Discussion Paper Series

IZA DP No. 18736

June 2026

The Short- and Long-Run Effects of Railroads on Mexico-US Migration

David Escamilla-Guerrero
University of St Andrews
and IZA@LISER

Giovanni Peri
UC Davis, CEPR, NBER
and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18736 June 2026

The Short- and Long-Run Effects of Railroads on Mexico-US Migration*

Abstract

This paper leverages variation in the access to the Mexican railroad network in the early 1900s to estimate its impact on migration to the United States and evaluate its long-run persistence after passenger rail service became obsolete. Using an IV strategy based on least-cost paths between historical cities, we find that locations with railroad access had migration rates four times higher than those without in the early twentieth century. Sequential migration was the key mechanism: railroads first facilitated internal mobility toward railroad hubs, then onward migration to the US. Railroad access also contributed to structural transformation, raising urbanization and local economic development. In terms of persistence, locations with historical railroad access show weakly lower total migration rates to the US in the early 21st century, consistent with local economic growth reducing the incentive to migrate. Yet destination-specific patterns prove remarkably durable: locations that disproportionately sent migrants to California, Arizona, or Texas in the 1900s continued to do so in the 2000s, reflecting the persistence of migrant networks.

JEL classification

J61, N36, N76, O31, R42

Keywords

railroads, Mexico-US migration, migration networks, economic growth

Corresponding author

David Escamilla-Guerrero
david.escamilla-guerrero@st-andrews.ac.uk

* *Acknowledgments*: Alfonso Marin Cano and Victor Ortega Le Hénanff provided excellent research assistance. For helpful comments, we thank Vellore Arthi, Dan Bogart, Leah Boustan, Leander Heldring, Leticia Arroyo-Abad, Mushfiq Mobarak, and Martín Fernández. We also thank seminar and conference participants at UC Merced, UC Irvine, UC Davis Global Migration Center, Yale University, the Rockwool Foundation Migration Forum, the Bristol Applied Economics Meetings, the HUMAN LACEA Migration Network, and the European Society for Population Economics Annual Meeting.

1. Introduction

The construction of railroads during the 19th century dramatically reduced transport costs and connected countries in unprecedented ways, with deep and broad economic consequences. By facilitating travel and integrating markets, railroads shaped economic growth across many dimensions. A large literature has evaluated the short- and long-run effects of railroad construction on market integration, productivity, and growth (e.g., Banerjee, Duflo and Qian, 2020; Berger and Enflo, 2017; Donaldson, 2018; Fogel, 1962; Jedwab and Moradi, 2016). Less studied are the effects on population mobility—particularly the short- and long-run international migration patterns generated by railroads connecting different countries.

This paper examines the short- and long-run effects of Mexico's national railroad network on Mexico-to-US migration, the largest and most persistent bilateral migration flow of the twentieth century (Borjas, 2007). Mexican migration to the US first became significant in the early 1900s (Gratton and Merchant, 2015), coinciding with a dramatic expansion of Mexico's railroad network and its connection to the American system.[1] Between 1880 and 1910, Mexico's network grew from approximately 900 km to over 20,000 km, making it the third-longest in Latin America (Bignon, Esteves and Herranz-Loncán, 2015). Railroad expansion was then abruptly disrupted by the Mexican Revolution (1910–1920) and never resumed at scale until very recently. In the subsequent decades, negligible investment was directed toward rail infrastructure, leaving the network virtually unchanged for nearly a century until 2023.[2] More consequential still was the automobile revolution, which caused rail's share of total passenger transportation to collapse from over 70% in 1940 to nearly zero by the 1990s.

These unique historical conditions allow us to ask three central questions about the railroad's impact on local Mexican communities and migration to the US. First, (i) how did railroad construction in 1880–1910 affect migration to the US and local development in connected communities relative to unconnected ones? Second, (ii) did the railroad channel Mexican migration to the US directly, or primarily through sequential internal migration toward railroad hubs? Third, (iii) did the effects of railroad access on total and destination-specific migration persist into the early 21st century, long after paved roads and motor vehicles had displaced rail passenger transportation?

[1] The Mexican and American networks were first connected at El Paso, Texas in 1884.

[2] Two passenger railroads were inaugurated in 2023: a 1,550 km line connecting cities across the Yucatan Peninsula, and a 50 km line linking Toluca to Mexico City—the first projects of such scale in almost one hundred years.

The first question contributes to understanding the role of railroads in mobility and local economic development. The second focuses on whether railroads enhanced long-distance migration through a two-step process—first attracting internal migrants, then facilitating onward movement to the US. The third sheds light on the long-run persistence of early migration patterns along the world's most important bilateral migration corridor. This last contribution speaks directly to the origin and durability of migrant networks and point-to-point connections between the two countries, which remain highly relevant for transmitting economic conditions across local economies (Caballero, Cadena and Kovak, 2018, 2023).

Answering these questions requires overcoming three empirical challenges. The first is collecting, geo-referencing, and organizing data on all trunk and branch lines constructed in Mexico before 1910, which allow us to identify which districts had railroad access by that date.

The second is assembling a dataset spanning more than a century. We collect and digitize micro-level data from historical US border-crossing records and tabulated data from historical censuses to capture migration patterns to the United States and within Mexico around 1910. We also compile district-level variables for 1880–1910 covering geographic, demographic, and socioeconomic characteristics, as well as measures of local development and structural change. In addition, we harmonize administrative migration data from the *Matrícula Consular de Alta Seguridad* (MCAS) identification card program (2006-2019) to historical district boundaries, allowing us to measure present-day Mexico-to-US migration for the same spatial units observed in 1910.

Third, and most importantly, is addressing the endogeneity of railroad placement. We use an IV strategy based on least-cost paths—a hypothetical network connecting Mexico's principal cities to one another and to the United States while minimizing construction costs.[3] Distance to the nearest least-cost path serves as an instrument for railroad access. The identifying intuition is that two otherwise similar districts—one closer to a least-cost path, one farther—differ in their likelihood of railroad access solely by virtue of their location, allowing us to isolate the effect of access while controlling for other geographic characteristics.

We find three main results. First, railroad access had a significant positive effect on Mexican emigration to the United States around 1910. Districts with railroad access had 1.4–1.6 more migrants per thousand inhabitants than those without—roughly four times the mean migration rate in unconnected locations.

[3] This approach follows recent work on the developmental effects of railroads; see, for example, Atack et al. (2010); Banerjee, Duflo and Qian (2020); Berger and Enflo (2017); Fenske, Kala and Wei (2023), discussed further below.

This result is robust to controlling for location fundamentals, demographic structure, labor market characteristics, living standards, landholding patterns, and information access.

The second result concerns the mechanism. We examine whether increased emigration to the US was accompanied by internal in-migration—a two-step process in which workers first relocated to railroad hubs before crossing the border. To do so, we decompose Mexico-US migration into *direct* migration (individuals whose place of birth and last residence coincide) and *sequential* migration (individuals who had relocated within Mexico before crossing the border).

Prior literature argues that a direct train journey to the US border was unaffordable for the average laborer (Coatsworth, 1979), making sequential migration a practical alternative: migrants could work along the way, finance their journey, and gather information (Artuc and Ozden, 2018). Our analysis shows that sequential (two-step) migration—not direct migration—largely accounts for higher emigration rates in railroad-connected districts. People born elsewhere moved to railroad hubs and subsequently emigrated to the US. We also show that this transitory population contributed to local economic development, acting as consumers and temporary workers and amplifying structural transformation in railroad locations.

The third result concerns long-run consequences. We find that the relationship between 1910 railroad access and total emigration reversed. Locations with historical railroad access had *lower* migration rates in 2006–2019. This finding holds after controlling for contemporary transportation infrastructure as well as historical factors, including networks, land reforms, and immigration policy. The result is consistent with these locations having developed faster, reaching income levels by the 2000s at which emigration and local development are inversely related (Dao et al., 2018). Over the long run, early industrialization spurred by railroad access reduced the economic incentive to emigrate.

In contrast, the most enduring legacy of the railroad is in destination-specific migration patterns. Locations that sent a disproportionate share of migrants to California, Arizona, or Texas in 1910—owing to their position along specific trunk lines—continue to send a disproportionate share to those same destinations in 2006–2019. This persistence survives the automobile revolution and holds after controlling for paved road connections. Specifically, districts connected in 1910 to a given trunk line send significantly fewer migrants today to destinations served by other trunk lines, even when these alternative lines are geographically close and accessible by road. The railroad infrastructure built over a century ago created persistent migration corridors that continue to shape destination choices today.

We contribute to a large literature on the effects of transportation infrastructure on economic development (e.g., Atack et al., 2010; Donaldson and Hornbeck, 2016; Donaldson, 2018; Fogel, 1962). Most of this research focuses on outcomes other than migration, and the few studies examining migration responses to transportation infrastructure consider mobility within countries (e.g., Black et al., 2015; Fenske, Kala and Wei, 2023; Morten and Oliveira, 2024). Our contribution—estimating the short- and long-run effects of railroad access on cross-border, Mexico-to-US migration—is new. So too is our finding that railroad access increased emigration primarily through sequential internal migration rather than direct border crossings, a mechanism that Artuc and Ozden (2018) and a handful of others have documented in related contexts.

We contribute to the migration economics literature in two additional ways. First, a large body of work documents that international migration is organized around networks: early migrants establish corridors that reduce costs for subsequent migrants, generating strong persistence (e.g., Hatton, 1995; McKenzie and Rapoport, 2007; Spitzer and Zimran, 2024). What generates these early networks, however, is less well understood. While Abramitzky and Boustan (2017) and others emphasize migrant-built institutions as a source of network persistence, we are among the first to show that the specific geographic direction of early railroad construction had an enduring effect on stepwise migration to the United States, creating self-reinforcing corridors that long outlasted the infrastructure itself. Second, we document, for the Mexico-US corridor, the non-monotonic relationship between economic development and emigration. Higher emigration was associated with faster local development in the early phase (circa 1910), while lower emigration characterized more developed districts in the later period, when Mexico had become a middle-income country. This inverse U-shaped relationship between emigration and income per capita documented in Dao et al. (2018) is directly confirmed in our results.

Finally, we contribute to the history of Mexico-US migration. Extensive research examines contemporary drivers of Mexican migration (e.g., Angelucci, 2015; Chort and De La Rupelle, 2016; Hanson and Spilimbergo, 1999; Munshi, 2003; Caballero, Cadena and Kovak, 2018), but causal evidence on the early period is scarce and contested. Some historians argue that railroads were fundamental to the onset of mass migration (Cardoso, 1980; Durand, 2016; Fogel, 1978); others contend that ticket prices for medium- and long-distance rail travel were prohibitive for most workers, limiting the railroad's role (Coatsworth, 1979). Our findings reconcile these views: railroads were indeed heavily used to reach the border, but primarily through sequential migration that allowed workers to make shorter, affordable trips and to accumulate the resources needed for onward movement. Sequential migration was a mechanism for relaxing credit constraints on international migration, and this paper is the first to document its quantitative importance in this context.

The remainder of the paper is organized as follows. Section 2 provides historical background on railroad construction in Mexico and the conceptual framework. Section 4 describes the data. Section 5 presents the empirical strategy and identification. Sections 6 and 7 report short-run effects on migration and other outcomes, and Section 8 presents long-run migration effects. Section 9 concludes.

