ISSN No. 2454 - 1427

CDE
June 2026

Human capital shock and economic development: Evidence from Partition of India

Akansha Gupta
Email: akansha@econdse.org
Department of Economics,
Delhi School of Economics,

Working Paper No. 365

Centre for Development Economics
Delhi School of Economics
Delhi- 110007

Human capital shock and economic development: Evidence from Partition of India

Akansha Gupta*

June 29, 2026

Abstract

This paper studies the socio-economic effects of skilled refugees inflows under limited state capacity using the 1947 Partition of India as a historical natural experiment. The Partition triggered one of the largest forced migrations in history, displacing approximately 15-18 million people within a few years. Refugees were culturally and linguistically similar to natives but significantly more literate, creating a positive human capital shock. Using a newly constructed village-level panel dataset spanning six decades, I examine how this sudden influx of skilled refugees shaped long-run economic outcomes. To address potential endogeneity in settlement patterns, I employ an instrumental variable strategy that exploits variation in distance from village centers to the India–Pakistan border. The results reveal large and persistent gains in literacy lasting several decades, with particularly pronounced effects for women. In contrast, local employment structures adjust slowly, with limited evidence of rapid structural transformation. Despite this, refugee-exposed villages experience higher consumption levels and higher incomes. The effects are especially strong in districts historically linked to canal colonies, suggesting that the interaction between refugee inflows and skill composition of migrants together played an important role in shaping long-run outcomes.

Keywords: Skilled refugees; Historical persistence; Forced migration; Partition of India; State capacity

JEL classification: O15, N35, D74

I would like to thank Prof. Anirban Kar, Dr. Sandip Datta, and Dr. Sourav Sarkar for their invaluable guidance, constructive feedback, and encouragement throughout the development of this paper.

*PhD Scholar, Department of Economics, Delhi School of Economics. Email: akansha@econdse.org

1 Introduction

In the tumult of the 1947 Partition of British India, a family fleeing from Lahore to Amritsar might have lost their land, their home, and all their physical possessions. But they did not lose their knowledge. The skills of a doctor, the training of a teacher, or simply a family's deep-seated norm of valuing education were assets that could be carried across the new border when everything else was left behind.

A growing body of research examines the effects of skilled migration on host regions, yet there remains considerable debate over how such inflows shape long-run economic development. Skilled migrants may influence host economies through multiple channels, including innovation (Hunt and Gauthier-Loiselle, 2010; Akcigit et al., 2017; Bernstein et al., 2022), productivity and capital accumulation (Huber et al., 2010; Peri, 2012; Hornung, 2014), knowledge spillovers to native workers (Ottaviano and Peri, 2012; Moser et al., 2014), and aggregate economic growth (Rocha et al., 2017; Borjas, 2019). At the same time, they may compete with native workers, particularly those with similar skills, in local labor markets (Braun and Mahmoud, 2014; Morales, 2018). Despite these insights, much of the existing evidence focuses narrowly on the labor market effects of skilled migration in advanced economies. Much less attention has been paid to the development effects of skilled migration in low and middle-income countries, particularly in rural settings that migrants do not typically choose as destinations.

This distinction matters because the effects of human capital shock may depend fundamentally on state capacity. In advanced economies, migrants enter settings with established public institutions, broad access to education and relatively developed markets. In contrast, in economies with limited administrative reach, scarce public resources and weak institutional capacity, human capital brought by migrants may play a larger role in shaping local development outcomes. Migrant-borne skills, knowledge, and preferences may either compensate for weak institutions or, alternatively, their benefits may not be fully realised in low-capacity settings. However, this dimension remains largely unexplored: whether skilled migrant inflows can generate persistent development gains under limited state capacity.

Identifying the causal impact of migration is challenging due to non-random selection into migration, endogenous destination choices, and the limited availability of micro-level data on host communities. These concerns are particularly acute in the context of voluntary migration, where migrants exercise control over the timing, destination, and circumstances of movement. Forced migration, by contrast, substantially weakens these sources of bias and provides a more credible setting for causal analysis.

This paper exploits the Partition of India in 1947 as a quasi-natural experiment to examine how a large and sudden influx of literate refugees shaped the economic trajectories of receiving regions with a limited institutional reach. The Partition divided British India into two independent states- India and Pakistan (with East Pakistan later becoming Bangladesh) and triggered one of the largest forced migrations in human history, displacing approximately 15–18 million people within four years (Bharadwaj et al., 2008). Figure 3 illustrates the territorial division of British India following Partition. Movement along the western frontier involved a near-total population exchange and was roughly three times larger than along the eastern frontier, where movement into India was smaller and largely one-sided. I focus on the western front, particularly the regions of *East Punjab and the Patiala and East Punjab States Union (PEPSU)*, which experienced the most intense refugee inflows and violence. Although the overall population size in these areas changed only modestly, the demographic composition shifted sharply leading to homogenization of population in terms of religion.

Studying this setting is particularly valuable because India at Independence was characterized by severe administrative and fiscal constraints. Public resources remained limited and state-led development institutions were only beginning to emerge. In 1947, literacy rate in India was 14 percent while female literacy stood at 8 percent, reflecting extremely weak human capital formation and limited educational access. Only one out of three children had access to primary schooling and educational inequality remained closely tied to income, gender and social stratification (Ministry of Human Resource Development, 2008). Educational infrastructure was limited: in 1947–48, Punjab (including present-day Haryana) had only about 5,588 educational institutions (Census (1951)). More broadly, Foster and Rosenzweig (2000) document that even by 1960 fewer than one-third of Indian villages were located within 10 kilometers of a secondary school. This motivates a question at the center of this paper: how does human capital affect regional development when state capacity is weak and do those effects attenuate as state capacity expands?

I exploit four distinctive features of this historical episode to study the persistence of an exogenous human capital shock and trace its dynamic effects on education, labor composition, and welfare. First, refugees brought significantly higher human capital to the destination regions relative to their local counterparts. While forced migration destroys physical and social capital, displaces communities, and disrupts livelihoods, it does not erase the education, skills, and values individuals carry with them. As Bharadwaj et al. (2015) documents and I corroborate using Census (1951) in this paper, refugees were on average more literate than native residents (Figure 7). Historical accounts reinforce this picture: refugees were disproportionately skilled, entrepreneurial and engaged in non-agricultural occupations. Bharadwaj et al. (2015) states that migrants were more likely to work outside agriculture, while Bharadwaj and Mirza (2016) documents a higher concentration of refugees in commercial activities. The exodus of Hindu and Sikh communities from Lahore also triggered the collapse of banks, factories, and financial institutions these groups had dominated (Kudaisya and Yong, 2004). Together, this evidence establishes that refugees from West Pakistan possessed substantially higher human capital than local residents, generating a large and plausibly exogenous human capital shock.

Second, refugees were culturally and linguistically similar to the native population, reducing concerns related to cultural heterogeneity. Third, selection at origin was minimal because displacement was driven by religious identity rather than economic choice. According to Census of Pakistan, 1951, Hindus and Sikhs constituted only 2 percent of West Pakistan's population in 1951, down from 17 percent in 1931, reflecting the near-complete displacement of these groups. This religious basis of migration reduces the scope for positive self-selection that typically plagues studies of voluntary migration. Fourth, and most important for identification, refugee resettlement was administered by the state rather than driven by migrant choice. Receiving regions were broadly comparable except for differences in refugee inflows. Using data from the 1931 Census, I show that areas receiving larger inflows did not exhibit differential pre-Partition literacy or employment levels. Archival records further confirm that settlement was centrally coordinated by the Ministry of Rehabilitation based on the availability of evacuee land, thereby limiting destination-based sorting. This setting provides a unique opportunity to study such a shock in a low-income, historically constrained setting where rigorous identification is otherwise difficult.

To examine this empirically, I construct a novel village-level panel dataset that links the spatial distribution of refugees recorded in the 1951 Census to subsequent Census rounds from 1961 to 2011. This unique matched dataset, digitized and harmonized using archival census records, maps, and village identifiers, enables identification of within-district heterogeneity in refugee presence and reduces aggregation bias that earlier district-level studies (e.g., Bharadwaj et al., 2008; Bharadwaj and Fenske, 2012; Bharadwaj et al., 2015) could not capture. On average, refugee exposure varies by approximately 21 percentage points across villages within the same district, indicating substantial variation in refugee exposure. In addition, the inclusion of district fixed effects mitigates concerns that the findings are confounded by differences in formal political institutions across regions. The final sample consists of 15,212 villages matched across Census rounds. The availability of Census data over several decades permits an analysis of the dynamic evolution of the refugee effects, shedding light on whether refugee-induced differences persist or converge over time, an aspect largely unexplored in the existing literature.

In 1951, the average literacy rate in the sample villages was 8.1 percent, compared with 12.1 percent in rural India, reflecting low initial levels of human capital. Over the next six decades, literacy increased to 64.6 percent by 2011, gradually converging to the national rural average of 66.8 percent. Refugee exposure in the sample is substantial, as an average village hosted approximately 85 displaced persons, accounting for 16.8 percent of the village population in 1951. Throughout the paper, I classify villages above the 75th percentile (22.5 percent) of the refugee share distribution as high-inflow villages and compare them to villages that received no refugees. In 1951, no-inflow villages had literacy rates of 6.15 percent whereas high-inflow villages had significantly higher literacy rates of 9.13 percent. Disaggregating by gender, I also find higher literacy rates in high-inflow villages for both males and females.

Even though historical accounts state that refugee resettlement was exogenously administered by the Ministry of Rehabilitation, there might still be concerns about unobserved village-level characteristics such as proximity to railway stations, pre-Partition Muslim population shares, etc. that could simultaneously influence refugee inflows and long-run economic outcomes. To account for the underlying endogeneity, I instrument the share of refugees in each village with the shortest distance from the village centroid to the India–Pakistan border, which strongly predicts the intensity of refugee arrival but is plausibly unrelated to economic outcomes. I provide evidence in support of this identifying assumption by showing that distance to the border is largely unrelated to observable pre-Partition (1931) characteristics such as literacy and employment. A remaining concern is that proximity to the border could affect the composition of arriving refugees. However, archival records indicate that resettlement decisions were guided primarily by matching origin and destination districts based on evacuee land availability rather than migrant-specific characteristics, thus limiting systematic sorting on unobservables. To examine the dynamic effects of the refugee shock over time and to avoid weak instrument concerns associated with interacting a time-invariant instrument with Census years, I estimate regressions separately for each Census round.

The results reveal large and persistent effects on human capital accumulation. Villages with higher refugee presence in 1951 exhibit significantly higher literacy rates for nearly five decades following Partition. The OLS estimates indicate a persistent gap in literacy rates of approximately 1 to 4 percentage points, peaking around 1971 and gradually converging thereafter. The IV estimates suggest even larger effects — for instance, against a sample mean of 24.02 percent, refugee-affected villages exhibit literacy rates more than 20 percentage points higher than villages without refugees in 1971. The effect is dampened in the twenty-first century with the expansion of nationwide literacy programs such as the National Literacy Mission (1988) and the near-universalization of elementary education. This gradual expansion of state capacity is important for interpretation of the results. In the early decades after Partition, refugee-borne human capital may have substituted for weak public provision and generated large spillovers in literacy. As state provision of education and social infrastructure expanded over time, however, the importance of these initial private human capital shocks likely diminished, contributing to the convergence in literacy outcomes in later decades. These findings are robust to the inclusion of district fixed effects and controls for geography, population density, and climatic conditions.

I then examine if the results are driven by the supply-side expansion of educational infrastructure in refugee-receiving villages. While OLS estimates suggest that high-inflow villages have a greater number of primary, middle, and secondary schools across Census years, IV estimates reveal either null or negative effects. This ambiguity indicates that supply-side expansion alone is unlikely to explain the persistent gains in literacy.

In contrast to educational outcomes, I find little evidence of sustained improvements in local employment rates or occupational structure. High-inflow villages experience a decline in the share of cultivators and a rise in agricultural laborers, consistent with land redistribution following Partition that left many refugee households with smaller or less productive landholdings. This suggests that while human capital accumulated significantly, the local economy did not undergo a parallel structural transformation capable of fully absorbing more educated workers. I then examine welfare outcomes. After accounting for endogenous settlement patterns, high-inflow villages exhibit higher per capita consumption and lower poverty rates. Moving from no refugees to the 75th percentile of refugee exposure increases per capita annual consumption by ₹13,663, representing a 55.8 percent increase over the sample mean in 2011.

Next, I examine heterogeneity in the effects of the human capital shock across gender. The results reveal disproportionately larger and more persistent improvements in female literacy in high-inflow villages, suggesting that refugee households may have transmitted norms that particularly improved girls' schooling opportunities. One possible explanation could be that lower baseline literacy among women, 2.6 percent in 1951, allowed refugee exposure to generate larger gains. Another potential channel relates to the demographic disruptions caused by Partition, which created a sizable population of *"unattached women,"*[1] which may have raised the perceived returns to female education. This mechanism is in line with the *"uprootedness hypothesis"*, which posits that

[1]This group included (i) women separated from their families or abducted during flight and (ii) widowed or destitute women who lost male household members in the violence. In a note dated December 1949, Rameshwari Nehru stated that the number of "unattached" women looked after by the government in October 1948 was 45,374. (Menon and Bhasin, 1998)

displacement induces greater investment in education as a means of restoring security and mobility (Becker et al., 2020). Despite these gains in education for females, I find no systematic differences in overall employment rates of men and women across high and low inflow villages. However, occupational structure differs by gender. High-inflow villages have fewer cultivators among both men and women but men experience an increase in agricultural labor and a modest shift toward other services whereas women are disproportionately more likely to be engaged in the services sector only.

Historical evidence indicates that a substantial share of refugees who settled in East Punjab were return migrants, commonly referred to as colonists (*abadkars*), from the canal colonies of West Punjab (Singh, 1952; Kaur, 2007). These canal colonies were established in the late nineteenth century under a large-scale irrigation and settlement program to expand agricultural production and fiscal revenues by allocating irrigated land to cultivators from East Punjab. Following Partition, Hindu and Sikh colonists were forcibly displaced from these regions and returned to their districts of origin. Having lived in relatively more developed agricultural and institutional environments prior to Partition, these households plausibly carried higher levels of human capital and stronger preferences for schooling, which may have shaped how refugee inflows translated into local development outcomes after resettlement.

I examine this hypothesis through a heterogeneity analysis that compares the effects of refugee exposure across districts that historically supplied colonists to the canal colonies and those that did not. The results show that villages located in colonist-origin districts exhibit significantly higher levels of human capital, and that the interaction between refugee exposure and colonist-origin districts is positive and statistically significant for literacy. This indicates that the literacy gains documented in the baseline analysis are concentrated in districts with historical exposure to canal colonization. Consistent with this pattern, high-inflow villages in colonist-origin districts display a greater shift toward service-sector employment.

The paper also explores whether the long-run effects of refugee inflows differed across regions with distinct colonial histories, comparing districts under direct British rule with those formerly under princely states (PEPSU). The results show larger and more persistent literacy gains in directly ruled districts, whereas shifts toward non-agricultural and permanent employment are broadly similar across both institutional settings, suggesting that these historical institutions mattered for how local economies responded.

Finally, I conduct several robustness checks. First, I replace the binary treatment variable with (i) the share of displaced persons in a village and (ii) an alternative dummy comparing villages with no refugees to those above the 90th percentile of refugee exposure. The results remain consistent in both direction and significance. Second, I control for 1931 tehsil-level literacy rates to assess whether the observed effects simply reflect initial differences rather than divergence following refugee inflows. The findings continue to show a positive and persistent impact on literacy over subsequent decades. Third, I exclude villages located close to the border to ensure that border villages are not driving the results. Fourth, for the period 1991–2011, I compute literacy rates using the population aged six and above [2] and the results remain robust. Next, I address concerns regarding changes in village boundaries over time or errors in matching villages

[2] Data for village population above 6 years of age is not available for the preceding Census rounds.

across years. The main analysis restricts the sample to villages experiencing less than a 10 percent change in area relative to 1951, and there is no substantive change in findings when tightening this threshold further to 5 percent and 1 percent. Finally, I also include villages that became towns in later decades and the results do not change much.

This paper makes several contributions to the literature. By exploiting the Partition of India as a historical natural experiment, it contributes to research on skilled migration, forced displacement, and the persistence of historical shocks in a developing economy. It traces the dynamic evolution of a refugee-induced human capital shock over six decades in a low income country, a dimension that has received limited attention in existing work. It also highlights the interaction between human capital and state capacity, showing that migrant-borne human capital may matter most when institutions are weak. Methodologically, the paper introduces a newly digitized village-level panel dataset spanning 1951–2011, which allows the exploitation of within-district variation in refugee presence that earlier district-level studies could not capture and helps rule out confounding institutional differences. The analysis further documents gender-disaggregated effects, revealing heterogeneous responses across men and women that remain underexplored in the forced migration literature. Finally, the paper uncovers important heterogeneity linked to historical canal colonization highlighting how pre-existing regional characteristics shape the long-run impacts of refugee shocks. Taken together, the findings demonstrate how a major demographic upheaval can leave a lasting imprint on human capital formation and development trajectories.

The remainder of the paper proceeds as follows. The following section discusses the relevant literature. Section 3 provides historical background on the Partition and the refugee resettlement process. Section 4 presents data summary and the construction of the panel dataset. Section 5 outlines the empirical strategy and discusses identification. Section 6 reports the main results followed by heterogeneous effects in section 7. Section 8 presents robustness checks. Section 9 provides discussion and concludes.

2 Literature Review

A broad literature seeks to explain why some regions attain higher levels of development than others. Much of this work emphasizes geography, institutions, and human capital as central determinants of long-term economic outcomes. This paper focuses on how the movement of human capital shapes economic development. Human capital migration can either be voluntary or forced depending on the motivations underlying migration decisions. Voluntary migrants typically retain agency over their decision to migrate, their choice of destination, and the timing of departure, making migration decisions potentially correlated with unobserved individual characteristics and local economic opportunities. Forced migrants, by contrast, are displaced by conflict, persecution or political upheaval with limited choice over destination. Although both forms of migration can transmit skills and knowledge across regions, forced migration settings provide cleaner identification strategy to estimate the causal effects of human capital movements because migration decisions are less likely to be driven by economic incentives.

This paper contributes to the growing literature on the economic consequences of forced displacement. Existing studies examine the effects of forced migration on employment (Kondylis, 2010; Bauer et al., 2013; Braun and Mahmoud, 2014; Ruiz and Vargas-Silva, 2015; Braun and Stuhler, 2024), education (Becker et al., 2020; Van der Werf, 2021; Morales, 2022), and long-run development (Acemoglu et al., 2011; Arbatli and Gokmen, 2018; Nakamura et al., 2022). Much of this research focuses on post–World War II expulsions and their long-run impact on host communities. For instance, Becker et al. (2020) document that descendants of migrants attained higher levels of education, consistent with the *uprootedness hypothesis*. Similarly, Braun and Mahmoud (2014) and Bauer et al. (2013) analyze labor market outcomes for displaced and native populations respectively. They found that there is a significant decline in native employment in the short run, however, it was transitory and in the long run there were negative consequences for the displaced. Much of this literature, however, focuses on advanced economies and contemporary economic outcomes. By studying the Partition of India, this paper combines the identification advantages of a forced migration setting with the economic significance of a large inflow of relatively skilled migrants in a developing-country context. Unlike much of the existing literature, which primarily examines refugee inflows as labor supply shocks, this paper studies forced migration as a human capital shock and investigates how refugee-borne skills and literacy shaped the development trajectory of receiving regions.

Another strand of literature explores the effects of voluntary skilled migration and the net impact is ambiguous: while it may generate substantial positive spillovers through the transfer of knowledge and human capital, it can also impose adjustment costs on native workers in the form of labor market competition (Maystadt and Verwimp (2014); Morales (2018); Braun and Mahmoud (2014)). Borjas (2019) argues that high-skilled immigration contributes more strongly to growth, while Rocha et al. (2017) finds that more educated immigrants in Brazil improved local income and educational attainment. A key empirical challenge throughout is separating the effects of migrant human capital from positive selection bias, since skilled migrants tend to choose destinations with better institutions and greater economic opportunity, making causal identification difficult. Only a handful of papers exploit historical migration shocks to address this challenge, most notably Rocha et al. (2017) and Hornung (2014). This paper contributes to this literature by examining a large inflow of relatively skilled refugees into a low-income agrarian economy. Since resettlement during the Partition of India was centrally administered and driven by religious identity rather than economic choice, the setting provides plausibly exogenous variation in the spatial distribution of human capital.

Whether migrant human capital translates into persistent gains depends critically on local institutions and state capacity. Besley and Persson (2009) show that low-capacity states may fail to provide the complementary public inputs necessary for private skills to generate sustained development. In addition, Acemoglu et al. (2014) emphasize that institutions and human capital evolve jointly, making it difficult to separate their effects on development. Evidence from forced migration settings also suggests that refugee impacts vary with local institutional conditions. Using the Syrian refugee influx into Turkey, Aracı et al. (2022) find that adverse labor market effects are weaker in more developed regions. In this paper, I find that right after Independence, when India suffered from limited state capacity particularly in rural areas, skilled migrants played a major role in the improvement of literacy rates. However, these gains gradually started converging as provision of public education expanded.

This paper also relates to studies that examine the persistence of historical shocks and their long-term effects on contemporary economic outcomes. Acemoglu et al. (2001) show that colonial institutions shaped long-run income differences whereas Dell (2010) documents persistent negative effects of extractive institutions in Peru and Bolivia. Other studies highlight persistent positive legacies of historical interventions (Dell and Olken, 2020) and the intergenerational transmission of cultural and social norms (Alesina et al., 2020). While this literature typically evaluates outcomes long after the historical shock has occurred, less is known about how such effects evolve over time. This paper contributes by tracing the dynamic effects of a large human capital shock from its onset in 1951 through 2011, allowing me to examine not only whether the effects persisted, but also how they changed across successive decades.

