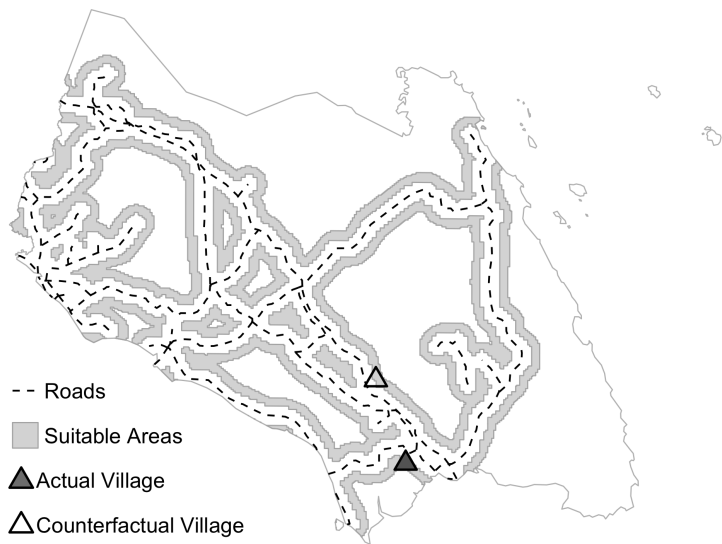
Figure 1. The New Villages and Transportation Network



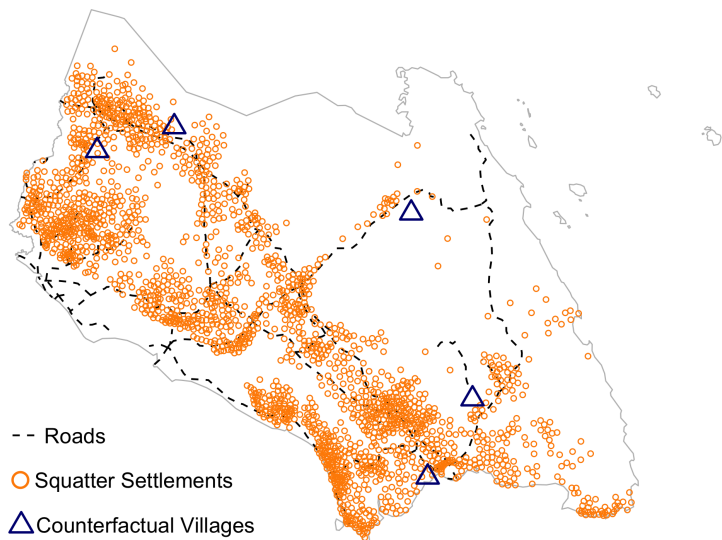
Notes: This figure shows the location of the New Villages (circles) and the roads and railway network in 1942 (lines). Gray polygons indicate state boundaries. New Village data are from the Corry report; road and railway data from the G8031 road map series (U.S. Office of Strategic Services, 1942).

Figure 2. Counterfactual Site Selection and Relocation

Panel A. Counterfactual Site Selection



Panel B. Counterfactual Relocation



Notes: This figure illustrates counterfactual site selection and relocation in state Johor. Panel A shows the selection of counterfactual sites: the solid triangle marks an actual New Village; dashed lines indicate the road and rail network; gray shaded areas represent regions equidistant from the actual village and equally suitable for resettlement. The hollow triangle denotes a counterfactual village location, randomly drawn from these suitable areas. Panel B shows the relocation of squatters to the counterfactual sites, with orange circles indicating initial squatter settlements.

Figure 3. County Resettlement Density, Expected and Residualized

Panel A. County Resettlement Density,
Expected

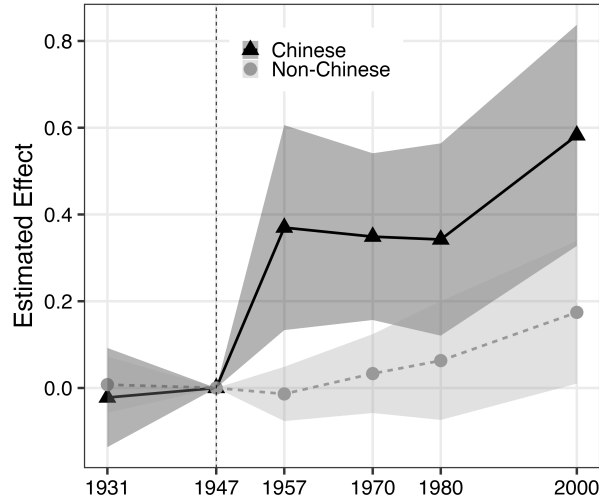
Panel B. County Resettlement Density,
Residualized



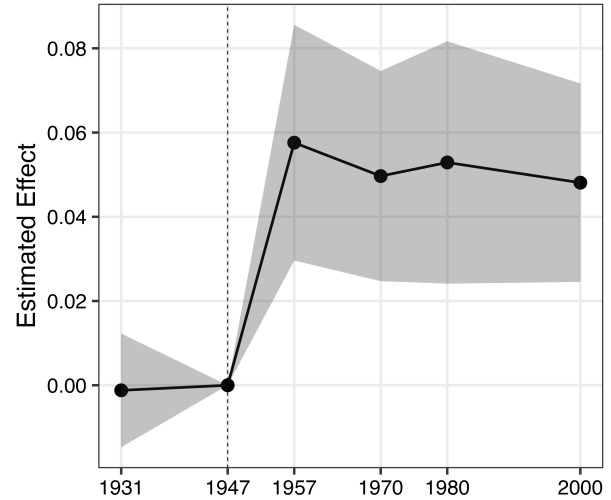
Notes: This figure maps expected and residualized county resettlement density for the 249 counties with at least one New Village. Darker shades indicate higher deciles of resettlement density. White bubbles denote New Villages, with sizes proportional to the resettled population. Panel A shows expected resettlement density, calculated using Equation (A-1). Panel B shows residualized resettlement density, controlling for state fixed effects and the main controls: expected resettlement density; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population in 1947; and the shares of land used for rubber and mining in 1943. Resettlement data 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



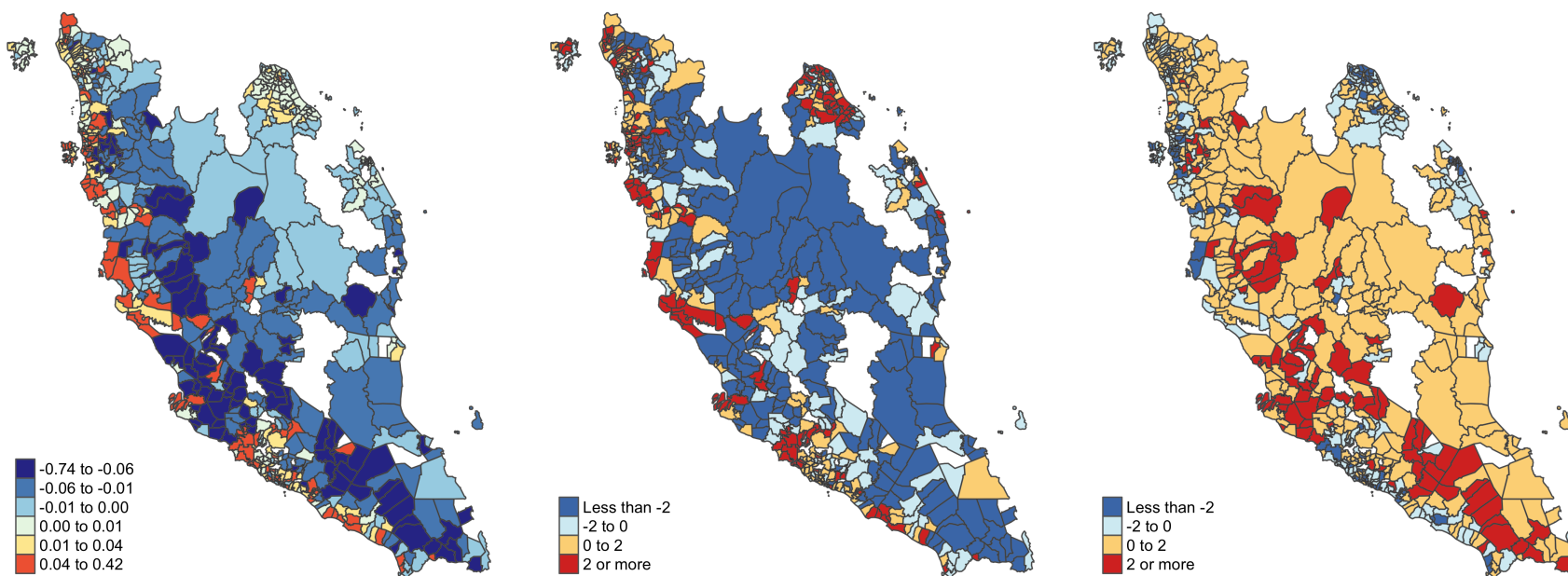
Notes: This figure shows regression estimates of county resettlement density on population growth by ethnic group (Panel A) and changes in Chinese population share (Panel B) from 1931 to 2000. All regressions include state fixed effects and the main controls: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population in 1947; and the shares of land used for rubber and mining in 1943. Shaded areas indicate 95% confidence interval based on Conley standard errors with a 30-kilometer distance cutoff.

Figure 5. Distributional Effects of Emergency Resettlement

Panel A. Change in Chinese
Population Share

Panel B. Percent Change in Chinese
Income Per Capita

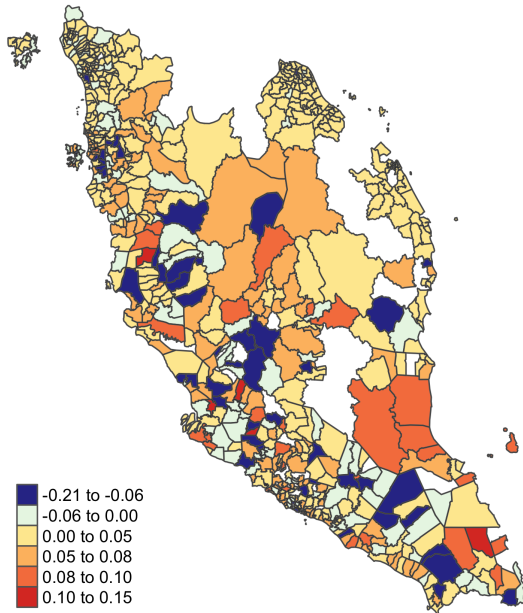
Panel C. Percent Change in Malay
Income Per Capita



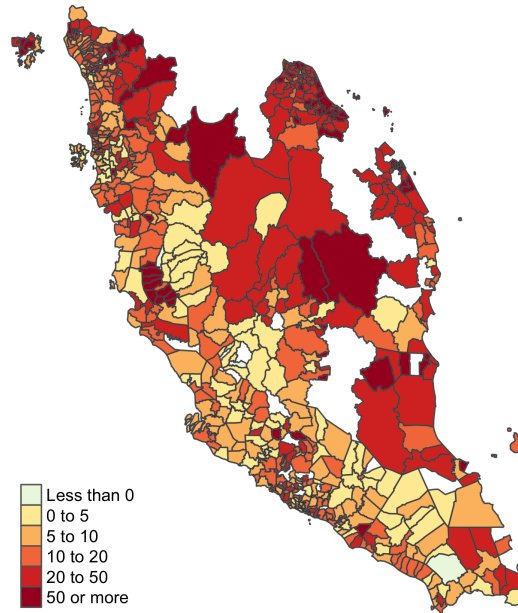
Notes: This figure shows changes in county outcomes from the “no resettlement” baseline to the observed 1980 economy. The no-resettlement baseline simulates a counterfactual 1980 equilibrium starting from the 1947 population distribution instead of the post-resettlement 1957 distribution. Panel A shows the change in Chinese population share. Panel B shows the percent change in Chinese income per capita. Panel C shows the percent change in Malay income per capita.

Figure 6. Distributional Effects of Reducing Cross-Ethnic Barriers to Productivity Spillovers

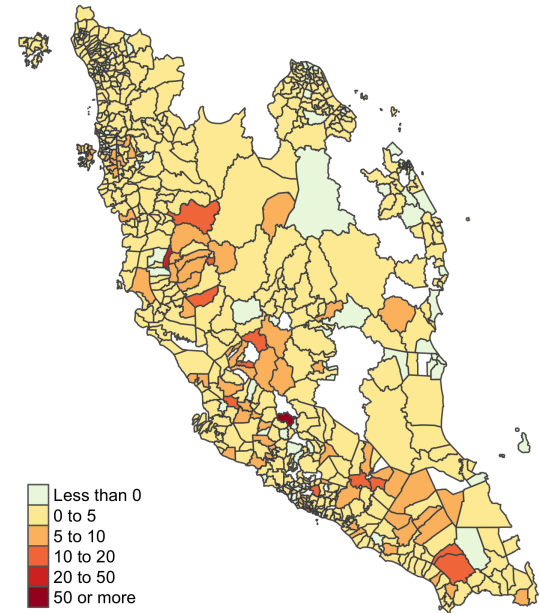
Panel A. Change in Chinese Population Share



Panel B. Percent Change in Chinese Income Per Capita



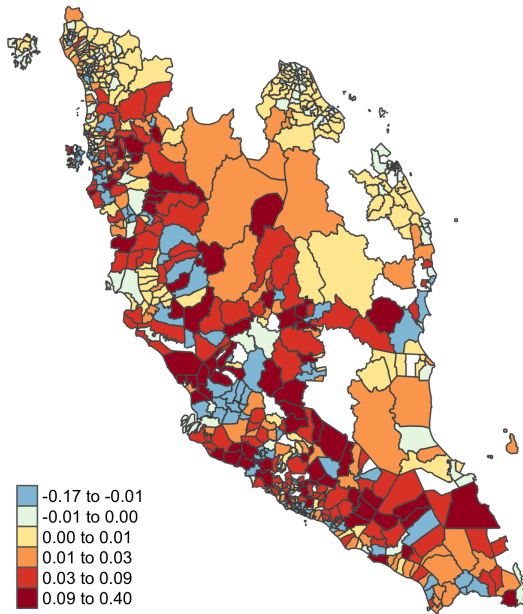
Panel C. Percent Change in Malay Income Per Capita



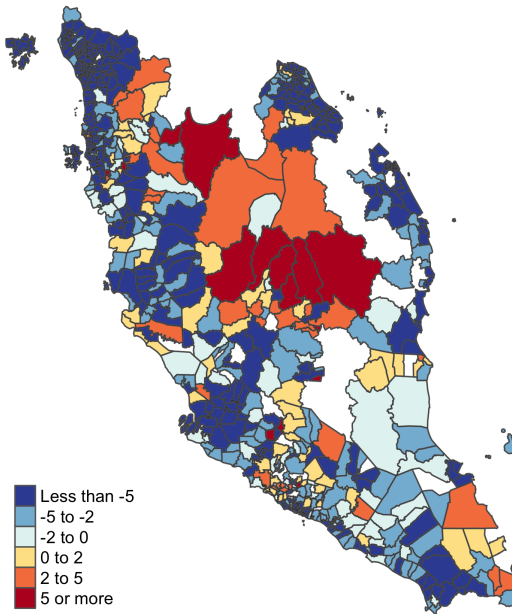
Notes: This figure shows changes in county outcomes from the baseline economy to a counterfactual equilibrium in which the barrier to cross-ethnic productivity spillovers (γ^e) is reduced by half. Panel A shows the change in the Chinese population share. Panel B shows the percent change in Chinese income per capita. Panel C shows the percent change in Malay income per capita.