2. Historical Background

2.1 Mexico-US Migration

From a historical perspective, the presence of Mexicans in the United States started around 1848, when the US-Mexican War ended and Mexico conceded more than half of its territory.[4] As a consequence, thousands of Mexicans who lived in the newly-acquired American territories became the first Mexican "immigrants" in the US, without ever leaving their home (Henderson, 2011; Kosack and Ward, 2020). While the population of Mexican origin grew slowly afterward, it was not until the 1900s that Mexican immigration increased sharply and expanded its geographic range of settlement in the United States (Durand, Massey and Zenteno, 2001; Feliciano, 2001; Gratton and Gutmann, 2000).

Sustained migration from Mexico to the US was maintained during the 1910's and 1920's up to 1929, when the Great Depression and a campaign of deportations of Mexicans from the US stopped the flow (see Lee, Peri and Yasenov (2022)). During the 40's the introduction of the Bracero Program re-started the migration of Mexicans to the US, especially as workers in Agriculture (see Clemens, Lewis and Postel (2018)). The end of this program in 1962, marked an initial decline in the US-Mexico migration, but then the Immigration and Nationality Act of 1965, opened the door to the large migration of Mexicans to the US, in the 1980's and 1990's, to become the largest immigrant group in the US by year 2000 (Borjas (2007)). Most of the focus of economists has been in studying the determinants and effects of these migration flows starting in the 1920's. However, their persistence for more than a century implies that it is important to understand how these flows started and what features persisted over time.

To explain the beginning of Mexican mass migration (before 1920), previous literature focused on the asymmetric economic conditions between Mexico and the United States at the time. The economic

[4]The conceded territories were California, Utah, Nevada, and most part of Arizona, New Mexico, and Colorado. The US-Mexican War also formalized the loss of Texas, which had been admitted to the American Union in 1845.

boom of the American Southwest (Gold Rush, railroad building, industrialization) represented greater employment opportunities in agriculture, mining, and railroads for the Mexican population, whose living standards in Mexico had deteriorated since the 1870s (Gratton and Merchant, 2015; Oñate, 1991; López-Alonso, 2007). Mexicans could earn at least four times more in the United States due to both the stagnation of wages in Mexico and the depreciation of the peso (Fogel, 1978; Rosenzweig, 1965).[5] In addition, high land concentration by a small elite had perpetuated economic and political inequality since colonial times (Chevalier, 1970; Florescano, 1987; Sellars and Alix-Garcia, 2018; Sokoloff and Engerman, 2000), implying that for most agricultural workers, migration to the US in search of a more prosperous life was an extremely attractive option. These incentives were reinforced by an open-border policy, which allowed Mexicans to cross the border without major immigration restrictions.[6] These elements, however, had been in place since the mid-nineteenth century, raising questions on their ability to explain the surge and induced mass migration to the United States, beginning around 1900.

Similar to European mass migration, improvements in transportation technology likely played a pivotal role in the history of Mexico-US migration.[7] With the dramatic expansion of railroads during the late 19th century, travel times and migration costs fell, making it feasible for populations farther from the border to migrate to the United States. Given this transportation improvement, many have argued that access to railways was indispensable for the surge of Mexican mass migration (Cardoso, 1980; Feliciano, 2001; Durand, 2016; Gamio, 1930). This argument is supported by the fact that the Mexican and American railroad networks were first connected during the 1880s and 1890s, shortly before Mexican migration to the United States became very significant. An important objection to the role of railways as a fundamental catalyst for Mexico-US migration was the observation that stagnant wages and liquidity constraints made train tickets unaffordable for the majority of the Mexican population (Coatsworth, 1979). Most Mexican population was not able to save enough to afford such costs, questioning the impact of the railway on mass migration. An effective explanation of the role of Railway on mass migration requires to explain how Mexican overcame the liquidity barrier. The evidence we show on sequential migration as a way to overcome this constraint is a key contribution of our analysis.

[5] The exchange rate at the time was 2 pesos per US dollar (Clark, 1908).

[6] Before 1910, Mexicans were not considered immigrants who sought to settle permanently, but temporary aliens who moved back and forth supplying labor without facing immigration restrictions (Fogel, 1978; Samora, 1982).

[7] See Abramitzky and Boustan (2017) and Hatton and Williamson (1998) for reviews on Europe-to-US mass migration.

2.2 The Mexican Railroad Network

In Mexico, the construction of railroads was significantly affected by episodes of war and political instability. It was not until 1873 that the first major rail line connecting the port of Veracruz to Mexico City was inaugurated. The country's rugged terrain proved exceptionally challenging and costly for railroad development. To finance the construction of additional railroads, the Mexican government issued concessions and granted subsidies to international companies, predominantly American (Donly, 1920). This policy resulted in the construction of three trunk lines from Central Mexico to the US border (Woodruff and Zenteno, 2007), reducing travel times from several weeks to 45-60 hours depending on the route (Coatsworth, 1979).[8] Multiple branch lines were also constructed by local elites and entrepreneurs to serve individual interests (Coatsworth, 1974).[9] These shorter railroads connected large estates, mines, and small towns to the trunk lines, primarily to facilitate the transport of goods, though some also served for mail delivery and passenger transportation (Pletcher, 1950). The construction of railroads was remarkably rapid. In less than 40 years, Mexico's railroad network increased from about 500 km in 1873 to 20,000 km in 1910 (Moreno-Brid and Ros, 2009), making it the third-longest railroad network in Latin America (Bignon, Esteves and Herranz-Loncán, 2015). In the year 1909 alone, the network transported about 18 million passengers (1.3 passengers per Mexican individual), demonstrating its importance and broad diffusion in allowing individual mobility.

2.2.1 The Decline of Railroad Passenger Transportation

The development and expansion of railroads, however, was interrupted by the Mexican Revolution (1910-1920), one of the deadliest civil wars in modern history (Escamilla-Guerrero, Kosack and Ward, 2025). Since then, investments in railroad transportation were marginal for over a century, until 2023, when a new railroad in the Yucatan peninsula was inaugurated. Furthermore, as described and documented in Figure 1, railroad passenger transportation experienced a dramatic decline over the mid-to-late 20th century (Panel A). The railroads' share of total passenger transportation plummeted from over 70% in 1940 to less than 5% during the 1990s (Panel B).

Two closely linked transformations occurring by 1950 explain the decline of railroad passenger transportation. First, the increasing substitution of passenger services for freight operations (Panel C), as

[8]The three trunk lines are the following. The Mexican Central Railway, connecting Mexico City with El Paso, Texas. The Mexican International Railroad, connecting Durango City to Eagle Pass via Torreon. The Mexican National Railroad, connecting Mexico City to Laredo via San Luis Potosi (Donly, 1920). See Figure H.1 for guidance.

[9]This contrasts with settings such as Sweden, where railroad routes were designed by state planners with broader economic development and connectivity objectives (Cermeño, Enflo and Lindvall, 2022).

freight transport proved more profitable and less sensitive to competition from emerging transportation technologies. Second, the exponential expansion of paved roads and the increasing use of motor vehicles (cars and buses) for medium- and long-distance passenger transportation (Panel D). While the railroad infrastructure remained essentially unchanged after 1910, with no net investment to increase its size and reach, paved roads for automobiles and Buese expanded from 13,500 km in 1950 to over 100,000 km by 2000 (INEGI, 1950–2000).[10] The development of Mexico's road system aimed to increase market integration, stimulate economic growth, and promote nation-building (Elizalde et al., 2026). This technological transition fundamentally restructured migration costs to the United States. By 1990 migrants from Mexico could travel by bus or private vehicle along flexible routes, and reach the border with the US from any Mexican location in a much more direct way, rather than being constrained to fixed railroad routes.

3. Conceptual Framework and Related Literature

Railroads revolutionized land transportation around the world in the 19th century by massively reducing transport costs and time of travel. From Europe and Asia to the Americas, railroads promoted market integration and economic development, and substantially improved passenger transportation, with impacts on travel and mobility.

A large body of literature evaluated the short- and long-run effects of railroads on diverse economic outcomes. These include trade (Bignon, Esteves and Herranz-Loncán, 2015; Donaldson, 2018), innovation (Andersson, Berger and Prawitz, 2023), value of land (Donaldson and Hornbeck, 2016), economic growth (Banerjee, Duflo and Qian, 2020; Fogel, 1962; Herranz-Loncán, 2006), urbanization (Atack et al., 2010; Berger and Enflo, 2017), and state capacity (Cermeño, Enflo and Lindvall, 2022), among others. The findings of this literature range from negligible to very large effects depending on settings, periods and countries.

The impact of railroads on migration is far from being fully understood. On the one hand, railroads can substantially reduce the costs, risks and uncertainty associated with migration by shortening journey times and facilitating information access about destinations (Hatton and Williamson, 1998; Leunig, 2006), encouraging a larger share of the population to migrate, as cost of migration may drop below the potential gains or below the potential migrants' liquidity constraints. On the other hand, by expanding local

[10] See Figure H.2 for Mexico's road system.

markets railroad access contributes to industrialization, economic transformation, it potentially enhances urbanization, producing higher living standards and better economic prospects (Berger, 2019; Fenske, Kala and Wei, 2023). These evolutions, in turn may reduce incentives to emigration while also increasing the benefits of immigration (Sequeira, Nunn and Qian, 2020).

Additionally if fixed and variable costs of migration (transportation and unemployment risk) remain high even after the building of (and gaining access to) railroads, as in (Imbert and Papp, 2020), railroad access may have limited effects on migration. The presence of these differential effects, and potential other hurdles to migration, make the impact of railway access on migration an empirical question with no clear ex-ante prediction.

Additionally, the effects of railroads on emigration may change between the short and long run. Their persistent effect in the long run, depends on the interplay between their impact on local economic development (which may first increase and then decrease the propensity to migrate as shown in Faini and Venturini (1993); Hatton and Williamson (1993); Dao et al. (2018)), their effect on costs of mobility as more transportation networks emerge and the path dependence of emigration networks.

Previous literature demonstrates that railroads continue shaping the distribution of economic activity long after their collapse in developing countries (Jedwab and Moradi, 2016), suggesting that historical railroad access may have enduring effects on development and possibly migration. However, the emergence of new transportation technologies and networks can reallocate economic activity, attenuating railroad-induced migration patterns and producing new equilibria (Donaldson, 2018; Faber, 2014). Nevertheless, the transition to new equilibria may be less likely if institutions, inter-generational persistence, and other migration-facilitating structures (e.g., immigrant aid societies) perpetuate the initial railroad-induced migration patterns. In the long run those are no longer linked to migration costs but they are solidified by path-dependent connections, information and personal linkages (Abramitzky and Boustan, 2017).

There is very little hard empirical evidence on how specific networks of migration arose and how transportation infrastructures, in general, and railroads, in particular, shaped such networks and their persistence. Morten and Oliveira (2024) show that access to highways increased migration rates within Brazil from 1960 to 2000, which explain about one fourth of the welfare gains associated with the expansion of highways. Black et al. (2015) also find that during the Great Migration (1910-1970), railroad access significantly influenced migration of African Americans from the US South, who often moved to locations along the rail lines. In contrast, Fenske, Kala and Wei (2023) shows that migration did not respond to railroad access in India over the period 1881-1931. Note that these findings capture effects on internal migration, leaving unexplored the impact of transportation infrastructure on migration between countries, when the railways of different countries are connected together. This study is the first to look directly at the effect of transportation improvements on cross-country migration.

4. Data

In this section, we describe the main variables used in the empirical analysis, as well as their underlying sources. We employ data at the district level for two separate time periods.[11] Our short-run analysis examines the effects of railroad access on Mexico-US migration patterns in the early 20th century—namely around 1910, shortly after Mexican and American railroads were connected. Our long-run analysis evaluates whether these effects persisted by the early 21st century—almost one century after we observe the short-run effects and when railroad passenger transportation had been negligible for decades.