In the Indian context, a number of studies have examined how colonial institutions shaped long-run development. Banerjee and Iyer (2005) show that districts where the British introduced landlord-based property rights exhibit persistently lower agricultural productivity and investment compared to regions where rights were granted directly to cultivators. Similarly, Iyer (2010) finds that areas under direct British rule experienced significantly lower levels of socio-economic development in the post-colonial period relative to regions governed by princely states. In contrast, I find that areas under direct British rule experienced stronger and more persistent gains in literacy and welfare relative to princely states. These results suggest that pre-existing institutional structures shape how regions absorb and respond to migration shocks.

Building on this role of historical context, I also document substantial persistence in the effects of the refugee shock over time. The effects for females persist even after six decades, suggesting that the transmission of human capital may operate differently across gender. Further, I examine heterogeneity across districts with historical exposure to canal colonies established in the late nineteenth century. I show that the positive effects of refugee inflows are majorly concentrated in districts that historically sent colonists to these canal colonies and subsequently received them back after Partition. This finding highlights how historical events shape the long-run impact of subsequent demographic shocks.

Finally, this paper contributes to the emerging literature on the economic consequences of the Partition of India. Bharadwaj et al. (2008),Bharadwaj and Fenske (2012), and Bharadwaj et al. (2015) document the scale of migration flows and their long-run demographic and economic consequences. Related work shows that refugee inflows affected agricultural productivity (Bharadwaj and Mirza, 2016) and jute cultivation (Bharadwaj and Fenske, 2012). More recent contributions by Mirza (2022) and Ayesh (2024), in the context of Pakistan, study the effect of mass displacement due to Partition on literacy rates and educational attainment of displaced children respectively. Other studies examine social cohesion (Bhattacharya and Mukhopadhyay, 2023); identity (Bhattacharya et al., 2023); marriage age and fertility (Lu et al., 2021); and violence (Jha and Wilkinson, 2012).

This paper differs from existing literature in three important ways. First, I focus on Punjab and Haryana, the regions that received the largest refugee inflows along the western border, thereby allowing for sharper identification of the effects of refugee settlement. Second, I construct a village-level panel dataset from 1951 to 2011, which captures localized variation in migration and enables the analysis of within-district differences and dynamic effects. Third, while existing work largely relies on correlations, this paper explicitly addresses endogeneity in migration flows using an instrumental variable strategy.

This study complements the existing literature in several ways. The Partition of India serves as a quasi-natural experiment, allowing us to identify the causal impact of skilled refugees shock on destination regions with limited state capacity. Prior research has documented that refugees from West Pakistan possessed higher literacy levels relative to natives (Bharadwaj et al., 2015), a finding that I corroborate using newly digitised data. This paper departs from the district-level analysis common in this literature by constructing a novel village-level panel dataset that allows me to trace the trajectory of the effects over time. The increased granularity also enables precise and reliable estimation of the economic effects of the refugee influx. A key contribution of this work is the digitization and matching of historical data spanning 1951–2011 for the states of Punjab and Haryana. This allows me to analyze both the immediate and long-term dynamic effects of the refugee shock, providing insights into the persistence and evolution of its impact. In addition, the paper uncovers important heterogeneity across gender and institutional contexts, providing new insights into how displacement shapes development trajectories over time.

3 Historical Background

The Partition of India resulted in one of the largest mass migrations in human history. For nearly two centuries before partition, the subcontinent was governed under the British rule. As the British Raj prepared to end colonial rule, political tensions between the Indian National Congress and the All-India Muslim League intensified over the future political structure of the subcontinent. The Muslim League, led by Muhammad Ali Jinnah, demanded a separate state for Muslims, arguing that they would face political marginalization in a unified India. In August 1947, the British administration under Lord Mountbatten partitioned British India into two independent dominions - India and Pakistan based largely on religious majorities (Figure 2). The hurried demarcation of borders by the Radcliffe Commission triggered widespread communal violence and mass displacement. Approximately 17.9 million people crossed the new borders over four years, however nearly 3.4 million individuals were unaccounted for at their intended destinations including those who died during migration or migrated elsewhere (Bharadwaj et al., 2008).

Partition generated migration along two fronts: the western border (between Pakistan and Indian Punjab) and the eastern border (between East Pakistan, now Bangladesh, and West Bengal) as depicted in Figure 3. The western migration was substantially larger and characterized by a near-complete exchange of populations, with Hindus and Sikhs moving into India and Muslims moving into Pakistan. In contrast, migration across the eastern border was more gradual and predominantly one-directional. This paper focuses on western border movement, particularly the refugee inflows into Punjab and PEPSU (Patiala and East Punjab States Union), where refugees constituted nearly 17 percent of the population in 1951. The scale and suddenness of this demographic shock make Partition an important setting for studying the long-run economic consequences of forced migration and migrant-borne human capital.

The partition of the Indian subcontinent was executed in a highly disorganized manner. Official boundaries were announced two days after independence by Sir Cyril Radcliffe, leading to many individuals unexpectedly finding themselves on the "wrong" side of the border. Although historians have documented Partition as a new constitutional or political arrangement, its impact went far beyond that, deeply reshaping the core structures of Indian society. For refugees, it marked a profound new beginning, leading to a radical reconstitution of communities that significantly altered the constitutional, political, and social landscapes of both northwestern and northeastern India (Pandey, 2001). The long-term effects on subsequent generations, however, remain underexplored.

3.1 Government Action in East Punjab

Understanding the resettlement process that determined the spatial distribution of refugees is central to this study, as this variation serves as the primary source of identification in the analysis of the impact on economic outcomes. The government of East Punjab constituted the Military Evacuation Organisation (MEO) on 1st September, 1947 to oversee refugee movement, protection during transit, and welfare in concentration camps (Singh, 1961). Simultaneously, a Rehabilitation Department was created to manage refugee resettlement and land allocation.

By mid-September 1947, evacuee agricultural land left behind by departing Muslims was distributed temporarily to incoming refugees for the 1947-48 agricultural seasons, irrespective of previous landholdings in Pakistan (Rural Rehabilitation, Punjab 1950). These allotments were made on the group basis, often allowing refugees from the same localities in West Punjab to settle together. This temporary system was later replaced by the *quasi-permanent allotment* scheme which linked refugee allotments to pre-Partition landholdings in Pakistan. Eligibility extended to former landowners, occupancy tenants, and certain categories of cultivators recognized under colonial land laws (Rural Rehabilitation, Punjab 1950). During the quasi-permanent allotment process, a large number of temporary allottees were confirmed in the villages where they had initially been settled.

To formalize these efforts, the East Punjab Refugees (Registration of Claims) Act, 1948 [Act 8 of 1948] and East Punjab Refugees (Registration of Land Claims) Act, 1948 [Act 12 of 1948] were enacted, allowing refugees to formally register losses incurred after March 1, 1947. Claims were often inflated by around 25 per cent, necessitating verification against revenue records received from Pakistan. By mid-1949, the administration finalized the framework for permanent land allocation. Land abandoned by Hindus and Sikhs in West Pakistan, along with evacuee land in East Punjab and PEPSU, was converted into 'Standard acres'[3] for allocation. Graded cuts were imposed so that larger landowners received proportionately smaller compensation (Singh, 1952).

By April 1950, quasi-permanent allotment had largely been completed (Randhawa, 1954). Consequently, when the Census enumeration began in February 1951, most refugees had taken possession of their allocated land. Consequently, the 1951 Census data provides a reliable estimate of the number of displaced persons in each village. This timing is important for the empirical strategy. The Census records refugee presence

[3]"Standard acres" was a unit of measurement adopted during the post-partition resettlement to account for variations in land fertility and irrigation, allowing for a fair redistribution or valuation of land.

soon after administrative allocation and before the economic outcomes examined in this paper, reducing concerns about reverse causality. The spatial variation in refugee exposure therefore reflects a coordinated state-led resettlement process rather than endogenous location choice alone.

3.2 Resettlement scheme

The resettlement scheme was designed with the guiding principle of relocating refugees from specific regions in Pakistan to designated districts in East Punjab and PEPSU, on the basis of availability of evacuee land (Singh, 1952, 1989). Flyers in Urdu and Punjabi were widely distributed in relief camps to inform refugees about the origin-destination allocation of land. This approach aimed to avoid overcrowding and to preserve social networks by allowing refugees to settle alongside people from similar backgrounds.This communal support played a crucial role in the psychological rehabilitation of refugees. (Randhawa, 1954)

Table 8 summarizes the intended allocation of refugees across districts. Detailed land allocation plans are outlined in table 9. To evaluate whether the allocation rule was implemented as planned, I use displaced persons origin (D-V) tables from the 1951 District Census Handbooks. These tables provide the number of displaced persons in each district, categorized by their district of origin and date of arrival in India. This allows me to directly compare planned allocations with realized settlement patterns. Although the allocation tables report land area while the Census reports population, the two measures align closely such that districts that allocate more land, in general, also received more refugees.

The data show that the allocation rule largely guided settlement patterns. For example, the highest number of refugees from Montgomery and Lahore went to Ferozepore whereas the maximum number of refugees from Sialkot went to Gurdaspur, as allocated. Figures 4, 5 and 6 illustrate these origin-destination patterns across districts.

Although some deviations from the allocation rule occurred, these too appear largely driven by administrative decisions and historical migration networks. Refugees from Bahawalpur and Multan, for instance, were officially assigned to Hissar but many settled in Ferozepore and Karnal instead. One explanation is that canal colony settlers returned to their ancestral districts after Partition. Since many canal colonies in West Punjab had originally been populated by migrants from districts such as Amritsar, Hoshiarpur, Ludhiana, and Ferozepore, these historical linkages likely influenced settlement choices (Randhawa, 1954). Similarly, refugees from Attock settled disproportionately in Ambala despite Ludhiana being the designated destination. The *Resettlement Manual* reports that larger land allotments were initially made available in Ambala, which likely shaped these outcomes (Singh, 1952).

These patterns suggest that government's land allocation policy, rather than individual choice, played a central role in determining refugee locations. Hence, concerns regarding selection bias are alleviated to an extent. The main treatment variable, i.e., refugee presence is exogenously determined by government rule.

However, several other factors influenced the flow of displaced persons. One key factor was the distance from the Partition border, as noted by Bharadwaj et al. (2008). Figure 1 shows that the proportion of displaced persons in a village are inversely proportional to the distance of village to the border. Another critical determinant was proximity to railway stations, with refugees who travelled by rail more likely to settle in areas closer to railway stations.

Additionally, the resettlement of agricultural migrants was influenced by the presence of their *village biradaris* (kinship groups). In many cases, relatives played a guiding role, directing migrants to settle in villages where evacuee property was available. Finally, the pull of large cities also shaped resettlement patterns. Bharadwaj et al. (2008) found that the presence of a major city in a district increased refugee settlement by 0.5 percent. Taken together, these factors may have introduced some deviation from the allocation rule but do not overturn the dominant role of state-led resettlement.

3.3 Government assistance

In addition to land allotment, the government provided refugees with financial support to restore refugee livelihoods. Refugees received loans for seeds, bullocks, agricultural implements, food and small business activity. In January 1948, the government introduced subsidized loans for repairing houses and wells Department, 1950. By September 1948, more than ₹1.45 crore had been distributed as loans and grants in rural areas Department, 1950.

Education posed a particularly serious challenge after Partition. The number of displaced Hindu and Sikh students in East Punjab far exceeded the educational capacity left behind by departing Muslims (Kaur, 2010). In response, the government appointed education officers for refugee camps, rehabilitated displaced teachers and introduced flexible admission procedures that allowed students to enroll without formal certificates. Refugee students were exempted from tuition fees, while scholarships awarded in West Punjab continued after migration (Nair, 1950). By December 1949, the government had distributed more than ₹23 lakh in educational loans and ₹35 lakh in grants (Nair, 1950).

The government also extended targeted support to disadvantaged groups. Under the *Harijan Welfare Scheme*, it provided stipends to students from marginalized communities enrolled in recognized institutions, exempted them from tuition fees, and covered the costs of public examinations (Kaur, 2010). These measures indicate that rehabilitation policy not only addressed displacement but also incorporated redistributive objectives.

A limitation of this study is that I cannot separately identify the effects of government assistance from the direct effects of refugee human capital. However, these programs largely aimed to restore baseline living conditions rather than provide persistent advantages to specific regions, suggesting that they are unlikely to fully account for the long-run differences observed in the analysis.

3.4 Refugees v/s Natives

This section examines how refugees differed from the native population along key socio-economic dimensions. Historical accounts consistently describe refugees as hardworking, entrepreneurial, and economically dynamic. Randhawa (1954) emphasized the resilience and adaptability of displaced Punjabi peasants, while Yadav (2012) document how refugee farmers introduced improved irrigation and cultivation practices. Refugees also worked longer hours and displayed strong entrepreneurial tendencies despite facing adverse conditions (Kumari, 2013; Yadav, 2012). These patterns suggest that refugees not only increased labour supply but also brought valuable skills and work practices that shaped local economies.

The occupational composition of refugees differed substantially from that of natives. About 70 percent of natives remained in agriculture, in contrast, 70 percent of refugees worked in manufacturing and services (Mehra, 2020). Similarly, Bharadwaj et al. (2015) show that displaced were substantially less likely to work in agriculture relative to residents. Refugee traders and shopkeepers also expanded commercial activity in rural areas (Randhawa, 1954). These patterns suggest that refugee inflows may have accelerated structural transformation in receiving regions. I also find that treated villages have significantly lower number of cultivators and a gradual shift towards services sector.

Turning to human capital, Bharadwaj et al. (2015) provide anecdotal evidence that refugees from West Punjab had higher literacy rates than natives. Building on this insight, I construct direct comparisons of refugee and native literacy rates using 1951 Census data at the *tehsil* level for 12 districts.[4] Figures 7 and 8 show that refugees had substantially higher literacy rates than natives in rural areas across all districts and in most urban areas as well. These findings remain robust when I extend the comparison to districts in Madhya Pradesh.[5]

The higher literacy and skill levels among refugees likely reflect their origins. Most refugees came from the northern and central parts of West Punjab (now in Pakistan), which had relatively higher human capital endowments relative to other parts of West Punjab (Cheema et al., 2008). In contrast, western and southern regions of Punjab were characterized by higher poverty and greater income inequality (Afzal and Malik, 2010). The concentration of canal colonies in the northern and central regions may have further contributed to stronger human capital in these areas. Motivated by this historical context, I later examine heterogeneity across districts historically linked to canal colonies.

3.4.1 Gender Differences

Partition also altered gender norms and female economic participation. Historical accounts suggest that refugee women, particularly from urban areas of West Punjab, introduced more liberal social norms into relatively conservative communities in East Punjab (Randhawa, 1954). Chatterji (2007) argue that displacement weakened rigid caste and gender norms and increased households' reliance on women's income. This shift encouraged female labor force participation and improved educational outcomes for women. Institutional reforms reinforced these changes, including the removal of restrictions on women appearing for competitive administrative examinations after 1950.

Consistent with this, Menon and Bhasin (1998) note that displacement "paved the way for women's abrupt entry into the economic life of the country," with women taking up roles as workers, professionals, and students. They emphasize that economic necessity and the breakdown of traditional family structures accelerated this transition. These qualitative accounts point to a structural shift rather than a temporary adjustment.

[4]Separate tables for natives and refugees literacy levels were not available for all the districts.

[5]Results are available upon request.

The data support these historical narratives. Figures 9 and 10 show that refugee literacy exceeded native literacy for both men and women in rural areas. Table 1 further confirms that these differences are statistically significant in rural areas, although they are not significant in urban settings. Importantly, the literacy gap is more pronounced for females, suggesting that refugee inflows may have had stronger effects on female human capital. Consistent with this, I later find persistent positive effects on female literacy and occupational outcomes in section 7.1. Together, these findings suggest that refugee inflows not only increased average human capital but also reshaped gender roles and opportunities in receiving regions.

Table 1: Difference in Literacy Rates by Displacement Status

	Obs.	Mean	Difference	Std. Error	t-statistic	p-value
Panel A: Rural areas						
Female Literacy Rate						
Refugees	43	0.105	0.086	0.011	7.90	0.000
Natives	43	0.020				
Male Literacy Rate						
Refugees	43	0.254	0.153	0.025	6.25	0.000
Natives	43	0.101				
Panel B: Urban areas						
Female Literacy Rate						
Refugees	14	0.287	0.089	0.05	1.75	0.088
Natives	14	0.198				
Male Literacy Rate						
Refugees	14	0.415	0.033	0.05	0.65	0.515
Natives	14	0.382				

Notes: This table reports two-sample t-tests with equal variances comparing literacy rates between refugees and natives at sub-district (tehsil) level. Panel A reports results for rural literacy rates and Panel B reports results for urban literacy rates.

4 Data

The primary source of data used in this paper is the Census of India (1951-2011). This long-run panel enables an analysis of the dynamic effects of Partition over seven decades. The analysis focuses on the former East Punjab and PEPSU regions, corresponding to present-day Punjab and Haryana.[6] These regions experienced the highest inflows of displaced from West Pakistan, enabling a precise estimation of displacement's impact on economic outcomes. Restricting the analysis to these areas also avoids introducing unnecessary noise from regions unaffected by Partition-related migration.

[6]The "East Punjab and PEPSU" designation here refers specifically to the Punjab plains division and PEPSU, excluding Delhi, Bilaspur, Himachal Pradesh, and the Himalayan Punjab Division because of their distinct geographical and economic conditions.

A key contribution of this paper is the construction of a novel village-level panel dataset for these states spanning 1951–2011. While previous studies have examined the effects of Partition flows at the district level (Bharadwaj et al., 2015; Bhattacharya and Mukhopadhyay, 2023; Jha and Talathi, 2024), I utilize village-level data to do the analysis at a more granular level. The data shows substantial dispersion in the share of displaced persons within a district. On average, the standard deviation of refugee exposure within districts is 21 percentage points indicating that some villages received almost no refugees while others experienced substantial inflows. This localized variation underscores the importance of conducting the analysis at a finer spatial scale.

The sample is restricted to rural areas, as urban areas could not be consistently matched across Census rounds.[7] As an additional exercise, I include villages that later transition into towns and the results remain unchanged. The final matched sample consists of 15,214 villages distributed across 70 tehsils and 19 districts in 1951, resulting in a panel of 106,498 village-year observations. To minimize concerns arising from administrative boundary changes, the baseline analysis uses villages with less than a 10 percent change in geographical area relative to 1951 (N = 94,268). Results remain robust under stricter thresholds of 5 percent and 1 percent area change. Table 10 in the Appendix shows that the final sample is broadly representative of the underlying population.

The dataset was constructed by digitizing village-level Primary Census Abstracts and matching villages across Census rounds using village names, village codes, and geographical area. Detailed information on digitization, matching procedures, data cleaning, and linkage with the SHRUG database is provided in Appendix 9.1.

4.1 Outcome variables

In this paper, I investigate the long-term impact of population inflows following the Partition of India on educational, labor market and welfare outcomes. My key education variable is literacy rate, defined as the proportion of literates in the total population of a village in percentage terms.[8] Literacy is the only education-related indicator consistently available at the village level for all seven census rounds between 1951 and 2011, making it particularly suitable for studying long-run educational change. To complement literacy-based measures, I also examine access to schools and colleges from the Census and educational qualifications from the Socio-Economic Caste Census (SECC) 2011.

In 1951, the average literacy rate in sample villages was 8.07 percent, compared to the national average of 18.32 percent. By 2011, literacy in sample villages had increased to 64.58 percent, relative to a national average of 72.98 percent (Ministry of Home Affairs). Even relative to rural average, the sample villages remained slightly below the national rural average over time. Figure 12 illustrates literacy rates across districts in 1951, while Figure 13 presents the literacy trends in the sample over seven decades.

[7] Urban areas use ward numbers instead of village names and these ward numbers are inconsistent across years, preventing their inclusion in the sample.

[8] Since information on age-specific populations (5 years and above) is not available at the village level for all the years, literacy is measured using the total population as the denominator.

For labor market outcomes, I use Census data from 1991–2011[9] to construct measures of overall employment, employment type (main versus marginal workers)[10] and sectoral employment composition. Sectoral categories include cultivators, agricultural laborers, household industry workers, and service-sector workers. These measures allow me to study both labor force participation and structural transformation.

To assess long-run welfare outcomes, I draw on data from the Socio-Economic Caste Census (SECC) 2011. The SECC provides information on rural household consumption, per capita consumption, poverty rates, and the share of households with monthly income above ₹5,000 and ₹10,000. These measures enable an examination of whether refugee inflows were associated with improvements in living standards and upward mobility in receiving regions. In addition, as a proxy for regional economic activity, I also use satellite-based nightlights data to capture spatial variation in economic development across receiving regions .

4.2 Refugee shock

The Census of India (1951) recorded information about displaced persons and their districts of origin. Displaced persons were defined as: *"Any person who has entered India having left or been compelled to leave his home in Western Pakistan on or after 1 March 1947, or his home in Eastern Pakistan on or after 15 October 1946, on account of civil disturbances or the fear of such disturbances, or on account of the setting up of the two Dominions of India and Pakistan."*

The principal explanatory variable is the *refugee shock*. The baseline measure is a binary indicator, *treatment_75*, that classifies villages into "high inflow" and "no inflow" categories. High-inflow villages are defined as those in which refugees constitute at least 22.5 percent of the total population (75th percentile), while no-inflow villages are those that received no refugees. Villages with intermediate levels of refugee exposure are excluded from this definition by construction, resulting in a reduced sample of 8,854 villages for analysis using *treatment_75*. [11] Figure 11 shows the spatial distribution of treated and control villages. For robustness, I also employ alternative measures of refugee inflows: (i) the proportion of displaced persons, defined as a continuous variable, and (ii) a binary variable *treatment_90*, which equals one for villages with refugee share greater than or equal to 63 percent of the village population (90th percentile) and zero for villages with no displaced persons.