Figure 7. Distributional Effects of Wage Subsidies for Malays in Manufacturing

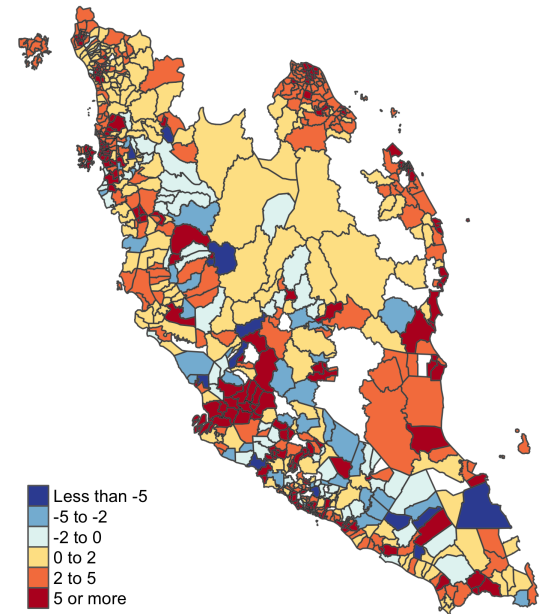
Panel A. Change in Chinese
Population Share



Panel B. Percent Change in Chinese
Income Per Capita



Panel C. Percent Change in Malay
Income Per Capita



Notes: This figure shows changes in county outcomes from the baseline economy to a counterfactual equilibrium with an 18% wage subsidy for Malays in non-agriculture, financed by a uniform income tax. Panel A shows the change in Chinese population share. Panel B shows the percent change in Chinese income per capita. Panel C shows the percent change in Malay income per capita. All income measures are net of subsidies and taxes.

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 (0.12)	3.86 (3.19)	−0.04 (0.01)	−0.01 (0.01)	−0.46 (0.42)	−0.35 (0.34)	−1.60 (0.69)	0.00 (1.37)	0.07 (0.02)	0.01 (0.01)	−0.97 (0.85)	−5.79 (1.67)
Panel B. Baseline Controls												
Higher Resettlement	0.07 (0.16)	−2.36 (4.92)	−0.05 (0.02)	−0.01 (0.02)	0.40 (0.49)	0.47 (0.41)	0.55 (0.93)	3.18 (1.96)	0.05 (0.01)	0.02 (0.01)	−0.69 (1.01)	−1.05 (1.90)
Panel C: Expected Resettlement												
Higher Resettlement	0.16 (0.27)	−7.35 (5.84)	−0.05 (0.03)	−0.02 (0.03)	0.43 (0.61)	0.74 (0.54)	0.29 (0.95)	2.09 (2.21)	0.03 (0.03)	0.01 (0.01)	−0.99 (1.10)	0.52 (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 examine geographic characteristics: elevation (Column 1), ruggedness (Column 2), padi rice suitability (Column 3), and coconut (Column 4). Columns 5–8 examine access to amenities as of 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 examine pre-period economic activity: 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 (Column 12). Panel A include state fixed effects. Panel B adds baseline controls (excluding land use shares for rubber and mining): an indicator for any resettlement; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; and log population in 1947. Panel C additionally controls for expected resettlement density. The unit of observation is the county. See Appendix Table A.1 for the data sources. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

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

	1957 (1)	1980 (2)	2000 (3)
Panel A. Log Total Population			
Higher Resettlement	0.094 (0.034)	0.109 (0.062)	0.179 (0.075)
Panel B. Chinese Population Share			
Higher Resettlement	0.048 (0.012)	0.050 (0.011)	0.041 (0.011)
# Counties	777	777	777

Notes: This table shows the relationship between county resettlement density and population distribution from 1957 to 2000. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A reports effects on log total population in 1957 (Column 1), 1980 (Column 2), and 2000 (Column 3). Panel B reports effects on Chinese population share in these years. 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the county. Data from the tabulated Census of Population. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

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

	Agriculture (1)	Non- Agriculture (2)	Difference (2) – (1) (3)
Panel A. Total Employment			
Higher Resettlement	0.111 (0.037)	0.291 (0.129)	0.180 (0.139)
# County-Years	1,554	1,554	
Panel B. Chinese Employment			
Higher Resettlement	0.272 (0.059)	0.351 (0.180)	0.079 (0.176)
# County-Years	1,516	1,502	
Panel C. Non-Chinese Employment			
Higher Resettlement	–0.006 (0.045)	0.244 (0.107)	0.250 (0.124)
# County-Years	1,516	1,502	

Notes: This table shows the relationship between county resettlement density and sectoral employment in 1980–1991. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A reports effects on total employment in the agricultural sector (“Agriculture”, Column 1), the non-agricultural sector (“Non-Agriculture”, Column 2), and the difference between the two estimates (Column 3). Panels B and C report effects on Chinese and non-Chinese employment, respectively. The agricultural sector includes agriculture, hunting, forestry, fishing, mining, and quarrying. The non-agricultural sector includes 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-by-year fixed effects and the main controls interacted with year: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the county-year. Data from the Census of Population in 1980 and 1991. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table 4. 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	0.110 (0.049)	0.039 (0.031)	0.070 (0.034)
# Households	9,634	20,549	
Panel B. Log Earnings, Agriculture			
Higher Resettlement	0.067 (0.041)	–0.006 (0.037)	0.073 (0.042)
# Households	2,197	9,359	
Panel C. Log Earnings, Non-Agriculture			
Higher Resettlement	0.125 (0.044)	0.051 (0.030)	0.074 (0.028)
# Households	7,437	11,190	

Notes: This table shows the relationship between county resettlement density and household income in 1980. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A reports effects on log household earnings for Chinese households (Column 1), non-Chinese households (Column 2), and the difference between the two estimates (Column 3). Panel B restricts the sample to households whose head is employed in the agricultural sector (agriculture and mining). Panel C restricts the sample to households whose head is employed in the non-agricultural sector. 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the household. Data from the 2% individual-level Census of Population microdata in 1980. Conley standard errors with a 30-kilometer distance cutoff 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 (0.008)	0.002 (0.007)	0.014 (0.010)
Mean of Outcome	0.578	0.569	
# Counties	524	745	
Panel B. Industry Specialization Index			
Higher Resettlement	0.017 (0.008)	–0.006 (0.009)	0.023 (0.009)
Mean of Outcome	0.306	0.266	
# Counties	752	776	
Panel C. Occupation Specialization Index			
Higher Resettlement	0.016 (0.008)	–0.006 (0.008)	0.022 (0.009)
Mean of Outcome	0.255	0.257	
# Counties	752	776	
Panel D. Share Employed in Managerial Occupations			
Higher Resettlement	0.012 (0.006)	–0.001 (0.000)	0.012 (0.006)
Mean of Outcome	0.056	0.025	
# Counties	752	776	

Notes: This table shows the relationship between county resettlement density and labor market outcomes in 1980–1991. "Higher Resettlement" is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel reports effects on a different labor market outcome: labor force participation rate in 1980 (Panel A); concentration of employment across industries in 1991, measured by the Herfindahl-Hirschman Index (HHI) (Panel B); concentration of employment across occupations in 1991, also measured by the HHI (Panel C); and the share of workers employed in managerial occupations (legislators, senior officials, and managers) in 1991 (Panel D). Column 1 shows estimates for Chinese individuals, Column 2 for non-Chinese individuals, and Column 3 the difference between the two. 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the individual for Panel A and the county for Panels B and C. Data from the 2% Census of Population microdata in 1980 and tabulations in 1991. Conley standard errors with a 30-kilometer distance cutoff 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	0.434 (0.231)	0.106 (0.123)	0.328 (0.161)
Mean of Outcome	5.39	5.12	
Panel B. Primary Education			
Higher Resettlement	0.037 (0.018)	0.018 (0.012)	0.019 (0.010)
Mean of Outcome	0.56	0.51	
Panel C. Secondary Education			
Higher Resettlement	0.041 (0.023)	0.013 (0.012)	0.028 (0.018)
Mean of Outcome	0.29	0.26	
# Individuals	31,507	57,345	

Notes: This table shows the relationship between county resettlement density and educational attainment in 1980. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel reports effects on a different education outcome: years of schooling (Panel A); completion of primary education (Panel B); and completion of secondary education (Panel C). Column 1 shows estimates for Chinese households, Column 2 for non-Chinese households, and Column 3 the difference between the two. 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the individual. The sample includes individuals aged 20 or above from the 2% Census of Population microdata in 1980. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table 7. Manufacturing Activity in 1970, by County Resettlement Density and Pre-Period Industry Share of Chinese Employment

	Number of Manufacturing Establishments (1)	Share of Employer Establishments (2)
Higher Resettlement	0.022 (0.125)	0.001 (0.020)
Higher Resettlement $\times$ Chinese Industries	0.217 (0.082)	0.020 (0.009)
# County-Industries	15,540	2,142

Notes: This table shows how the relationship between county resettlement density and manufacturing activity in 1970 varies by industries with high pre-period Chinese employment. “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 where more than 80% of employment in 1947 was Chinese. These include all manufacturing industries except food products, wood products, textiles, and other miscellaneous manufacturing (see Appendix Figure A.7). Column 1 reports the effect on the number of manufacturing establishments, estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator. Column 2 reports OLS estimates for the share of establishments with at least one full-time employee, weighted by the number of establishments in the county-industry. All regressions include 2-digit industry fixed effects, 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the county-industry. Data from the Directory of Manufacturing in 1970 and the tabulated Population Census in 1947. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table 8. Firm Ownership and Revenue in 2011–2015, by County Resettlement Density

	Chinese Owned Firms (1)	Non-Chinese Owned Firms (2)	Difference (1) – (2) (3)
Panel A. Number of Firms			
Higher Resettlement	0.518 (0.301)	0.306 (0.097)	0.212 (0.128)
# County-Industry-Years	93,240	93,240	
Panel B. Log of Average Firm Revenue			
Higher Resettlement	0.189 (0.078)	0.082 (0.104)	0.107 (0.126)
# County-Industry-Years	12,608	9,468	

Notes: This table shows the relationship between county resettlement density and firm activity in 2011–2015. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A reports effects on the number of firms, estimated using Poisson pseudo-maximum-likelihood (PPML). Panel B reports OLS estimates of log average firm revenue, weighted by the number of establishments by ownership in the county-industry-year. Column 1 shows estimates for Chinese-owned firms, Column 2 for non-Chinese owned firms, and Column 3 the difference between the two. Firm ownership is based on the ethnicity of ultimate owners, as described in Appendix A.2. All regressions include 2-digit NAICS industry-by-year fixed effects and their interactions with 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1943. The unit of observation is the county-industry-year. Data from the Orbis Historical Disk. Conley standard errors with a 30-kilometer distance cutoff 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	0.517	(0.003)
α	Expenditure share on agriculture	0.310	(0.001)
θ	Skill dispersion	3.327	(0.035)
γ_A	Productivity spillover w.r.t. size, agr.	-0.120	(0.067)
γ_M	Productivity spillover w.r.t. size, non-agr.	0.223	(0.052)
γ^e	Productivity spillover w.r.t. ethnic share	0.132	(0.039)
β	Amenity spillover w.r.t. size	-0.005	(0.044)
β^e	Amenity spillover w.r.t. ethnic share	0.125	(0.091)
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 reports parameter estimates from the model. Panel A reports parameters estimated from the data, as described in Section VII. Panel B reports three parameters that are assumed or calibrated using external moments: σ and ν from Balboni (2025), and ξ from Monte, Redding and Rossi-Hansberg (2018). Column 1 lists the parameter symbols, Column 2 provides descriptions, Column 3 reports estimates, and Column 4 shows bootstrap standard errors in parentheses.

Table 10. Aggregate Impact of the Emergency Resettlement

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Non-Agricultural Employment Share	1.03	−0.95	−0.28
Output per Capita	1.54	2.37	2.01
For Agriculture	−0.13	1.28	0.38
For Non-Agriculture	1.61	3.58	2.84
Aggregate Output (% Baseline Output)	0.67	1.34	2.01
From Reallocation of Labor	0.45	1.03	1.48
From Changes in Productivity	−0.19	0.08	−0.10
From Joint Changes	0.40	0.23	0.63
Average Utility	1.21	6.68	4.82

Notes: This table shows changes in economic outcomes from the “no resettlement” baseline to the observed 1980 economy. The baseline equilibrium uses the 1947 population distribution as the initial condition, as opposed to the post-resettlement 1957 distribution. Results are reported for Chinese (Column 1), Malays (Column 2), and the overall economy (Column 3). The first panel reports percentage point changes in the share of non-agricultural employment. The second panel reports percent changes in output per capita—overall and by sector. The third panel reports changes in aggregate output as a share of baseline output, decomposed into contributions from labor reallocation, productivity changes, and their interaction. The final panel reports percent changes in average utility.

Table 11. Aggregate Impact of Reducing Cross-Ethnic Frictions

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Non-Agricultural Employment Share	−1.72	0.81	−0.05
Output per Capita	4.35	3.99	4.14
For Agriculture	4.52	2.70	3.98
For Non-Agriculture	4.89	4.22	4.24
Aggregate Output (% Baseline Output)	1.89	2.26	4.14
From Reallocation of Labor	−0.50	0.52	0.02
From Changes in Productivity	2.09	1.58	3.67
From Joint Changes	0.29	0.16	0.45
Average Utility	1.24	4.63	3.47

Notes: This table shows changes in economic outcomes from the baseline economy to a counterfactual equilibrium where the cross-ethnic barrier to productivity spillovers γ^e is reduced by half. The baseline equilibrium uses the 1947 population distribution as the initial condition, as opposed to the post-resettlement 1957 distribution. Results are reported for Chinese (Column 1), Malays (Column 2), and the overall economy (Column 3). The first panel reports percentage point changes in the share of non-agricultural employment. The second panel reports percent changes in output per capita—overall and by sector. The third panel reports changes in aggregate output as a share of baseline output, decomposed into contributions from labor reallocation, productivity changes, and their interaction. The final panel reports percent changes in average utility.

Table 12. Aggregate Impact of Wage Subsidies for Malays in Non-Agriculture

	Chinese (1)	Malays (2)	Total (3)
Changes in Outcomes Relative to Baseline:			
Non-Agricultural Employment Share	−8.14	9.17	3.29
Output per Capita	−2.12	3.36	0.98
For Agriculture	1.39	9.07	10.67
For Non-Agriculture	−0.60	−3.46	−4.20
Aggregate Output (% Baseline Output)	−0.92	1.90	0.98
From Reallocation of Labor	−0.86	1.21	0.35
From Changes in Productivity	−0.20	1.30	1.10
From Joint Changes	0.14	−0.60	−0.46
Average Utility	−10.03	8.24	2.03

Notes: This table shows changes in economic outcomes from the baseline economy to a counterfactual equilibrium with an 18% wage subsidy for Malays in non-agriculture. The baseline equilibrium uses the 1947 population distribution as the initial condition, as opposed to the post-resettlement 1957 distribution. Results are reported for Chinese (Column 1), Malays (Column 2), and the overall economy (Column 3). The first panel reports percentage point changes in the share of non-agricultural employment. The second panel reports percent changes in output per capita—overall and by sector. The third panel reports changes in aggregate output as a share of baseline output, decomposed into contributions from labor reallocation, productivity changes, and their interaction. The final panel reports percent changes in average utility.