4.1 Railroad Access

Our key variable of interest is a measure of the district's railroad access in 1910. To construct it, we use a detailed map—published by Mexico's state-owned railroad company (*Ferrocarriles Nacionales de México*)—displaying the routes of all trunk and branch lines constructed before and still existing in 1910 (FNM, 1914). More specifically, we estimate the Euclidean distance from the centroid of each district to the nearest railroad. We then create an indicator variable that equals one if such a distance is less than 40 km and zero otherwise. This criterion is based on Coatsworth (1979), who documents that people were used to walking 30 km per day at the time, but about twice this distance could be covered using stagecoaches or pack animals. Therefore, back-of-the-envelope calculations suggest that a 40 km distance could have been covered in 3 to 5 hours through a combination of walking and stagecoach travel. Considering that the average district area was 5.2 thousand km^2, a 40 km criterion should be a sensible proxy for direct access to railroads before 1910. In the empirical analysis we will also show the results considering smaller and larger distances from the railway to define access. We also illustrate that the impact of railway access decays with distance, becoming insignificant for distances above 70 km. Given that Mexico's railroad network remained virtually unchanged throughout the 20th century, as documented in the historical overview above, we use the same map and criterion to identify districts with historical railroad access in our long-run analysis.

[11] A district is a group of contiguous municipalities (similar to State Economic Areas in the United States).

4.2 Outcome Variables

Early 20th Century Migration Rates

We estimate emigration rates to the US, by Mexican district, using micro data digitized by Escamilla-Guerrero (2020) and Escamilla-Guerrero, Kosack and Ward (2025). These data consist of 35,159 individual records of Mexican-born immigrants arriving at 12 ports of entry along the Mexico-US border between October 1906 and September 1910.[12] The original source of these data are the Mexican Border Crossing Records (MBCRs), immigration forms used by American officials to record the flow of immigrants entering the US and their characteristics. The MBCRs report the place of birth and the last permanent residence for each crossing immigrant. This information enables us to identify the sequential migration patterns of migrants as they distinguish between the place of birth and the last residence.

We classify these locations into districts using the 1910 Census Catalog of Localities and the Mexican Historical Archive of Localities (AHL), both maintained by Mexico's National Institute of Statistics and Geography (INEGI). We collapse the individual-level data to calculate migration rates by district for each of the four 12-month periods covered by the data: October 1906 - September 1907 ... October 1909 - September 1910.[13] We calculate migration rates (per thousand inhabitants) as the number of migrants over the Mexican district population level measured in the 1910 census. We then compute the mean migration rate across years. We implement this procedure separately for locations of birth and last residence. The difference in locations of birth and last residence allows us to identify "direct" and "sequential" migrants and to compute migration rates for each of them.

Early 21st Century Migration Outcomes

For the more recent period (post 2000), we use the *Matrícula Consular de Alta Seguridad* (MCAS) identification card program (Caballero, Cadena and Kovak, 2018) to capture Mexico-US migration patterns. This administrative dataset reports the origin (municipality of birth in Mexico) and destination (county in the United States) of 11,374,440 Mexican immigrants for the period 2006-2019. The large sample size, the comprehensive geographic coverage and the fact that it includes undocumented immigrants imply that MCAS is likely the best data source for studying recent Mexican migration to the United States

[12] The data exclude duplicate entries and records with incomplete geographic information. Border crossings were not systematically recorded before this period (Escamilla-Guerrero, 2020). Although the available data extend to December 1920, we limit the sample to September 1910 to avoid capturing migration responses induced by conflict events during the Mexican Revolution (November 1910 - February 1917).

[13] Aggregating migration rates for 12-month periods (starting in October) avoids issues of seasonal migration patterns. Clark (1908) documents that border crossings increased during planting and harvest seasons.

when focusing on bilateral flows: from Mexican municipalities to US states (or counties). The MCAS data identify new registrations for each year in the sample, representing a good proxy for the flows of Mexicans, comparable to the historical data that we use. One limitation of the data is that we only observe location of birth, which does not allow the identification of sequential migration patterns within Mexico in the 2006-2019 sample. To maintain geographical comparability with the short-run analysis, we assign municipalities of birth to 1910 districts and collapse the data to compute district-level annual migration rates, measured as the number of migrants per thousand inhabitants using district population from the 2020 census. We then compute the mean migration rate over the 2006-2019 period.

4.3 Additional Outcome Variables

We collect data on a series of district characteristics that may have been directly or indirectly affected by the construction of the railroad network and co-vary with emigration rates. We observe these outcome variables in 1910 and thus capture the short-run effects of railroad access. The first set of variables relates to structural economic transformation. We construct measures of urbanization and agglomeration: population density (inhabitants per square kilometer), urban density (urban population per square kilometer), and housing density (dwellings per square kilometer). As indicators of demographic composition, we use data on the working-age population (population share 11-60 years old), working-age sex ratio (male-to-female sex ratio 11-60 years old), and child-woman ratio (children 0-5 years per woman 16-45 years). To measure social development, we use literacy (population share over 11 years old able to read) and health access (health workers per thousand inhabitants). We obtain the aforementioned data from 1910 census tabulates (DGE, 1918). We complement the economic development variables with data on land concentration in 1910 (population share living in large estates—haciendas or ranches) from Sellars and Alix-Garcia (2018).[14]

The second set of variables captures changes in the labor market, which we construct using census tabulates reporting employment counts by occupation for 1895 and 1910 (DGE, 1899, 1918). More specifically, we use data on changes in labor force participation (working-age population share employed or unemployed), employment (working-age population share employed), and unemployment (working-age population share unemployed), as well as on changes in sectoral employment (working-age share employed in extractive, industry, or service sector occupations). This set of variables allows us to capture variation in economic activity at the extensive margin and labor reallocation across sectors.

[14] Productivity in the vast majority of haciendas and ranches was based on labor exploitation and coercive mechanisms of debt bondage that restricted migration (Moreno-Brid and Ros, 2009; Sellars and Alix-Garcia, 2018).

4.4 Control Variables

We collect data on a rich array of district-level geographic characteristics that may have affected railroad construction and migration, as they are correlated with transportation costs, market access, and labor mobility. As these variables are exogenous to economic development, we include them as controls in most specifications to isolate more precisely the effect of railroad access. Specifically, they comprise distance from the US border and distance from the three principal historical seaports—Veracruz, Tampico, and Acapulco—all entered linearly and quadratically, as well as the mean altitude, the longitude-latitude interaction, and district surface area.[15] We also include an indicator for districts in the Yucatan Peninsula, as markets and transport networks in this region remained isolated from central Mexico until the mid-20th century.

We additionally include measures of agricultural productivity potential, related to land quality. We use land suitability indices for Mexico's main staple crops—beans, carrots, coffee, maize, potatoes, tomato, and wheat—from the Global Agro-Ecological Zones (GAEZ) project (FAO, 2021). Finally, to account for contemporaneous climate shocks, we use drought severity data for the years when migration is measured. These data come from the Mexican Drought Atlas and consist of reconstructions of the self-calibrating Palmer Drought Severity Index (PDSI) on a 0.5° latitude/longitude grid centered over Mexico (Stahle et al., 2016).

We use the same set of controls in both the short- and long-run analyses, with two adjustments. First, we replace the historical seaport distances with distances to the principal contemporary seaports (Altamira, Guaymas, Lázaro Cárdenas, Manzanillo, Salina Cruz, Tampico, Tuxpan, and Veracruz) and include the distance (linear and quadratic) to the nearest paved road from CONABIO (2008). The former controls for the evolution of market access over the 20th century, while the latter accounts for the presence of roads. As Mexico's railroad passenger transportation declined dramatically after 1950, controlling for road access ensures that any observed differences in migration patterns between historically railroad-connected and unconnected locations reflect persistence in migration patterns rather than effects of contemporary transportation networks built along the railroad lines. Second, we use the Palmer Drought Severity Index for the period 2006-2012, overlapping with the examined migration period. Additional control variables are described as we present the results.

[15] Distance to the US border is strongly correlated with latitude in Mexico, so we avoid including both simultaneously.

4.5 Summary Statistics

Panel A of Figure 2 displays the railroad network of Mexico ca 1906, which remained virtually unchanged throughout the 20th century. There are two aspects to note. First, the network was largely shaped by terrain ruggedness. Railroad routes tended to avoid the major mountain systems—the *Sierra Madre Occidental* (west) and *Sierra Madre Oriental* (east)—running instead along the coastal plains and central plateau before converging in central Mexico. Consequently, many regions of Mexico had no direct access to this transportation technology. Second, the network was designed with a clear northward orientation. The main trunk lines connect the Mexican interior to the US border, terminating at key crossing points such as Nogales, El Paso, Laredo, and Brownsville. Moreover, with few exceptions, all ports of entry into the US observed in the historical immigration data were terminus stations connected to the American railroad network, making railroads the fastest means of transportation to travel to the United States at the time (Coatsworth, 1979).

The map in Panel B of of Figure 2 adds to the location of the railroad network, information on the migration rates to the United States in 1910 for each district, with darker shades of blue implying larger rates. Migration rates varied significantly across districts, with rates ranging from 0 to 32 per thousand inhabitants. Migration was particularly intense near the border, as expected, but many districts in the West-Central region had equally high rates. There was also migration from the Center and Southeast regions, though relatively low and concentrated in specific districts (see Figure H.1 for guidance). A feature, visible even in a cursory look at the map, is that the intensity of migration rates is higher (sometimes much higher) in places along the railroad than in places without rail access, even in regions equally far from the border. Consider, for example, the state of Michoacan. Figure F.1 shows that districts crossed by or in proximity to the railroad network, such as La Piedad (0.92), Zamora (0.66), or Pátzcuaro (0.21), exhibit substantially higher migration rates than very close districts with no railroad access, such as Coalcomán (0.009), Ario de Rosales (0.013), or Huetamo (0.007).

Table 1 reports the means of the main variables used in the empirical analysis for districts with (column 2) and without (column 3) access to railroads, along with tests for unconditional and conditional differences in means. Panel A shows that districts with railroad access differed from those without access in 1910 (column 4), but several of these differences disappear when comparing districts within states. Within-state comparisons (column 5) show no economically meaningful differences in distance to the border, maize suitability, or drought severity (among geographical variables), nor significant differences in population of working age, gender ratio, unemployment, or average wage relative to the US. However, districts with railroad access had larger populations, higher urban density, fewer native-language speakers, a lower share of illiterates, and fewer large estates. These locations also had higher emigration rates to the United States: 0.45 more migrants per thousand inhabitants. These statistics reveal an association already after 1910 between railroad, migration and economic development. Panel B shows that locations with historical access to railroads continued to have larger populations, higher urban density, and better overall living standards in 1990—by which time railroads had become obsolete for passenger transportation. While municipalities with historical access to railroads are closer to paved roads, the main passenger transportation mode today, the difference is not economically meaningful (an average of only 0.73 kilometers). Strikingly, locations with historical access to railroads, by year 1990 had significantly lower emigration rates (-2.3 migrants per thousand inhabitants), capturing a reversal of the correlation observed in 1910. This preliminary analysis shows that, in the early 20th century, locations with railroad access had *higher* emigration rates. However, by the end of the 20th century, these same locations had *lower* emigration rates than those without a historical connection, consistent with railroad-induced economic gains first increasing and then decreasing emigration rates as oftent pointed out in the literature finding an inverted U relation between income and emigration (Dao et al. (2018)).