Figure 1 illustrates the spatial distribution of displaced persons across villages by proportion of displaced. As shown, refugee inflows were disproportionately concentrated in villages closer to the border. To address concerns of potential endogeneity, I leverage this geographical variation by constructing a measure of distance to the Indo–Pak border, defined as the shortest distance between the centroid of each village and the

[9] The employment data for preceding years is not available in the digitized format and hence have not been included in this paper.

[10] Main workers are individuals engaged in economically productive activity for six months or more during the reference year and marginal workers are employed for less than six months, often seasonal or part-time workers.

[11] The choice of this threshold reflects the idea that migration effects may arise only beyond a certain intensity. Small inflows can be absorbed with limited economic impact. In the sample, approximately 30% of villages received no refugees, indicating that many villages experienced little to no exposure to the shock.

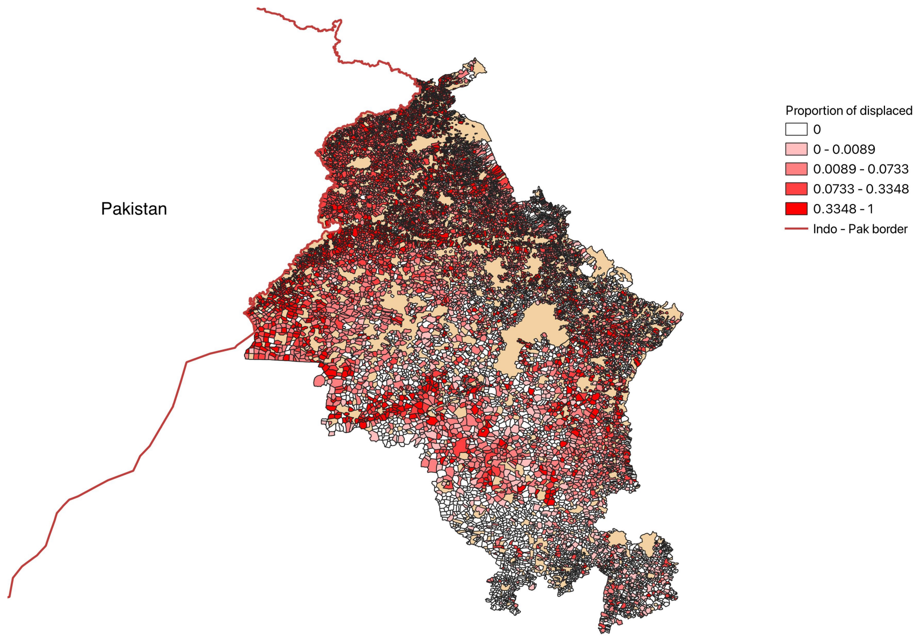


Figure 1: Spatial distribution of displaced persons
Notes: This map shows the spatial distribution of displaced persons across sample villages (N = 14,596) in Punjab and PEPSU in 1951. Refugee exposure is measured as the share of displaced persons in the village population and is grouped into quintiles. The color gradient represents increasing intensity of refugee exposure, as indicated in the legend.

border. This measure was computed in QGIS using a self-digitized shapefile of the Indo–Pak border. On average, villages in the sample are located 154 kilometers from the border.

Descriptive statistics in Table 4.2 indicate that, on average, a village hosted about 85 displaced persons, constituting 16.8 percent of the village population. The gender composition of refugees was relatively balanced, with males representing 54 percent and females 46 percent.

4.3 Control variables

The analysis includes a set of demographic and geographic controls that may influence refugee settlement and long-run development outcomes. These comprise demographic

Table 2: Summary statistics - Refugee shock

Variable	Obs	Mean	SD	Min	Max
Total displaced	15212	84.84	202.4	0	7281
Male displaced	15212	45.56	107.05	0	3771
Female displaced	15212	39.29	95.796	0	3510
Proportion of displaced in a village	15212	0.168	0.256	0	1
Treatment_75	8854	0.429	0.495	0	1
Treatment_90	6569	0.231	0.421	0	1
Distance to border (in metres)	14594	153596.2	109784.7	0	403616.8

Table 3: Summary statistics - Control variables

Variable	Obs	Mean	SD	Min	Max
Area of village (in sq. miles)	100,361	1.97	4.19	0	927.57
Total population	101,514	1347.13	4223.87	0	886,519
Population density	100,277	925.45	17,444.46	0	4,371,110.5
Sex ratio	100,926	883.48	111.64	0	8663.64
Total rainfall (in mm)	102,480	722.98	407.81	0	3245.21
Average elevation	102,494	245.03	40.70	174.79	859.22
Latitude	101,822	30.36	1.18	27.66	32.5
Longitude	101,822	75.94	0.81	73.91	77.59

characteristics (population density and sex ratio) and geographic attributes (average rainfall and average elevation). I extract rainfall data from Pai et al. (2014) and match villages to the nearest weather station using latitude–longitude coordinates in QGIS. Average elevation data were drawn from Farr and Kobrick (2000) and merged directly using the SHRUG identifier. Other control variables were constructed using the Census of India. To further absorb unobserved heterogeneity, I include district fixed effects in all specifications. I also considered to include latitude and longitude as controls, but dropped them from the main models because of collinearity with other geographic variables.

Table 4.3 shows the summary statistics for the control variables. On an average, villages in the sample have a population of approximately 1,350 individuals, with about 884 females per 1,000 males. In terms of geographic characteristics, the mean annual rainfall is 723 mm, ranging from 0 to 3,245 mm, reflecting substantial climatic variation across villages and decades. The average elevation is 245 meters, consistent with the topography of the northern plains.

4.4 State Capacity

State capacity refers to the ability of the state to implement policies, mobilize fiscal resources, and provide public goods such as education, healthcare, infrastructure and legal enforcement. At Independence, India faced severe administrative and fiscal constraints, particularly in rural areas. The disruption caused by Partition further strained the newly formed state through mass displacement, violence and the urgent need for rehabilitation, thus increasing pressure on already scarce public resources. Rural areas, in particular, suffered from limited schooling infrastructure, low literacy, inadequate transportation networks, and shortages of trained teachers and administrators. In such an environment, local human capital stocks likely played a central role in transmitting literacy and skills across generations because the state lacked the capacity to provide these inputs at scale.

Over time, however, the Indian state gradually expanded its institutional capacity. The early Five-Year Plans emphasized agricultural development, community development programmes, rural infrastructure and the expansion of primary education. The state later shifted its focus toward social sector development and universal access to education. In Punjab and Haryana, this process accelerated alongside rapid agricultural growth during the Green Revolution period, which increased fiscal capacity and enabled greater public investment in rural infrastructure and schooling.

Educational infrastructure expanded substantially over time. In 1947–48, Punjab (including present-day Haryana) had around 5,588 educational institutions (Vashishta, 1951). In Punjab alone, the number of schools increased from 3,819 in 1947–48 to 16,580 in 1990–91 and further to 23,399 by 2010–11 (Singh et al., 2015). Haryana also experienced a rapid expansion in educational infrastructure during the 1990s, with the number of educational institutions rising sharply from 10,174 in 1995–96 to 16,247 in 1998–99 (Statistical Abstracts). Public spending patterns also reflected this shift toward human capital formation. The share of education expenditure in total state budgets increased steadily, particularly from the mid-1970s onward in Punjab and during the 2000s in Haryana (Department of Education). This rapid expansion significantly reduced geographic barriers to schooling in rural areas.

The expansion of state capacity also transformed rural labor markets and household living conditions beyond education. Public investments in roads, electricity, transport networks and market infrastructure reduced transaction costs and integrated rural economies with urban and regional markets. At the same time, investments in irrigation and agricultural infrastructure contributed to productivity growth during the Green Revolution period.

The government also launched several literacy and educational programmes aimed at improving human capital in rural areas. Punjab participated in the National Literacy Mission (NLM), which expanded nationally in 1988 and targeted functional literacy among adults aged 15–35 years. During the 1990s, district-level Total Literacy Campaigns (TLCs), Post-Literacy Programmes (PLPs), and continuing education centres further strengthened literacy efforts in rural regions. These programmes focused on basic literacy, retention of reading and writing skills and community participation. From the early 2000s onward, school universalization programmes such as the *Sarva Shiksha Abhiyan* substantially expanded access to elementary education across villages.

The role of the state in social sector development became even more prominent during the Tenth Five-Year Plan (2002–2007), which explicitly emphasized literacy, gender equality, health outcomes, rural infrastructure, and decentralized governance through Panchayati Raj Institutions. By this period, Punjab and Haryana no longer represented low-capacity regions. Both states had relatively high literacy rates, extensive rural road connectivity, widespread school access, and stronger administrative capacity compared to the early post-Independence decades.

5 Empirical Strategy

This paper exploits spatial variation in the refugees presence across villages to examine how the influx of literate displaced persons affected local economic development. The analysis leverages a village-level panel dataset, which allows us to account for both cross-sectional and temporal variation in outcomes. Village-level analysis allows us to compare areas that are not only geographically and culturally similar but also governed by the same local administrative authority. This reduces the likelihood that omitted variables could confound the causal relationship between refugee presence and economic outcomes. The analysis is restricted to villages whose land area changed by no more than 10 percent relative to 1951, thereby minimizing errors arising from inconsistencies in village boundary matching across Census years. As a starting point, I estimate a baseline panel ordinary least squares (OLS) specification, where the unit of observation is village i, to quantify the effect of Partition-induced migration on key economic outcomes.

$$Y_{it} = \alpha + \beta\, Treatment\_75_i + \phi_t\, Treatment\_75_i * Year_t + \rho\, X_{it} + \gamma_d + \delta_t + \omega_{dt} + \varepsilon_{it} \quad (1)$$

In above specification, Y_{it} denotes the economic outcome for village i in year t, where t corresponds to Census years from 1951 to 2011. The key explanatory variable, $Treatment\_75_i$, is a binary indicator identifying whether village i experienced a high inflow of displaced persons following Partition. Since the data on refugee inflows are available only in the 1951 Census, this treatment variable is time-invariant. The interaction term coefficient, ϕ_t, captures the dynamic effect of displaced persons on outcomes such as literacy rates for each consecutive year relative to the base year.

The model further controls for relevant village-level characteristics, including average annual rainfall, population density, sex ratio and average elevation, as discussed in Section 4.3. District fixed effects (γ_d) and year fixed effects (δ_t) are included to absorb unobserved, time-invariant heterogeneity across districts and common shocks affecting all villages in a given year, respectively. ω_{dt} captures the district-year fixed effects, i.e., it removes the influence of any factor that affects all villages within a district in a given year (e.g., district-level policies, local infrastructure programs, rainfall shocks at district scale, etc.). A key advantage of using fixed effects at a granular level is that it minimizes the potential for selection on unobservables that bias the estimates. Standard errors are clustered at the 1991 panchayat level to account for potential serial correlation within panchayats over time. A *Panchayat Samiti*, also known as a Community Development Block (CDB), is the intermediate administrative unit (block-level) in the rural governance system of India. The sample comprises 270 such units, with an average of 56 villages per panchayat block.

5.1 Instrumental Variable Estimation

Although section 3.2 highlighted that the resettlement of refugees was largely involuntary and centrally managed by the government, there may still be some unobserved factors such as proximity to railway stations, pre-Partition Muslim population shares, etc. that influenced both the allocation of refugees across villages and subsequent patterns of economic development. It was also observed that refugees settled in smaller villages. The presence of such endogeneity would bias the estimates from the Ordinary Least Squares (OLS) framework discussed above. To address this concern, I employ an instrumental variable (IV) approach that exploits the distance of each village from the Indo–Pak border as a source of exogenous variation in refugee inflows.

The choice of this instrument is motivated by the strong spatial gradient in refugee settlement. Villages located closer to the border received a disproportionately higher inflow of displaced persons compared to those farther away, as evident from Figure 1. This pattern aligns with the well-documented "distance effect" in migration studies, where the volume of migration between two regions declines with greater distance (Ravenstein, 1889; Zipf, 1946; Karemera et al., 2000; Anderson, 2011). Consistent with Bharadwaj et al. (2008), the costs associated with longer travel distances from Pakistan played a central role in determining the spatial distribution of refugees in post-Partition India. In this context, costs refer not only to transportation costs but also to safety risks associated with longer journeys due to the fear of violence.

The instrument satisfies two key conditions. First, it is strongly correlated with the likelihood of a village receiving a high share of refugees, satisfying the relevance condition. Second, conditional on other controls, the distance of a village from the border is plausibly exogenous to its long-term economic outcomes, such as literacy rates or employment structures. Prior research has used similar geographic instruments to identify the causal effects of refugee or migrant inflows on local economies (Bharadwaj and Fenske, 2012; Miguel and Roland, 2011).

Formally, the first-stage equation is given by:

$$Treatment\_75_i = \pi_0 + \pi_1\, Distance_i + \pi_2\, X_i + \gamma_d + \varepsilon_i, \tag{2}$$

where $Treatment\_75_i$ denotes whether village i experienced a high refugee inflow (as defined earlier), and $Distance_i$ measures the shortest distance between the village centroid and the Indo–Pak border. The vector X_i includes village-level control variables, while γ_d represents district fixed effects. I use district-year fixed effects instead of tehsil-year fixed effects because the variation in distance within tehsils is likely to be minimal, and thus, district-level fixed effects capture the relevant geographic variation more effectively.

In addition to the treatment variable, its interactions with year dummies are also potentially endogenous. This generates multiple endogenous regressors in equation 1 corresponding to each census year. A common approach is to interact the instrument with year dummies to instrument these interaction terms. However, when the instrument is time-invariant, this strategy can substantially weaken the first stage as each additional interaction inflates the number of instruments without adding much identifying variation, leading to lower first-stage F-statistics and potential weak instrument concerns.

To address this issue, I estimate separate instrumental variable regressions for each census year, using distance to the border as the instrument for the treatment variable. This approach preserves instrument strength while allowing the estimated coefficients to capture the dynamic effects of refugee inflows over time. This approach has been widely used in the literature to study the temporal evolution of treatment effects (e.g., Kondylis (2010); Hornbeck (2012); Brun et al. (2005); Mariella (2023); Kremer et al. (2022)). All regression tables report the F-statistics, which are well above conventional thresholds, confirming that the instrument remains strong across all years.

The second-stage estimating equation for each year $t \in [1951, 2011]$ is specified as:

$$Y_i = \alpha + \beta\, \widehat{Treatment}\_75_i + \rho\, X_i + \gamma_d + \varepsilon_i. \tag{3}$$

In the above equation, I modified the baseline specification in Equation 1 to account for year-specific estimation. Although a panel framework generally provides more efficiency and variation, separate year-wise regressions offer a transparent way to capture the evolving impact of refugee inflows across decades without making the instruments weak. For comparison, I have also reported the corresponding OLS estimates for each year.

5.2 Threats to Identification

The key identifying assumption underlying the empirical framework is that, conditional on the inclusion of village controls and district-fixed effects, any remaining variation in the refugee shock is essentially random. This section discusses potential threats to this assumption and the strategies employed to address them.

5.2.1 Selection at Origin

A primary concern is the possibility of selection at origin, whereby only certain groups such as more educated or wealthier individuals chose to migrate from origin (West Pakistan) to destination (India). However, the nature of Partition migration substantially mitigates this concern. The displacement was sudden, large-scale, and largely involuntary, affecting nearly the entire Hindu and Sikh population residing in West Pakistan irrespective of their socioeconomic status. By 1951, Hindus and Sikhs constituted only 2% of West Pakistan's population, down from 17% in 1931 (Census of Pakistan, 1951), indicating near-complete out-migration of these groups.

Although refugees arriving in East Punjab were, on average, more literate than the native population in both rural and urban areas, this difference reflects pre-existing religious and regional disparities in literacy rather than selective migration (Figure 7). Hence, the higher literacy levels of refugees in 1951 are best interpreted as reflecting these pre-existing compositional differences rather than self-selection during displacement. This distinction is central to the paper as I aim to examine if the inflow of a disproportionately educated refugee population fundamentally altered the human capital composition of receiving regions and shaped their subsequent development trajectories.

5.2.2 Selection at Destination

There might be a presumption that regions receiving a higher influx of migrants during the Partition had higher levels of economic development even before the movement began. Yet, as outlined in Section 3.2 , the spatial distribution of refugees was primarily dictated by the government's land allocation policies, which depended on the availability of evacuee land. Thus, migrants relocated to areas previously vacated by outgoing populations, termed as the replacement effect by Bharadwaj et al. (2008)[12].

Table 12 presents estimates from regressions of pre-Partition (1931) demographic and economic characteristics on the post-Partition proportion of displaced persons. The coefficients on the refugee share are statistically insignificant across all outcomes, indicating that tehsils with higher post-Partition inflows were not systematically different from others in terms of pre-Partition demographic or literacy characteristics, mitigating concerns of selection at destination.

For outcomes such as employment rates that are available only at the district level, analogous regressions are run using district-level data (Table 13). The results show that male employment rates were higher in districts that later received a higher share of displaced persons. However, district fixed effects cannot be included in these regressions

[12]Using district-level data from 1931, I verify that areas with higher refugee shares in 1951 had a greater proportion of Muslims in 1931, consistent with this replacement mechanism. Results available upon request.

and any correlation could simply reflect historical differences across districts rather than within-district sorting related to refugee settlement.

Overall, these findings suggest that refugee settlement patterns were largely not driven by pre-existing economic conditions. Moreover, in robustness checks, I include 1931 tehsil-level literacy rates as an additional control and find that the main results remain robust.

5.2.3 Validity of the Instrument

A further concern pertains to the validity of the instrumental variable - distance to the Indo–Pak border. If the areas located closer to the border may have systematically different economic or geographic characteristics that independently influence long-term development outcomes, exclusion restriction would be violated. Exclusion condition requires that distance to the border affect present-day outcomes only through its impact on refugee settlement intensity.

To address this concern, I regress pre-Partition (1931) outcomes such as literacy or employment on distance to the border and find no significant relationship, suggesting that distance is not correlated with pre-existing development levels, thereby supporting the validity of the exclusion restriction. Table 14 reports these results. Although distance to the border exhibits a modest positive correlation with pre-Partition district employment, this relationship cannot account for the estimated effects on employment. If anything, higher baseline employment in villages closer to the border would bias estimates toward finding higher subsequent employment. Instead, the IV estimates indicate lower employment in refugee-receiving villages, suggesting that the estimated effects are unlikely to be driven by pre-existing employment differences.

In addition, the empirical specification includes district fixed effects, which restrict identification to within-district variation in refugee settlement intensity. This approach absorbs all unobserved, time-invariant district-level heterogeneity that might be correlated with both border proximity and subsequent economic outcomes. As an additional robustness check, I exclude villages located closest to the border to mitigate any potential direct effects of border proximity on outcomes. The estimated coefficients remain stable in magnitude and significance, indicating that the instrument captures exogenous variation in refugee inflows rather than pre-existing spatial disparities.

5.2.4 Selection in the type of migrants

The exclusion restriction could also fail if distance to the border influences not only the likelihood of refugee settlement but also the characteristics of migrants. However, as discussed earlier, resettlement of refugees was governed by the rehabilitation committee on the basis of origin and destination districts (Table 8) based primarily on the availability of evacuee land rather than migrant characteristics. Moreover, as discussed in Section 3.4, refugees had higher literacy rates than natives in rural areas across all districts, irrespective of their proximity to the border. While Bharadwaj et al. (2015) find that more literate refugees tended to move farther and to larger cities, I instead find stronger literacy effects in villages closer to the border. These patterns indicate that differential migrant selection by distance is unlikely to bias the results.

6 Results

6.1 Impact on Education

This section examines the dynamic effects of refugee inflows following the Partition of India on village-level literacy rates over the period 1951–2011. Panel A of Table 4 reports ordinary least squares (OLS) estimates. The specification controls for a rich set of village characteristics, including population density, sex ratio, total rainfall and average elevation along with the district fixed effects. Standard errors are clustered at the 1991 panchayat level throughout. Conditional on these controls, exposure to refugee inflows is assumed to be plausibly exogenous in OLS estimates.

Columns (1)–(7) present cross-sectional estimates for each Census year to trace out the dynamic effect and column (8) reports pooled estimates from a village-level panel spanning all seven Census rounds. The OLS results indicate that villages receiving a high influx of refugees experienced persistently higher literacy rates relative to other villages for more than six decades following Partition. The estimated literacy gap ranges between 0.6 and 4 percentage points across Census years. The divergence appears immediately in the post-Partition period, peaks around 1971, and gradually narrows thereafter as literacy rates improved nationwide. This convergence is consistent with the expansion of state-led educational programmes from the late 1970s onward, including literacy campaigns and rising public expenditure on education outlined in section 4.4, which reduced the relative importance of the initial refugee-driven human capital advantage. Nevertheless, the estimates remain positive and statistically significant even in 2001.

The pooled estimate in column (8) implies that high-inflow villages exhibit literacy rates that are, on average, 2.5 percentage points higher than those in zero-inflow villages. In 1951, the estimated coefficient of 3.53 percentage points corresponds to an increase of roughly 48 percent relative to the mean literacy rate of 7.4 percent. Importantly, these differences do not reflect pre-existing educational advantages. As documented in Section 5.2.2, refugee-receiving villages were not systematically different from others in terms of economic and educational conditions.

Panel B of Table 4 addresses potential endogeneity in refugee settlement patterns by instrumenting refugee shock with distance to the Indo–Pak border. The IV estimates identify a local average treatment effect (LATE) for villages whose refugee exposure was shaped by proximity to the border. The first-stage estimates are strong across all specifications, with F-statistics well above conventional thresholds, confirming instrument relevance.