Online Appendices

Coercive Growth: Forced Resettlement and Ethnicity-Based Agglomeration

Shanon Hsuan-Ming Hsu
University of Chicago

A Data Appendix	1
A.1 Other Supplemental Data	1
A.2 Assigning Firm Ownership by Ethnicity in Orbis Data	1
A.3 Data Imputation for Structural Estimation	2
B Empirical Results	2
B.1 Identification Assumptions	2
B.2 Supplementary Analysis	4
B.3 Robustness	5
C Theoretical Results	6
C.1 Sectoral Labor Supply	6
C.2 Migration	8
C.3 Iterative Procedure for Solving the Equilibrium	8
C.4 Sufficient Conditions for Uniqueness of Equilibrium	11
D Structural Estimation	13
D.1 Market Access Terms	13
D.2 Migration Cost Elasticity	15
D.3 Parameter Estimates	16
E Counterfactuals	18
E.1 Lower Bound for Utility Loss from Forced Resettlement	18
F Appendix Figures	23
G Appendix Tables	37

A Data Appendix

This section describes data sources and construction details not covered in the main text. Appendix Table A.1 lists each data source along with the variables extracted.

A.1 Other Supplemental Data

Schools. I collect data on Chinese vernacular and other national schools from Malaysia’s Ministry of Education. The dataset includes all primary and secondary schools in 2022, with information on school names, number of teachers and students, and geographic 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 geocode using the provided addresses. I use the location of Chinese vernacular schools as a proxy for Chinese settlements and that of other national schools for non-Chinese settlements to construct a measure of geographic segregation.

Buildings. I use data from the Global Human Settlement Layer (GHSL) to measure built-up volumes starting in 1975. The data combine surface and height measurements at 100-meter resolution, derived from Sentinel-2 and Landsat satellite imagery.

A.2 Assigning Firm Ownership by Ethnicity in Orbis Data

This section describes how I classify firms in the Orbis database as Chinese-owned or non-Chinese.

Orbis provides annual ownership “Links” datasets that report direct and indirect ownership shares at the subsidiary-shareholder level (see Bureau van Dijk (2019) for details). I trace each firm’s ownership structure from 2001 to 2021 to identify its ultimate owner, using the global ultimate owner (GUO) indicators: GUO50 and GUO25, which flag entities holding more than 50% or 25% control, respectively. I define the ultimate owner as follows:

- (i). If any GUO with at least 50% ownership is listed, I use that entity.
- (ii). If no GUO exceeds 50% but some exceeds 25%, I use the 25%-plus owner(s).
- (iii). If no GUO is listed, I use the largest immediate shareholder with over 25% ownership.⁸¹

I classify an ultimate owner as a natural person if Orbis lists its entity type as “I” (individual or family). Some firms lack a person owner in certain years due to dispersed ownership or missing data. To reduce measurement error, I use each firm’s full ownership history from 2001 to 2021. I classify a firm as Chinese-owned if any of its ultimate person owners at any

⁸¹Ownership thresholds in the 20–25% range are commonly used to define control (La Porta, Lopez-de Silanes and Shleifer, 1999; Claessens, Djankov and Lang, 2000).

point during this period has a name containing any Chinese characters (e.g., “Lee,” “Tan,” “Lim,” “Chen”).

A.3 Data Imputation for Structural Estimation

Because the 1980 population census microdata is a 2% sample, some counties lack ethnicity-by-sector income data even though the sample is representative. To ensure sufficient geographic coverage for structural estimation, I impute missing average incomes using the following two-step procedure.

First, if a county-sector-ethnicity cell is missing but the corresponding district-sector-ethnicity cell is observed, I impute income using the district-level average. This assumes that the average income for an ethnic group in a sector is similar across counties within the same district (districts average about ten counties). Appendix Figure A.10 supports this assumption: household earnings for an ethnic group in a sector are centered around the district average.

Second, if a county-sector-ethnicity cell is missing and no district-level value is available, but income data for the same county and ethnicity in the other sector exist, I impute income using that sector’s average. This assumes sectoral earnings are similar within a county-ethnicity group, consistent with the structural model’s assumption of Fréchet-distributed productivity. Appendix Figure A.11 shows that household earnings in both sectors are centered around the same county-ethnicity mean.

B Empirical Results

This section first formalizes the identification assumptions outlined in Section IV, then presents supplementary analyses and robustness checks.

B.1 Identification Assumptions

To illustrate the assumptions, I express county resettlement density in terms of 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 $i = 1, \dots, I$ indexes resettlement “sites,” each potentially hosting at most one New Village, with a total of I sites in the state. The indicator g_{1i} denotes whether site i was selected for resettlement, and g_{2i} is the number of people resettled to that site. $area_c$ denotes the area of county c .⁸²

⁸²The variables g_{1i} and g_{2i} are interdependent, as no resettlement occurs at a site unless it was selected ($g_{1i} = 0 \Rightarrow g_{2i} = 0$).

Based on historical accounts, I make two assumptions regarding the selection of sites ($g_1 \equiv \{g_{1i}\}_{i=1}^I$) and the relocation of population to sites ($g_2 \equiv \{g_{2i}\}_{i=1}^I$). First, I assume that site selection is orthogonal to county-level unobservables $\varepsilon \equiv \{\varepsilon_c\}_{c=1}^C$ (where C is the number of counties), conditional on a vector of site-level characteristics $w_1 \equiv \{w_{1i}\}_{i=1}^I$, including distance to the transportation network, land-use type, and the decile of the county's squatter population. Second, conditional on selected sites g_1 and a vector of characteristics $w_2 \equiv \{w_{2i}\}_{i=1}^I$ —including distances to initial squatter settlements and their populations—the number of people resettled to each site is also 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.⁸³
- (ii) (Number resettled) $g_2 \perp\!\!\!\perp \varepsilon \mid (g_1, w_2)$: Conditional on selected sites and the initial squatter distribution, the number resettled to a site was exogenous.

Under Assumption 1, the potential source of omitted variable bias is the conditional expectation of resettlement density $\mathbb{E}[f_c(g_1, g_2) \mid w]$, where $w \equiv (w_1, w_2)$. As shown by Borusyak and Hull (2023), the coefficient β is identified if this expected density is either controlled for or used to re-center the realized resettlement density.

To measure $\mathbb{E}[f_c(g_1, g_2) \mid w]$, I leverage institutional knowledge and impose two additional assumptions on 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) $G_1(g_1 \mid w_1)$ is uniform: All 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) $\mathbb{E}[f_c(g_1, g_2) \mid g_1, w] = f_c(g_1, \bar{g}_2(g_1, w))$: Conditional on selected village sites and the initial distribution of Chinese squatters, the average relocation followed a gravity-based model, where

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

Here, n_j is the population of Chinese squatters at origin j , d_{ji} is the distance between j and site i , and ψ is the elasticity of relocation costs with respect to distance.

⁸³A weaker assumption of mean independence between g_1 and ε , conditional on w , suffices for identification.

Assumption 2(i) states that the British considered observationally similar sites as equally suitable, while 2(ii) captures their stated objective of minimizing dislocation, albeit subject to idiosyncratic constraints. Together, these imply:

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

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

I approximate this conditional expectation using a permutation procedure, indexed by $i = 1, \dots, S$, and implemented independently for each state.

- (i). Randomly (and uniformly) permute counterfactual New Village sites $g_1^{(s)}$, conditional on w_1 .
- (ii). Calculate gravity-based counterfactual resettled populations $\bar{g}_2^{(s)}$.
- (iii). Calculate 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:

$$\overline{ResettleDensity}_c \equiv \frac{1}{S} \sum_{s=1}^S f_c(g_1^{(s)}, \bar{g}_2^{(s)}). \quad (\text{A-1})$$

Empirically, actual county resettlement densities center around their expected densities (Appendix Figure A.3), suggesting that the gravity-based model captures the main systematic component of resettlement patterns. While not required for identification, site-level resettled populations also align closely with model predictions.

B.2 Supplementary Analysis

Built-up volume. Higher population density in more resettled areas was accompanied by a substantial increase in built-up volume (Appendix Table A.4). By 1975, counties with higher resettlement had 33% more buildings. The larger percentage increase in buildings relative to population (around 11%) suggests a relatively elastic housing supply. At a finer scale, 1990 satellite imagery shows dense building clusters around New Villages, in contrast to the more uniform settlement patterns in surrounding areas before the Emergency (Appendix Figure A.5).

B.3 Robustness

This section shows that the main results are robust to alternative specifications of expected resettlement density, different covariate choices, and sample restrictions.

Alternative specifications of counterfactual resettlement. Appendix Table A.15 examines alternative specifications of counterfactual resettlement density. The baseline assumes the British prioritized siting near rivers when no roads were within 5 kilometers. Results are similar when assuming a preference for roads over rivers up to 10 kilometers (row 2). The baseline also allows counterfactual villages to be arbitrarily close to one another. While this is plausible—since the minimum distance between observed villages is only 200 meters—most are at least 1 kilometer apart. I show that results are similar when imposing a 1-kilometer minimum spacing between counterfactual villages (row 3).

The expected number of squatters resettled to each counterfactual site is also robust to different squatter definitions and resettlement cost elasticities. The baseline defines squatters as Chinese communities living within 5 kilometers of the forest. Estimates remain similar when using smaller or larger cutoffs (rows 4–5). The baseline resettlement cost elasticity with respect to distance is calibrated to 0.65 using observed villager populations. A higher elasticity implies that counterfactual resettlement density more closely mirrors the original squatter density. Results are robust to alternative elasticity values (rows 6–7).

Finally, the baseline population shifter is defined as the inverse hyperbolic sine of the number of resettled persons per unit area. As county resettlement density is used to shift log population density in the structural estimation, this log-like transformation improves the power of the first stage. Results are similar when using a logarithm transformation, restricting the sample to resettled counties.⁸⁴

Additional controls. Appendix Table A.16 considers additional county covariates. The estimates remain stable when additionally controlling for transportation and population characteristics of neighboring counties (rows 2–3). They are also robust to including features of productivity fundamentals, such as ruggedness (row 4), paddy rice and coconut suitability (row 5), distance to prewar industrial facilities (row 6), and distance to major cities (row 7). These patterns are not surprising, given the balance result established in Section IV.B.

Sample restrictions. Counties in the baseline sample vary in size, with some large, sparsely populated counties inland and smaller, denser counties along the coast. To address this, I control for county area in the main specification. Appendix Table A.17 shows

⁸⁴Since all regressions include an indicator for any resettlement, identification variation comes only from resettled counties. Non-resettled counties are included to improve precision in estimating the effects of covariates.

that results are not sensitive to excluding the largest and smallest counties (row 2), counties with extreme resettlement density (row 3), or the most densely populated prewar towns (row 4). Finally, since individual- and household-level outcomes come from the 2% microdata of the 1980 Population Census—covering only about two-thirds of the baseline counties with sampled Chinese—I restrict the sample to these counties and find similar results (row 5).

C Theoretical Results

C.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 $\Lambda^e = (\Lambda_A^e, \Lambda_M^e)$ from the joint distribution:

$$F_n^e(\Lambda_A, \Lambda_M) = \prod_{k=A,M} F_{nk}^e(\Lambda_k),$$

where the marginal probability distribution is Fréchet:

$$F_{nk}^e(\Lambda_k) = \exp(-\phi_{nk}^e \Lambda_k^{-\theta}).$$

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

$$\begin{aligned} y_{ink}^e &= w_{nk} \lambda_{ink}^e \\ &= w_{nk} \Lambda_{ink}^e f(L_{nk}^c, L_{nk}^m) \\ &= w_{nk}^e \Lambda_{ink}^e, \end{aligned}$$

where

$$w_{nk}^e \equiv w_{nk} f(L_{nk}^c, L_{nk}^m).$$

Function $f(L_{nk}^c, L_{nk}^m)$, which depends on local population distribution, captures human capital externalities.

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

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

k is

$$\pi_{nk}^e \equiv \mathbb{P}(y_{ink}^e = \max_s y_{ins}^e) = \frac{\phi_{nk}^e (w_{nk}^e)^\theta}{\sum_s \phi_{ns}^e (w_{ns}^e)^\theta} = \phi_{nk}^e \left(\frac{w_{nk}^e}{\bar{w}_n^e} \right)^\theta,$$

where

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

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

$$\mathbb{E}[\max_k y_{ink}^e] = \Gamma_\theta \left(\sum_k \phi_k^e (w_{nk}^e)^\theta \right)^{1/\theta} = \Gamma_\theta \bar{w}_n^e.$$

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

$$\mathbb{E}[\underbrace{y_{ink}^e / w_{nk}^e}_{\Lambda_{ink}^e} | y_{ink}^e = \max_s y_{ins}^e] f(L_{nk}^c, L_{nk}^m) = \Gamma_\theta \bar{w}_n^e w_{nk}^{-1}.$$

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

$$\Gamma_\theta (\phi_{nk}^e)^{1/\theta} (\pi_{nk}^e)^{-1/\theta} f(L_{nk}^c, L_{nk}^m),$$

where the neoclassical force $(\pi_{nk}^e)^{-1/\theta}$ 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 $f(L_{nk}^c, L_{nk}^m)$ 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 :

$$w_{nk} H_{nk}^e = L_n^e \pi_{nk}^e \Gamma_\theta \bar{w}_n^e.$$

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

$$\begin{aligned}
H_{nk} &= \Gamma_\theta \sum_e L_n^e \phi_{nk}^e (w_{nk}^e)^\theta w_{nk}^{-1} (\bar{w}_n^e)^{1-\theta} \\
&= \Gamma_\theta \sum_e L_n^e \phi_{nk}^e w_{nk}^{-1} w_{nk}^e (w_{nk}^e)^{1-\theta} (\bar{w}_n^e)^{1-\theta} \\
&= \Gamma_\theta \sum_e L_n^e \phi_{nk}^e (L_{nk})^{\gamma_k} \left(\frac{L_{nk}^e}{L_{nk}} \right)^{\gamma_e} \left(\frac{w_{nk}^e}{\bar{w}_n^e} \right)^{\theta-1}.
\end{aligned}$$

C.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_n^e is assumed to drawn from the following location-specific Fréchet distribution

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

where the scale $\bar{a}_n^e$ 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:

$$a_n^e = u_n^e (L_n)^\beta \left(\frac{L_n^e}{L_n} \right)^{\beta^e}.$$

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

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

C.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:

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

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:

$$\tilde{\pi}_{nk}^e \equiv \phi_{nk}^e \left(\frac{w_{nk}^e}{\bar{w}_n^e} \right)^\theta.$$

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

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

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 $H_{nk} = \sum_e H_{nk}^e$, where $H_{nk}^e = \Gamma_\theta L_{nk}^e \left(\frac{\bar{w}_n^e}{w_{nk}} \right)$.