5. Empirical Strategy

To evaluate the short-run effect of railroad access on migration to the United States, we estimate the following equation:

$$mig_i = \alpha + \beta \cdot rail_i + \mathbf{X}_i' \cdot \mathbf{B} + \lambda_{r(i)} + \varepsilon_i, \tag{1}$$

where mig_i is the average migration rate from 1906-1910 in district i; $rail_i$ is an indicator for districts located within 40 km of the nearest railroad as it existed ca. 1906; and $\mathbf{X}_i$ is a vector of district-level controls that accounts for geographical, land, and weather characteristics. $\lambda_{r(i)}$ represents region fixed effects (as defined in Durand (2016)) that control for any time-invariant, region-specific characteristics that may influence migration patterns, such as differences in preexisting migrant networks or cultural traits associated with migration.[16] The coefficient of interest is β, which captures the difference in migration rates between districts with direct access to railroads and those without access. For inference, we report robust standard errors and Conley (1999) standard errors that flexibly allow for spatial correlation within a 100 km cutoff window.

[16] McKenzie and Rapoport (2010) show that migrant networks have been historically stronger in the West-Central region, which today continue to be an important migrant source.

In the baseline specification, we only control for the main location characteristics, namely both linear and quadratic distances to the US border and nearest seaport, an indicator for districts in the Yucatan Peninsula, and region dummies. In more saturated specifications, we additionally control for other location fundamentals (mean altitude and the longitude-latitude interaction), land suitability for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), weather shocks (drought severity index), and district surface area. Note that we do not control for demographics, labor market indicators, living standards, or information access, as these factors are likely outcomes of railroad access rather than independent controls. Nonetheless, we show in a robustness check that our results are not sensitive to their inclusion.

The estimating equation for the long-run effects of railroad access on migration takes the same form as in Equation 1, with migration rates measured over 2006-2019 as the dependent variable. In these specifications, we include two key additional controls. First, distance (linear and quadratic) to the nearest paved road, which accounts for the effects of contemporary transportation networks on migration. Second, a set of covariates capturing historical factors that could confound our long-run estimates, which we discuss in detail when presenting the results. Given the granularity of the modern migration data, we also examine the effects of railroad access at a finer geographic level, using municipalities rather than districts as the unit of observation.

Regardless of the inclusion of control variables, our least square estimates are likely to suffer from endogeneity and omitted variable bias. For example, railroads may connect locations with a pre-existing higher demand for passenger transportation that attracted railroad investments in the first place, which would lead to upward-biased estimates. Conversely, railroad routes may have been systematically built in locations with low migration potential (with the policy goal of revitalizing them), leading to downward-biased estimates due to negative selection into railroad access. To address these issues we develop and use an Instrumental Variable strategy.

5.1 Instrumental Variables Strategy

The key endogeneity issue is that the location of the railroad lines was partly a consequence of unobserved and observed factors, including local economic conditions, population density, and market access. To address this concern we implement an instrumental variables strategy based on identifying the least cost paths (LCPs) connecting historical cities and border crossings. We first use the 1790 Colonial Population Census (Castro Aranda, 2010) and a map of the Mexico-US border published in 1847 (Disturnell, 1847) to identify target destinations with high market potential before the beginning of railroad construction: historical cities and historical border crossing points. We then connect the selected target destinations (nodes) using straight lines, as in Banerjee, Duflo and Qian (2020). We draw a straight line from each city to the nearest city or border crossing point. To account for high construction costs due to terrain ruggedness, we do not draw lines crossing mountain systems with an altitude greater than 1,500 meters. If there are two or more target destinations where the differences in distances is less than 100 km, we draw a line to all of them. We extend the line past the target until it hits a natural barrier to avoid endogeneity arising from stopping the line in termini cities. The resulting LCPs reflect a hypothetical network of transportation routes as they would have been designed by state planners in the late 1800's to connect Mexico's main cities to the United States, with the goal of minimizing construction costs. Crucially, the design ensures that among all districts outside of the nodes (historical cities and border crossings) those along (and close to) these routes are more likely to be traversed by a rail line than those farther from them. This variation in predicted railroad access among otherwise similar municipalities (and equally distant from cities and the border) is what generates the identifying variation, hardly correlated with other factors, hence quasi-random.

We then calculate the distance from the centroid of each district to the nearest LCP, and us this variable as instrument for railroad access. The first-stage regression equation for our IV strategy is:

$$rail_i = \gamma + \phi \cdot lcp_i + \mathbf{X}_i' \cdot \Pi + \lambda_{r(i)} + \mu_i, \tag{2}$$

which predicts the probability of railroad access for each observation. As the proximity to a LCP is mechanically correlated with distance to target destinations (historical cities and border crossings), we control for distance to the nearest target destination in all regressions.

We test several characteristics of the IV to verify that it is a strong and valid instrument. First, the distance to the LCPs is a strong predictor of the location of railroads. Figure 3 displays the LCPs design alongside the railroad network ca 1906, showing that railroads, particularly trunk lines, were constructed near the LCPs. A binscatter plot confirms a strong, positive and roughly linear correlation between distance to LCPs and distance to railroads (see Figure F.2).

The second crucial identifying assumption is that the distance to LCPs must not directly affect migration patterns other than through railroad presence. The instrument is built based on variables (the location of historical cities and border crossings) measured around or before 1860, fully 50 years earlier than the migration and economic outcomes we analyze. This temporal separation ensures that any migration patterns observed in either 1910 or post-2000 are unlikely to be correlated with proximity to LCPs except through railroad access.[17] Additionally, Table 2 shows that (conditional on fixed effects) proximity to LCPs is uncorrelated with pre-railroad characteristics that could independently influence future migration. Locations within 40 km of a LCP are statistically indistinguishable from those farther away across multiple dimensions measured ca. 1800, including population, state capacity (presence of customs offices and sales tax collection), and climatic conditions. Although altitude and distance to mines exhibit statistically significant differences, these are economically negligible (84 meters and 9.6 kilometers, respectively, when the average of those variables are 1,354 and 271) for the Mexican context.

Third, proximity to LCPs must not capture access to transportation infrastructure existing prior to railroad construction, which could have influenced future migration patterns. This concern is mitigated by Mexico's geographic and development context in the mid-19th century. Mexico lacks navigable rivers, and overland transportation was extremely precarious. Most roads were inaccessible to stagecoaches, with freight transport relying almost exclusively on pack animals (Bignon, Esteves and Herranz-Loncán, 2015, p. 1279). Hence, it is unlikely that proximity to LCPs capture access to meaningful pre-railroad transportation networks.

6. Short-Run Results

We begin by examining the effect of railroad access on migration to the US. Table 3 presents IV estimates of the short-run effect of railroad access on migration rates in 1906-1910. Migration rates are calculated using, alternatively, the number of immigrants by place of birth (Panel A) or by place of last residence (Panel B). Table A.1 and Table A.2 in the Appendix report first-stage and reduced form estimates for each of these regressions as well as OLS estimates for comparison purposes.

Column 1 shows the estimated coefficient and standard errors (as well as Conley standard errors accounting for spatial correlation in square brackets) for the baseline specification. Controls include linear and quadratic distances to the US border, historical seaports, and LCP target destinations, a Yucatan Peninsula indicator, and region fixed effects. Panel A shows that railroad access increases migration

[17] At the same time, a wide temporal separation can generate a LCPs design weakly correlated with the railroad network if urbanization patterns changed dramatically. Alix-Garcia and Sellars (2020) show that urbanization patterns in Mexico remained largely unchanged from colonial times until the mid-20th century. This evidence reduces concerns about weak instrument bias in our setting.

by 0.72 per thousand inhabitants (Column 1)—about 1.5 times the mean migration rate in districts without access (0.48). The effect is extremely robust to the inclusion of measures of location, agricultural productivity, weather shocks, and land area (columns 2 to 5), with the point estimate remaining close to 0.7 per thousand people across specifications.[18] The estimated effect associated with railroad access is economically meaningful and it could explain about 27 to 45% of the growth in Mexico-US migration during the period of analysis.[19] Panel B shows that using immigrants' last residence in calculating migration rates produces even larger effects that are also stable and significant across specifications. Conditional on geographic, agricultural, weather, surface area controls and region fixed effects, districts with railroad access had 1.54 more migrants per thousand inhabitants than those without access. Effects based on immigrants' last residence are higher because they capture sequential migration—that is, migrants that moved within Mexico to (and resided in) a location with railroad access before moving to the US. We will later show that railroads disproportionately increased such sequential migration towards the north, suggesting that railroads enabled a process of moving internally to a "railway location" closer to the border first and then to the US.

Table A.1 and Table A.2 confirm that the aforementioned findings are supported by strong first-stage results: distance to the nearest LCP significantly predicts actual railroad construction across all specifications, with Kleibergen-Paap F statistics consistently exceeding 78—well above the conventional threshold of 10 for weak instrument. They also show that OLS estimates for railroad access are smaller in magnitude. This is consistent with negative selection of districts into railroad access. Recall that many branch lines were constructed to serve particular interests or political constituents, sometimes connecting places with low demand for transportation and low migration potential. Contemporary assessments are consistent with this argument. In 1908, Mexico's Ministry of Finance published a report evaluating the state of the railroad network. It concluded that "Unnecessary railroads were constructed ... some railroads that start in the desert end in the same desert ... there are railroads that will never be fully exploited." Furthermore, the report also documents that "Hundreds of miles of parallel track were built, where there was not sufficient traffic for one line alone" (Cuéllar, 1936).

[18] In Table C.1, we show the estimates adding controls for demographic structure, labor market conditions, living standards, landholding patterns, and information access. While the coefficient is very stable, these covariates are likely "bad controls" as they are probably affected by the presence of the railroad.

[19] We calculate total migration induced by railroad access (7.2 thousand additional immigrants per year) as the product of the difference in migration rates (0.65) and total population in districts with railroad access. Previous literature documents that Mexican immigration ranged from 16 to 26 thousand per year (see Cardoso, 1980; Clark, 1908; US Immigration Commission, 1911).

6.1 Robustness Checks

A possible threat to our identification strategy is the potentially heterogeneous predictive power of the instrument in the first stage. Figure 3 shows that our LCPs design is less precise in predicting railroad construction in some regions, particularly in the Southeast. To address this, we interact the instrument with region indicators to account for variation in the strength of first-stage identification across regions. This alternative specification yields very similar estimates (see Table B.1), with railroad access increasing migration rates by 0.52 and 1.18 per thousand based on immigrants' place of birth and last residence, respectively.

Another concern is that the chosen distance cutoff (100 km) for computing Conley standard errors may be inappropriate in regions where districts are large, failing to adequately correct for spatial correlation. Table C.2 shows that this is not the case, as the statistical significance of our main findings holds when using 200 km or 300 km cutoffs. To further ensure that our results are not driven by omitted spatial linkages, we calculate a formal market access measure following Donaldson and Hornbeck (2016), which accounts for proximity to all potential markets weighted by their economic size (population) and bilateral trade costs.[20] To allow for non-linear effects of market access on migration, we use both quadratic and non-parametric specifications. Table C.3 and Table C.4 show that our findings remain robust after controlling for the full spatial distribution of economic opportunities available at each location.

We also perform a series of checks to ensure our results are not attributable to non-representative districts. One concern is that districts with historical cities could generate confounding effects given their role as LCP target destinations. We address this in two ways: first, by excluding districts with historical cities; and second, by excluding all districts within 20 km of these cities to rule out the possibility that our findings are driven by agglomeration effects around these locations. Another concern is that our estimates might be driven by regions with a high density of railroads such as Mexico City and neighboring states (Valley of Mexico), which also had large urban centers in 1910. We directly address this by excluding districts in the Mexico City area. We further exclude districts with high urban density (above the 80th percentile of the distribution) to verify that our findings apply to rural contexts. Finally, we exclude districts at the US border, where unique economic conditions might influence migration independently of railroad access.