The IV estimates are substantially larger than their OLS counterparts. For instance, in 1971, moving from no refugees to the 75th percentile of the refugee share distribution is associated with an increase of approximately 20 percentage points in village literacy rates within the same district, significant at the 1 percent level. Between 1961 and 1991, refugee exposure raises literacy by roughly 12–20 percentage points. Although the magnitude of the effect declines in later decades, it remains economically and statistically significant till 1991, with an estimated increase of 12.5 percentage points significant at 10 percent level. In the pooled panel specification, high-inflow villages exhibit literacy rates that are higher by 11.55 percentage points. Bharadwaj et al. (2015) also reported an increase in literacy of about 2.8 percentage points due to a 10 percent increase in

Partition inflows between 1931 and 1951 in western India.

The divergence between OLS and IV estimates suggests that OLS understates the true impact of refugee inflows. The large IV magnitudes are consistent with low initial literacy levels and the concentration of refugee settlement in relatively disadvantaged locations.

As noted above, literacy rates have been calculated using the total village population as the denominator due to data limitations. However, as a robustness check, for the period 1991–2011, I compute literacy rates using the population aged six and above, which aligns with the official definition. The results remain robust to this alternative construction. In addition, I estimate an alternate specification that controls for baseline literacy rates in 1951. This allows me to assess whether the observed effects simply reflect pre-existing differences across villages or capture dynamic changes following refugee exposure. The estimates remain qualitatively unchanged, suggesting that the results are not driven by initial literacy gaps but reflect sustained impacts over time.

To explore potential mechanisms, I examine whether higher literacy rates reflect improved access to educational infrastructure (Table 15). The evidence is mixed. Although the OLS estimates indicate that high-inflow villages had a larger number of primary, middle, and secondary schools across Census years, the IV estimates are either null or negative, suggesting that expansion in schooling infrastructure alone is unlikely to explain the observed literacy gains.

Table 4: Impact of Refugee Shock on Literacy Rates

	Dependent Variable: Literacy Rates							
	1951	1961	1971	1981	1991	2001	2011	Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS Estimates								
Treatment_75	3.529***	3.819***	3.845***	1.928***	1.996***	0.795*	0.641	2.494***
	(0.471)	(0.560)	(0.687)	(0.645)	(0.563)	(0.475)	(0.432)	(0.461)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818	52,867
R-squared	0.112	0.209	0.250	0.330	0.337	0.322	0.295	-
Within R-squared	–	–	–	–	–	–	–	0.912
Panel B: IV Estimates								
Treatment_75	7.543**	12.500***	19.680***	17.575***	12.448*	5.849	3.992	11.554**
	(3.216)	(4.545)	(6.716)	(6.606)	(7.228)	(7.660)	(6.046)	(5.351)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818	52,867
Cragg–Donald Wald F statistic	223.3	230.1	184.4	138.3	184.8	167.1	186.0	1323
Kleibergen–Paap rk Wald F statistic	32.84	35.79	26.60	24.35	24.89	27.92	25.68	29.74
Mean of outcome	7.363	15.99	24.02	34.16	42.62	54.91	63.96	34.52
No. of clusters	260	259	257	249	260	260	260	260
10% area change	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE clustered at panchayat level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS (Panel A) and IV (Panel B) estimates of the impact of refugee inflows on literacy rates. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion of displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall, and average elevation. District fixed effects are included in all specifications. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates use distance to the Indo–Pak border as an instrument for refugee exposure. The Stock–Yogo critical value for 10% maximal IV bias is 16.38; values above this threshold indicate that the instrument is not weak. Mean and standard deviation of the dependent variable are computed over the IV estimation sample.
*p<0.10, **p<0.05, ***p<0.01.

Finally, to assess whether the effects extend beyond basic literacy, particularly in light of nationwide education reforms such as the *Sarva Shiksha Abhiyan* launched in

2001, I examine educational attainment using SECC 2011 data. It is evident from table 16 that high-inflow villages do not exhibit any differential in educational attainment by 2011. This is consistent with my main findings that show convergence of two types of villages after 1991 in terms of educational outcomes.

Taken together, these findings provide strong evidence that the large-scale inflow of educated refugees following Partition generated substantial and persistent improvements in human capital in receiving villages. The effects are long-lasting, economically meaningful, and particularly pronounced in the early post-Partition decades, indicating the spillovers of refugee human capital on local educational outcomes under limited state capacity.

6.2 Impact on Labor Market Outcomes

Next, I explore whether these sustained improvements in literacy translated into long-term gains in labor market performance. Table 5 reports the OLS and IV estimates of the impact of refugee inflows on overall employment rates from 1991 to 2011. Across specifications, there is no consistent evidence of higher employment in high-influx villages. OLS estimates show small negative effects but IV results are inconclusive and do not reveal any clear pattern. These results suggest that higher literacy did not fully translate into improved aggregate employment outcomes at the village level.

To better understand labor market adjustment, Table 6 disaggregates employment by sector. According to Census, employment is divided into cultivators, agricultural laborers, household industry, and other services. Panel A presents OLS estimates, which reveal a consistent and sizable decline in the share of cultivators in high-influx villages, accompanied by a significant increase in the share of agricultural laborers across all census years. This suggests a shift in the nature of rural employment from self-employed cultivation toward wage-based agricultural work. One possible explanation relates to the post-Partition land redistribution process, under which many refugees received relatively small or less fertile land parcels, thus pushing them into agricultural wage labor. In addition, Bharadwaj and Mirza (2016) show that migrants were more likely to adopt advanced agricultural technologies during the Green Revolution. This raises the possibility that high-inflow regions experienced faster mechanization and commercialization of agriculture, which may have further reduced the demand for small cultivators and increased dependence on agricultural labor.

Table 5: Impact of refugee shock on Labour market, by year

	Dependent Variables: Employment Rate								
	1991			2001			2011		
	Total workers	Main workers	Marginal workers	Total workers	Main workers	Marginal workers	Total workers	Main workers	Marginal workers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS Estimates									
Treatment_75	-0.681**	-0.355*	-0.326*	-2.917***	-2.050***	-0.867***	-1.243***	-0.973***	-0.271
	(0.296)	(0.202)	(0.194)	(0.510)	(0.477)	(0.321)	(0.347)	(0.300)	(0.273)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
R-squared	0.136	0.231	0.098	0.175	0.145	0.197	0.146	0.153	0.113
Panel B: IV Estimates									
Treatment_75	-0.263	1.696	-1.959	-9.310	-9.626	0.316	3.155	9.148**	-5.993*
	(3.324)	(2.211)	(2.358)	(6.342)	(7.022)	(4.281)	(5.171)	(4.125)	(3.125)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
Cragg–Donald Wald F statistic	166.1	166.1	166.1	154.5	154.5	154.5	187.2	187.2	187.2
Kleibergen–Paap rk Wald F statistic	21.55	21.55	21.55	25.78	25.78	25.78	24.83	24.83	24.83
Mean of outcome	31.28	28.96	2.321	40.86	30.99	9.869	35.56	27.89	7.668
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS (Panel A) and IV (Panel B) estimates for labour market outcomes - employment rate. Main workers are workers engaged in economically productive activity for six months or more during the reference year and marginal workers worked for less than six months. The treatment variable is an indicator for villages without refugees and above the 75th percentile of the proportion displaced (Treatment_75 = 1). All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, irrigated area, total rainfall, and average elevation. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates use distance to the Indo–Pak border as an instrument for refugee exposure. The Stock–Yogo critical value for 10% maximal IV bias is 16.38; values above this threshold indicate that the instrument is not weak. Mean of the dependent variable is computed over the IV estimation sample.
*p<0.10, **p<0.05, ***p<0.01.

Table 6: Impact of Refugee Shock on Employment Rate by Sector and Year

	Employment Rate (%)		
Occupation Category	1991	2001	2011
Panel A: OLS Estimates			
Cultivators	-10.222***	-6.927***	-7.524***
	(0.966)	(0.670)	(0.703)
Agricultural labourers	7.348***	3.144***	2.600***
	(0.817)	(0.593)	(0.634)
Household industry	0.199	0.171	0.282
	(0.316)	(0.108)	(0.172)
Other services	3.413***	4.185***	4.156***
	(0.643)	(0.735)	(0.783)
Observations	7,710	7,651	7,818
Panel B: IV Estimates			
Cultivators	-25.727***	-16.254***	-13.952*
	(8.580)	(5.996)	(7.159)
Agricultural labourers	23.149***	17.443***	12.341*
	(8.665)	(6.436)	(7.475)
Household industry	-0.941	0.940	1.661
	(3.032)	(1.089)	(1.647)
Other services	7.075	-5.267	16.197**
	(6.291)	(10.083)	(8.212)
Observations	7,710	7,651	7,818
Cragg–Donald Wald F statistic	166.1	154.4	187.2
Kleibergen–Paap rk Wald F statistic	21.55	25.77	24.83
Mean of outcome (Cultivators)	46.92	33.13	30.98
Mean of outcome (Agricultural labourers)	25.80	12.59	15.04
Mean of outcome (Household industry)	5.225	1.790	2.045
Mean of outcome (Other services)	16.59	29.93	31.85
Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

Notes: This table presents OLS (Panel A) and IV (Panel B) estimates for labour outcomes - employment rate by sector. Census categorizes employment into cultivators, agricultural laborers, household industry and other services. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, irrigated area, total rainfall, and average elevation. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates use distance to the Indo–Pak border as an instrument for refugee exposure. The Stock–Yogo critical value for 10% maximal IV bias is 16.38; values above this threshold indicate that the instrument is not weak. Mean of the dependent variable is computed over the IV estimation sample.
* p<0.10, ** p<0.05, *** p<0.01.

At the same time, there is evidence of an expansion in employment in non-agricultural services, though the magnitudes are modest and significant only for OLS estimates. Household industry employment remains largely unchanged. The IV results reinforce these patterns, highlighting a gradual but incomplete structural shift away from agriculture. In the Indian context, Bharadwaj et al. (2015) report comparable results, showing that refugees were about 28 percentage points less likely to work in agriculture than natives.

Several mechanisms may underlie the observed patterns, though the analysis does not directly test them. First, limited non-agricultural employment opportunities in rural areas may have constrained the local labor market returns to rising literacy, inducing educated individuals to seek employment outside their villages. Second, the structure of landholdings in receiving regions may have shaped occupational outcomes: if cultivable land was concentrated among relatively few landowners, increased reliance on agricultural wage labor rather than self-cultivation would follow. Third, social frictions or incomplete integration of refugees in destination villages may have reduced access to land, credit, or local labor markets, potentially contributing to selective outmigration prior to 1991.

I also examine the effects on local economic activity using satellite-based nightlights data covering the period 1994–2013.[13] Specifically, I use three measures of nighttime luminosity—log of total light intensity, log of luminosity per capita, and maximum light within the village to capture different dimensions of economic activity. The estimates, however, are largely insignificant for IV specification across years. One possible explanation is that economic gains occurred through channels that are not well reflected in light intensity, particularly in rural areas. Moreover, nightlights data have well-known limitations, especially at the village level where measurement error and low light saturation can obscure variation in economic activity.

Overall, the results present a nuanced picture. Refugee inflows generated durable improvements in human capital, yet these gains only partially translated into labor market upgrading. Despite some diversification into services, receiving regions continued to rely heavily on low-productivity agricultural work. This suggests slower structural transformation and limited formal job creation, which may have constrained the returns to education.

6.3 Impact on Consumption and Poverty

Lastly, I examine whether the refugee inflow translated into long-run improvements in material well-being, focusing on rural consumption levels, poverty incidence, and income distribution using Socio-Economic Caste Census (SECC) 2011 data. Table 7 reports both OLS and IV estimates of the impact of refugee exposure on these outcomes.

The OLS estimates (Panel A) suggest that villages receiving higher refugee inflows exhibit lower average rural consumption and higher poverty rates, along with a modest decline in the share of households in higher income brackets. Taken at face value, these estimates would imply that refugee settlement was associated with adverse welfare outcomes in receiving regions. However, these correlations are likely confounded by non-random placement of refugees, as it was observed that displaced populations were disproportionately settled in smaller or less developed villages.

[13] Results are available upon request.

Table 7: Impact of Refugee Shock on Welfare

	Dependent Variables					
	Log(Rural per capita consumption)	Log(Rural HH consumption)	Rural poverty rate	Tendulkar poverty rate	Highest Income > ₹5,000	Highest Income > ₹10,000
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS Estimates						
Treatment_75	-0.021**	-0.032***	0.018***	0.012***	-0.017**	-0.002
	(0.010)	(0.007)	(0.004)	(0.003)	(0.008)	(0.006)
Observations	7,791	7,791	7,791	7,791	7,791	7,791
R-squared	0.173	0.116	0.123	0.101	0.021	0.032
Panel B: IV Estimates						
Treatment_75	0.533***	0.328***	-0.116*	-0.078	0.269***	0.216***
	(0.164)	(0.117)	(0.068)	(0.051)	(0.084)	(0.067)
Observations	7,791	7,791	7,791	7,791	7,791	7,791
Cragg–Donald Wald F statistic	183.8	183.8	183.8	183.8	183.8	183.8
Kleibergen–Paap rk Wald F statistic	25.38	25.38	25.38	25.38	25.38	25.38
Mean of outcome	10.08	11.72	0.143	0.0965	0.423	0.175
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents OLS (Panel A) and IV (Panel B) estimates of the impact of refugee inflows on consumption, poverty, and income distribution outcomes. The dependent variables in columns (5) and (6) refer to the share of households whose highest earning member has monthly income greater than ₹ 5,000 and ₹ 10,000, respectively. Data belong to the Socio-Economic Caste Census (SECC), 2011. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates use distance to the Indo–Pak border as an instrument for refugee exposure. The Stock–Yogo critical value for 10% maximal IV bias is 16.38; values above this threshold indicate that the instrument is not weak. Mean of the dependent variable is computed over the IV estimation sample.
*p<0.10, **p<0.05, ***p<0.01.

In contrast, the IV estimates (Panel B), which address endogenous settlement patterns, reveal a contrasting picture. High inflow villages have higher and statistically significant rural household consumption and lower poverty rates. There are also higher proportion of households whose highest earning member has monthly income greater than ₹5000 and ₹10,000 as shown in columns 5 and 6 of Table 7. The magnitudes are economically meaningful. For example, moving a village from the 25th to the 75th percentile of refugee exposure increases rural household consumption by approximately 41,105 rupees and reduces the rural poverty rate by more than 12 percentage points. These effects persist across alternative poverty measures and income thresholds, indicating broad-based improvements in material living standards.

Once endogeneity is addressed, the results provide strong evidence that refugee inflows generated sizable and persistent welfare gains in receiving villages. Notably, these improvements occur despite the absence of clear gains in aggregate employment outcomes, suggesting that higher human capital and productivity may have operated through channels such as improved agricultural efficiency, non-farm income diversification, remittances, or inter-generational mobility rather than local employment expansion alone.

7 Heterogeneous effects

The previous section documented average village-level effects of Partition-induced refugee inflows. I find that refugee exposure is associated with higher literacy rates, gradual occupational reallocation away from cultivation toward wage labor and services, and improvements in welfare outcomes, but it does not raise aggregate employment rates. To better understand the underlying mechanisms, this section examines heterogeneity in these effects across gender, colonists-origin districts and princely states. I also briefly discuss differences across former princely states.

7.1 Across gender

Using village-level census data disaggregated by sex, I compare outcomes in high refugee inflow (treated) villages to those with no refugee inflow (control) villages. Existing evidence on the gendered impacts of refugee inflows is mixed: some studies document improvements in female outcomes following displacement (Franz, 2003; Braun and Stuhler, 2024; Chilton et al., 2025), while others find limited or adverse effects (Bauer et al., 2013; Morales, 2018; Lu et al., 2021). This ambiguity motivates a closer examination in the context of the Partition of India.

7.1.1 Education

Table 17 reports the effects of refugee exposure on literacy separately for men and women across Census years. Refugee-exposed villages exhibit significantly higher literacy rates for both genders, especially in the decades immediately following Partition. However, the magnitude and persistence of these differences vary sharply by gender.

For males (Panels A and B), literacy gains are concentrated in the earlier decades after Partition. Both OLS and IV estimates show positive effects in 1961 and 1971, but these differences decline steadily over time and become statistically insignificant by 2001 and 2011. This pattern suggests convergence in male literacy as overall literacy expanded more broadly over time.

For females (Panels C and D), the effects are substantially larger and more persistent. Female literacy rates remain significantly higher in refugee-exposed villages across all Census rounds, and IV estimates indicate especially strong effects between 1981 and 1991 that continue through 2011. The magnitude of these estimates suggests that refugee exposure generated durable improvements in female educational attainment, far exceeding the corresponding differences for males in the same periods. This result is consistent with existing evidence on gendered effects of forced migration. Bauer et al. (2013) show that once they control for parental pre-war characteristics, the positive effect on men's education disappears, whereas the effect for women persists and is statistically significant.

Consistent with this pattern, Table 21 shows that the gender parity index remains significantly higher in refugee villages even in later decades, indicating a narrowing of the gender gap in education in high-inflow villages relative to non-exposed villages. Figure 9.1 illustrates the declining gender gap across the two types of villages. While male literacy gaps largely disappear after 1981, female literacy advantages persist for over six decades.

Several mechanisms may explain these findings. Since male literacy levels were already relatively higher at baseline, the scope for long-run divergence was limited. Lower baseline literacy among women allowed refugee exposure to generate larger and more sustained gains. In addition, refugee households may have transmitted norms or preferences that favored girls' education. Finally, the demographic disruption caused by Partition—including the emergence of a large population of "unattached women" may have increased perceived returns to female self-reliance and schooling. This interpretation aligns with the *"uprootedness hypothesis"*, which posits that displacement induces greater investment in education as a means of restoring security and mobility (Becker et al., 2020).

7.1.2 Employment and Occupational Structure

Table 18 reports gender-specific employment outcomes. Refugee-exposed villages exhibit higher male employment rates but lower female employment rates, thus explaining the absence of a significant aggregate employment effect in the baseline results. Among men, refugee exposure also increases the share of main (permanent) workers relative to marginal workers.

Further, tables 19 and 20 examine occupational structure by gender. Among men, refugee exposure reduces the share of cultivators and increases agricultural labour and service-sector employment, suggesting a gradual shift away from self-cultivation toward wage labor and non-agricultural activities. IV estimates report large and statistically significant reduction in male cultivators and a corresponding rise in agricultural labour. However, these effects were strongest in 1991 and weakened over time.

Female occupational responses differ in both magnitude and composition. Refugee-exposed villages show persistent decline in female cultivators alongside increases in agricultural labour and services, particularly in 1991 and 2011. These patterns indicate a reallocation of female labor away from family-based cultivation toward wage and service-sector activities. IV estimates reveal especially strong growth in female service-sector employment. In contrast, household industry employment remains largely unaffected for both men and women.

The gender heterogeneity analysis thus reveals that refugee inflows generated large and persistent gains in female human capital relative to men. They also primarily altered labor allocation within agriculture and services rather than triggering rapid industrialization. These findings align with Menon and Bhasin (1998), who argue that Partition accelerated women's entry into education and economic activity by disrupting traditional social arrangements and increasing the need for female economic participation.

7.2 Across colonists origin districts

I next examine whether the effects of refugee inflows differed across districts that historically supplied migrants to the British canal colonies of West Punjab. During the late nineteenth and early twentieth centuries, the colonial state established canal colonies as part of a large irrigation and settlement project.

The administration primarily allotted land to Punjabi peasant castes such as Jats, Arains, and Rajputs and ex-soldiers, encouraging large-scale migration from eastern Punjab. More than one million Punjabis settled in these colonies. These settlers, known as *abadkars* (colonizers), differed sharply from the existing *maqami* (local) population. Colonial records describe them as relatively skilled cultivators, selected from the most productive farming households. The new colony towns and villages followed systematic planning: markets and railway access lay within short distances, and settlements included schools, dispensaries, cooperative and agricultural staff. As a result, social and economic interactions increasingly revolved around education and work rather than caste or kinship ties, fostering higher levels of social and political mobilization than in other parts of Punjab. Six districts namely Amritsar, Gurdaspur, Jalandhar, Hoshiarpur, Ludhiana and Ambala supplied the bulk of migrants to the canal colonies

Kaur (2007).[14] Partition dramatically altered this institutional setting. Hindu and Sikh colonists were forced to leave West Punjab and migrate back to India, while Muslim refugees (*Muhajirs*) from India were resettled in their place. Importantly, returning colonists received preferential access to land and rehabilitation in their origin districts. Given their prior exposure to relatively developed agrarian institutions, infrastructure, and schooling, these refugees may have carried higher levels of human capital and agricultural knowledge than other displaced populations.

This raises the possibility that the long-run effects of refugee inflows differ systematically across districts depending on their historical links to canal colonies.

To examine this channel, I interact refugee exposure with an indicator for colonists-origin districts. One limitation of this analysis is that district fixed effects cannot be included, since the colonist-origin indicator varies only at the district level. Table 22 reports heterogeneous effects on literacy. The results show that the positive literacy effects of refugee exposure concentrate almost entirely in colonists' origin districts. The interaction between refugee exposure and the colonist-origin indicator is positive and statistically significant from 1961 to 2001. In contrast, the direct effect of refugee exposure outside these districts remains small and imprecisely estimated. These patterns suggest that refugee inflows alone are insufficient to raise literacy; instead, gains emerge when refugee exposure interacts with high skilled colonists' presence.

I next examine heterogeneity in labor market outcomes. Table 23 shows that refugee exposure increases the share of main workers and reduces the share of marginal workers across all years, indicating a shift toward more stable employment relationships. Refugee inflows thus appear to reshape the composition of employment rather than expand employment outright. However, these effects differ sharply across district types. In colonists-origin districts, the interaction between refugee exposure and historical canal-colony linkage is negative for main employment and positive for marginal employment. This suggests that labor market adjustment in these districts relied more heavily on casual and flexible forms of work, possibly reflecting land constraints.