(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 :

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

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:

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

- (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:

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

where

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

3. Calculate the implied population distribution: $\tilde{L}_n^e = \sum_r \tilde{L}_r^e m_{rn}^e$.

4. Update population iteratively until convergence:

$$L_{n,new}^e \equiv \iota L_n^e + (1 - \iota) \tilde{L}_n^e.$$

C.4 Sufficient Conditions for Uniqueness of Equilibrium

This section applies Theorem 1 from Allen, Arkolakis and Li (2024) to derive sufficient conditions for the uniqueness of equilibrium. I rewrite the system of equations that characterizes the equilibrium in terms of 15 unknowns

$$\{w_{nA}, w_{nM}, P_{nA}, P_{nM}, H_{nA}, H_{nM}, \bar{w}_n^c, \bar{w}_n^m, L_{nA}, L_{nM}, L_n^c, L_n^m, L_n, \Pi_n^c, \Pi_n^m\}$$

and 15 equations:

$$\begin{aligned} w_{nk}^\sigma H_{nk} &= \sum_r \alpha_k \tau_{nr}^{1-\sigma} (w_{rA} H_{rA} + w_{rM} H_{rM}) P_{rk}^{\sigma-1} \\ P_{nk}^{1-\sigma} &= \sum_r \tau_{rn}^{1-\sigma} w_{rk}^{1-\sigma} \\ w_{nk} H_{nk} &= \sum_e \Gamma_\theta (\phi_{nk}^e)^{\frac{1}{1-\gamma^e\theta}} \bar{w}_n^e (w_{nk})^{\frac{\theta}{1-\gamma^e\theta}} (L_n^e)^{\frac{1}{1-\gamma^e\theta}} (L_{nk})^{\frac{\theta(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\ (\bar{w}_n^e)^\theta &= \sum_k (\phi_{nk}^e)^{\frac{1+\gamma^e\theta}{1-\gamma^e\theta}} (w_{nk})^{\frac{\theta}{1-\gamma^e\theta}} (L_n^e)^{\frac{\gamma^e\theta}{1-\gamma^e\theta}} (L_{nk})^{\frac{\theta(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\ L_{nk}^{1-(\gamma_k-\gamma^e)\theta} (w_{nk})^{-\theta} &= \sum_e (\phi_{nk}^e)^{\frac{1+\gamma^e\theta}{1-\gamma^e\theta}} (\bar{w}_n^e)^{-\theta} (w_{nk})^{\frac{\gamma^e\theta^2}{1-\gamma^e\theta}} (L_n^e)^{\frac{1}{1-\gamma^e\theta}} (L_{nk})^{\frac{\gamma^e\theta^2(\gamma_k-\gamma^e)}{1-\gamma^e\theta}} \\ (L_n)^{(\beta^e-\beta)\nu} (L_n^e)^{1-\beta^e\nu} (\bar{w}_n^e)^{-\nu} P_{nA}^{\nu\alpha} P_{nM}^{\nu(1-\alpha)} &= \sum_r \bar{a}_n^e \check{L}_r^e \eta_{rn}^{-\nu} (\Pi_r^e)^{-\nu} \\ (\Pi_n^e)^\nu &= \sum_r \eta_{nr}^{-\nu} \bar{a}_r^e (L_r)^{(\beta-\beta^e)\nu} (L_r^e)^{\beta^e\nu} (\bar{w}_r^e)^\nu P_{rA}^{-\nu\alpha} P_{rM}^{-\nu(1-\alpha)} \\ L_n &= \sum_e L_n^e. \end{aligned}$$

The equilibrium contains a set of $\mathcal{N} = \{1, \dots, N\}$ locations and a set of $\mathcal{H} = 1, \dots, H$ economic interactions (or endogenous variables), where $H = 15$. The $H \times H$ matrices B and

Γ , as in Allen, Arkolakis and Li (2024), are given by

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

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

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, Arkolakis and Li (2024), 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, Arkolakis and Li (2024), 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.

D Structural Estimation

D.1 Market Access Terms

I derive four underlying conditions involving the trade and migration market access terms from the equilibrium conditions (15)–(17).

- (i). Total sales equals payments to labor: $w_{nk}H_{nk} = \sum_r X_{nrk}$. Using Equation (12), 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 (5), this can be written as

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

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

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

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

$$\mathcal{P}_{nk}^{1-\sigma} = \frac{\alpha_k}{\Omega_{nk}} \sum_r \tau_{nr}^{1-\sigma} Y_r P_{rk}^{\sigma-1}, \quad (\text{A-2})$$

$$P_{rk}^{1-\sigma} = \sum_n \tau_{nr}^{1-\sigma} Y_n \mathcal{P}_{nk}^{\sigma-1}, \quad (\text{A-3})$$

$$(\mathcal{V}_n^e)^{-\nu} = \sum_r \eta_{rn}^{-v} \check{L}_r^e (\Pi_r^e)^{-v}, \quad (\text{A-4})$$

$$(\Pi_r^e)^v = \sum_n \eta_{rn}^{-v} L_n^e (\mathcal{V}_n^e)^\nu, \quad (\text{A-5})$$

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-2)–(A-5) 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:

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

The migration market access equations are:

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

I can rewrite the first set of equations as:

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

where $x_{nk} \equiv \mathcal{P}_{nk}^{\sigma-1}$ and $y_{nk} \equiv P_{nk}^{\sigma-1}$. Using Allen, Arkolakis and Li (2024), 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, Arkolakis and Li (2024), 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:

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

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:

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

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.

D.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:

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

(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:

$$loss \equiv \frac{1}{N_d^2} \sum_{j,h} (\ln m_{jh} - \ln \hat{m}_{jh})^2, \quad (\text{A-6})$$

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.3 Parameter Estimates

This section discusses the parameter estimates and compares them with the literature.

Migration cost elasticity. The estimated elasticity of migration costs with respect to distance, κ , is 0.52.⁸⁵ 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 1955.⁸⁶

Skill dispersion. The estimated Fréchet shape parameter, $\theta = 3.3$, lies 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 2012.⁸⁷

⁸⁵Estimates are similar across ethnic groups: $\kappa = 0.54$ for Chinese and 0.51 for non-Chinese.

⁸⁶Bryan 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 $\partial \ln \eta_{nr} / \partial \ln d_{nr} \approx 0.37$.

⁸⁷One 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.

Productivity spillovers. I estimate that local employment in the non-agricultural sector increases labor productivity with an elasticity of $\gamma_M = 0.22$, while the agricultural sector shows a smaller, negative elasticity of $\gamma_A = -0.12$. My estimate for non-agricultural sectors is similar to the 0.2 estimated by Kline and Moretti (2014) but lower than the 1.25–3.1 range reported by Greenstone, Hornbeck and Moretti (2010).⁸⁸ 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.⁸⁹ 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, $\gamma^e = 0.13$. 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 (9) 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 (2004) estimates wage elasticities of 0.14 for college graduates and 0.21 for high school graduates with respect to college share in a city.⁹⁰ Rossi-Hansberg, Sarte and Schwartzman (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.

Amenity spillovers. I estimate the amenity spillover elasticity with respect to local population size at $\beta = -0.005$. 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 the amenity spillover can be decomposed as $\beta = \beta_a - \delta\beta_r$, where β_a represents the pure

⁸⁸See a discussion in Kline and Moretti (2014).

⁸⁹See Melo, Graham and Noland (2009), Combes and Gobillon (2015), and Ahlfeldt and Pietrostefani (2019) 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.

⁹⁰Moretti (2004) 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.

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.19). This elasticity is higher than U.S. estimates, which range from 1 to 3 (Gyourko, Saiz and Summers, 2008; Saiz, 2010).⁹¹ In 1980, housing expenditure accounted for 17.6% of total spending, implying a pure amenity spillover of $\beta_a = \beta + \delta\beta_r = 0.05$.⁹²

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 (2022) 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).⁹³

E Counterfactuals

E.1 Lower Bound for Utility Loss from Forced Resettlement

This section describes how I derive a lower bound for the utility loss due to forced resettlement.

I first recover the location amenity fundamentals that rationalize the 1947 population distribution as a steady state such that, in the absence of resettlement, it would have persisted until 1957. Let $\tilde{a}_n^e$ denote these 1947 amenity fundamentals. The migration shares for group

⁹¹In Indonesia, Bryan and Morten (2019) estimate a value of 4, though with limited statistical power.

⁹²The 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).

⁹³There 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: $(\gamma_{UU}^A, \gamma_{SU}^A, \gamma_{US}^A, \gamma_{SS}^A) = (-0.43, 0.18, -1.24, 0.77)$, where γ_{SU}^A denotes the marginal amenity spillover of a college graduate (S) on the utility of a non-college graduate (U), and so on.

e from region r from 1947 to 1957 are given by:

$$\tilde{m}_{rn}^e = \frac{\tilde{a}_n^e \left(\eta_{rn}^{-1} (L_{n,47})^\beta (L_{n,47}^e / L_{n,47})^{\beta^e} \bar{w}_n^e P_n^{-1} \right)^\nu}{\sum_{l=1}^N \tilde{a}_l^e \left(\eta_{rl}^{-1} (L_{l,47})^\beta (L_{l,47}^e / L_{l,47})^{\beta^e} \bar{w}_l^e P_l^{-1} \right)^\nu}.$$

where $\{L_{n,47}^e\}$ denotes the population distribution in 1947.

Using the balance of migration flows, $L_{n,47}^e = \sum_r L_{r,47}^e \tilde{m}_{rn}^e$, we obtain:

$$L_{n,47}^e = \sum_r L_{r,47}^e \frac{\tilde{a}_n^e \left(\eta_{rn}^{-1} (L_{n,47})^\beta (L_{n,47}^e / L_{n,47})^{\beta^e} \bar{w}_n^e P_n^{-1} \right)^\nu}{\sum_{l=1}^N \tilde{a}_l^e \left(\eta_{rl}^{-1} (L_l)^\beta (L_{l,47}^e / L_{l,47})^{\beta^e} \bar{w}_l^e P_l^{-1} \right)^\nu},$$

which can be arranged as

$$\tilde{a}_n^e = \frac{1}{L_{n,47}^e} \sum_r L_{r,47}^e \frac{\eta_{rn}^{-\nu} \left((L_{n,47})^\beta (L_{n,47}^e / L_{n,47})^{\beta^e} \bar{w}_n^e P_n^{-1} \right)^\nu}{\sum_{l=1}^N \tilde{a}_l^e \eta_{rl}^{-\nu} \left((L_{l,47})^\beta (L_{l,47}^e / L_{l,47})^{\beta^e} \bar{w}_l^e P_l^{-1} \right)^\nu}. \quad (\text{A-7})$$

This system (A-7) provides $N - 1$ equations, allowing me to solve for $\tilde{a}_n^e$ up to a scale.

Using these recovered amenity fundamentals, I then solve for the ethnicity- and place-specific wage subsidies required to voluntarily relocate Chinese and Malays from the 1947 population distribution to the 1957 resettled distribution.

Let ϵ_n^e denote the ad-valorem subsidy for group e in region n . Again, using the balance of migration flows, $L_{n,57}^e = \sum_r L_{r,47}^e m_{rn}^e$, we obtain:

$$L_{n,57}^e = \sum_r L_{r,47}^e \frac{\tilde{a}_n^e \left(\eta_{rn}^{-1} (L_{n,57})^\beta (L_{n,57}^e / L_{n,57})^{\beta^e} (1 + \epsilon_n^e) \bar{w}_n^e P_n^{-1} \right)^\nu}{\sum_{l=1}^N \tilde{a}_l^e \left(\eta_{rl}^{-1} (L_{l,57})^\beta (L_{l,57}^e / L_{l,57})^{\beta^e} (1 + \epsilon_l^e) \bar{w}_l^e P_l^{-1} \right)^\nu}.$$

Rearranging gives:

$$(1 + \epsilon_n^e)^{-\nu} = \frac{1}{L_{n,57}^e} \sum_r L_{r,47}^e \frac{\tilde{a}_n^e \eta_{rn}^{-\nu} \left((L_{n,57})^\beta (L_{n,57}^e / L_{n,57})^{\beta^e} \bar{w}_n^e P_n^{-1} \right)^\nu}{\sum_{l=1}^N \tilde{a}_l^e \eta_{rl}^{-\nu} \left((L_{l,57})^\beta (L_{l,57}^e / L_{l,57})^{\beta^e} (1 + \epsilon_l^e) \bar{w}_l^e P_l^{-1} \right)^\nu}. \quad (\text{A-8})$$

Since migration depends only on relative wages across regions, Equation (A-8) implies that ϵ_n^e is determined only up to a scale. To calculate the least-cost, weakly positive subsidies, I scale the solution vector so that the minimum subsidy across locations is zero. Specifically, let $\{\epsilon_n^e\}$ be any solution to Equation (A-8). Then, the minimum wage subsidies for group e ,

denoted $\{\varepsilon_n^e\}$, are given by:

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

Finally, I calculate the total real subsidies received by the resettled population as:

$$\sum_e \sum_n \varepsilon_n^e L_{n,resettled}^e \frac{y_n^e}{P_n},$$

where $y_n^e = \sum_k w_{nk} H_{nk} / L_n^e$ denotes average output per capita for group e in region n , and $L_{n,resettled}^e$ is the number of resettled persons from group e in region n , which I approximate by assuming that 90% of the resettled population was Chinese based on historical accounts.

Appendix References

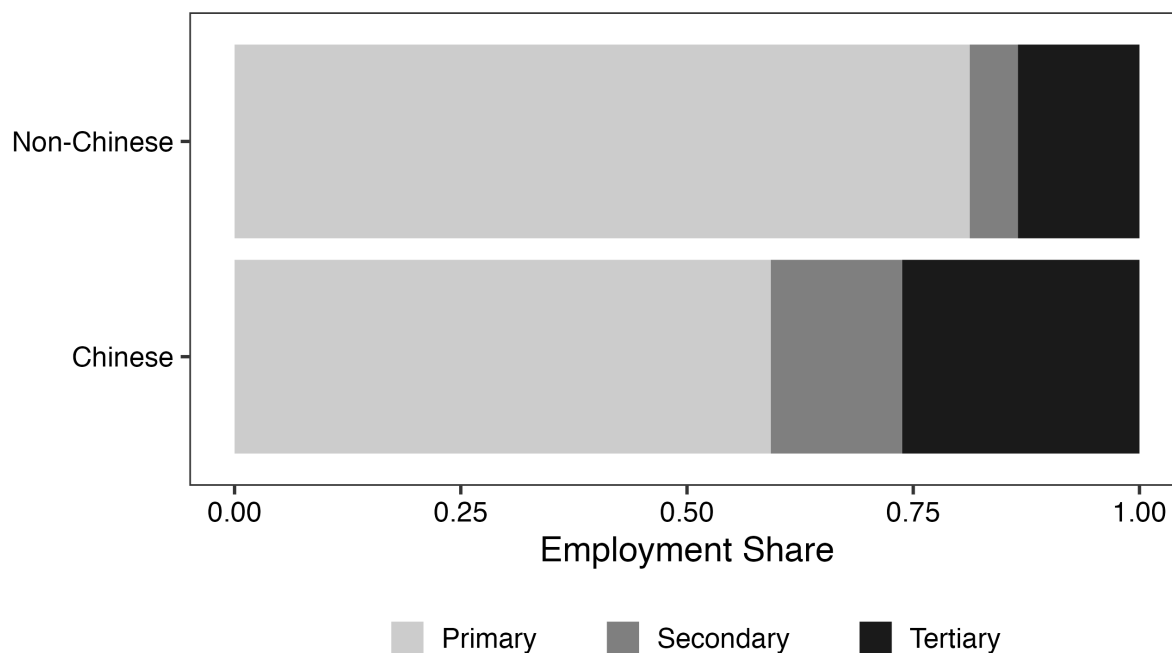
- Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani.** 2019. “The Economic Effects of Density: A Synthesis.” *Journal of Urban Economics*, 111: 93–107.
- Bureau van Dijk.** 2019. “Bureau van Dijk Historical Ownership Databases User Guide.” Bureau van Dijk, Version Q2 2019.
- Claessens, Stijn, Simeon Djankov, and Larry H.P Lang.** 2000. “The Separation of Ownership and Control in East Asian Corporations.” *Journal of Financial Economics*, 58(1): 81–112. Special Issue on International Corporate Governance.
- Combes, Pierre-Philippe, and Laurent Gobillon.** 2015. “Chapter 5 - The Empirics of Agglomeration Economies.” In *Handbook of Regional and Urban Economics*. Vol. 5 of *Handbook of Regional and Urban Economics*, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, 247–348. Elsevier.
- 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.
- Department of Statistics Malaysia.** 1980. “Report of the Household Expenditure Survey: Peninsular Malaysia (1980), Sabah & Sarawak (1982).”
- 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.
- 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.
- Kalemli-Özcan, Şebnem, Bent E. Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcen Yeşiltaş.** 2024. “How to Construct Nationally Representative Firm-Level Data from the Orbis Global Database: New Facts on SMEs and Aggregate Implications for Industry Concentration.” *American Economic Journal: Macroeconomics*, 16(2): 353–74.
- Lagakos, David, and Michael E Waugh.** 2013. “Selection, Agriculture, and Cross-Country Productivity Differences.” *American Economic Review*, 103(2): 948–980.
- La Porta, Rafael, Florencio Lopez-de Silanes, and Andrei Shleifer.** 1999. “Corporate Ownership Around the World.” *The Journal of Finance*, 54(2): 471–517.
- 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.
- Moretti, Enrico.** 2004. “Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data.” *Journal of Econometrics*, 121(1-

2): 175–212.