[20]Market access is estimated as $MA_o \approx \sum_{d \neq o} \tau_{od}^{-\theta} N_d$, with $\theta = 5.03$—the median value of trade elasticity in the literature Head and Mayer (2014).

Table C.5 and Table C.6 show that based on immigrants' place of birth and last residence, respectively, our findings hold after performing these robustness checks, confirming that our results are not driven by migration patterns of specific regions. The only exception is estimates excluding border districts when using immigrants' last residence to compute migration rates, which are smaller in magnitude than our main results. Excluding border districts reduces the effect of railroad access because these locations were among the main destinations for sequential migrants before moving to the United States. As we demonstrate next, this two-step migration pattern is the primary mechanism through which railroads increased international migration in the short run.

6.2 Short-Run Mechanism: Sequential Migration

Railroads transformed economies worldwide by increasing market access due to reductions in trade costs, reflected in greater interregional and international trade (Donaldson and Hornbeck, 2016; Donaldson, 2018). Similarly, railroads expanded labor market options for workers, as they reduced travel times and transportation costs (Cermeño, Enflo and Lindvall, 2022), increasing migration when mobility constraints were absent (Forero et al., 2021). In the case of Mexico, previous literature argues that while railroads reduced travel times dramatically, stagnant wages and underdeveloped credit markets made railroads unaffordable for the majority of the population (Coatsworth, 1979), which questions its impact on population mobility.

We now show that sequential migration is the main mechanism underlying our short-run results. We proceed in three steps. First, we show that railroad access induced greater population mobility within Mexico, specifically toward the north. Second, we show that this northward mobility pattern is strongly correlated with a second step involving migration to the United States. Third, we demonstrate that sequential migration—rather than direct migration—accounts for most of the migration effects associated with railroad access.

Table 4 presents estimates of our IV strategy using the number of internal migrants (per thousand inhabitants) as dependent variable, which we calculate using district-level census tabulates reporting population counts by state of birth. More specifically, this variable captures the number of residents per thousand who were born outside each district's state, which can be aggregated at different geographic levels to identify internal migration patterns. Column 1 shows that districts with railroad access had more internal migrants from other regions than districts without access, with economically and statistically significant estimates equal to 72.35 migrants per thousand inhabitants. This pattern holds when examining mobility across states (column 2). Districts with railroad access attracted about 72 more internal migrants, per thousands inhabitants from other states, confirming that railroads significantly reduced transportation costs reflected in greater internal mobility. Columns 3 and 4 reveal the direction of these migration flows within Mexico. Railroad access significantly increased the number of internal migrants from states located to the south (northward mobility), with an effect equal to 46.7 migrants per thousand inhabitants across specifications. In contrast, the effects on internal migration from states located to the north (southward mobility) are much smaller and less significant—25 per thousand inhabitants. This pattern indicates that railroads primarily facilitated mobility toward the north, drawing migrants from all regions toward districts closer to the US border.

To further support our findings on northward mobility, we conduct two robustness checks. First, we examine to what extent this pattern captures internal migration to Mexico City, the largest market during our period of analysis.[21] Column 5 of Table 4 shows that districts with railroad access did not have more migration to Mexico City (column 5), suggesting that migration to Mexico City is unlikely to have driven railroad-induced northward mobility. Second, we test whether south-north rather than east-west oriented rail lines influenced northward mobility. We operationalize this analysis by calculating the azimuth angle of each rail segment and classifying segments as east-west (within ±15° of horizontal) or north-south otherwise. We then compute the share of rail segments that are north-south oriented for each district. If this share falls below the 25th percentile of the distribution, we classify the district as having access to predominantly east-west oriented railroads. Intuitively, these districts were less clearly on a route toward the north and thus should have attracted fewer northward internal migrants. Table C.7 shows that the effect of railroad access on northward internal migration (column 1) is similar when excluding districts with predominantly east-west oriented railroads (column 2), while access to east-west oriented railroads has no effect on northward mobility (column 3).

How did this northward internal mobility relate to migration to the United States? A binscatter plot shows that districts with higher rates of residents from states located to the south exhibit systematically higher shares of sequential migrants to the US (see Figure F.4). This provides evidence that railroads induced northward mobility within Mexico, which served as a stepping stone for subsequent migration to the United States. Building on this evidence, we decompose total Mexico-US migration into direct migration (individuals whose place of birth and last residence are the same) and stage migration (individ-

[21] We calculate emigration rates to Mexico City using counts of Mexico City's population by state of birth, which we allocate to origin districts proportionally to each district's share of state population.

uals whose place of birth and last residence differ). Table 5 shows that the effect of railroads on total migration operates almost exclusively through sequential migration. While railroad access increases overall Mexico-US migration rates by 1.54 per thousand inhabitants (Column 1), the effect on direct migration is economically small and statistically insignificant, at 0.31 per thousand (Columns 2). In stark contrast, railroad access increases sequential migration rates by 1.22 per thousand inhabitants (Columns 3), accounting for about 85% of the railroad effect on total migration. This finding is reinforced by examining the share of stage migrants among total migrants: railroad access increases this share by 17.8 percentage points (Column 4). These results provide compelling evidence that railroads increased migration to the United States primarily by enabling stage migration.

Overall, our short-run analysis shows that railroad access had a large impact on Mexican migration to the United States, supporting the view that railroads were indispensable to the surge of mass migration (Durand, 2016; Gamio, 1930). Our findings on the role of sequential migration explain how railroads were intensively used despite train tickets being unaffordable to the average laborer (Coatsworth, 1979). While a direct journey to the border was probably too expensive, migrating in stages helped overcome income and credit constraints, allowing individuals to work to finance their journey while working, receiving support and gathering useful information at locations along the railway, and then making the final migration move to the US.

7. Additional Short-Run Effects on Local Economic Activity

We now turn to studying the implications that railroad access and stage migration had for local economies. Table 6 presents IV estimates of the effects of railroad access on a battery of district-level economic outcomes. Panel A shows that districts with railroad access had higher population, urban density, and housing densities than districts without access, consistent with evidence that railroads promoted urbanization in other contexts (e.g., Atack et al., 2010; Berger, 2019; Donaldson and Hornbeck, 2016). Panel B shows that railroad access increased health access (health workers per thousand inhabitants) by 0.53 SD. Additionally, railroad access reduced land concentration (population share living in large estates) by 1.35–1.42 SD, suggesting that railroads disrupted traditional landholding structures by expanding market access and economic opportunities beyond haciendas. Panel C shows that districts with railroad access also had significantly higher shares of working-age population (0.83–0.80 SD) and lower child-woman ratios (0.93–0.88 SD) compared to districts without access, indicating that these districts attracted prime-age workers—potentially internal migrants seeking employment in expanding economic sectors—rather than families with young children, with potential implications for employment and labor allocation.

To examine the effects of railroad access on local labor markets, we collect additional district-level data on labor force participation and employment by occupation from the 1895 and 1910 population censuses. We use these data to calculate the difference between 1895 and 1910 in labor force participation and the share of the working-age population employed in each economic sector: extractive (agriculture and mining), industry, and services. Table 7 shows that railroad access induced significant labor reallocation across sectors. Column 1 shows that railroads increased the share of working-age population employed or unemployed by 13.5 percentage points, confirming that districts with railroad access attracted more labor than those without access due to both increased northward population mobility and ongoing structural transformation. Columns 4-6 show that the share of the working-age population employed in extractive activities fell by 7.3 percentage points in districts with railroad access, while employment in industry and services rose by 1.6 and 5.9 percentage points, respectively—changes that sum to zero, reflecting no overall employment creation (column 2). However, unemployment increased by 5.3 percentage points (column 2), suggesting that railroads drew previously inactive individuals into the labor force—potentially including internal migrants—who could not immediately find employment, but searched for jobs increasing the size of the labor force.

The combined effects described in this section suggest the following narrative. The building of railroads in districts that were not previously connected to the rail network, produced two important effects in the early 20th century in Mexican districts. First, it attracted sequential migrants from the rest of Mexico, who resided there for a while and then migrated to the US, increasing total migration to the US. Second, by attracting people (even temporarily) and expanding market potential, it also transformed the local economies, making them more populated, more industrial, and (likely) richer as evidenced by better health conditions and a larger share of workers in the service sector. Our next question is how these transformations affected, if at all, long-run migration outcomes to the US.

8. Long-Run Results

As argued previously, railroad passenger transportation in Mexico declined rapidly after the 1950s with the expansion of paved roads and the emergence of commercial aviation. We now examine whether the migration patterns observed in the early 20th century persisted into the early 21st century, long after the transportation infrastructure that shaped them became obsolete. In particular, we distinguish between the magnitude of the migration flow (the district aggregate emigration rate) from its specific *direction*. The total emigration rate is a result of local aggregate factors, affected by economic development. The relative direction of it (i.e. what share of migrants go to specific US destinations) is instead more affected by the initial network formation.

8.1 Reversal of Migration Patterns

We first examine whether historical railroad access influences the level of emigration rates in the 21st century. There are two competing forces affecting this variable. On one hand, railroad-induced migration networks established in the early 20th century could persist and reduce the costs of migration much later if they continued to convey a large migration cost advantage, as networks reduce the costs and risks for subsequent migrants (Hatton, 1995; McKenzie and Rapoport, 2010; Munshi, 2003). On the other hand, however, the railroad-induced industrialization and urbanization documented above could have reduced emigration pressures over time by increasing local income and opportunities. If the aggregate cost-reduction channel prevailed, we would expect districts with historical railroad access to still have higher migration rates in recent years. If the economic growth effect prevailed, we would expect it to attenuate or even reverse that pattern.

Table Table 8 presents IV estimates of the effect of historical railroad access on contemporary, mean annual emigration rates using MCAS data for the period 2006–2019. The point estimate in column 1 is negative (not statistically significant), consistent with an attenuation/reversal of migration patterns: districts with historical railroad access have lower migration rates (1.03 fewer migrants per thousand inhabitants) to the United States in the early twenty-first century, suggesting that railroad-induced development dominated network effects in the long run. One concern is that our result may be driven by access to paved roads—the infrastructure that increasingly replaced railroads from the 1950s—which could have increased market potential and economic growth. We address this concern by controlling for distance (linear and quadratic) to the nearest paved road. Another concern is that historical factors, other than railroad access, could influence our long-run estimates. We address this by controlling for three potential confounders. First, historical migration networks could have reduced contemporary migration through persistent remittance flows.[22] Second, Mexico's major land reform—which redistributed more than 50% of all arable land throughout the twentieth century (Murphy and Rossi, 2016)—could have

[22]We control for historical networks using mean migration rates at the district level measured in 1906–1910 from Escamilla-Guerrero, Kosack and Ward (2025).

shaped migration patterns by altering labor and land allocations (De Janvry et al., 2015).[23] Third, immigration policies, particularly the Bracero Program (1942–1964) could have generated development spillovers through seasonal return migration.[24] Our finding on the reversal in migration patterns remains virtually unchanged after controlling for all these factors (columns 2-5). The stability of the coefficient across specifications indicates that the weak but negative relationship between historical railroad access and contemporary migration is not driven by modern-day infrastructure or historical confounders.