Table 24 documents heterogeneity in occupational structure. Refugee exposure increased the share of agricultural labour, consistent with the baseline findings. However, colonists-origin districts with high-inflow experienced an increase in service-sector employment. This suggests that districts with historical links to canal colonies followed a different adjustment path, characterized by greater occupational diversification outside agriculture.

Finally, Table 25 reports the welfare outcomes across colonists origin districts. It is evident that refugee exposure significantly increases rural household consumption and per capita consumption. While the average effect on poverty rates is negative, these coefficients are not statistically significant. The effects are substantially amplified in colonist-origin districts for upper-income outcomes. Refugee exposure in these districts leads to statistically significant increases in the share of households earning above Rs. 5,000 and Rs. 10,000 per month. The magnitude of these interaction effects suggests that colonists-origin districts were better positioned to translate refugee inflows into upward income mobility.

Overall, the heterogeneity analysis shows that the long-run effects of refugee inflows

[14] Ferozepore also supplied migrants, but primarily to later colonies such as Upper Bari Doab and Nili Bar. The Chenab Colony, the most extensive and successful, drew mainly from the six districts listed above.

depended critically on local historical conditions and on the composition of the migrants themselves. Districts that had historically supplied settlers to the canal colonies appear better able to convert refugee inflows into sustained gains in literacy, occupational mobility and welfare.

7.3 Across princely states

I also examine whether the effects of refugee inflows differed across areas formerly under princely states (PEPSU) and those under direct British administration (Tables 26 - 29). These regions differed historically in administrative capacity, public investment and institutional development, which may have shaped their ability to absorb refugee inflows and translate them into long-run development outcomes. The results suggest significant heterogeneity in long-run literacy and welfare outcomes across these institutional settings. The long-run literacy improvements associated with refugee exposure are concentrated in districts that were previously under direct British administration. In contrast, the employment and occupational structure results show somewhat stronger shifts toward permanent and non-agricultural employment in both direct British ruled and PEPSU districts following refugee settlement. Overall, these findings suggest that historical governance structures influenced how local economies absorbed refugee inflows and converted them into different dimensions of development. This underscores the importance of pre-existing institutions in shaping the long-run economic consequences of human capital shocks.

8 Conclusion

This paper exploits the Partition of India in 1947 as a quasi-natural experiment to examine the long-run consequences of a large human capital shock i a limited state capacity setting to begin with. I show that villages receiving a higher share of educated refugees from West Pakistan experienced persistently higher literacy rates, higher household consumption, and lower poverty in subsequent decades. While the effects on overall employment remain limited, refugee-receiving villages exhibit a reallocation from self-cultivation toward agricultural wage labor. These findings suggest that human capital gains did not immediately translate into structural transformation but nevertheless generated sustained improvements in welfare.

This study contributes to the migration and development literature by tracing the dynamic effects of skilled refugee inflows over six decades in a low-income setting. By focusing on forced migration, the analysis minimizes concerns about migrant self-selection and endogenous destination choice that typically complicate causal inference in voluntary migration contexts. In addition, the village-level design exploits within-district variation in refugee exposure, reducing the likelihood that differences in formal institutions or regional policies drive the results.

I construct a novel village-level panel by digitizing and harmonizing Census data from 1951 to 2011, allowing a consistent comparison of long-run outcomes across villages in Punjab and PEPSU. Historical and archival records further support the identification strategy by documenting that refugees were more literate than natives and that the government centrally administered their rehabilitation according to origin–destination rules rather than migrant characteristics.

The analysis also uncovers important heterogeneity. Villages with high refugee inflows experienced disproportionately larger gains in female literacy. Women in these villages were more likely to transition out of self-cultivation and into service-sector activities, whereas men shifted towards agricultural wage labor. These gendered patterns suggest that displacement shocks can reshape social norms and educational investments in ways that affect men and women differently. The stronger female gains may reflect lower baseline literacy, changes in attitudes toward girls' schooling, or increased investment in education as a strategy to restore economic security after displacement.

I further examine heterogeneity across districts that historically supplied settlers to the canal colonies of West Punjab. The positive literacy effects concentrate in villages located in these colonist-origin districts. Many refugees in these areas were return migrants who had lived for decades in relatively more developed canal colony regions before Partition. Their higher skills and exposure to more advanced agricultural and institutional environments likely amplified the impact of refugee inflows. These return migrants contributed to higher levels of development through improved educational outcomes, higher consumption, increased incomes, and a shift toward the services sector. Together, these findings highlight that both the magnitude and the skill composition of migrant inflows shape long-run development trajectories.

However, I cannot directly observe the mechanisms through which refugee exposure generated persistent gains. Possible channels include stronger preferences for education, out-migration of educated individuals due to limited local employment opportunities, land redistribution that pushed households into wage labor, or remittance flows from migrants leaving high-inflow villages. The precise mechanisms through which refugee inflows affected economic outcomes and sustained these effects over six decades remain an important area for future research. Data constraints prevent a direct test of these channels. In addition, I do not distinguish between outcomes for refugees and natives after 1951 and therefore do not claim to identify spillover effects separately. The estimates capture village-level impacts associated with the share of displaced persons rather than individual-level effects.

Overall, the findings show that forced displacement need not generate only destruction. When refugee inflows deliver a positive human capital shock, they can generate substantial and persistent development gains in receiving regions, even in weak institutional settings. However, these gains diminish as the state builds its capacity and implements nationwide policies. Understanding how the scale and composition of migrant inflows interact with local institutions remains central to assessing the long-run development consequences of displacement.

References

(1948a). The east punjab refugees (registration of claims) act, 1948 (east punjab act no. viii of 1948). Received the assent of the Governor on 1st April, 1948.

(1948b). The east punjab refugees (registration of land claims) act, 1948 (east punjab act no. xii of 1948). Received the assent of the Governor on 1st April, 1948.

Acemoglu, D., Gallego, F. A., and Robinson, J. A. (2014). Institutions, human capital, and development. *Annu. Rev. Econ.*, 6(1):875–912.

Acemoglu, D., Hassan, T. A., and Robinson, J. A. (2011). Social structure and development: A legacy of the holocaust in russia. *The Quarterly Journal of Economics*, 126(2):895–946.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401.

Afzal, U. and Malik, A. (2010). The unchanging profile of development: A historical study of the punjab 1961–2008 [with comments]. *The Pakistan Development Review*, pages 841–862.

Akcigit, U., Grigsby, J., and Nicholas, T. (2017). Immigration and the rise of american ingenuity. *American Economic Review*, 107(5):327–331.

Alesina, A., Seror, M., Yang, D. Y., You, Y., Zeng, W., et al. (2020). *Persistence through revolutions*. National Bureau of Economic Research.

Anand, R. L. (1961). Census of india, 1961. Director of Census Operations, Punjab.

Anderson, J. E. (2011). The gravity model. *Annu. Rev. Econ.*, 3(1):133–160.

Aracı, D. T., Demirci, M., and Kırdar, M. G. (2022). Development level of hosting areas and the impact of refugees on natives' labor market outcomes in turkey. *European Economic Review*, 145:104132.

Arbatli, C. E. and Gokmen, G. (2018). Ethnic capital and sub-national development: Armenian and greek legacy in post-expulsion turkey. *Available at SSRN 3273964*.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review*, 35(4).

Ayesh, A. (2024). Mass involuntary migration and educational attainment. *World Development*, 181:106677.

Banerjee, A. and Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in india. *American economic review*, 95(4):1190–1213.

Bauer, T. K., Braun, S., and Kvasnicka, M. (2013). The economic integration of forced migrants: Evidence for post-war germany. *The Economic Journal*, 123(571):998–1024.

Becker, S. O., Grosfeld, I., Grosjean, P., Voigtländer, N., and Zhuravskaya, E. (2020). Forced migration and human capital: Evidence from post-wwii population transfers. *American Economic Review*, 110(5):1430–1463.

Bernstein, S., Diamond, R., Jiranaphawiboon, A., McQuade, T., and Pousada, B. (2022). The contribution of high-skilled immigrants to innovation in the united states. Technical report, National Bureau of Economic Research.

Besley, T. and Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American economic review*, 99(4):1218–1244.

Bharadwaj, P. and Fenske, J. (2012). Partition, migration, and jute cultivation in india. *Journal of Development Studies*, 48(8):1084–1107.

Bharadwaj, P., Khwaja, A., and Mian, A. (2008). The big march: migratory flows after the partition of india. *Economic and Political Weekly*, pages 39–49.

Bharadwaj, P., Khwaja, A. I., and Mian, A. (2015). Population exchange and its impact on literacy, occupation and gender–evidence from the partition of india. *International Migration*, 53(4):90–106.

Bharadwaj, P. and Mirza, R. A. (2016). Displacement and development: Long term impacts of the partition of india. *Unpublished Manuscript*.

Bhattacharya, N., Pakrashi, D., Saha, S., and Wang, L. C. (2023). Identity assimilation: Impact of conflict and partition on the giving behaviors of refugees and natives in west bengal. Technical report, GLO Discussion Paper.

Bhattacharya, P. S. and Mukhopadhyay, A. (2023). Forced displacement and social capital: long-run impact of the indian partition. *Available at SSRN 4498918*.

Borjas, G. J. (2019). Immigration and economic growth.

Braun, S. and Mahmoud, T. O. (2014). The employment effects of immigration: evidence from the mass arrival of german expellees in postwar germany. *The Journal of Economic History*, 74(1):69–108.

Braun, S. T. and Stuhler, J. (2024). The economic consequences of being widowed by war: A life-cycle perspective. *Journal of Public Economics*, 239:105241.

Brun, J.-F., Carrère, C., Guillaumont, P., and De Melo, J. (2005). Has distance died? evidence from a panel gravity model. *The World Bank Economic Review*, 19(1):99–120.

Chatterji, J. (2007). *The Spoils of Partition: Bengal and India, 1947–1967*. Cambridge University Press, Cambridge.

Cheema, A., Khalid, L., Patnam, M., et al. (2008). The geography of poverty: Evidence from the punjab. *The Lahore Journal of Economics*, 13(1):163–188.

Chilton, A., Linos, K., Malani, A., and Tabassum, Z. (2025). Forced migration and women's empowerment: evidence from rohingya refugees. *Journal of Ethnic and Migration Studies*, pages 1–30.

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica*, 78(6):1863–1903.

Dell, M. and Olken, B. A. (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. *The Review of Economic Studies*, 87(1):164–203.

Department, P. I. P. R. (1950). *Panjab digital library - digitization of rural rehabilitation in east Punjab*. The Publicity Department, Government of East Punjab, Shimla.

Department of Education (2026). State-wise budgeted expenditure on education and state domestic product in India. Indiastat. Data compiled for multiple years. Retrieved May 13, 2026, from https://www.indiastat.com.

Farr, T. G. and Kobrick, M. (2000). Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, 81(48):583–585.

Foster, A. D. and Rosenzweig, M. R. (2000). Does economic growth increase the demand for schools? evidence from rural india, 1960-1999. *Center for Economic Development and Policy Reform and University of Pennsylvania, mimeo*.

Franz, B. (2003). Bosnian refugee women in (re) settlement: Gender relations and social mobility. *Feminist review*, 73(1):86–103.

Government of India (1971). Census of india, 1971. Office of the Registrar General and Census Commissioner, India.

Government of India (1981). Census of india, 1981. Office of the Registrar General and Census Commissioner, India.

Government of India (1991). Census of india, 1991. Office of the Registrar General and Census Commissioner, India.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2011). A Bright Idea for Mesuring Economic Growth. *American Economic Review*.

Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review*, 102(4):1477–1507.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in prussia. *American Economic Review*, 104(1):84–122.

Huber, P., Landesmann, M., Robinson, C., and Stehrer, R. (2010). Migrants' skills and productivity: A european perspective. *National Institute Economic Review*, 213:R20–R34.

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2):31–56.

Iyer, L. (2010). Direct versus indirect colonial rule in india: Long-term consequences. *The Review of Economics and Statistics*, 92(4):693–713.

Jha, P. and Talathi, K. (2024). Impact of colonial institutions on economic growth and development in india: Evidence from night-lights data. *Economic Development and Cultural Change*, 72(4):000–000.

Jha, S. and Wilkinson, S. (2012). Does combat experience foster organizational skill? evidence from ethnic cleansing during the partition of south asia. *American Political Science Review*, 106(4):883–907.

Karemera, D., Oguledo, V. I., and Davis, B. (2000). A gravity model analysis of international migration to north america. *Applied economics*, 32(13):1745–1755.

Kaur, K. (2010). *Riots, Refugees and Rehabilitation: A Case Study of Punjab (1946-56).* Phd thesis, Department of History, Punjabi University, Patiala.

Kaur, R. (2007). The second migration: Displacement and refugees from rawalpindi during partition. *JPS*, 14(1):90.

Kondylis, F. (2010). Conflict displacement and labor market outcomes in post-war bosnia and herzegovina. *Journal of Development Economics*, 93(2):235–248.

Kremer, M., Willis, J., and You, Y. (2022). Converging to convergence. *NBER macroeconomics annual*, 36(1):337–412.

Kudaisya, G. and Yong, T. T. (2004). *The aftermath of partition in South Asia*. Routledge.

Kumari, A. (2013). Delhi as refuge: Resettlement and assimilation of partition refugees. *Economic and Political Weekly*, pages 60–67.

Lu, F., Siddiqui, S., and Bharadwaj, P. (2021). Marriage outcomes of displaced women. *Journal of Development Economics*, 152:102684.

Mariella, V. (2023). Landownership concentration and human capital accumulation in post-unification italy. *Journal of Population Economics*, 36(3):1695–1764.

Maystadt, J.-F. and Verwimp, P. (2014). Winners and losers among a refugee-hosting population. *Economic development and cultural change*, 62(4):769–809.

Mehra, A. (2020). Economic impact of the partition on india and pakistan. *Int J Adv Res*, 8:982–8.

Menon, R. and Bhasin, K. (1998). *Borders & boundaries: women in India's partition*. Rutgers University Press.

Miguel, E. and Roland, G. (2011). The long-run impact of bombing vietnam. *Journal of development Economics*, 96(1):1–15.

Ministry of Rural Development, G. o. I. (2011). Socio economic and caste census.

Mirza, R. (2022). The long-term effect of refugee resettlement: evidence from pakistan through partition. Technical report, Technical Report.

Morales, C. (2022). Do refugee students affect the academic achievement of peers? evidence from a large urban school district. *Economics of Education Review*, 89:102283.

Morales, J. S. (2018). The impact of internal displacement on destination communities: Evidence from the colombian conflict. *Journal of Development Economics*, 131:132–150.

Moser, P., Voena, A., and Waldinger, F. (2014). German jewish émigrés and us invention. *American Economic Review*, 104(10):3222–3255.

Nair, L. (1950). *Relief and Rehabilitation in Punjab*. The Publicity Department, Government of Punjab, Shimla. Accessed: 2024-9-4.

Nakamura, E., Sigurdsson, J., and Steinsson, J. (2022). The gift of moving: Intergenerational consequences of a mobility shock. *The Review of Economic Studies*, 89(3):1557–1592.

Office of the Registrar General and Census Commissioner, India (2011). Census of india. Accessed via SHRUG https://www.devdatalab.org/shrug_download/, v2.0.pakora.

Ottaviano, G. I. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European economic association*, 10(1):152–197.

Pai, D., Rajeevan, M., Sreejith, O., Mukhopadhyay, B., and Satbha, N. (2014). Development of a new high spatial resolution (0.25× 0.25) long period (1901-2010) daily gridded rainfall data set over india and its comparison with existing data sets over the region. *Mausam*, 65(1):1–18.

Pandey, G. (2001). *Remembering partition: Violence, nationalism, and history in India*, volume 7. Cambridge University Press Cambridge.

Peri, G. (2012). The effect of immigration on productivity: Evidence from us states. *Review of Economics and Statistics*, 94(1):348–358.

Randhawa, M. S. (1954). *Out of the ashes*. Public Relations Department, Punjab.

Ravenstein, E. G. (1889). The Laws of Migration. *Journal of the Royal Statistical Society*, 52(2):241.

Rocha, R., Ferraz, C., and Soares, R. R. (2017). Human capital persistence and development. *American Economic Journal: Applied Economics*, 9(4):105–136.

Ruiz, I. and Vargas-Silva, C. (2015). The labor market impacts of forced migration. *American Economic Review*, 105(5):581–586.

Singh, K. (1989). *The Partition of the Punjab*. Registrar, Punjabi University, Patiala, Punjab.

Singh, R. (1961). *The Military Evacuation Organisation, 1947-48*. Manager Press.

Singh, S., Monga, T., and Kaur, G. (2015). Implications of privatization of school education in rural areas of punjab: Some field level observations. *Indian Journal of Economics and Development*, 11(2):533–541.

Singh, T. (1952). *Land resettlement manual*. Controller of Printing and Stationery, Ambala.

Slade, E. H. (1951). Census of pakistan, 1951. Director of Census Operations, Pakistan.

Statistical Abstracts (2026). Total number of recognised educational institutions in India. Compiled by Indiastat. Data compiled for multiple years. Retrieved May 13, 2026, from https://www.indiastat.com.

Van der Werf, C. (2021). The impact of refugees on native students' educational outcomes. *Descargado de https://www. cynthiavanderwerf. com/uploads/4/3/5/0/43507971/van der werf*.

Vashishta, L. C. (1951). Census of india, 1951. Director of Census Operations, Punjab.

Yadav, S. (2012). Partition and displacement: A study of rural rehabilitation in punjab(1947-1958). *International Journal of Advanced Research in Management and Social Sciences*.

Zipf, G. K. (1946). The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, 11(6):677–686. Publisher: [American Sociological Association, Sage Publications, Inc.].

9 APPENDIX

9.1 Data construction

The village-level panel dataset was constructed by digitizing the Primary Census Abstracts for Punjab and PEPSU from 1951 to 1981, as these records were not available in machine-readable format. After digitization, villages had to be manually matched across years. Villages were matched using a combination of village codes, village names, and geographical area. The matching process was complicated by spelling variations, renaming and occasional village mergers or splits. For example, Akbalpur Nangli in 1951 became Iqbal Nangli in 1961.

An additional challenge arose in linking the historical Census data to the 1991 SHRUG dataset. To address this, I digitized the 1991 Census and created a consistent identifier using state, district, tehsil and village codes. This unique ID was crucial for creating a consistent link between the digitized 1991 data and the existing SHRUG dataset, which uses its own unique identifier (shrid) consistent across 1991, 2001, and 2011. This process produced a consistent village-level panel spanning 1951–2011.

From an initial set of approximately 20,024 villages, I successfully digitised about 86 percent (17,554) for 1951 after accounting for missing data from certain areas. Data were not available for the Patiala tehsil of Patiala district and a few other villages from other districts. Furthermore, Kandaghat, Nalagarh, and Una were later transferred to Himachal Pradesh on linguistic grounds and are thus excluded from the sample. After incorporating and cleaning data from later years and addressing inconsistencies (e.g., literacy rates or proportion of displaced individuals exceeding 100%), I was able to match 15,214 villages, accounting for approximately 87 percent of the available villages.[15]

If a village was not matched then it means either of the following options: (1) village was renamed between 1951 and the given year; (2) village ceased to exist or was merged to another existing village; (3) village was founded after 1951. The sample villages are distributed across 70 tehsils within 19 districts in 1951. This results in a final panel dataset of 106,498 observations.

For precise estimation, I have used sub-sample where the percentage change in the area of village with respect to 1951 is less than 10 per cent (N = 94,268).[16] For robustness I have also used sub-sample where the percentage change in the area of village with respect to 1951 is less than 5 per cent (N = 93,446) and 1 per cent (N = 80,786). Results remain robust under alternative area thresholds.

[15] Number of villages matched: 1951- 15214, 1961 - 15214, 1971 - 14673, 1981- 12157, 1991 - 15214, 1991 (shrug) - 14185, 2001 - 14676, 2011- 14600

[16] Percentage change in area has been defined as

$$\frac{\text{(Area of village i in 1961 - Area of village i in 1951) * 100}}{\text{Area of village i in 1951}}$$

Table 8: Origin-Destination resettlement rule

Origin District	**Destination District**
Attock	Ludhiana
Bahawalpur	Hissar
Gujranwala	Karnal
Gujrat	Ambala
Jhang	Rohtak
Jhelum	Ambala
Lahore	Ferozepur
Lyallpur	Jullundur
Mianwali	Gurgaon
Montgomery	Ferozepur
Multan	Hissar
Muzaffargarh	Rohtak
Rawalpindi	Ambala
Shahpur	Ambala
Sheikhpura	Karnal
Sialkot	Gurdaspur
Sind	Kapurthala
Colonists	Home districts in East Punjab or East Punjab States

Notes: This table reports, for each origin district, the destination district that allocated maximum land to that district's refugees. Colonists were given preference in their home districts and the surplus was moved to other districts.