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

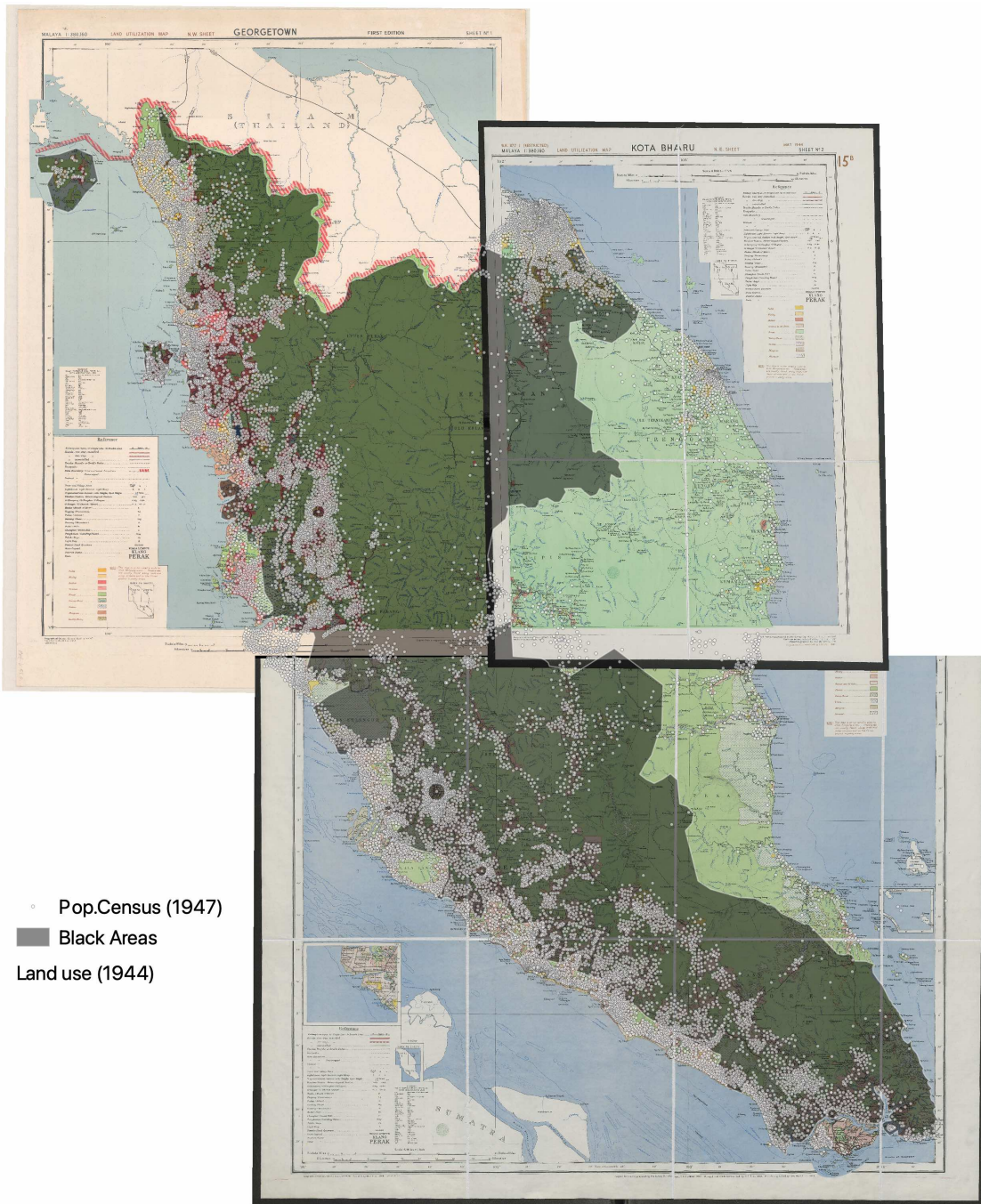
F Appendix Figures

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



Notes: This figure shows the distribution of employment across primary, secondary, and tertiary sectors for Chinese and non-Chinese populations in 1947. The primary sector (also referred to as “agriculture” in the main text) includes agriculture, hunting, forestry, fishing and mining. The secondary sector includes manufacturing, utilities, and construction. The tertiary sector includes storage, transport, communication, commerce, finance, business, and other services. Data from the 1947 Census of Population (Del Tufo, 1947).

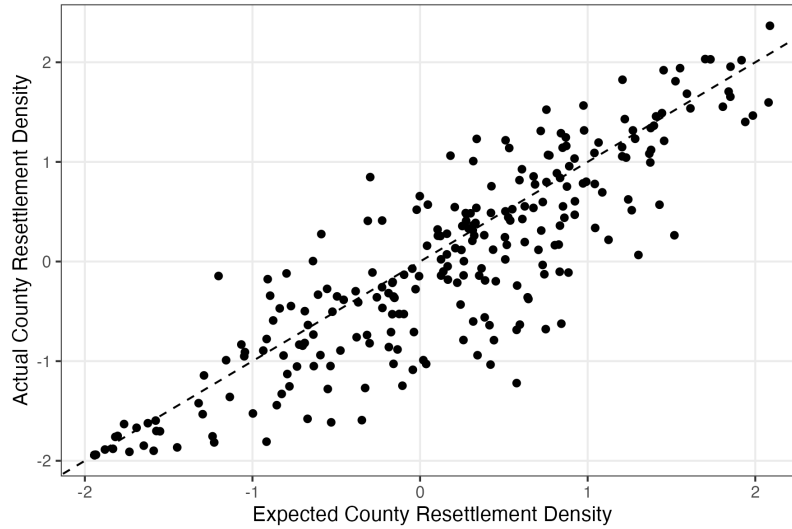
Figure A.2. Distribution of Squatter Settlements in 1947



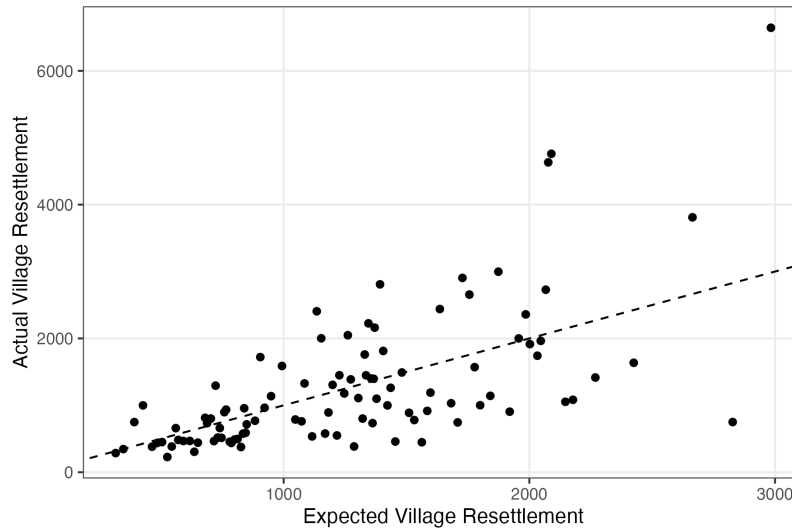
Notes: This figure maps the distribution of squatter settlements based on the intersection of three historical sources. Gray dots indicate population clusters from the 1947 Census of Population. Dark-shaded areas represent “Black areas” under Emergency regulations due to communist insurgency. Green-shaded areas are forests, based on land utilization maps from 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



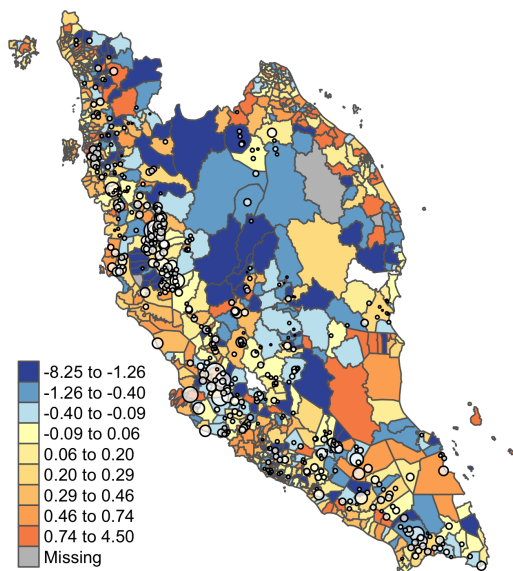
Panel B. Village Resettled Population,
Compared to Expected Resettled Population



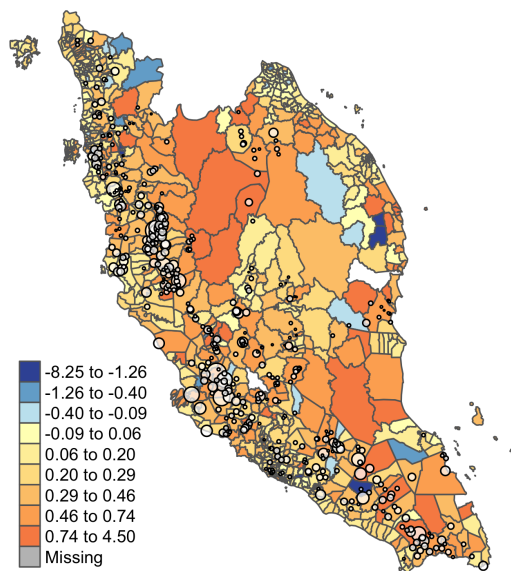
Notes: This figure compares actual resettlement outcomes with predictions from the gravity model described in Appendix Section B.1. Panel A plots actual county resettlement density against expected density, calculated from Equation (A-1), conditional on the actual locations of New Villages. Panel B compares the actual resettled population in each village with the counterfactual population predicted by the dislocation-minimizing plan in Equation (2), also conditional on village locations. Data from the Corry report.

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

Panel A. Log Change in Population,
Chinese

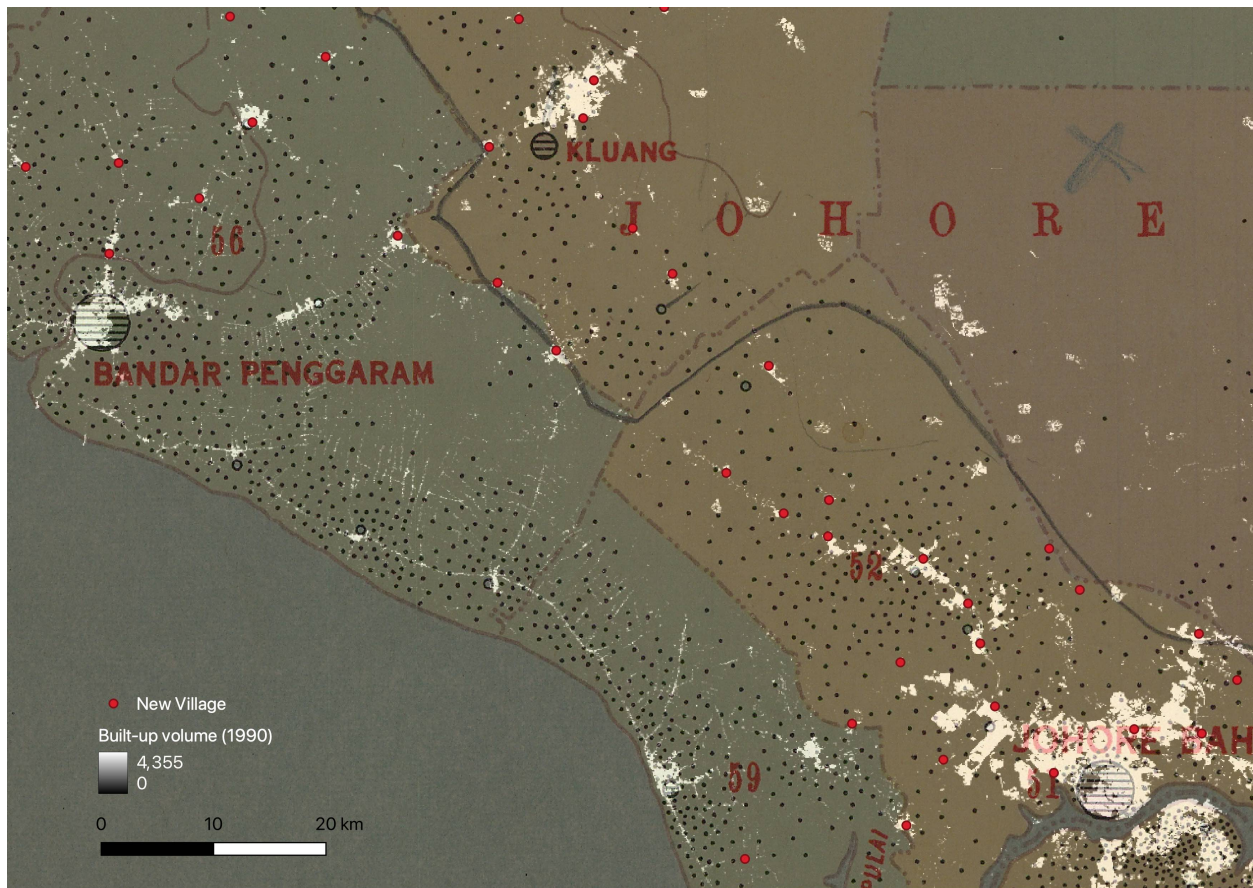


Panel B. Log Change in Population,
Non-Chinese



Notes: This figure maps county population growth from 1947 to 1957 by ethnic group. Panel A shows log changes in Chinese population. Panel B shows log changes in non-Chinese population. White bubbles denote New Villages, with size proportional to the log resettled population. Counties with missing data are shaded in gray. Data from the tabulated Census of Population and the Corry report.