Note that none of the estimates reach conventional levels of statistical significance, suggesting more an attenuation than a complete reversal. This could be consistent with reaching a part of the relation between emigration and economic development that is relatively flat, before becoming downward sloping. Alternatively, attenuation can be due to lack of precision due to error or to local averaging among heterogeneous districts. As districts are rather large, within-district variation in local economic conditions induced by closer railroad access is averaged, and this could attenuate the estimated effect.

One feature of the MCAS data is that it reports immigrants' place of birth by municipality, which allows us to study the effects of historical railroad access at a finer level of geography. The estimates in Table E.1 in the Appendix show that municipalities with historical access to railroads have significantly lower emigration rates today. The point estimates range from -3.65 to -3.75 and are statistically significant at the 1% level. These estimates are robust to the inclusion of state fixed effects and to correcting for spatial correlation using Conley (1999) standard errors with a 100 km threshold. Moreover, they remain stable and significant across specifications that control for access to paved roads and historical factors. Overall, the attenuation and likely reversal of historical migration patterns suggests that the development channel, and its inverted U relationship with emigration, dominated network effects in the long run: municipalities with railroad access developed stronger local economies that reduced emigration pressures over the following century.

8.2 Persistence in Destination Choices

The second long-term effect we analyze, rather than on total emigration rates is on the specific direction of migration from one district to US states. Using MCAS data, we calculate for each district the share of migrants settling in California, Arizona, New Mexico, and Texas—states that together receive

[23] We control for land access using the share of land redistributed in each district through outright grants and restitutions in 1916–1976 from Sanderson (2013).

[24] Under this program, approximately half a million Mexican seasonal workers crossed the border each year to work on US farms (Clemens, Lewis and Postel, 2018). We control for exposure to the program using an indicator variable for districts with bracero recruiting centers from Kosack (2021).

approximately 78% of contemporary Mexican migrants. We then compare these shares to historical railroad destination patterns observed in the early 20th century. Specifically we ask whether a district with a large share of migrants moving to California in 1910, which was importantly affected by being on a railway line towards California back then, is still more likely to show a larger share of migrants to California in the recent period, even if it is far from the California border (and maybe in terms of distance closer, say, to Texas).

Figure 4 presents binscatter plots of the relationship between historical and recent migrant destination shares, controlling for region fixed effects to account for geographic proximity to the border. Destination-specific migration patterns show remarkable persistence over a century: districts that sent a higher share of migrants to California, Arizona, or Texas in 1906–1909 continue to send a higher share to those same destinations in 2006–2019. The persistence is strongest for California (Panel A) and Texas (Panel D), the two largest destination states, suggesting that network effects are most persistent where migrant communities are largest and most established. Importantly, these correlations hold within regions, meaning the persistence reflects more than mechanical proximity to border crossings—that is, municipalities with similar distances to the border but different historical destination patterns continue to exhibit those differences today. This persistence is remarkable considering that since 1950, subsequent transportation improvements—including commercial aviation and the expansion of paved roads—have fundamentally altered mobility costs, providing ample opportunity for destination choices to change.

The persistence in destination choices follows a clear geographic segmentation. Figure F.5 displays the spatial distribution of migrant destination shares to each US border state and the historical railroad network. California receives migrants primarily from municipalities in western states (Sonora, Sinaloa, Nayarit, Jalisco, Colima, Michoacan and Guerrero). Arizona draws migrants primarily from the northwestern state of Sonora (Panel B), while New Mexico receives migrants mainly from a central narrow corridor crossing Chihuahua and Durango (Panel C). Texas dominates migration from northeastern Mexico, with the highest shares originating in Coahuila, Nuevo León, San Luis Potosí, and Veracruz (Panel D). This spatial pattern closely mirrors the historical railroad network: municipalities send migrants to the US states that were most accessible via the trunk lines that connected them to the border a century ago. Notice for instance, that in terms of simple distance from the border, many districts in Jalisco or Michoacan would be closer to Texas than to California.

To formally test whether access to specific trunk lines shape migrant destination choices in recent time, we classify the 1910 railroad network into three trunk lines based on the border crossings they connected to: the western trunk line (leading to Nogales and serving California and Arizona), the central trunk line (leading to El Paso and serving New Mexico), and the eastern trunk line (leading to Laredo, Eagle Pass, and Brownsville and serving Texas). Railroads south of Guanajuato—where the three trunk lines converge—are classified as ambiguous since they do not lead to a specific border crossing. For each municipality, we create indicator variables for access to each trunk line, defined as being located within 40 km of the nearest rail segment. We then regress contemporary destination shares (to California-Arizona, New Mexico and Texas, respectively) on these trunk-line indicators, with municipalities without railroad access as the omitted category. The estimated coefficients capture the difference in destination shares between municipalities with access to a given trunk line (Western, Central, Eastern, or Ambiguous) and those without any railroad access.

Table 9 shows the results. Access to the western trunk line increases share of migrants to California-Arizona by 14.9 percentage points, while decreasing the probability of migrating to the other destinations. Access to the eastern trunk line increases the share of migrants to Texas by 12 percentage points, while decreasing the probability to migrate to California-Arizona. Similarly, access to the central trunk line raises New Mexico's share by 2.4 percentage points. All effects remain significant after correcting for spatial correlation. The off-diagonal coefficients are negative and significant: municipalities connected to one trunk line send fewer migrants to destinations served by other trunk lines. Access to the ambiguous network south of Guanajuato shows less systematic effect on destination shares, confirming that trunk lines with direct connections to the US border shaped migration corridors. Table E.2 in the Appendix shows very similar results in terms of magnitude and significance when leveraging variation at the municipality level. Together, these results demonstrate that the railroad infrastructure built over a century ago created persistent migration corridors that continue to channel migrants toward specific US destinations today—an example of how historical transportation networks can leave lasting imprints on population movements long after the infrastructure itself becomes obsolete.

9. Conclusion

This paper investigates the impact of railroad access on Mexican migration to the United States. In the 1900s, when Mexico-US mass migration surged, railroads were the main mode of passenger transportation in Mexico. By the end of the century, the railroads' share of total passenger transportation had fallen virtually to 0%, due to the emergence of new transportation technologies that fundamentally changed migration costs after 1950. This setting provides an excellent environment to examine the short-run effects of transportation infrastructure on international migration—as well as on other local economic outcomes—and the long-run effects. We evaluate the persistence of the effect over one century, focusing on both the magnitude (total emigration rates) and direction (destination choices) of migration.

The short-run results show that railroad access approximately quadrupled migration rates in the early 20th century. The evidence indicates that railroads increased migration to the United States primarily by enabling northward internal mobility, with people moving within Mexico to locations with railroad access first and then migrating to the US. This process of sequential migration not only contributed to the Mexico-to-US migration flow, but also to the development and structural transformation of the local economies where the railroad was built.

The long-run results are also very interesting. They reveal an attenuation/reversal of total emigration rates: locations with historical railroad access have lower emigration rates to the United States in recent years, likely a consequence of their more rapid development and industrialization induced by railroad access. This development effect outweighed that of reduced mobility costs for subsequent migrants that would otherwise have perpetuated high emigration rates in the long run.

Remarkably, however, we find a strong persistence in the *direction* of migration i.e. the relative destination choices of Mexican migrants, long after railroad passenger transportation became obsolete for recent migration flows. Locations that sent a higher share of migrants to California, Arizona, or Texas in the 1900's continue to send a higher share to those same destinations in the 2000's. This persistence occurred despite fundamental changes in transportation technology and immigration policies that provided ample opportunity for destination choices to reach new equilibria. The main long-run implication of historical railroad access is thus the creation of enduring migration corridors, likely operating by reducing the relative costs of migration across destinations, even if, in absolute terms, the faster development of districts with historical railroad access reduced total emigration rates.

Overall, the results highlight that investments in transportation infrastructure can induce international migration by enabling people to move in steps, and that the patterns established by these initial flows can persist for generations, long after the infrastructure itself becomes obsolete. Our study provides the first systematic evidence of, and an explanation for, the evolution of the migration between Mexico and the US over the past century, both in terms of magnitude and its direction.

References

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature*, 55(4): 1311–45.

Agencia Reguladora del Transporte Ferroviario. 2023. "Mapas del Sistema Ferroviario Mexicano." *Gobierno de Mexico*.

Alix-Garcia, Jennifer, and Emily A. Sellars. 2020. "Locational fundamentals, trade, and the changing urban landscape of Mexico." *Journal of Urban Economics*, 116: 103213.

Andersson, David, Thor Berger, and Erik Prawitz. 2023. "Making a market: Infrastructure, Integration, and the Rise of Innovation." *Review of Economics and Statistics*, 105(2): 258–274.

Angelucci, Manuela. 2015. "Migration and Financial Constraints: Evidence from Mexico." *Review of Economics and Statistics*, 97(1): 224–228.

Arias, Luz Marina, and Luis De la Calle. 2021. "The Legacy of Civil War Dynamics: State Building in Mexico, 1810–1910." *Latin American Research Review*, 56(4): 814–830.

Artuc, Erhan, and Caglar Ozden. 2018. "Transit Migration: All Roads Lead to America." *The Economic Journal*, 128(612): F306–F334.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A Margo. 2010. "Did railroads induce or follow economic growth?: Urbanization and population growth in the American Midwest, 1850–1860." *Social Science History*, 34(2): 171–197.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2020. "On the road: Access to transportation infrastructure and economic growth in China." *Journal of Development Economics*, 145: 102442.

Berger, Thor. 2019. "Railroads and Rural Industrialization: Evidence from a Historical Policy Experiment." *Explorations in Economic History*, 74: 101277.

Berger, Thor, and Kerstin Enflo. 2017. "Locomotives of local growth: The short- and long-term impact of railroads in Sweden." *Journal of Urban Economics*, 98: 124–138.

Bignon, Vincent, Rui Esteves, and Alfonso Herranz-Loncán. 2015. "Big push or big grab? Railways, government activism, and export growth in Latin America, 1865-1913." *The Economic History Review*, 68(4): 1277–1305.

Black, Dan A, Seth G Sanders, Evan J Taylor, and Lowell J Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review*, 105(2): 477–503.

Borjas, George J. 2007. *Mexican immigration to the United States.* University of Chicago Press.

Caballero, Maria Esther, Brian C. Cadena, and Brian K. Kovak. 2018. "Measuring Geographic Migration Patterns Using Matrículas Consulares." *Demography*, 55(3): 1119–1145.

Caballero, María Esther, Brian C. Cadena, and Brian K. Kovak. 2023. "The international transmission of local economic shocks through migrant networks." *Journal of International Economics*, 145: 103832.

Cardoso, Lawrence A. 1980. *Mexican emigration to the United States, 1897-1931.* Tucson:University of Arizona Press.

Castro Aranda, Hugo. 2010. *Primer censo de población de la Nueva España 1790. Censo de Revillagigedo "Un Censo Condenado".* Mexico City:Instituto Nacional de Estadística y Geografía.

Cermeño, Alexandra L, Kerstin Enflo, and Johannes Lindvall. 2022. "Railroads and Reform: How

Trains Strengthened the Nation State." *British Journal of Political Science*, 52(2): 715–735.

Chevalier, François. 1970. *Land and Society in Colonial Mexico: The Great Hacienda.* University of California Press.

Chort, Isabelle, and Maëlys De La Rupelle. 2016. "Determinants of Mexico-US Outward and Return Migration Flows: A State-Level Panel Data Analysis." *Demography*, 53(5): 1453–1476.

Clark, Victor. 1908. "Mexican Labor in the US." *Bulletin of the United States Bureau of Labor*, 78: 466–522.

Clemens, Michael A, Ethan G Lewis, and Hannah M Postel. 2018. "Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion." *American Economic Review*, 108(6): 1468–1487.