Table 9: Allocation of Refugee Land Across Destination Districts by Origin Tehsil

	Destination Districts																			
Origin	**Ambala**	**Amritsar**	**Barnala**	**Bhatinda**	**Fatehgarh**	**Ferozepur**	**Gurdaspur**	**Gurgaon**	**Hissar**	**Hoshiarpur**	**Jullundur**	**Kangra**	**Kapurthala**	**Karnal**	**Kohistan**	**Ludhiana**	**Mohindergarh**	**Patiala**	**Rohtak**	**Sangrur**
Attock	4231					56				1			19		13	5908	6	2170		2
Bahawalpur				186	677				7146				91							143
D_G_Khan				25				40393	1151								5456			37
Gujranwala	34		2492	19390	2918	5138	38		671	2228	8661		5195	61860		109		33071		40885
Gujrat	17806		14	10	3	88	24	13	31	3309	708		1858			6282		2705		44
Jhang						20			12657	72			6			52		49	50638	3
Jhelum	15971					158			326	18			15		1080	290	7	11		1
Kohat															75					
Lahore	14	43384	2416	19271	989	233533	6		3500	214	269		14897	8936	136	18		4106		635
Lyallpur	150		2992	311	1972	380			4834	196	17335		3954	12691	28	2992		1840		1379
Mianwali	4359			172	2	21		19856		3			81	698	1		1588	66		
Mirpur																				42
Montgomery	79		171	2893	253	142020			13623	20	4199		3198	9827	20		173	1420		355
Multan	54		537	176	159	81		28	136133	42	640		1438	58685				1072		141
Muzaffargarh								3024	13321	105			172	14871		238	3409	237	24910	
Narowal		7429																		
Rawalpindi	16217		1	23	23	1538				5			7		953	255		347		9
Sargodha					857					115			945							
Shahpur	25504		194	335		317			75						3460	9672		5110		864
Sheikhpura		144	3112	2277	4201	9097			1615	5499	3314		13264	68946	87	2		33914		6812
Sialkot	49		6130	508	16043	557	124980		487	51958	6527		36909		23	48		2576		1428
Sind				14	155				280				507		48		292	129		
Total colonists	29106	112252	13955	29343	12974	77376	45142	2964	37086	83876	163375	2952	15783	1637	245	81499	1840	2431	10475	3014

Notes: This table reports the allocation of land (in standard acres) across destination districts by origin tehsil. Each cell represents the area allocated to refugees from a given origin in a specific destination district. The final row aggregates allocations across all colonist-origin tehsils.

Table 10: Comparison of Sample and Population Mean

Variable	Sample	Population	t-value
Area of village (sq. miles)	1.97	1.94	-1.07
	(2.71)	(2.65)	
Total population	671.39	664.96	-0.79
	(728.81)	(732.52)	
Males population	360.22	356.67	-0.81
	(397.43)	(395.70)	
Females population	311.13	308.29	-0.75
	(337.86)	(341.51)	
Number of literates	59.04	59.82	0.60
	(115.12)	(118.84)	
Male literates	49.74	50.21	0.47
	(89.63)	(92.03)	
Female literates	9.30	9.60	0.86
	(30.37)	(31.88)	
Number of displaced	84.85	86.73	0.80
	(202.40)	(222.42)	
Male displaced	45.56	46.58	0.82
	(107.05)	(117.86)	
Female displaced	39.29	40.15	0.77
	(95.80)	(105.08)	
Treatment_75	0.429	0.425	-0.56
	(0.495)	(0.494)	

Table 11: Comparison of Control and Treated Groups, 1951 (Treatment_75)

	Control mean	Treated mean	t-statistic	p-value
Literacy rate	6.145	9.128	-16.291***	0.000
Male literacy rate	10.298	13.929	-13.909***	0.000
Female literacy rate	1.260	3.596	-18.800***	0.000
Total population	494.513	491.576	0.249	0.804
Area	1.785	1.557	5.049***	0.000
Population density	421.156	412.343	0.137	0.891
Sex ratio	863.173	866.363	-0.716	0.474
PEPSU	0.132	0.090	6.199***	0.000
Total rainfall	477.597	569.453	-14.714***	0.000
Mean elevation	250.323	240.111	11.059***	0.000
Colonists origin districts	0.220	0.583	-37.577***	0.000
Observations		8855		

Notes: This table compares means between control and treated groups in 1951 where Treatment_75 = 1 indicates refugee exposure greater than or equal to 75th percentile. t-statistics test equality of means. *** denotes significance at the 1% level.

Table 12: Pre-Partition Characteristics and Refugee Inflows (Tehsil-Level, 1931)

	Dependent Variable: Pre-Partition (1931) Characteristics					
	Area (sq. miles)	Total population	Sex ratio	Literacy rate	Male literacy rate	Female literacy rate
Proportion of displaced	-49.19	-82,046.23	-138.62**	-0.071	-1.548	0.754
	(597.25)	(135,538.10)	(60.52)	(4.85)	(7.43)	(1.66)
Area (sq. miles)		157.45***	0.034	-0.002	-0.003	-0.001**
		(50.86)	(0.031)	(0.002)	(0.003)	(0.001)
Total population (1931)			0.000	0.000	0.000	0.000***
			(0.000)	(0.000)	(0.000)	(0.000)
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
SE clustered at district level	Yes	Yes	Yes	Yes	Yes	Yes
Observations	54	54	54	54	54	54
Mean dependent variable	528.78	201,131.61	828.89	4.67	7.71	0.969

Notes: Each column reports results from a separate OLS regression of the pre-Partition (1931) tehsil-level characteristic on the proportion of displaced persons resettled in each sub-district (*tehsil*) post-Partition. All regressions include district fixed effects. Standard errors, clustered at the district level, are reported in parentheses.
*p<0.10, **p<0.05, ***p<0.01.

Table 13: Pre-Partition Labour Outcomes and Refugee Inflows (District-Level, 1931)

	Employment rate	Male employment rate	Female employment rate
Proportion of displaced	9.23	13.91***	-4.68
	(10.36)	(4.35)	(6.29)
Observations	16	16	16
Controls	Yes	Yes	Yes
Mean dependent variable	41.97	33.8	8.16

Notes: Each column presents results from a separate OLS regression of pre-Partition (1931) district-level employment rates on the proportion of displaced persons resettled post-Partition. All regressions control for total population and district area (both measured in 1931). Robust standard errors are reported in parentheses.
*p<0.10, **p<0.05, ***p<0.01.

Table 14: Pre-event characteristics regressed on instrument (Distance to border)

	Area (sq. miles)	Total population	Sex ratio	Literacy rate	Male literacy rate	Female literacy rate	Employment rate	Male emp. rate	Female emp. rate
Distance to border (m)	0.003	-0.399	0.000	0.000	0.000	0.000	0.0000265**	0.000	0.0000253***
	(0.003)	(0.524)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total population 1931			0.000	0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Area (sq. miles)			0.002	-0.001***	-0.001**	0.000***	0.001	0.000	0.001
			(0.006)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Observations	13	13	13	13	13	13	13	13	13
Mean dependent variable	2589.46	913554.46	828.04	4.99	8.14	1.15	38.13	32.32	5.81
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column presents results from a separate OLS regression of pre-Partition (1931) district-level economic and demographic outcomes on the instrument (distance to the border). All regressions control for total population and district area (both measured in 1931). Robust standard errors are reported in parentheses.
*p<0.10, **p<0.05, ***p<0.01.

Table 15: Did refugee villages have better educational infrastructure?

	OLS				IV			
	Primary School	Middle School	High School	College	Primary School	Middle School	High School	College
Treatment_75	0.136***	0.074***	0.085***	0.000	-0.249	-0.449***	-0.536***	-0.003
	(0.020)	(0.017)	(0.019)	(0.003)	(0.205)	(0.162)	(0.181)	(0.011)
Observations	30,860	30,918	30,898	23,098	30,860	30,918	30,898	23,098
R-squared	0.303	0.259	0.161	0.022	-0.013	-0.090	-0.090	-0.000
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE clustered at panchayat level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of clusters	260	260	260	260	260	260	260	260
Cragg–Donald Wald F					754.6	757.1	756.2	570.8
Kleibergen–Paap rk Wald F					28.82	28.89	28.91	29.98
Mean of outcome	1.110	0.381	0.269	0.00355	1.110	0.381	0.269	0.00355

Notes: The table reports the impact of refugee exposure on educational infrastructure across villages. The dependent variables are indicators for the number of primary schools, middle schools, high schools and colleges at the village level. Columns (1)–(4) present OLS estimates and columns (5)–(8) report instrumental-variable (IV) estimates, where refugee exposure is instrumented using distance to the Indo–Pak border. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. District, year, and district×year fixed effects are included. Standard errors clustered at the 1991 panchayat level are reported in parentheses. The Stock–Yogo critical value for 10% maximal IV bias is 16.38.
* p<0.10, ** p<0.05, *** p<0.01.

Table 16: Impact of Refugee Exposure on Educational Attainment

	Dependent Variables					
	Primary school	Some education	Middle school	Secondary school	Senior secondary school	Graduate school
Panel A: OLS Estimates						
Treatment_75	0.004	0.007	-0.002	-0.007	-0.003	-0.000
	(0.007)	(0.006)	(0.006)	(0.006)	(0.003)	(0.002)
R-squared	0.140	0.125	0.145	0.128	0.088	0.052
Panel B: IV Estimates						
Treatment_75	0.054	0.061	0.056	0.078	0.045	0.026
	(0.090)	(0.082)	(0.087)	(0.072)	(0.041)	(0.020)
Cragg–Donald Wald F	193.2	193.2	193.8	191.6	191.6	191.6
Kleibergen–Paap rk Wald F	26.94	26.94	26.35	26.37	26.37	26.37
Observations	7,705	7,705	7,766	7,748	7,748	7,748
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
SE clustered at panchayat level	Yes	Yes	Yes	Yes	Yes	Yes
No. of clusters	260	260	260	260	260	260
Mean of outcome	0.537	0.658	0.378	0.249	0.119	0.0399

Notes: The table reports the impact of refugee exposure on educational attainment outcomes at the village level. The dependent variables are the shares of village population that have completed a given education level or above. Panel A presents OLS estimates and Panel B reports instrumental-variable (IV) estimates, where refugee exposure is instrumented using distance to the Indo–Pak border. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. Standard errors clustered at the 1991 panchayat level are reported in parentheses. The Stock–Yogo critical value for 10% maximal IV bias is 16.38.
* p<0.10, ** p<0.05, *** p<0.01.

Table 17: Gender Heterogeneity in the Impact of Refugee Inflows on Literacy

	Dependent Variable: Literacy Rate						
	1951	1961	1971	1981	1991	2001	2011
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: OLS Estimates – Males							
Treatment_75	4.709***	4.054***	3.170***	0.943	0.998*	0.002	0.003
	(0.645)	(0.661)	(0.705)	(0.619)	(0.534)	(0.438)	(0.402)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818
R-squared	0.108	0.182	0.213	0.268	0.298	0.285	0.269
Panel B: IV Estimates – Males							
Treatment_75	10.445**	13.922**	13.416*	3.908	-2.100	-6.272	-5.007
	(4.803)	(5.982)	(7.639)	(7.214)	(7.993)	(7.531)	(6.064)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818
Cragg–Donald Wald F	223.3	230.1	184.4	138.3	184.8	167.1	186.0
Kleibergen–Paap rk Wald F	32.84	35.79	26.60	24.35	24.89	27.92	25.68
Mean of outcome	11.79	24.22	33.09	43.93	51.82	62.28	70.24
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel C: OLS Estimates – Females							
Treatment_75	2.277***	3.715***	4.516***	3.065***	3.033***	1.687***	1.397***
	(0.320)	(0.479)	(0.724)	(0.746)	(0.654)	(0.578)	(0.512)
Observations	8,267	7,730	7,569	5,893	7,900	7,651	7,816
R-squared	0.084	0.240	0.325	0.462	0.432	0.401	0.362
Panel D: IV Estimates – Females							
Treatment_75	3.619**	10.614***	27.083***	33.267***	28.975***	19.866**	13.789**
	(1.666)	(3.287)	(6.769)	(8.003)	(8.035)	(8.786)	(6.730)
Observations	8,267	7,730	7,569	5,893	7,900	7,651	7,816
Cragg–Donald Wald F	223.3	233.6	184.6	138.5	186.0	167.1	186.7
Kleibergen–Paap rk Wald F	32.77	36.15	26.60	24.35	24.94	27.91	25.78
Mean of outcome	2.232	6.506	13.70	23.15	32.24	46.58	56.96
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports gender-disaggregated OLS and IV estimates of the impact of refugee inflows on literacy rates. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. District fixed effects are included in all specifications. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 16.38. Mean outcomes are computed over the IV estimation sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Gender Heterogeneity in the Impact of Refugee Exposure on Employment Outcomes

	Dependent Variables: Employment Rates								
	1991			2001			2011		
	Total workers	Main workers	Marginal workers	Total workers	Main workers	Marginal workers	Total workers	Main workers	Marginal workers
Panel A: OLS Estimates – Males									
Treatment_75	-0.507	-0.433	-0.074**	-1.064***	-1.153***	0.089	-0.698***	-1.215***	0.516*
	(0.318)	(0.319)	(0.036)	(0.296)	(0.380)	(0.242)	(0.245)	(0.371)	(0.292)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
R-squared	0.380	0.383	0.014	0.189	0.261	0.088	0.286	0.233	0.043
Panel B: IV Estimates – Males									
Treatment_75	12.510***	12.791***	-0.282	6.086*	5.968	0.118	11.141***	16.555***	-5.414**
	(4.751)	(4.892)	(0.419)	(3.502)	(4.692)	(2.684)	(3.954)	(5.242)	(2.366)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
Cragg–Donald Wald F	166.1	166.1	166.1	154.5	154.5	154.5	187.2	187.2	187.2
Kleibergen–Paap rk Wald F	21.55	21.55	21.55	25.78	25.78	25.78	24.83	24.83	24.83
Mean of outcome	51.52	51.28	0.240	52.14	45.42	6.720	52.45	45.14	7.310
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel C: OLS Estimates – Females									
Treatment_75	-0.894*	-0.308	-0.585	-5.034***	-3.099***	-1.935***	-1.773***	-0.620	-1.153***
	(0.489)	(0.254)	(0.403)	(0.919)	(0.772)	(0.504)	(0.637)	(0.415)	(0.373)
Observations	7,708	7,708	7,708	7,651	7,651	7,651	7,816	7,816	7,816
R-squared	0.180	0.117	0.098	0.250	0.093	0.213	0.159	0.055	0.155
Panel D: IV Estimates – Females									
Treatment_75	-16.869**	-13.068***	-3.801	-27.175**	-27.233**	0.057	-6.261	0.102	-6.363
	(6.960)	(4.245)	(4.873)	(11.325)	(11.624)	(6.707)	(8.913)	(4.995)	(5.085)
Observations	7,708	7,708	7,708	7,651	7,651	7,651	7,816	7,816	7,816
Cragg–Donald Wald F	167.2	167.2	167.2	154.4	154.4	154.4	187.9	187.9	187.9
Kleibergen–Paap rk Wald F	21.60	21.60	21.60	25.77	25.77	25.77	24.92	24.92	24.92
Mean of outcome	8.541	3.891	4.651	28.14	14.72	13.41	16.87	8.784	8.090
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports gender-disaggregated OLS and IV estimates for employment rates. Main workers are individuals engaged in economically productive activity for six months or more in the reference year and marginal workers are those employed for less than six months. The treatment variable is an indicator for villages without refugees and above the 75th percentile of refugee exposure (Treatment_75 = 1). All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall, average elevation and irrigated area. District fixed effects are included in all specifications. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 16.38. Mean outcomes are computed over the IV estimation sample.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 19: Occupational Heterogeneity in the Impact of Refugee Exposure: Males

	1991	2001	2011
Panel A: OLS (Treatment_75)			
Cultivators	-10.623***	-8.897***	-8.459***
	(0.977)	(0.770)	(0.765)
Agricultural Labour	7.311***	3.035***	2.606***
	(0.822)	(0.649)	(0.649)
Household Industry	0.190	0.144	0.188
	(0.311)	(0.094)	(0.137)
Other Services	3.246***	5.320***	4.443***
	(0.649)	(0.694)	(0.821)
Observations	7,710	7,651	7,818
Panel B: IV (Treatment_75)			
Cultivators	-26.815***	-14.871*	-14.989*
	(9.404)	(7.769)	(8.680)
Agricultural Labour	24.687***	20.671***	13.524*
	(8.876)	(7.251)	(7.680)
Household Industry	-1.567	-0.073	1.614
	(3.032)	(0.713)	(1.293)
Other Services	4.101	-5.021	12.917
	(6.804)	(7.923)	(8.635)
Observations	7,710	7,651	7,818
Cragg–Donald Wald F	166.1	154.4	187.2
Kleibergen–Paap rk Wald F	21.55	25.77	24.83
Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

Notes: The table presents OLS (Panel A) and IV (Panel B) estimates for male labour outcomes by occupation. Census categorizes employment into cultivators, agricultural labourers, household industry and other services. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall, average elevation and irrigated area. District fixed effects are included in all specifications. Standard errors clustered at the 1991 panchayat level are reported in parentheses. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 16.38.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 20: Occupational Heterogeneity in the Impact of Refugee Exposure: Females

	1991	2001	2011
Panel A: OLS (Treatment_75)			
Cultivators	-5.338***	-4.451***	-4.669***
	(1.011)	(0.687)	(0.676)
Agricultural Labour	6.111***	2.573***	2.644***
	(1.498)	(0.596)	(0.689)
Household Industry	-1.411*	0.181	0.265
	(0.830)	(0.254)	(0.300)
Other Services	3.398*	1.644	2.396***
	(1.733)	(1.178)	(0.853)
Observations	6,260	7,568	7,743
Panel B: IV (Treatment_75)			
Cultivators	-67.703***	-35.736***	-25.727***
	(17.555)	(9.946)	(6.968)
Agricultural Labour	5.002	9.206*	2.852
	(11.473)	(5.509)	(6.739)
Household Industry	4.801	4.260	1.993
	(5.116)	(2.640)	(2.894)
Other Services	49.113***	-0.729	40.138***
	(17.410)	(19.197)	(10.514)
Observations	6,260	7,568	7,743
Cragg–Donald Wald F	131.1	155.1	183.9
Kleibergen–Paap rk Wald F	17.49	25.56	24.72
Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

Notes: The table presents OLS (Panel A) and IV (Panel B) estimates for female labour outcomes by occupation. Census categorizes employment into cultivators, agricultural labourers, household industry and other services. The treatment variable is an indicator for villages without refugees and above 75th percentile (Treatment_75 = 1) of proportion of displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall, average elevation and irrigated area. District fixed effects have also been included. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 16.38.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 21: Impact of Refugee Inflows on Gender Gap in Literacy

	1951	1961	1971	1981	1991	2001	2011
Panel A: OLS Estimates							
Treatment_75	0.099***	0.124***	0.111***	0.066***	0.054***	0.027***	0.019***
	(0.014)	(0.013)	(0.015)	(0.013)	(0.009)	(0.007)	(0.005)
Observations	6,942	7,640	7,548	5,868	7,898	7,650	7,816
Controls	YES	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES	YES
R-squared	0.082	0.074	0.279	0.498	0.443	0.479	0.455
No. of clusters	255	259	257	249	260	260	260
Panel B: IV Estimates							
Treatment_75	0.038	0.319***	0.580***	0.753***	0.589***	0.415***	0.263***
	(0.068)	(0.080)	(0.127)	(0.165)	(0.130)	(0.103)	(0.068)
Observations	6,942	7,640	7,548	5,868	7,898	7,650	7,816
Controls	YES	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES	YES
No. of clusters	255	259	257	249	260	260	260
Cragg–Donald Wald F statistic	191.3	232.1	184.4	137.2	186.6	166.9	186.7
Kleibergen–Paap rk Wald F statistic	27.50	35.91	26.61	24.17	25.03	27.88	25.78

Notes: The table presents OLS (Panel A) and IV (Panel B) estimates for the impact of refugee exposure on gender gap in literacy. Gender parity index has been calculated as Female literacy rate/Male literacy rate wherein a value less than 1 indicates disadvantageous position of females and value greater than 1 indicates disparity in favour of women. The treatment variable is an indicator for villages without refugees and above 75th percentile (Treatment_75 = 1) of proportion of displaced. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. District fixed effects have also been included. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 16.38.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 22: Heterogeneous Effects of Refugee Exposure on Literacy Rates by Canal Colonists origin districts

	Literacy rates						
	1951	1961	1971	1981	1991	2001	2011
IV Estimates							
Treatment_75	-0.815	1.787	1.232	-0.689	-0.380	0.787	-1.213
	(1.293)	(1.572)	(2.162)	(2.917)	(3.337)	(3.445)	(3.135)
Colonists origin districts	6.821***	3.587**	-0.084	1.963	-2.687	-2.856	-0.659
	(1.888)	(1.710)	(2.628)	(2.853)	(2.545)	(2.743)	(2.251)
Treatment_75 × Colonists origin districts	-4.088*	4.362*	9.122**	7.192*	8.412**	8.479*	4.810
	(2.344)	(2.428)	(3.706)	(4.234)	(3.747)	(4.416)	(3.826)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818
No. of clusters	260	259	257	249	260	260	260
Cragg-Donald Wald F statistic	884.4	894.1	791.4	584.6	851.9	817.0	831.8
Kleibergen–Paap rk Wald F statistic	46.30	58.80	43.72	43.43	50.73	43.28	48.49

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on literacy rates by canal colonists origin districts. The treatment variable is an indicator for villages without refugees and above 75th percentile (Treatment_75 = 1) of proportion of displaced. Colonists origin districts indicate districts that supplied migrants to canal colonies namely Amritsar, Gurdaspur, Jalandhar, Hoshiarpur, Ludhiana and Ambala. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 23: Heterogeneous Effects of Refugee Exposure on Employment Outcomes by Canal Colonists Origin Districts

	Employment Outcomes								
	1991			2001			2011		
	Total	Main	Marginal	Total	Main	Marginal	Total	Main	Marginal
IV Estimates									
Treatment_75	1.727	5.923***	-4.196***	-5.840***	7.443***	-13.282***	5.195**	9.902***	-4.708***
	(1.331)	(1.032)	(1.067)	(1.565)	(1.878)	(1.844)	(2.070)	(1.987)	(1.140)
Colonists origin districts	-0.407	1.695**	-2.102***	-4.975*	4.141	-9.116***	-0.766	2.314	-3.080***
	(1.178)	(0.770)	(0.771)	(2.706)	(3.020)	(1.601)	(1.781)	(1.496)	(1.188)
Treatment_75 × Colonists origin districts	-2.103	-5.165***	3.062***	5.577*	-9.177**	14.754***	-3.587	-9.078***	5.491***
	(1.498)	(1.228)	(1.030)	(3.239)	(3.743)	(2.247)	(2.672)	(2.482)	(1.609)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
No. of clusters	247	247	247	260	260	260	260	260	260
Cragg-Donald Wald F statistic	906.7	906.7	906.7	822.6	822.6	822.6	832.3	832.3	832.3
Kleibergen–Paap rk Wald F statistic	54.58	54.58	54.58	43.83	43.83	43.83	48.79	48.79	48.79