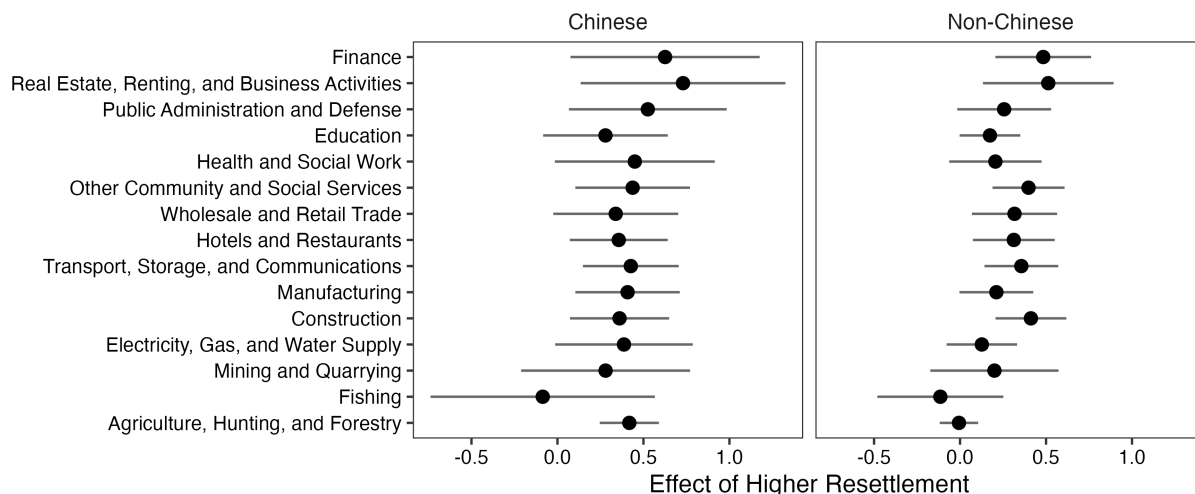
Figure A.5. Built-up Volume in 1990, Johor



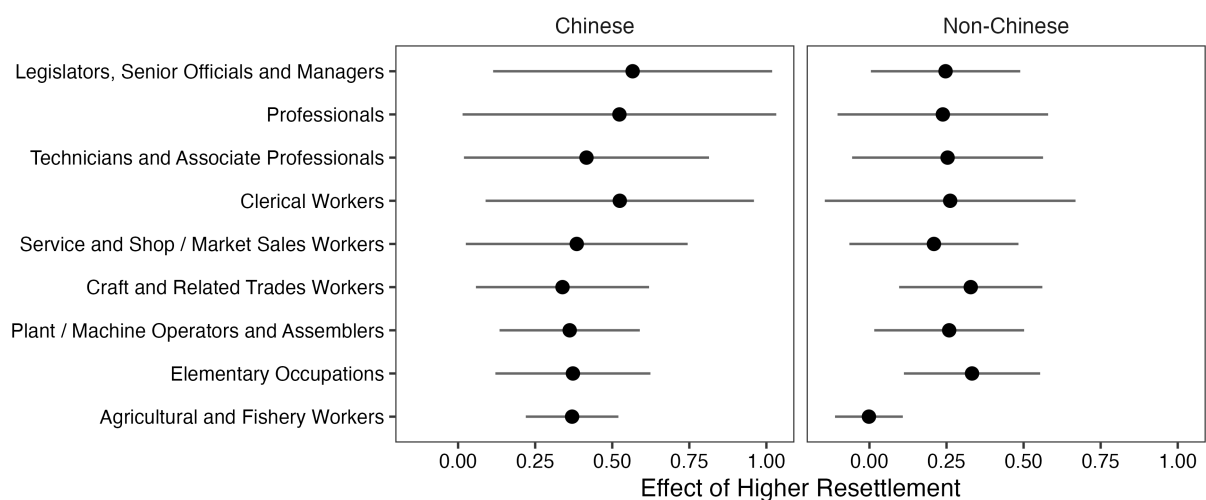
Notes: This figure maps built-up volumes in 1990 for a region in Johor. Built-up volumes are calculated from 100-meter resolution surface and height data based on Sentinel-2 and Landsat satellite imagery, with higher volumes shaded in white. Red dots indicate the locations of New Villages, and black dots represent population clusters from the 1947 Census. Built-up volume data from the GHSL project; New Village locations from the Corry report.

**Figure A.6. Effect of Higher Resettlement on Employment Size in 1991,
by Industry and Occupation**

Panel A. By Industry

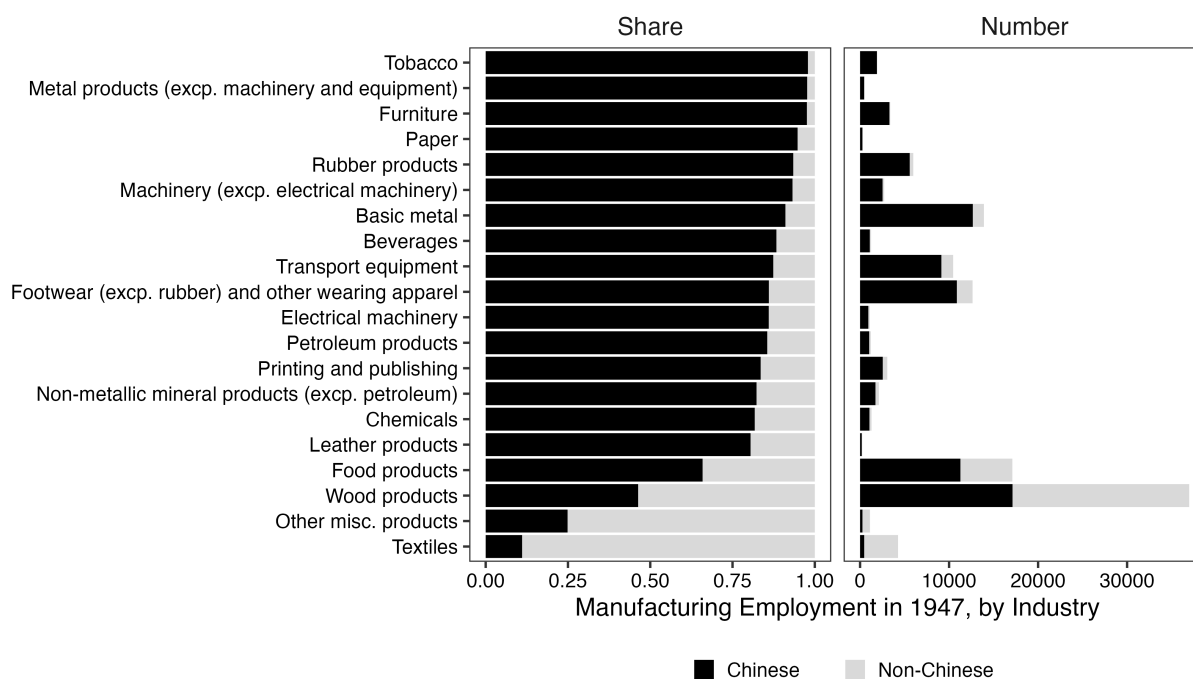


Panel B. By Occupation



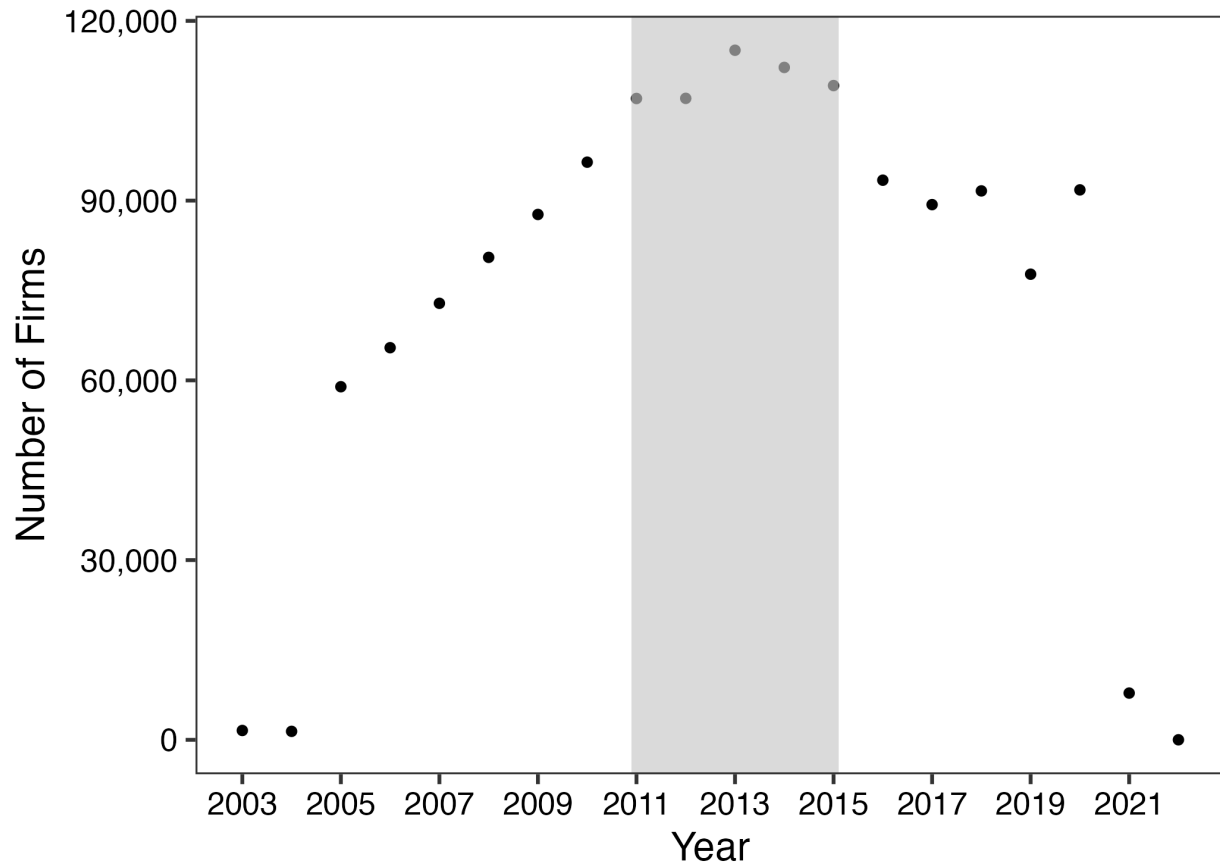
Notes: This figure shows the relationship between county resettlement density and employment size by industry (Panel A) and occupation (Panel B) in 1991, separately for Chinese and non-Chinese workers. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. All regressions are estimated using the Poisson pseudo-maximum-likelihood (PPML) estimator 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county. Data from the tabulated Census of Population in 1991. Error bars show 95% confidence intervals based on Conley standard errors with a 30-kilometer cutoff.

Figure A.7. Chinese Manufacturing Employment Share and Number in 1947, by Industry



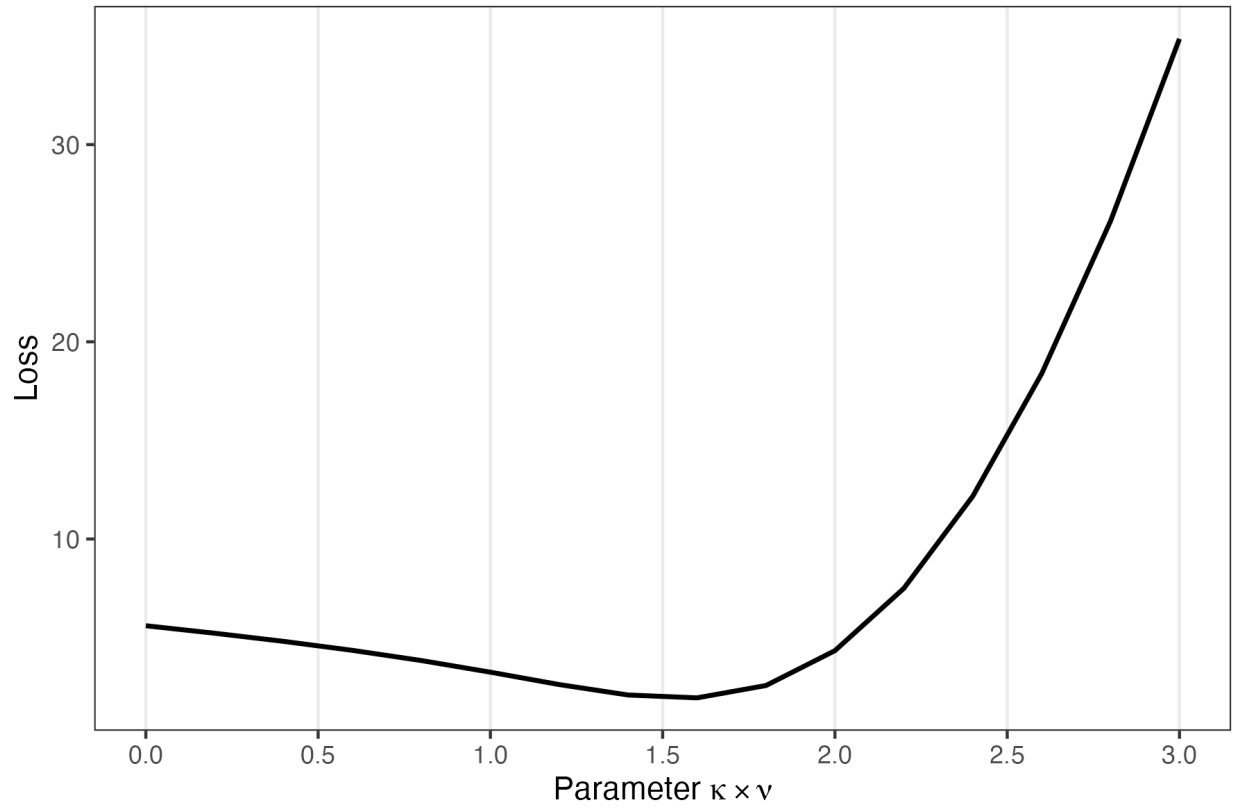
Notes: This figure shows the distribution of Chinese and non-Chinese employment across manufacturing industries in 1947. The left panel shows the share of employment by ethnicity within each industry. The right panel shows the total number of workers by ethnicity. Black bars denote Chinese employment, and gray bars denote non-Chinese employment. Data from the 1947 Census of Population (Del Tufo, 1947).