Coatsworth, John. 1974. "Railroads, landholding, and agrarian protest in the early porfiriato." *Hispanic American Historical Review*, 54(1): 48–71.

Coatsworth, John H. 1979. "Indispensable railroads in a backward economy: The case of Mexico." *The Journal of Economic History*, 39(4): 939–960.

CONABIO. 2008. "Carreteras Estatales y Federales. Catálogo de metadatos geográficos." [shapefile].

CONAPO. 2023. "Bases de datos de la Conciliación Demográfica 1950 a 2019 y Proyecciones de la población de México 2020 a 2070." Consejo Nacional de Población. [dataset].

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

Cuéllar, Alfredo B. 1936. "Railroad Problems of Mexico." *The ANNALS of the American Academy of Political and Social Science*, 187(1): 193–206.

Dao, Thu Hien, Frédéric Docquier, Chris Parsons, and Giovanni Peri. 2018. "Migration and development: Dissecting the anatomy of the mobility transition." *Journal of Development Economics*, 132(C): 88–101.

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

DGE. 1899. "I Censo de Poblacion de los Estados Unidos Mexicanos de 1895. Tabulados Básicos." Dirección General de Estadística. [*dataset digitized by INEGI*].

DGE. 1918. "III Censo de Poblacion de los Estados Unidos Mexicanos de 1910. Tabulados Básicos." Dirección General de Estadística. [*dataset digitized by INEGI*].

Disturnell, J. 1847. *Mapa de los Estados Unidos de Méjico.* Disturnell.

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

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *The Quarterly Journal of Economics*, 131(2): 799–858.

Donly, A.W. 1920. "The Railroad Situation in Mexico." *The Journal of International Relations*, 11: 234–251.

Durand, Jorge. 2016. *Historia Mínima de la Migración México-Estados Unidos.* Mexico City:El Colégio de México.

Durand, Jorge, Douglas S Massey, and Rene M Zenteno. 2001. "Mexican immigration to the United States: Continuities and Changes." *Latin American Research Review*, 107–127.

Elizalde, Aldo, Eduardo Hidalgo, Nayeli Salgado, and Sotiris Kampanelis. 2026. "Public good or public bad? Nation-building and indigenous institutions." *Journal of Development Economics*, 179: 103652.

Escamilla-Guerrero, David. 2020. "Revisiting Mexican migration in the Age of Mass Migration: New evidence from individual border crossings." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(4): 207–225.

Escamilla-Guerrero, David, Edward Kosack, and Zachary Ward. 2025. "The impact of violence on the dynamics of migration: Evidence from the Mexican Revolution." *Journal of Development Economics*, 176: 103515.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *The Review of Economic Studies*, 81(3): 1046–1070.

Faini, Riccardo, and Alessandra Venturini. 1993. "Trade, aid and migrations: Some basic policy issues." *European Economic Review*, 37(2): 435–442.

FAO. 2021. "Agro-Ecological Suitability and Potential Yields. Global Agro-Ecological Zones (GAEZ) v4." [dataset].

Feliciano, Zadia M. 2001. "The skill and economic performance of Mexican immigrants from 1910 to 1990." *Explorations in Economic History*, 38(3): 386–409.

Fenske, James, Namrata Kala, and Jinlin Wei. 2023. "Railways and Cities in India." *Journal of Development Economics*, 161: 103038.

Florescano, Enrique. 1987. "The Hacienda in New Spain." In *Colonial Spanish America.* , ed. Bethell Leslie, 250–285. Cambridge University Press.

FNM. 1914. *Mapa de los Ferrocarriles Nacionales de Mexico.* Ferrocarriles Nacionales de Mexico.

Fogel, Robert William. 1962. "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings." *The Journal of Economic History*, 22(2): 163–197.

Fogel, Walter. 1978. *Mexican Illegal Alien Workers in the United States.* Los Angeles:Institute of Industrial Relations, University of California.

Forero, Andrés, Francisco A Gallego, Felipe González, and Matías Tapia. 2021. "Railroads, specialization, and population growth: evidence from the first globalization." *Journal of Population Economics*, 34(3): 1027–1072.

Gamio, Manuel. 1930. *Mexican immigration to the United States.* Chicago:University of Chicago Press.

Gratton, Brian, and Emily Merchant. 2015. "An Immigrant's Tale: The Mexican American Southwest 1850 to 1950." *Social Science History*, 39(4): 521–50.

Gratton, Brian, and Myron P Gutmann. 2000. "Hispanics in the United States, 1850-1990: Estimates of Population Size and National Origin." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 33(3): 137–153.

Hanson, Gordon H, and Antonio Spilimbergo. 1999. "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the US-Mexico Border." *American Economic Review*, 89(5): 1337–1357.

Hatton, Timothy J. 1995. "A model of UK emigration, 1870-1913." *The Review of Economics and*

Statistics, 407–415.

Hatton, Timothy J., and Jeffrey G. Williamson. 1993. "After the Famine: Emigration from Ireland, 1850–1913." *The Journal of Economic History*, 53(3): 575–600.

Hatton, T. J, and J. G. Williamson. 1998. *The Age of Mass migration: Causes and Economic Impact.* Oxford University Press.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*. Vol. 4, 131–195. Elsevier.

Henderson, Timothy J. 2011. *Beyond borders: A history of Mexican migration to the United States.* John Wiley & Sons.

Herranz-Loncán, Alfonso. 2006. "Railroad Impact in Backward Economies: Spain, 1850-1913." *The Journal of Economic History*, 66(4): 853–881.

Imbert, Clément, and John Papp. 2020. "Costs and benefits of rural-urban migration: Evidence from India." *Journal of Development Economics*, 146: 102473.

INEGI. 1950–2000. "Censos de Transportes de México, 1950-2000." Various years: 1950, 1960, 1970, 1980, 1990, 2000.

INEGI. 1991. "XI Censo de Poblacion de los Estados Unidos Mexicanos de 1990. Tabulados Básicos." Instituto Nacional de Estadística y Geografía. [dataset].

INEGI. 2010. "Estadísticas Históricas de México." Edición 2009.

Jedwab, Remi, and Alexander Moradi. 2016. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics*, 98(2): 268–284.

Kosack, Edward. 2021. "Guest Worker Programs and Human Capital Investment: The Bracero Program in Mexico, 1942–1964." *Journal of Human Resources*, 56(2): 570–599.

Kosack, Edward, and Zachary Ward. 2020. "El Sueño Americano? The Generational Progress of Mexican Americans Prior to World War II." *The Journal of Economic History*, 80(4): 961–995.

Lee, Jongkwan, Giovanni Peri, and Vasil Yasenov. 2022. "The labor market effects of Mexican repatriations: Longitudinal evidence from the 1930s." *Journal of Public Economics*, 205: 104558.

Leunig, Timothy. 2006. "Time is Money: A Re-Assessment of the Passenger Social Savings from Victorian British Railways." *The Journal of Economic History*, 66(3): 635–673.

López-Alonso, Moramay. 2007. "Growth with inequality: Living standards in Mexico, 1850–1950." *Journal of Latin American Studies*, 39(1): 81–105.

McKenzie, David, and Hillel Rapoport. 2007. "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico." *Journal of Development Economics*, 84(1): 1–24.

McKenzie, David, and Hillel Rapoport. 2010. "Self-selection Patterns in Mexico-US Migration: The Role of Migration Networks." *The Review of Economics and Statistics*, 92(4): 811–821.

Moreno-Brid, Juan Carlos, and Jaime Ros. 2009. *Development and Growth in the Mexican Economy: A Historical Perspective.* Oxford University Press.

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

Munshi, Kaivan. 2003. “Networks in the modern economy: Mexican migrants in the US labor market.” *The Quarterly Journal of Economics*, 118(2): 549–599.

Murphy, Tommy E, and Martín A Rossi. 2016. “Land reform and violence: Evidence from Mexico.” *Journal of Economic Behavior & Organization*, 131: 106–113.

Oñate, A. 1991. “La política agraria del estado mexicano durante el Porfiriato.” In *Cincuenta Años de Historia de México*. , ed. A. Hernández and M. Miño, 293–314. El Colégio de México.

Pletcher, David M. 1950. “The Building of the Mexican Railway.” *The Hispanic American Historical Review*, 30(1): 26–62.

Rosenzweig, Fernando. 1965. “El Desarrollo Económico de México de 1877 a 1911.” *El Trimestre Económico*, 32(127(3)): 405–454.

Samora, J. 1982. *Los mojados: The wetback story.* University of Notre Dame Press.

Sanderson, Susan Walsh. 2013. “Land Reform in Mexico 1910-1976.” *ICPSR34388-v1*, distributor.

Sellars, Emily A, and Jennifer Alix-Garcia. 2018. “Labor scarcity, land tenure, and historical legacy: Evidence from Mexico.” *Journal of Development Economics*, 135: 504–516.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian. 2020. “Immigrants and the Making of America.” *The Review of Economic Studies*, 87(1): 382–419.

Sokoloff, Kenneth L, and Stanley L Engerman. 2000. “History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World.” *Journal of Economic Perspectives*, 14(3): 217–232.

Spitzer, Yannay, and Ariell Zimran. 2024. “Like an ink blot on paper: Testing the diffusion hypothesis of mass migration, Italy 1876-1920.” National Bureau of Economic Research Working Paper 30847.

Stahle, David W., Edward R. Cook, Dorian J. Burnette, Jose Villanueva, Julian Cerano, Jordan N. Burns, Daniel Griffin, Benjamin I. Cook, Rodolfo Acuña, and Max Torbenson. 2016. “The Mexican Drought Atlas: Tree-ring reconstructions of the soil moisture balance during the late pre-Hispanic, colonial, and modern eras.” *Quaternary Science Reviews*, 149: 34–60.

Tanck de Estrada, Dorothy. 2005. *Atlas Ilustrado De Los Pueblos De Indios: Nueva España, 1800.* Mexico City:El Colegio de México.

US Immigration Commission. 1911. *Abstracts of Reports of the Immigration Commission: With Conclusions and Recommendations and Views of the Minority (In Two Volumes).* Vol. 2, Washington DC:US Government Printing Office.

Woodruff, Christopher, and Rene Zenteno. 2007. “Migration networks and microenterprises in Mexico.” *Journal of Development Economics*, 82(2): 509–528.