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on employment outcomes by canal colonists origin districts. The treatment variable is an indicator for villages without refugees and above 75th percentile (Treatment_75 = 1) of proportion of displaced. Colonists origin districts indicate districts that supplied migrants to canal colonies namely Amritsar, Gurdaspur, Jalandhar, Hoshiarpur, Ludhiana and Ambala. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 24: Heterogeneous Effects of Refugee Exposure on Occupational Structure by Canal Colonists Origin Districts

	Occupation Shares											
	1991				2001				2011			
	Cultivators	Agricultural Labour	HH Industry	Other Services	Cultivators	Agricultural Labour	HH Industry	Other Services	Cultivators	Agricultural Labour	HH Industry	Other Services
IV Estimates												
Treatment_75	1.124	22.019***	-1.365	-10.474***	3.949	19.116***	0.728**	3.369	6.691**	17.238***	0.256	-7.980***
	(3.427)	(4.091)	(0.944)	(2.117)	(2.542)	(2.572)	(0.352)	(3.188)	(2.696)	(2.721)	(0.457)	(3.090)
Colonists origin districts	-1.925	6.914*	1.465	-0.724	-0.324	7.543***	0.356	9.086*	-1.763	12.567***	-0.414	-3.192
	(3.830)	(3.855)	(1.500)	(3.420)	(2.787)	(2.128)	(0.401)	(4.706)	(2.756)	(2.636)	(0.577)	(4.175)
Treatment_75 × Colonists origin districts	-1.903	-13.385**	-1.792	8.211*	-7.142	-18.048***	0.072	-3.754	-7.140	-22.919***	0.985	11.941**
	(5.710)	(5.544)	(1.599)	(4.241)	(4.452)	(3.321)	(0.590)	(6.250)	(4.629)	(3.979)	(0.754)	(5.895)
Observations	7,710	7,710	7,710	7,710	7,651	7,651	7,651	7,651	7,818	7,818	7,818	7,818
No. of clusters	247	247	247	247	260	260	260	260	260	260	260	260
Cragg-Donald Wald F statistic	906.7	906.7	906.7	906.7	822.2	822.2	822.2	822.2	832.3	832.3	832.3	832.3
Kleibergen–Paap F statistic	54.58	54.58	54.58	54.58	43.81	43.81	43.81	43.81	48.79	48.79	48.79	48.79

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on occupational structure by canal colonists origin districts. The treatment variable is an indicator for villages without refugees and above 75th percentile (Treatment_75 = 1) of proportion of displaced. Colonists origin districts indicate districts that supplied migrants to canal colonies namely Amritsar, Gurdaspur, Jalandhar, Hoshiarpur, Ludhiana and Ambala. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall, average elevation and irrigated area. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 25: Heterogeneous Effects of Refugee Exposure on Welfare by Canal Colonists Origin Districts

	SECC Rural Economic Outcomes (2011)					
	Log(Rural per capita consumption)	Log(Rural HH consumption)	Rural poverty rate	Tendulkar poverty rate	Highest Income > Rs. 5,000	Highest Income > Rs. 10,000
IV Estimates						
Treatment_75	0.211***	0.096***	-0.028	-0.014	-0.003	0.039**
	(0.056)	(0.036)	(0.029)	(0.023)	(0.029)	(0.017)
Colonists origin districts	-0.074	-0.097**	0.044*	0.029*	-0.080***	-0.069***
	(0.057)	(0.042)	(0.023)	(0.016)	(0.028)	(0.021)
Treatment_75 × Colonists origin districts	0.013	0.067	-0.050	-0.035	0.119***	0.059**
	(0.079)	(0.057)	(0.034)	(0.025)	(0.040)	(0.025)
Observations	7,791	7,791	7,791	7,791	7,791	7,791
No. of clusters	260	260	260	260	260	260
Cragg–Donald Wald F statistic	813	813	813	813	813	813
Kleibergen–Paap F statistic	46.14	46.14	46.14	46.14	46.14	46.14
Mean of outcome	10.08	11.72	0.143	0.0965	0.423	0.175

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on SECC rural economic outcomes (2011) by canal colonists origin districts. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1) of proportion displaced. Colonists origin districts indicate districts that supplied migrants to canal colonies namely Amritsar, Gurdaspur, Jalandhar, Hoshiarpur, Ludhiana and Ambala. All regressions impose a 10% area-change restriction and include village-level controls such as population density, sex ratio, total rainfall and average elevation. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 26: Heterogeneous Effects of Refugee Exposure by Princely States (PEPSU)

	Literacy Rates						
	1951	1961	1971	1981	1991	2001	2011
IV Estimates							
Treatment_75	0.319	6.951***	7.075***	6.272***	4.045*	4.766*	1.412
	(0.880)	(1.365)	(1.814)	(1.956)	(2.417)	(2.592)	(2.402)
PEPSU	-2.161***	-1.755	0.317	-1.800	3.115	1.622	1.965
	(0.757)	(1.232)	(1.477)	(1.626)	(2.082)	(2.003)	(1.751)
Treatment_75 × PEPSU	5.583***	2.648	-2.309	0.817	-10.737*	-5.002	-11.734*
	(1.831)	(4.066)	(4.182)	(4.430)	(6.481)	(5.427)	(6.508)
Observations	8,284	7,743	7,573	5,895	7,902	7,652	7,818
No. of clusters	260	259	257	249	260	260	260
Cragg–Donald F	761.1	732.3	704.0	716.7	707.5	715.2	523.5
Kleibergen–Paap F	9.435	8.546	8.798	11.73	9.322	9.720	8.550

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on literacy rates across regions that were part of the princely state of PEPSU and those under direct British administration. The treatment variable is an indicator for villages without refugees and above the 75th percentile (Treatment_75 = 1). All regressions impose a 10% area-change restriction and include village-level controls. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 27: Heterogeneous Effects of Refugee Exposure on Employment by Princely States (PEPSU)

	Employment Outcomes								
	1991			2001			2011		
	Total Work	Main Work	Marginal Work	Total Work	Main Work	Marginal Work	Total Work	Main Work	Marginal Work
IV Estimates									
PEPSU	-0.081	-0.837	0.756	1.890	-1.972	3.862*	2.834**	0.029	2.805**
	(0.711)	(0.660)	(0.742)	(1.323)	(1.505)	(1.755)	(1.005)	(1.148)	(1.087)
Treatment_75	-0.139	3.120***	-3.258***	-5.196***	3.182***	-8.378***	2.889**	5.869***	-2.980***
	(0.896)	(0.475)	(0.684)	(1.069)	(0.983)	(0.949)	(1.373)	(1.177)	(0.786)
Treatment_75 × PEPSU	2.790	7.425**	-4.635	-5.822	15.162**	-20.983**	-3.699	8.915*	-12.613**
	(1.773)	(3.336)	(2.884)	(5.151)	(6.501)	(8.981)	(3.117)	(5.252)	(5.187)
Observations	7,710	7,710	7,710	7,652	7,652	7,652	7,818	7,818	7,818
No. of clusters	247	247	247	260	260	260	260	260	260
Cragg–Donald F	690.3	690.3	690.3	713.7	713.7	713.7	522.0	522.0	522.0
Kleibergen–Paap F	9.405	9.405	9.405	9.800	9.800	9.800	8.558	8.558	8.558

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on employment outcomes across PEPSU and non-PEPSU regions. The treatment variable is an indicator for villages without refugees and above the 75th percentile of refugee exposure. All regressions impose a 10% area-change restriction and include village-level controls. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 28: Heterogeneous Effects of Refugee Exposure on Occupational Structure by PEPSU

	Occupation Shares											
	1991				2001				2011			
	Cultivators	Agricultural Labour	HH Industry	Other Services	Cultivators	Agricultural Labour	HH Industry	Other Services	Cultivators	Agricultural Labour	HH Industry	Other Services
IV Estimates												
PEPSU	-0.421	-1.765	-0.962	0.637	3.428**	-1.065	-0.468**	-8.497***	3.539**	-5.951***	0.088	-2.513
	(1.787)	(2.444)	(0.641)	(1.592)	(1.432)	(1.775)	(0.223)	(2.514)	(1.574)	(1.986)	(0.274)	(1.958)
Treatment_75	-1.464	16.451***	-1.835***	-4.837***	-0.260	11.385***	0.874***	4.762**	1.754	10.457***	0.599*	-3.033
	(2.505)	(1.711)	(0.635)	(1.431)	(2.164)	(1.198)	(0.266)	(2.047)	(2.204)	(1.688)	(0.315)	(2.622)
Treatment_75 × PEPSU	0.760	19.097	3.879**	-9.785*	-10.027**	15.057	2.081**	33.884***	1.381	28.008***	-0.337	0.529
	(5.506)	(12.779)	(1.799)	(5.850)	(4.989)	(9.209)	(0.999)	(12.964)	(4.811)	(10.034)	(0.930)	(6.095)
Observations	7,710	7,710	7,710	7,710	7,651	7,651	7,651	7,651	7,818	7,818	7,818	7,818
No. of clusters	247	247	247	247	260	260	260	260	260	260	260	260
Cragg–Donald F	690.3	690.3	690.3	690.3	713.6	713.6	713.6	713.6	522.0	522.0	522.0	522.0
Kleibergen–Paap F	9.405	9.405	9.405	9.405	9.800	9.800	9.800	9.800	8.558	8.558	8.558	8.558

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on occupational structure across PEPSU and non-PEPSU regions. The treatment variable is an indicator for villages without refugees and above the 75th percentile of refugee exposure. All regressions impose a 10% area-change restriction and include village-level controls. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 29: Heterogeneous Effects of Refugee Exposure on Welfare Outcomes by PEPSU

	Welfare Outcomes					
	Log(Rural per capita consumption)	Log(Rural HH consumption)	Rural poverty rate	Tendulkar poverty rate	Highest Income > Rs. 5,000	Highest Income > Rs. 10,000
IV Estimates						
PEPSU	0.040	0.019	-0.030**	-0.024**	0.007	0.024*
	(0.032)	(0.021)	(0.015)	(0.012)	(0.019)	(0.014)
Treatment_75	0.177***	0.079***	-0.034	-0.019	0.021	0.036***
	(0.040)	(0.027)	(0.022)	(0.017)	(0.022)	(0.013)
Treatment_75 × PEPSU	0.152*	0.132**	0.015	0.019	-0.032	-0.033
	(0.090)	(0.058)	(0.043)	(0.035)	(0.052)	(0.038)
Observations	7,791	7,791	7,791	7,791	7,791	7,791
No. of clusters	260	260	260	260	260	260
Cragg–Donald F	523.2	523.2	523.2	523.2	523.2	523.2
Kleibergen–Paap F	8.550	8.550	8.550	8.550	8.550	8.550
Mean of outcome	10.08	11.72	0.143	0.0965	0.423	0.175

Notes: The table reports IV estimates of the heterogeneous effects of refugee exposure on welfare outcomes across PEPSU and non-PEPSU regions. The treatment variable is an indicator for villages without refugees and above the 75th percentile of refugee exposure. All regressions impose a 10% area-change restriction and include village-level controls. IV estimates instrument refugee exposure using distance to the Indo–Pak border. The Stock–Yogo critical value for 10% maximal IV bias is 7.03. Standard errors are clustered at the 1991 panchayat level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

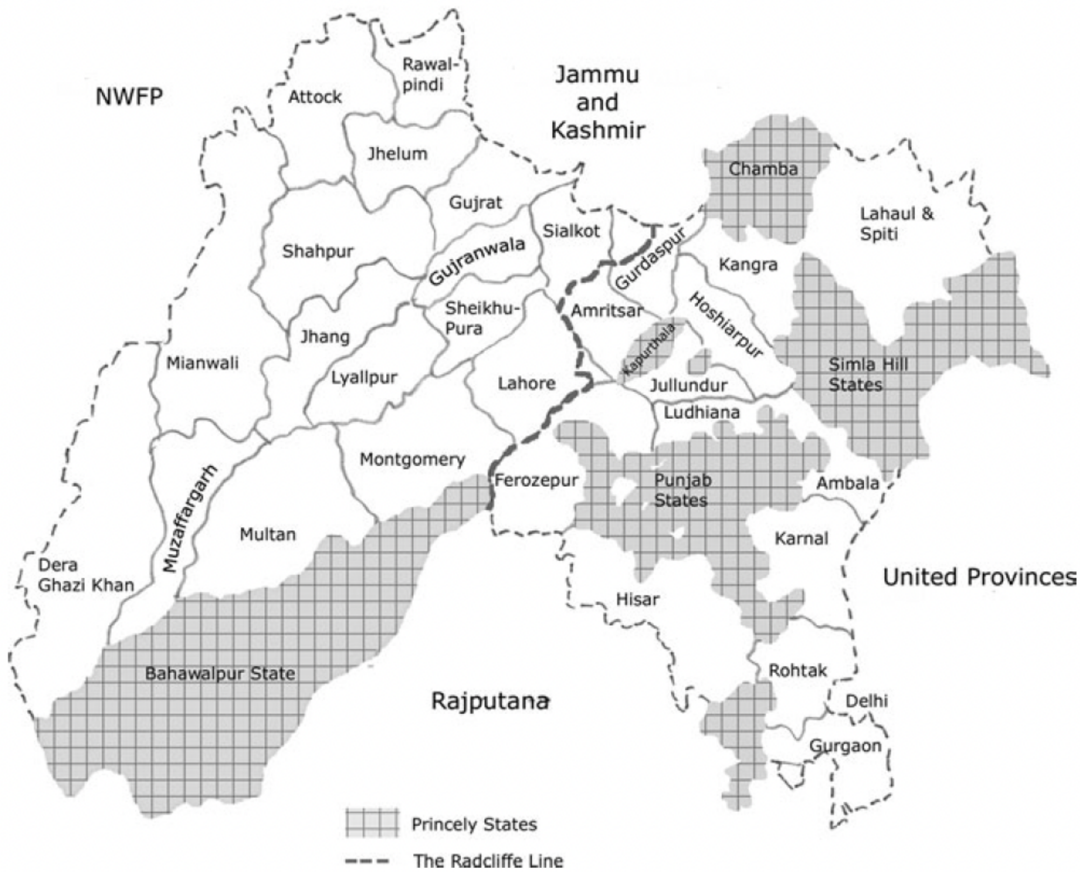


Figure 2: Map of Punjab in 1947

Notes: This district-level map depicts the division of British India along the western border in 1947. The Radcliffe Line demarcated the two newly formed dominions—India and Pakistan. Districts to the right of the Radcliffe Line fall within India, while those to the left correspond to West Pakistan. *Source:* Virdee, Pippa. *From the Ashes of 1947: Reimagining Punjab.*

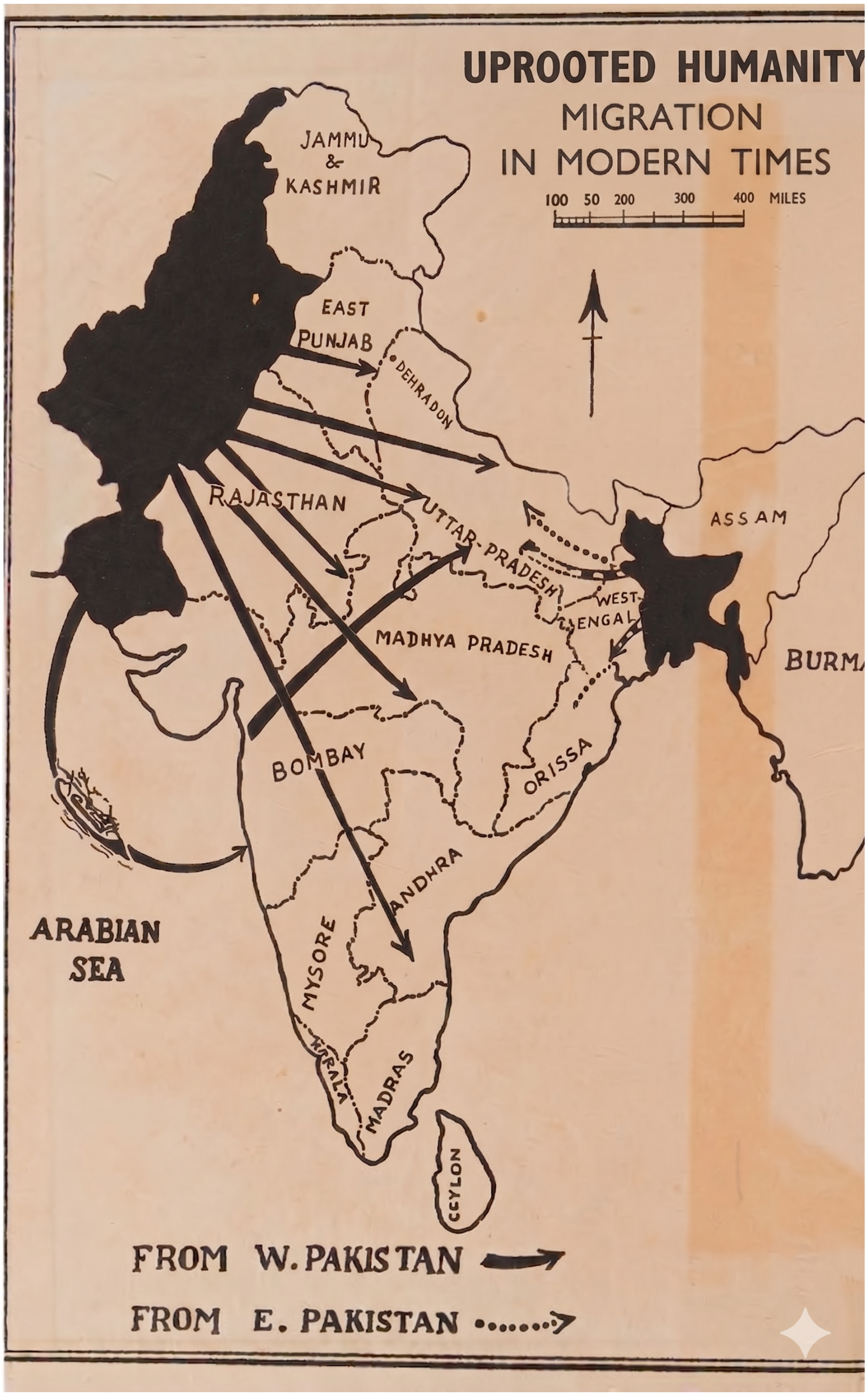


Figure 3: Migratory flows during Partition, 1947

Notes: This stylized historical map illustrates the general direction of refugee migration into India following the 1947 Partition of British India. Solid black arrows indicate flows from West Pakistan and dotted arrows indicate flows from East Pakistan. This map is for illustrative purposes only, depicting the direction of migratory flows and the division of British India.