Figure A.8. Number of Malaysian Firms in Orbis Meeting Sample Criteria, 2003–2022



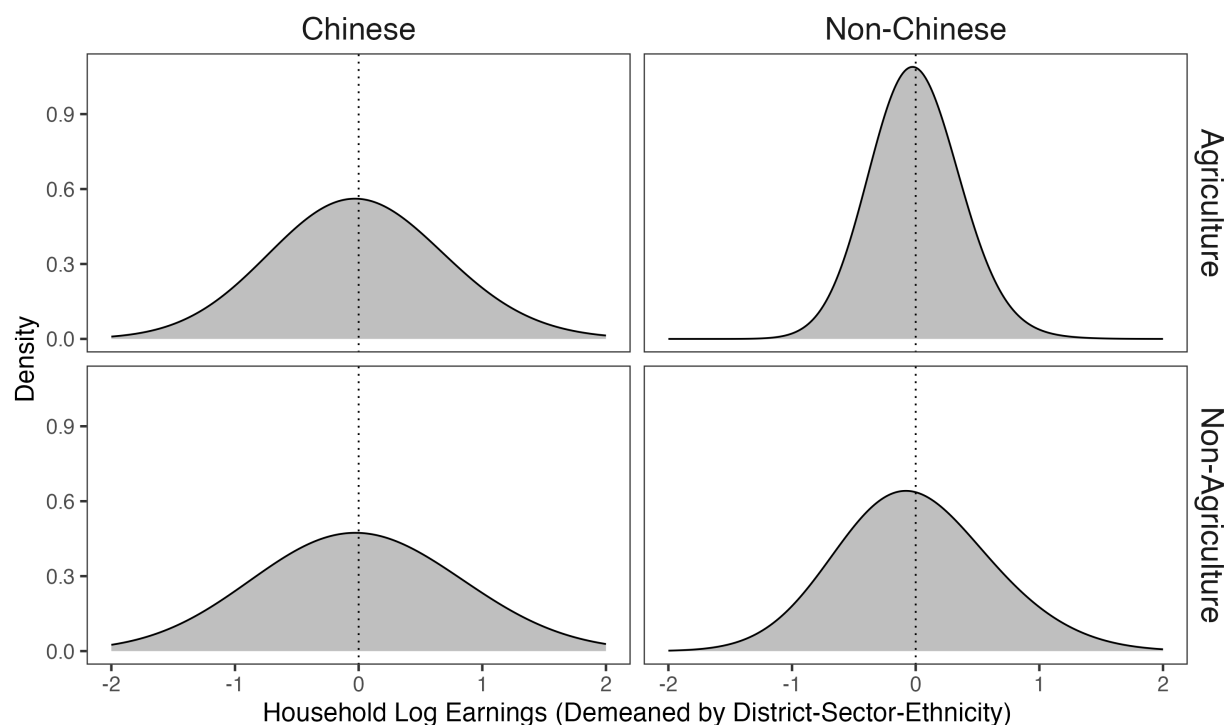
Notes: This figure shows the number of firms in the Orbis database from 2003 to 2022 that meet the main sample criteria: positive revenue, a Malaysian ultimate person owner, a non-missing NAICS industry code, and a location in Peninsular Malaysia. The analysis focuses on the 2011–2015 period, shaded in gray. The classification of ultimate person owners is described in Appendix A.2. Data from the Orbis Historical Disk, accessed in 2023.

Figure A.9. Convexity of the Loss Function in Estimating Migration Costs



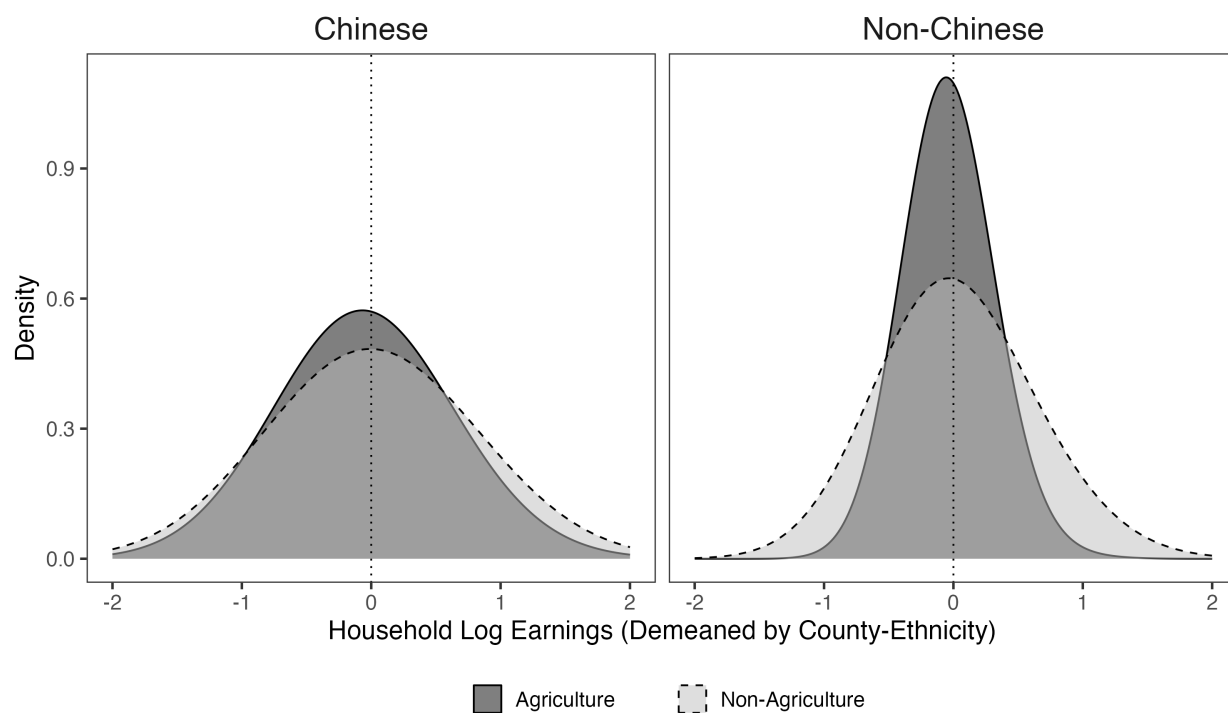
Notes: This figure shows the convexity of the loss function used to estimate migration costs. The y-axis plots the loss from Equation (A-6). The x-axis shows the product of κ and ν , the parameter being estimated. Data from the tabulated Census of Population in 1980.

Figure A.10. Distribution of Demeaned Household Log Earnings, Relative to District-Sector-Ethnicity Average



Notes: This figure shows the distribution of household log earnings after demeaning by the district-sector-ethnicity average, separately for Chinese and non-Chinese households. The top panel shows households in agriculture, and the bottom panel shows households in non-agriculture. The vertical line marks the group mean of zero. Data from the 2% Census of Population microdata in 1980.

Figure A.11. Distribution of Demeaned Household Log Earnings, Relative to County-Ethnicity Average



Notes: This figure shows the distribution of household log earnings after demeaning by the county-ethnicity average, separately for Chinese and non-Chinese households. Within each group, distributions are plotted for households in agriculture (solid line) and non-agriculture (dashed line). The vertical line marks the group mean of zero. Data from the 2% Census of Population microdata in 1980.

Figure A.12. Estimated Production Fundamentals in 1980, by Sector

Panel A. Agriculture

Panel B. Non-Agriculture



Notes: This figure maps the estimated log production fundamentals by county, with darker shades indicating higher productivity deciles. Panel A shows agricultural fundamentals. Panel B shows non-agricultural fundamentals. Values represent ethnic-population-weighted averages within each county-sector. Counties not included in the sample are shown in white.

Figure A.13. Estimated Amenity Fundamentals in 1980, by Ethnic Group

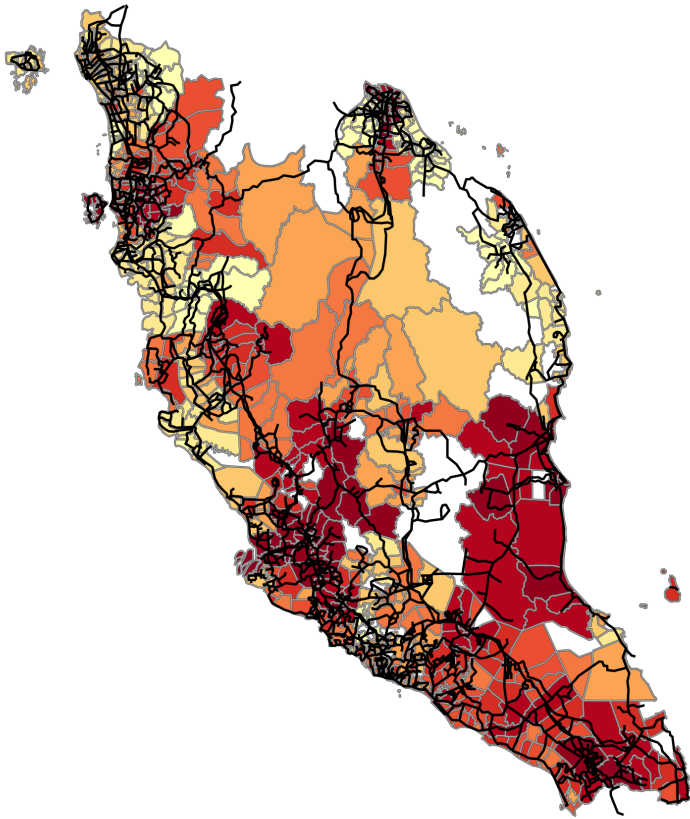
Panel A. Chinese

Panel B. Non-Chinese



Notes: This figure maps the estimated log amenity fundamentals by county, with darker shades indicating higher amenity deciles. Panel A shows amenity fundamentals for the Chinese population; Panel B for the non-Chinese population. Counties not included in the sample are shown in white.

Figure A.14. County Productivity and Transport Network in 1980 Equilibrium



Notes: This figure maps county productivity (real wages) in the 1980 equilibrium, averaged across ethnic groups and sectors. Darker red shades indicate higher productivity. Black lines show the road and railway network in 1983. Transport data from G8031 road map series. (U.S. Office of Strategic Services, 1942).

G Appendix Tables

Table A.1. Data Sources

Data Source	Publication Year(s)	Variables
Population Census (Tabulated)	1911, 1921, 1931, 1947, 1957, 1970, 1980, 1991, 2000	Population by ethnicity (1911, 1921); population by county and ethnicity (1931–2000); employment by industry and ethnicity (1947); population map (1947); migration flows by district and ethnicity (1980); employment by county, ethnicity, and industry/occupation (1991)
Population Census (Microdata)	1980	Household assets (house, vehicle, phone, etc.); household size; number of children born; educational attainment; years of schooling; duration of residence in present locality; migration status; employment status and industry; language spoken
Second Malaysian Family Life Survey	1989	Household assets; annual earnings; characteristics of resettled Chinese: first job, schooling, land ownership
Directory of Manufacturing	1970	Establishments by county and industry; share employing full-time workers
Orbis Historical Disk	2003–2022	Firms by ownership, county, and industry; average annual revenue
Ministry of Education Malaysia	2022	By school type (Chinese/non-Chinese): number of schools, distance to nearest school, teacher-student ratio
Directory of Singapore and Malaya	1959	Distance to Chinese schools; number of Chinese schools
G8031 Road Maps	1942, 1961, 1983	Distance to roads and railroads
HIND 1076 Topographical Maps	1945	Distance to rail station, police station, post/telegraph office, hospital, Chinese temple
GS GS 4474 Land Use Maps	1943	Land share in rubber and mining; forest areas
US National Archives, RG226	1944	Distance to industrial facilities
Global Human Settlement Layer	1975, 1990, 2005	Built-up volumes
Shuttle Radar Topography Mission (SRTM)	2000	Elevation
Nunn and Puga (2012)	2012	Terrain ruggedness
DIVA-GIS	2011	Distance to rivers
FAO GAEZ v4	2022	Rice suitability; coconut suitability
Galor and Özak (2016)	2016	Caloric suitability index
A General Survey of New Villages (Corry, 1954)	1954	Resettled population by county
The National Archives in the UK (CO 1030/1)	1957	Black Areas
Author’s Calculations	N/A	Distance to coastline and major cities; squatter population distribution

Table A.2. Population in British Malaya from 1911 to 1957, by Ethnic Group

Year	Chinese		Malays		Indians and Others	
	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.3. Predicting Log Household Income in 1988

	Log Household Earning, 1988	
	(1)	(2)
Vehicle	0.567 (0.288)	0.608 (0.319)
Motorcycle	0.002 (0.157)	0.098 (0.159)
Bicycle	−0.113 (0.154)	−0.055 (0.159)
Phone	1.006 (0.400)	0.907 (0.398)
Refrigerator	0.424 (0.176)	0.253 (0.179)
Television	0.182 (0.142)	0.117 (0.144)
Household Size	0.050 (0.010)	0.054 (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.4. Build-up Volumes, by County Resettlement Density

	Log Build-up Volumes, by Year:		
	1975	1990	2005
	(1)	(2)	(3)
Higher Resettlement	0.335	0.259	0.199
	(0.110)	(0.086)	(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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county. Data from the Global Human Settlement Layer (GHSL) project. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.5. 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	0.047 (0.029)	0.004 (0.017)	0.043 (0.023)
Mean of Outcome	0.39	0.47	
# Individuals	38,390	71,234	
Panel B. Internal Migrant After 1960			
Higher Resettlement	0.051 (0.032)	0.012 (0.019)	0.039 (0.025)
Mean of Outcome	0.30	0.40	
# Individuals	38,258	70,976	
Panel C. Number of Children Born			
Higher Resettlement	–0.020 (0.113)	–0.143 (0.061)	0.123 (0.107)
Mean of Outcome	4.01	4.16	
# Women	12,259	24,158	
Panel D. Household Size			
Higher Resettlement	0.076 (0.106)	–0.060 (0.048)	0.137 (0.125)
Mean of Outcome	5.81	5.20	
# Households	9,820	21,238	

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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.6. 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	0.300 (0.124)	0.282 (0.139)	–0.018 (0.064)
# County-Years	1,554	1,554	
Panel B. Chinese Employment			
Higher Resettlement	0.357 (0.166)	0.352 (0.203)	–0.005 (0.064)
# County-Years	1,400	1,476	
Panel C. Non-Chinese Employment			
Higher Resettlement	0.235 (0.105)	0.239 (0.117)	0.004 (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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.7. Educational Attainment in 1980, by Age Cohorts and County Resettlement Density

	Chinese Individuals, by Age Cohort:			Non-Chinese Individuals, by Age Cohort:		
	20–35 (1)	36–50 (2)	>50 (3)	20–35 (4)	36–50 (5)	>50 (6)
Panel A. Years of Schooling						
Higher Resettlement	0.436 (0.239)	0.337 (0.193)	0.068 (0.135)	0.128 (0.113)	0.099 (0.131)	−0.102 (0.097)
Panel B. Primary Education						
Higher Resettlement	0.031 (0.014)	0.030 (0.017)	0.007 (0.016)	0.016 (0.011)	0.025 (0.014)	−0.004 (0.012)
Panel C. Secondary Education						
Higher Resettlement	0.046 (0.028)	0.038 (0.018)	0.000 (0.007)	0.014 (0.012)	0.006 (0.011)	0.000 (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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.8. 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 (0.016)	0.020 (0.017)	0.026 (0.019)
Panel B. Have Vehicle			
Higher Resettlement	0.059 (0.027)	0.023 (0.012)	0.036 (0.021)
Panel C. Have Fridge			
Higher Resettlement	0.043 (0.029)	0.037 (0.021)	0.006 (0.025)
Panel D. Have TV			
Higher Resettlement	0.039 (0.012)	0.006 (0.018)	0.033 (0.017)
Panel E. Have Phone			
Higher Resettlement	0.038 (0.021)	0.014 (0.008)	0.024 (0.015)
# Households	10,593	23,269	