Source: Saksena R. N. *Refugees : A study in changing attitudes.*

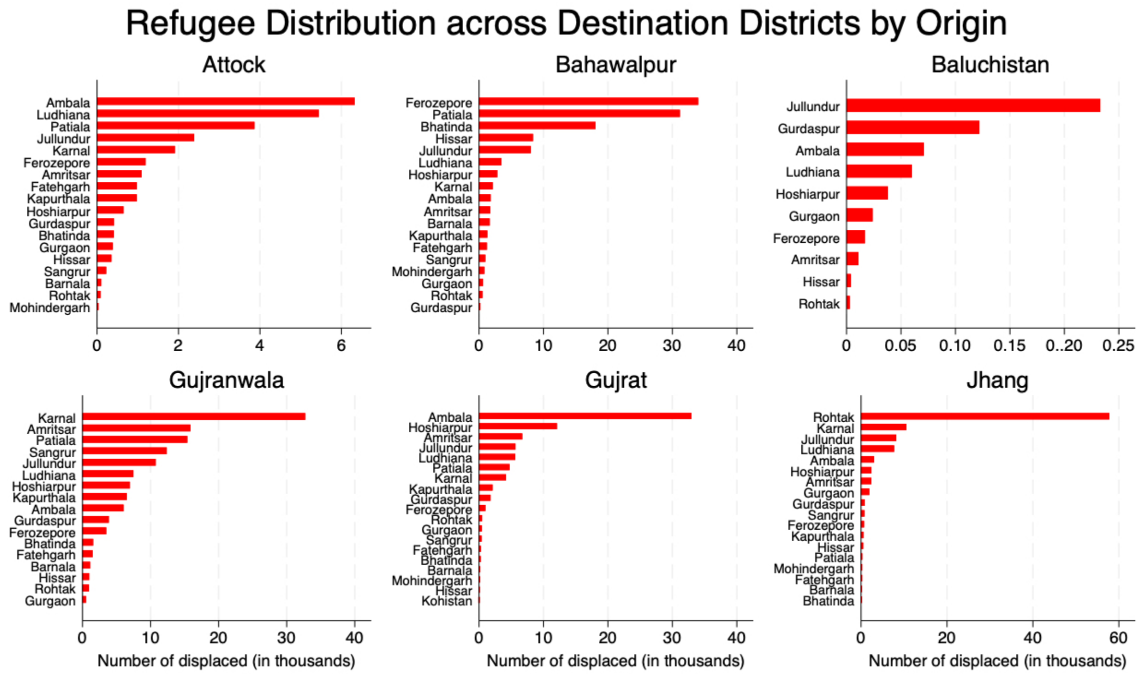


Figure 4: Resettlement of refugees by district of origin

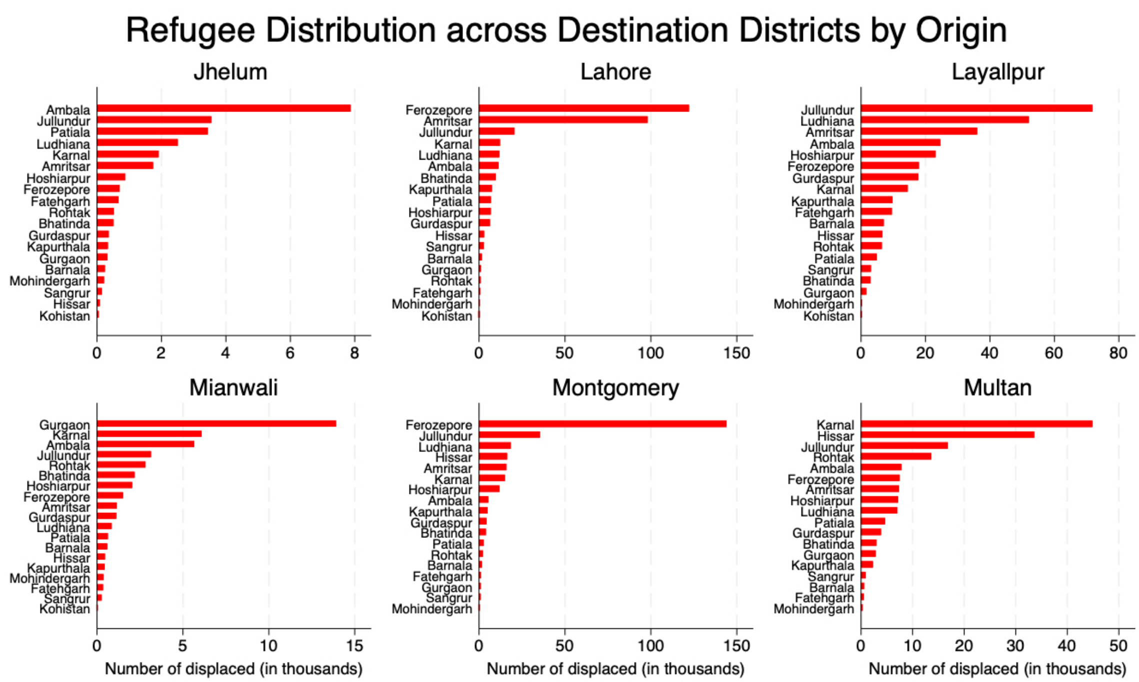


Figure 5: Resettlement of refugees by district of origin

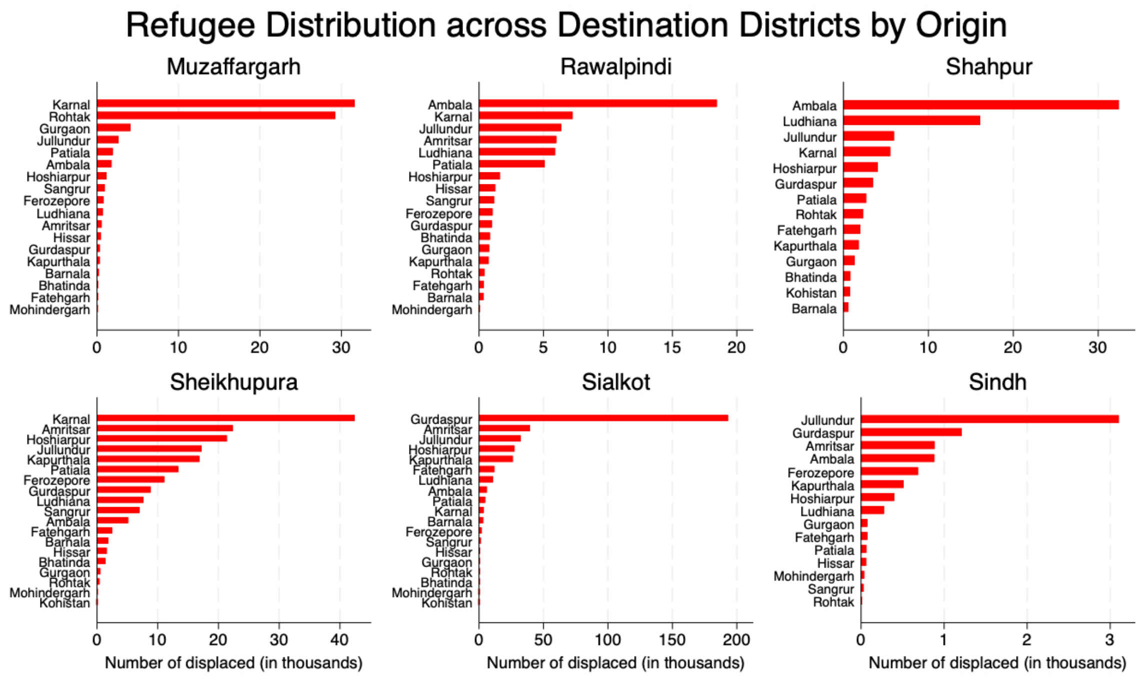


Figure 6: Resettlement of refugees by district of origin

Notes: Figures 3–5 illustrate the distribution of refugees across Indian destination districts by their district of origin. Each graph plots the number of displaced persons (in thousands) from a given origin district across destination districts, arranged in descending order of refugee inflows. These graphs help assess the extent to which refugee resettlement followed the official allocation policy and provide evidence on whether variation in refugee presence across districts was plausibly exogenous.

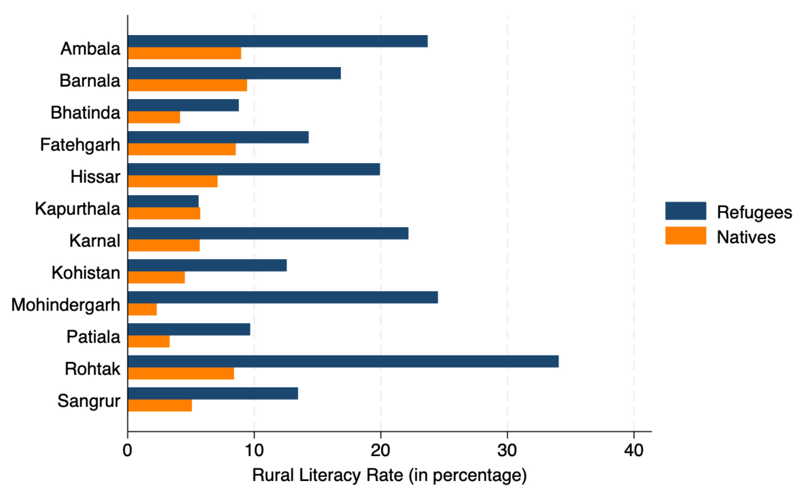


Figure 7: Literacy rates among refugees and natives across districts - Rural

Notes: This figure plots rural literacy rates for natives and refugees using data from the 1951 District Census Handbooks (Table C-IV). It shows that refugees had substantially higher literacy levels than the native population in rural areas in 1951. The number of districts is smaller than in the main sample because these tables are not available for all districts.

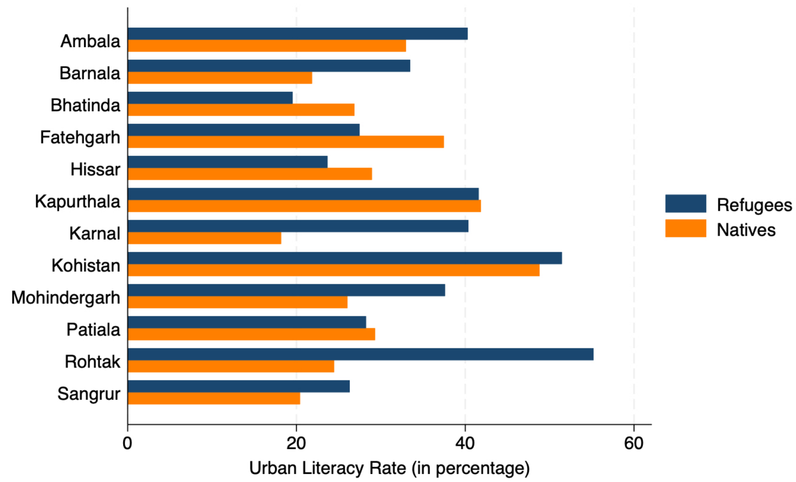


Figure 8: Literacy rates among refugees and natives across districts - Urban
Notes: This figure plots urban literacy rates for natives and refugees using data from the 1951 District Census Handbooks (Table C-IV). In more than half of the districts, refugees exhibit substantially higher literacy rates than the native population in urban areas in 1951. The number of districts is smaller than in the main sample because these tables are not available for all districts.

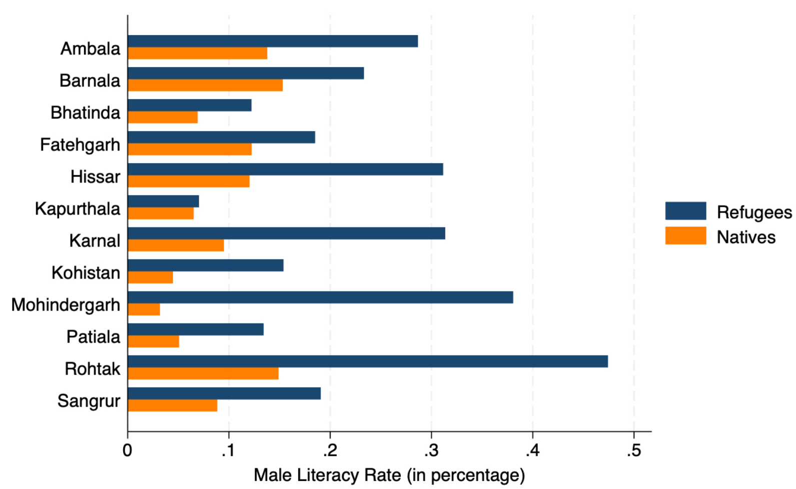


Figure 9: Comparison of male Literacy rates - Rural
Notes: This figure compares rural male literacy rates for natives and refugees using data from the 1951 District Census Handbooks (Table C-IV). It shows that refugee males had higher literacy rates than native males across all districts. The number of districts is smaller than in the main sample due to limited availability of these tables.

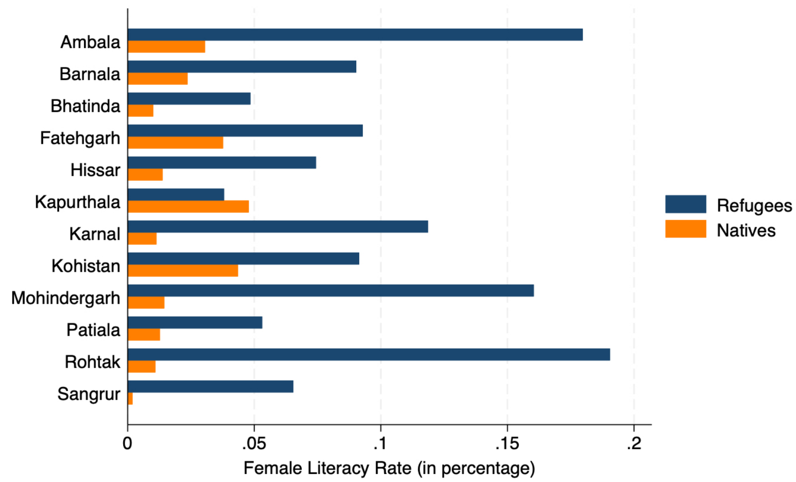


Figure 10: Comparison of female Literacy rates - Rural

Notes: This figure compares rural female literacy rates for natives and refugees using data from the 1951 District Census Handbooks (Table C-IV). It shows that refugee females had higher literacy rates than native females across all districts except Kapurthala. The literacy gap is more pronounced for females than for males. The number of districts is smaller than in the main sample due to limited availability of these tables.

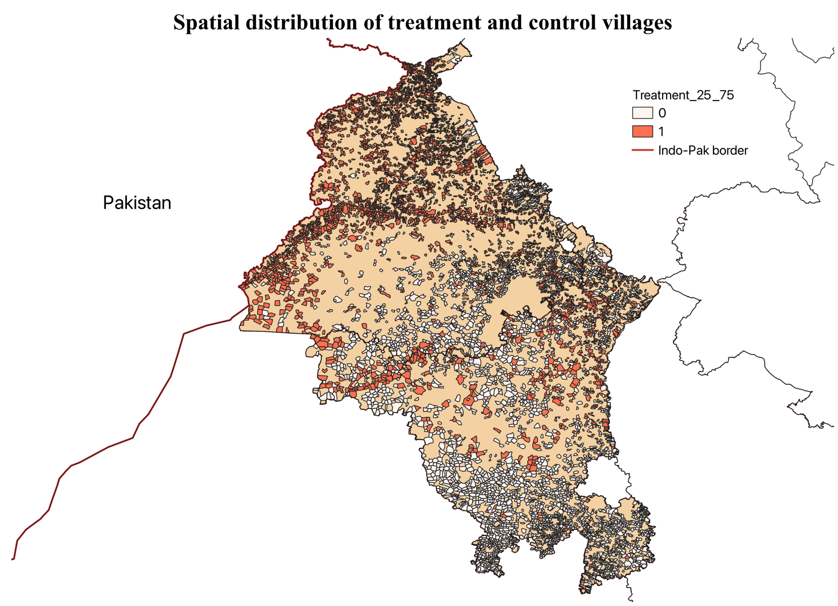


Figure 11: Spatial distribution of displaced persons (treatment_75)

Notes: This map shows the spatial distribution of villages classified under the *treatment_75* definition. Villages shaded in orange (dummy = 1) are high-inflow villages, defined as those with refugee shares greater than or equal to 22.5 percent of the village population (75th percentile). Villages shown in white (dummy = 0) received no refugees. Villages with intermediate levels of refugee exposure are excluded by construction.

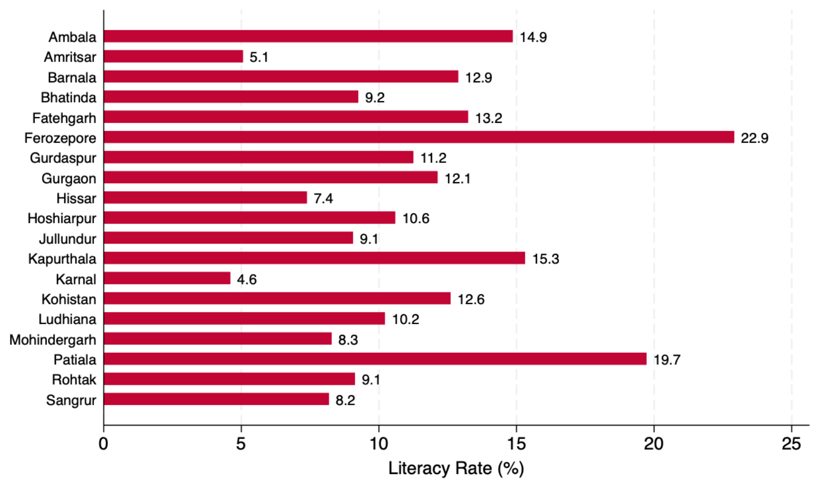


Figure 12: 1951 literacy rates across districts
Notes: This figure shows district-level literacy rates in 1951 using Census data. It highlights substantial spatial variation in initial literacy levels across districts at the time of Partition.

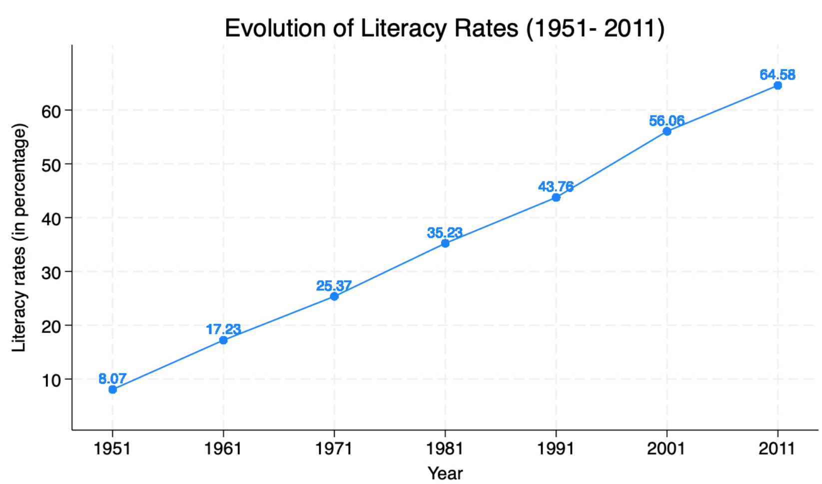
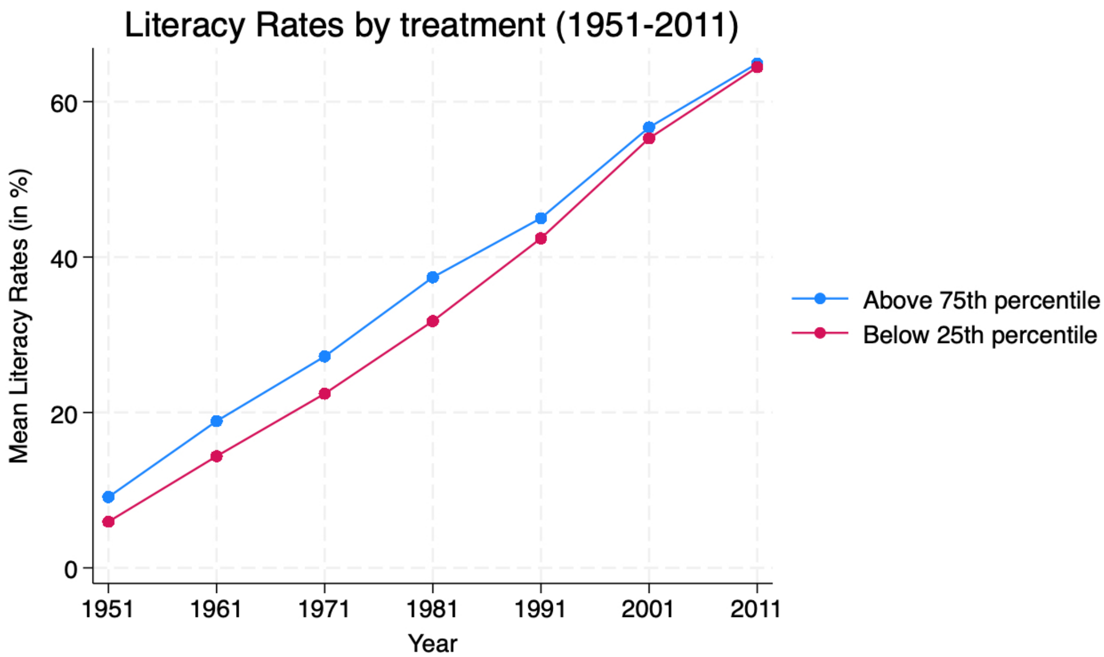
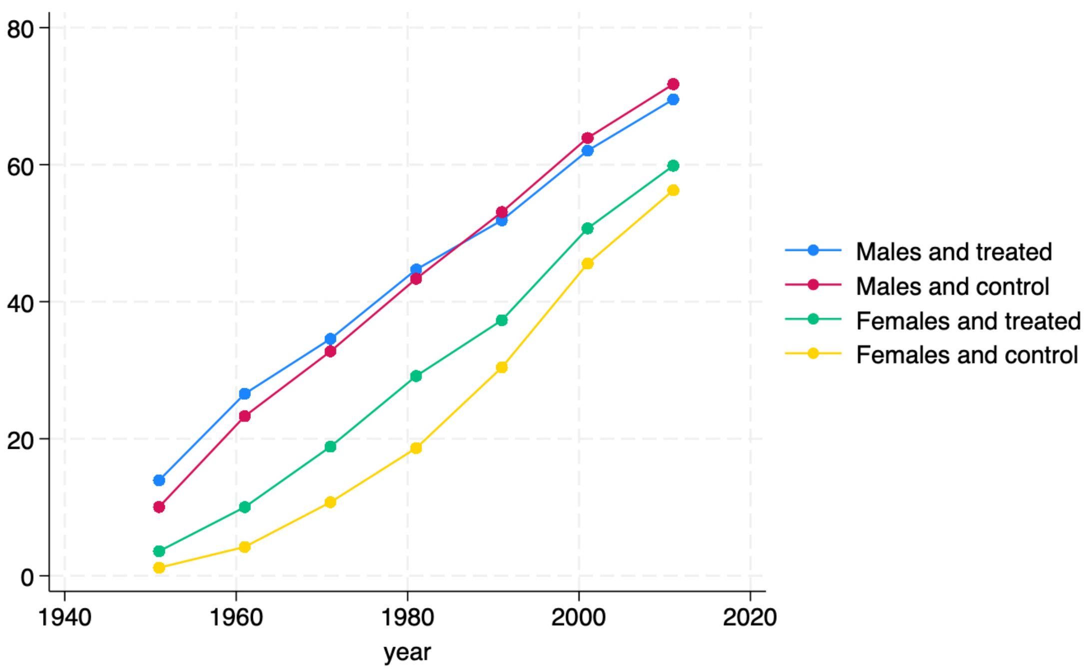


Figure 13: Evolution of literacy rates
Notes: This figure plots the evolution of average literacy rates from 1951 to 2011 using Census data. We can see a steady increase in literacy over time, with levels gradually converging toward the national rural average. In 1951, the average literacy rate in the sample villages was 8.07 percent, compared with 12.1 percent in rural India. Over the next six decades, literacy increased to 64.58 percent by 2011, gradually converging to the rural national average of 66.8 percent.


Literacy Rates by treatment (1951-2011)
Mean Literacy Rates (in %)
0
20
40
60
1951
1961
1971
1981
1991
2001
2011
Year
Above 75th percentile
Below 25th percentile


0
20
40
60
80
1940
1960
1980
2000
2020
year
Males and treated
Males and control
Females and treated
Females and control