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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.9. Heterogeneous Effects of Higher Resettlement on Household Income in 1980, by County Characteristics

	Log Household Earnings		
	Chinese Households (1)	Non-Chinese Households (2)	Difference (1) – (2) (3)
Panel A. Total Population Size			
Higher Resettlement	0.019 (0.050)	−0.037 (0.044)	0.056 (0.040)
Higher Resettlement × Larger Population	0.094 (0.027)	0.080 (0.039)	0.014 (0.029)
Panel B. Ethnic Segregation Distance (km)			
Higher Resettlement	0.117 (0.044)	0.061 (0.032)	0.057 (0.027)
Higher Resettlement × Segregation Distance	−0.006 (0.004)	−0.013 (0.004)	0.007 (0.003)
Panel C. Chinese Population Share			
Higher Resettlement	0.127 (0.062)	0.022 (0.053)	0.104 (0.051)
Higher Resettlement × Higher Chinese Share	−0.015 (0.053)	0.018 (0.051)	−0.032 (0.041)
Panel D. Presence of Non-Chinese Chinese Speakers			
Higher Resettlement	0.080 (0.052)	0.007 (0.032)	0.073 (0.038)
Higher Resettlement × Chinese-Speaking	0.038 (0.032)	0.048 (0.033)	−0.010 (0.029)
# Households	9,634	20,549	

Notes: This table shows how the effect of county resettlement density on household income in 1980 varies with receiving-county characteristics. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Panel A interacts it with above-median Chinese share in 1947 (“Larger Population”); Panel B with above-median Chinese share in 1947 (“Higher Chinese Share”); Panel C with presence of any non-Chinese that could speak Chinese in 1980 (“Chinese-Speaking”); and Panel D with average distance between Chinese and non-Chinese primary/secondary schools in 2022 (“Segregation Distance”). Columns 1 and 2 report estimates for Chinese and non-Chinese households, respectively; column 3 reports the difference. All regressions are estimated by OLS and include the main controls: state fixed effects, expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the household. Data from the Census of Population and the Ministry of Education. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.10. Differences Between Resettled Chinese and Other Residents

	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 (0.069)	−0.125 (0.063)	−0.151 (0.023)	−105.390 (78.932)
# Observations	896	989	989	395
Panel B. Comparison with Non-Chinese				
Resettled Chinese	−0.155 (0.062)	−0.049 (0.053)	−0.080 (0.057)	−46.747 (16.517)
# Observations	3,153	3,627	3,627	1,299
Mean of Resettled Chinese	0.36	0.47	0.17	3.03

Notes: This table reports differences in characteristics between resettled Chinese and other residents living in the same state in 1988, using data from the Second Malaysian Family Life Survey. Resettled Chinese are identified as individuals not born in a New Village but who had “migrated” to one before 1960. Panel A compares them with other Chinese in the same state; Panel B compares them with non-Chinese in the same state. Each column reports the estimated difference in a separate outcome: first job in agriculture (Column 1); completion of primary education (Column 2); completion of secondary education (Column 3); and acres of land owned (Column 4). All regressions include state-by-gender fixed effects and control for age and age squared. The unit of observation is the individual for columns 1–3 and the household for column 4. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.11. Non-Migrant 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	0.084 (0.046)	–0.007 (0.033)	0.091 (0.039)
# Households	4,192	9,116	
Panel B. Log Earnings, Agriculture			
Higher Resettlement	0.053 (0.040)	–0.031 (0.040)	0.084 (0.050)
# Households	1,009	5,159	
Panel C. Log Earnings, Non-Agriculture			
Higher Resettlement	0.105 (0.048)	0.003 (0.035)	0.102 (0.039)
# Households	3,183	3,957	

Notes: This table shows the relationship between county resettlement density and non-migrant household income. “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 agriculture, comprised of agriculture and mining. Panel C restricts the sample to households whose head is employed in non-agriculture. All regressions are estimated by OLS and include the main controls: state fixed effects, expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.12. 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	−0.165 (0.090)	0.041 (0.030)
# Counties	777	777
Panel B. Negative Log Distance to Schools		
Higher Resettlement	−0.034 (0.032)	0.053 (0.026)
# Counties	777	777
Panel C. Teacher-to-Student Ratio		
Higher Resettlement	−0.032 (0.018)	−0.002 (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: expected resettlement density; an indicator for any resettlement in the county; log county area; distance to the nearest road; road density; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county. School data from the Ministry of Education. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.13. 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	0.097 (0.042)	0.038 (0.029)	0.059 (0.032)
# Households	9,634	20,549	
Panel B. Log Earnings, Agriculture			
Higher Resettlement	0.064 (0.039)	–0.005 (0.037)	0.069 (0.041)
# Households	2,197	9,359	
Panel C. Log Earnings, Non-Agriculture			
Higher Resettlement	0.112 (0.039)	0.052 (0.028)	0.059 (0.025)
# Households	7,437	11,190	

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 agricultural sector, comprised of agriculture and mining. Panel C restricts the sample to households whose head is employed in the non-agriculture sector. 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and 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 30-kilometer distance cutoff are reported in parentheses.

Table A.14. Road Access in 1961 and 1983, by County Resettlement Density

	1961 (1)	1983 (2)	Difference (2) – (1) (3)
Panel A. Distance to Main Roads			
Higher Resettlement	0.210 (0.450)	–0.189 (0.276)	–0.398 (0.382)
Mean of Outcome	5.99	3.53	–2.46
Panel B. Distance to Other Roads			
Higher Resettlement	–0.386 (0.658)	–0.402 (0.673)	–0.016 (0.533)
Mean of Outcome	10.87	7.80	–3.07
Panel C. Main Road Density			
Higher Resettlement	–0.001 (0.004)	0.016 (0.014)	0.017 (0.013)
Mean of Outcome	0.03	0.12	0.10
Panel D. Other Road Density			
Higher Resettlement	0.007 (0.003)	0.012 (0.008)	0.005 (0.006)
Mean of Outcome	0.02	0.05	0.03
# Counties	777	777	

Notes: This table shows the relationship between county resettlement density and road access in 1961 and 1983. “Higher Resettlement” is the county resettlement density defined in Section IV, standardized to have a standard deviation of 1. Each panel reports the effect of resettlement density on a different road variable: distance (km) to the nearest main roads (Panel A); distance to other roads (Panel B); main road density (Panel C); and other road density (Panel D). Column 1 reports estimates for roads measured in 1961; column 2 reports estimates in 1983; and column 3 reports their difference. All regressions are estimated using 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county. Road data from the G8031 maps. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.15. Robustness to Specifications of Counterfactual Resettlement Density

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total Population (1)	Share of Chinese (2)	Agriculture (3)	Non- Agriculture (4)	Chinese Households (5)	Non-Chinese Households (6)
1. Baseline	0.109 (0.062)	0.050 (0.011)	0.111 (0.037)	0.291 (0.129)	0.110 (0.049)	0.039 (0.031)
2. Prioritize roads over river up to 10 km	0.093 (0.067)	0.053 (0.011)	0.126 (0.040)	0.298 (0.145)	0.105 (0.050)	0.033 (0.030)
3. Minimum distance of 1 km between villages	0.093 (0.066)	0.054 (0.011)	0.126 (0.040)	0.298 (0.142)	0.105 (0.050)	0.033 (0.030)
4. Squatters within 2.5 km of forest	0.107 (0.062)	0.053 (0.011)	0.124 (0.039)	0.289 (0.133)	0.105 (0.048)	0.035 (0.032)
5. Squatters within 10 km of forest	0.100 (0.065)	0.054 (0.012)	0.129 (0.040)	0.298 (0.143)	0.099 (0.050)	0.029 (0.030)
6. Lower resettlement cost elasticity ($\psi = 0.5$)	0.095 (0.066)	0.054 (0.011)	0.122 (0.039)	0.292 (0.142)	0.105 (0.050)	0.034 (0.030)
7. Higher resettlement cost elasticity ($\psi = 0.8$)	0.094 (0.066)	0.054 (0.011)	0.126 (0.040)	0.292 (0.145)	0.103 (0.050)	0.031 (0.031)
8. Log resettlement density (resettled counties)	0.096 (0.049)	0.042 (0.013)	0.085 (0.033)	0.260 (0.155)	0.131 (0.048)	0.028 (0.036)

Notes: This table shows robustness to alternative specifications of counterfactual resettlement density in 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 agriculture in 1980–1991 (column 3), total employment in non-agriculture 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). Row 1 reports the baseline specification, which includes 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county in columns 1–4 and the households in columns 5–6. Rows 2–7 additionally control for a variant specification of expected resettlement density. Row 8 uses a log transformation for both actual and expected resettlement density, restricting the sample to resettled counties. Data from the tabulated population Census in 1980 and 1991, and the 2% sample of microdata in 1980. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.16. Robustness to Controls of Location Fundamentals

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total Population (1)	Share of Chinese (2)	Agriculture (3)	Non- Agriculture (4)	Chinese Households (5)	Non-Chinese Households (6)
1. Baseline	0.109 (0.062)	0.050 (0.011)	0.111 (0.037)	0.291 (0.129)	0.110 (0.049)	0.039 (0.031)
2. Neighboring roads	0.111 (0.061)	0.050 (0.011)	0.119 (0.037)	0.295 (0.118)	0.116 (0.046)	0.049 (0.031)
3. Neighboring population	0.116 (0.065)	0.051 (0.012)	0.111 (0.036)	0.297 (0.131)	0.111 (0.049)	0.043 (0.033)
4. Ruggedness	0.106 (0.062)	0.052 (0.011)	0.096 (0.036)	0.297 (0.128)	0.105 (0.047)	0.039 (0.032)
5. Rice and coconut suitability	0.112 (0.061)	0.051 (0.011)	0.112 (0.034)	0.264 (0.136)	0.101 (0.054)	0.047 (0.033)
6. Distance to prewar industrial sites	0.109 (0.063)	0.050 (0.011)	0.116 (0.036)	0.278 (0.131)	0.101 (0.046)	0.034 (0.031)
7. Distance to major cities	0.110 (0.063)	0.050 (0.011)	0.110 (0.037)	0.278 (0.129)	0.093 (0.048)	0.042 (0.032)
8. Road access in 1961	0.106 (0.062)	0.050 (0.011)	0.103 (0.037)	0.298 (0.138)	0.108 (0.046)	0.036 (0.031)
9. All above (rows 2–8)	0.123 (0.062)	0.053 (0.012)	0.092 (0.030)	0.290 (0.129)	0.111 (0.049)	0.058 (0.036)

Notes: This table shows robustness to adding controls for location fundamentals in 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 agriculture in 1980–1991 (column 3), total employment in non-agriculture 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). Row 1 reports the baseline specification, which includes 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. Rows 2–8 add additional controls to this baseline. The unit of observation is the county in columns 1–4 and the households in columns 5–6. Data from the tabulated population Census in 1980 and 1991, and the 2% sample of microdata in 1980. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.17. Robustness to Sample

	Population, 1980		Employment, 1980–1991		Log Earnings, 1980	
	Log Total	Share of	Agriculture	Non-	Chinese	Non-Chinese
	Population	Chinese		Agriculture	Households	Households
	(1)	(2)	(3)	(4)	(5)	(6)
1. Baseline	0.109 (0.062)	0.050 (0.011)	0.111 (0.037)	0.291 (0.129)	0.110 (0.049)	0.039 (0.031)
2. Exclude top and bottom 1% counties by area	0.101 (0.059)	0.059 (0.014)	0.112 (0.036)	0.254 (0.130)	0.112 (0.055)	0.027 (0.035)
3. Exclude top and bottom 1% resettled counties	0.103 (0.063)	0.048 (0.011)	0.118 (0.037)	0.285 (0.131)	0.110 (0.050)	0.032 (0.031)
4. Exclude 10 most densely populated towns	0.108 (0.060)	0.052 (0.011)	0.098 (0.038)	0.260 (0.146)	0.090 (0.050)	0.036 (0.035)
5. Only counties with sampled Chinese	0.077 (0.061)	0.047 (0.013)	0.101 (0.038)	0.271 (0.135)	0.110 (0.049)	0.053 (0.033)

Notes: This table shows robustness to alternative county sample restrictions in 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 agriculture in 1980–1991 (column 3), total employment in non-agriculture 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). Row 1 reports the baseline sample. Row 2 excludes the top and bottom 1% of counties by area. Row 3 excludes the top and bottom 1% by resettlement density. Row 4 excludes counties containing the ten most densely populated towns in 1947. Row 5 restricts the sample to counties with observed Chinese individuals in 2% census microdata. All regressions 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; distance to the nearest rail station; distance to coastline; Chinese population share in 1947; log population density in 1947; and the shares of land used for rubber and mining in 1944. The unit of observation is the county in columns 1–4 and the households in columns 5–6. Data from the tabulated population Census in 1980 and 1991, and the 2% sample of microdata in 1980. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.18. Over-identification Check: Correlates of Estimated Fundamentals

	Production Fundamental		Amenity Fundamental	
	Agriculture (1)	Non- Agriculture (2)	Chinese (3)	Non-Chinese (4)
Ruggedness	−0.005 (0.002)	0.000 (0.002)		
Rice Suitability	0.060 (0.181)	0.039 (0.185)		
Caloric Suitability	0.003 (0.002)	−0.001 (0.002)		
Mfg. Estab. Density	0.001 (0.019)	0.094 (0.036)		
Police Station Density			7.753 (1.637)	1.413 (1.312)
Post Office Density			8.504 (5.420)	5.683 (4.744)
Distance to Coast			−0.006 (0.004)	−0.007 (0.004)
Distance to School			−0.298 (0.064)	−0.128 (0.062)
School Density			0.684 (0.320)	−0.047 (0.329)
# Counties	685	685	685	685

Notes: This table shows how estimated production and amenity fundamentals correlate with various productivity and amenity measures listed in each row. Columns 1 and 2 use log production fundamentals as the dependent variable by sector: agriculture (Column 1) and non-agriculture (Column 2), both averaged across ethnic groups. Columns 3 and 4 use log amenity fundamentals as the dependent variable by ethnic group: Chinese (Column 3) and non-Chinese (Column 4). Ruggedness is the terrain ruggedness index (Nunn and Puga, 2012). Rice Suitability is suitability for padi rice cultivation from FAO GAEZ v4. Caloric Suitability is average agricultural suitability in terms of calories (Galor and Özak, 2016). Mfg. Estab. Density is the number of manufacturing establishments per square kilometer in 1970. Police Station Density is the number of police stations per square kilometer in 1945. Post Office Density is the number of post or telegraph office per square kilometer in 1945. Distance to Coast is the average distance to the coastline. Distance to School is the average distance to the nearest school in 2022. School Density is the number of schools per square kilometer in 2022. All regressions control for log county area. The unit of observation is the county. Conley standard errors with a 30-kilometer distance cutoff are reported in parentheses.

Table A.19. Housing Elasticity in 1989

	Log Rents (1989)	
	OLS (1)	IV (2)
Panel A. Year 1980		
Log Population	0.266 (0.054)	0.325 (0.154)
F-stat (1st Stage)		64.5
Panel B. Year 2000		
Log Population	0.267 (0.059)	0.270 (0.122)
F-stat (1st Stage)		105.7
# 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 30-kilometer distance cutoff are reported in parentheses.