Table 1: Summary Statistics

	1	2	3	4	5
	All Units	Railroad Access	No Railroad Access	Difference	Conditional Difference
Panel A. District Characteristics in 1910					
Geography					
Distance to border (km)	651.47	607.60	737.85	-130.26^{***}	-11.64
Altitude (m)	1,228.84	1,381.49	928.24	453.26^{***}	209.01^{***}
Maize suitability index	922.72	1,010.09	750.65	259.44^{***}	62.42
Drought index	-0.57	-0.51	-0.69	0.18	0.05
Population Structure					
Population (thousands)	39.28	43.84	30.28	13.56^{***}	12.96^{***}
Population share aged 16-35 years	0.34	0.34	0.34	-0.00	0.00
Population share of native-language speakers	0.13	0.11	0.17	-0.06^{**}	-0.06^{**}
Sex ratio	1.02	0.99	1.07	-0.08	-0.01
Labor Markets					
US-Mexico wage gap	3.92	3.91	3.93	-0.02	-0.02
Working-age population share of agric. peons	0.31	0.30	0.32	-0.02	-0.03^{***}
Working-age population share of unemployed	0.02	0.02	0.03	-0.01	0.00
Living Standards and Land Ownership					
Urban density	6.64	9.65	0.72	8.93^{***}	3.93^{***}
Population share of illiterate	0.69	0.67	0.73	-0.06^{***}	-0.05^{***}
Dwellings share of huts	0.52	0.52	0.53	-0.01	-0.10^{***}
Number of large estates	96.96	90.23	110.21	-19.97	-18.04^{*}
Population share living in large estates	0.31	0.30	0.32	-0.01	-0.05^{***}
Migration to the United States (1906-1910)					
Mean migration rate per thousand (place of birth)	0.91	1.13	0.48	0.65^{***}	0.45^{**}
Observations	386	256	130	386	386
Panel B. District Characteristics in 1990					
Geography					
Drought index	-0.25	-0.35	-0.07	-0.28^{**}	-0.28^{***}
Population Structure					
Population (thousands)	210.49	243.84	144.81	99.03^{**}	114.56^{***}
Population share aged 15-29 years	0.27	0.28	0.26	0.01^{***}	0.02^{***}
Population share of native-language speakers	0.12	0.09	0.17	-0.08^{***}	-0.07^{***}
Sex ratio	0.99	0.98	1.00	-0.02^{***}	-0.02^{***}
Labor Markets					
US-Mexico wage gap	8.56	8.58	8.52	0.05	0.86
Employed share of agric. peons	0.18	0.17	0.18	-0.00	-0.02^{**}
Working-age population share of unemployed	0.03	0.03	0.03	-0.01^{***}	-0.01^{***}
Employed share earning minimum wage	0.24	0.22	0.26	-0.04^{***}	-0.02
Living Standards and Land Ownership					
Distance to paved road (km)	3.49	3.28	3.92	-0.64^{**}	-0.73^{**}
Urban density	125.05	181.38	14.12	167.26^{***}	97.25^{***}
Population share of illiterate	0.19	0.17	0.23	-0.06^{***}	-0.05^{***}
Dwellings share of huts	0.11	0.10	0.13	-0.02^{**}	-0.02^{**}
Share of dwellings owned	0.85	0.84	0.87	-0.03^{***}	-0.03^{***}
Migration to the United States (2006-2019)					
Mean migration rate per thousand (place of birth)	5.19	4.84	5.88	-1.04^{*}	-2.30^{***}
Observations	386	256	130	386	386

Source: Mexican Border Crossing Records (MBCRs), DGE (1918), INEGI (1991), FAO (2021), and Sellars and Alix-Garcia (2018).

Note: The table presents means of observable characteristics. I estimate differences (column 4) and differences conditional on state fixed effects (column 5) between units with and without railroad access. Units with railroad access are those within 40 km distance of the nearest railroad line. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table 2: Balance in Pre-Treatment Characteristics

	1	2	3	4	5
	All Locations	Closer to LCPs	Farther from LCPs	Difference	Conditional Difference
Panel A. Pueblos ca. 1800					
Geography					
Altitude (m)	1,473.54	1,752.71	1,174.19	578.52***	108.38***
Drought index	0.34	0.37	0.30	0.07***	0.03**
Maize index	1166.51	1447.03	865.70	581.33***	92.30**
Population					
Total population	459.91	516.90	398.85	118.05***	42.12
Indios population	373.53	434.22	308.45	125.77***	−4.15
Population share of indios	0.62	0.68	0.55	0.12***	0.01
Observations	4470	2313	2157	4470	4470
Panel B. Municipalities ca. 1800					
Geography					
Altitude (m)	1,354.29	1,607.82	1,107.68	500.14***	84.80***
Distance to mine (km)	271.73	206.69	335.00	−128.32***	−9.69***
Drought index	0.36	0.40	0.33	0.06***	−0.01
Maize index	977.48	1125.20	833.79	291.41***	54.40
State capacity					
Customs office	0.18	0.21	0.15	0.06***	0.04
Sales tax collected	0.90	1.36	0.45	0.91***	0.30
Observations	2456	1211	1245	2456	2456

Source: Tanck de Estrada (2005) - Panel A and Arias and De la Calle (2021) - Panel B.

Note: The table reports means of location characteristics for *pueblos de indios* (Panel A) and areas mapped to municipalities as in 2010 (Panel B) by proximity to least cost paths (LCPs). Column 2 includes locations within 40 kilometers of a LCP. Column 3 includes locations more than 40 kilometers away from a LCP (see Figure 3). Column 4 reports differences in means between these location groups. Column 5 reports differences conditional on *subdelegación* (Panel A) and *district* fixed effects (Panel B). The drought index is the Palmer Drought Severity Index from the Mexican Drought Atlas (Stahle et al., 2016), averaged over 1790–1799. In Panel A, altitude is calculated using each pueblo's latitude and longitude. Distance variables are measured from each pueblo to the target destination. In Panel B, altitude is the median elevation calculated from raster data. Distances are measured from municipality centroids to the target destinations. *Military presence* indicates whether insurgents, royalists, or local militias were present in at least one locality of the municipality. *Customs office* indicates whether the municipality had a *receptoría* or *subreceptoría*. *Sales tax collected* refers to the *alcabala*, expressed in thousands of pesos and averaged over 1800–1810. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table 3: Railroad Access and International Migration

	1	2	3	4	5
Panel A. Place of Birth					
railroad access = 1	0.721**	0.586**	0.731**	0.748**	0.696**
	(0.293)	(0.268)	(0.322)	(0.318)	(0.294)
	[0.302]	[0.271]	[0.375]	[0.369]	[0.297]
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88
Panel B. Place of Last Residence					
railroad access = 1	1.445***	1.408**	1.605***	1.613***	1.540**
	(0.548)	(0.615)	(0.620)	(0.620)	(0.666)
	[0.532]	[0.566]	[0.615]	[0.604]	[0.630]
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface Area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic) and an indicator for observations in the Yucatan Peninsula. *Location fundamentals* includes mean altitude and the interaction between latitude and longitude. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the observation's surface area (km^2). All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 4: Railroad Access and Internal Mobility

	1	2	3	4	5
	In-migration from				Out-migration to
	other regions	other states	states to the south	states to the north	Mexico City
railroad access = 1	72.353***	72.061***	46.702***	25.359*	-0.012
	(17.677)	(22.663)	(17.822)	(13.000)	(3.752)
	[22.459]	[29.223]	[26.519]	[14.203]	[7.557]
Observations	386	386	386	386	385
Kleibergen-Paap F-statistic	48.82	48.82	48.82	48.82	48.54
Controls					
Baseline	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Dependent variables in columns 1 to 4 are the rate of internal migrants per thousand inhabitants, classified by origin within Mexico. The dependent variable in column 5 is the rate of migrants to Mexico City per thousand inhabitants. We calculate migration rates to Mexico City using counts of Mexico City's residents by state of birth, which we allocate to origin districts proportionally to each district's share of state population. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations and distance to the nearest historical seaport (both linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). Distance to the US border is excluded from these regressions and replaced by latitude, as the two variables are highly collinear by construction: districts further south are further from the border. Latitude is included instead because the dependent variables in columns 3 and 4 are constructed as functions of district latitude, and controlling for it directly absorbs the mechanical north-south gradient in the outcome variable. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 5: Railroad Access and Sequential Migration

	1	2	3	4
	Mexico-US migration rate	Direct migration rate	Sequential migration rate	Share of stage migrants
railroad access = 1	1.540**	0.316	1.223**	0.178***
	(0.666)	(0.249)	(0.621)	(0.063)
	[0.630]	[0.243]	[0.628]	[0.056]
Observations	386	386	386	386
Kleibergen-Paap F-statistic	78.88	78.88	78.88	78.88
Controls				
Baseline	✓	✓	✓	✓
Region FE	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Dependent variables are migration rates (per thousand inhabitants). These are calculated based on the immigrants' place of residence. Direct migrants are individuals whose reported place of birth and residence is the same. Sequential immigrants are individuals whose reported place of birth and place of residence is different. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 6: Short-Run Implications. Structural Transformation

	1	2	3	4	5	6
Panel A. Urbanization						
	Population Density		Urban Density		Housing Density	
railroad access = 1	0.300**	0.246**	0.209*	0.143***	0.320**	0.251***
	(0.135)	(0.097)	(0.114)	(0.054)	(0.136)	(0.089)
	[0.202]	[0.133]	[0.180]	[0.042]	[0.186]	[0.124]
Observations	386	356	386	356	386	356
Kleibergen-Paap F-statistic	93.62	82.80	93.62	82.80	93.62	82.80
Panel B. Economic Development						
	Literacy		Health Access		Land Concentration	
railroad access = 1	0.277	0.259	0.536**	0.492**	-1.351***	-1.443***
	(0.178)	(0.183)	(0.239)	(0.214)	(0.257)	(0.268)
	[0.200]	[0.194]	[0.267]	[0.221]	[0.297]	[0.309]
Observations	386	356	386	356	386	356
Kleibergen-Paap F-statistic	78.88	71.71	78.88	71.71	78.88	71.71
Panel C. Demographic Composition						
	Working-Age Population		Working-Age Sex Ratio		Child-Woman Ratio	
railroad access = 1	0.836***	0.801***	0.020	0.036	-0.932***	-0.880***
	(0.222)	(0.225)	(0.093)	(0.120)	(0.255)	(0.260)
	[0.232]	[0.238]	[0.126]	[0.148]	[0.237]	[0.249]
Observations	386	356	386	356	386	356
Kleibergen-Paap F-statistic	78.88	71.71	78.88	71.71	78.88	71.71
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Excluding Historical Cities		✓		✓		✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Regressions in columns 1, 3, and 5 include all districts. Regressions in columns 2, 4, and 6 exclude districts with historical cities. Dependent variables are calculated as follows: population density (inhabitants per km^2), urban density (urban inhabitants per km^2), housing density (total dwellings per km^2), literacy (population share over 11 years old able to read), health access (health workers per thousand inhabitants), land concentration (population share living in large estates—haciendas or ranches), working-age population (population share 11-60 years old), working-age sex ratio (male to female sex ratio 11-60 years old), and child-woman ratio (children 0-5 years per woman 16-45 years). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). All continuous variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 7: Short-Run Implications. Labor Reallocation

	1	2	3	4	5	6
	Labor Force Participation	Labor Force Employed	Labor Force Unemployed	Employed by Sector Extractive	Industry	Service
railroad access = 1	0.135***	0.022	0.055***	-0.073***	0.016**	0.059***
	(0.034)	(0.023)	(0.013)	(0.016)	(0.008)	(0.022)
	[0.060]	[0.023]	[0.024]	[0.020]	[0.007]	[0.027]
Observations	386	386	386	386	386	386
Kleibergen-Paap F	76.36	79.93	74.78	81.90	78.93	78.70
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Dependent variables are differences in shares of working-age population (1895-1910) by labor market status. They are calculated as follows: labor force participation (working-age share employed or unemployed), employed (working-age share employed), unemployed (working-age unemployed), extractive (working-age share employed in primary sector occupations), industry (working-age share employed in industry sector occupations), services (working-age share employed in service occupations). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 8: Long-Run Effects of Railroad Access. District level

	1	2	3	4	5
railroad access = 1	-1.033	-0.934	-1.112	-1.114	-1.265
	(1.229)	(1.270)	(1.272)	(1.256)	(1.307)
	[1.561]	[1.614]	[1.672]	[1.655]	[1.724]
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	53.65	51.60	46.50	46.93	42.68
Controls					
Baseline	✓	✓	✓	✓	✓
Roads		✓	✓	✓	✓
Land reform			✓	✓	✓
Bracero program				✓	✓
Historical networks					✓
Region FE	✓	✓	✓	✓	✓

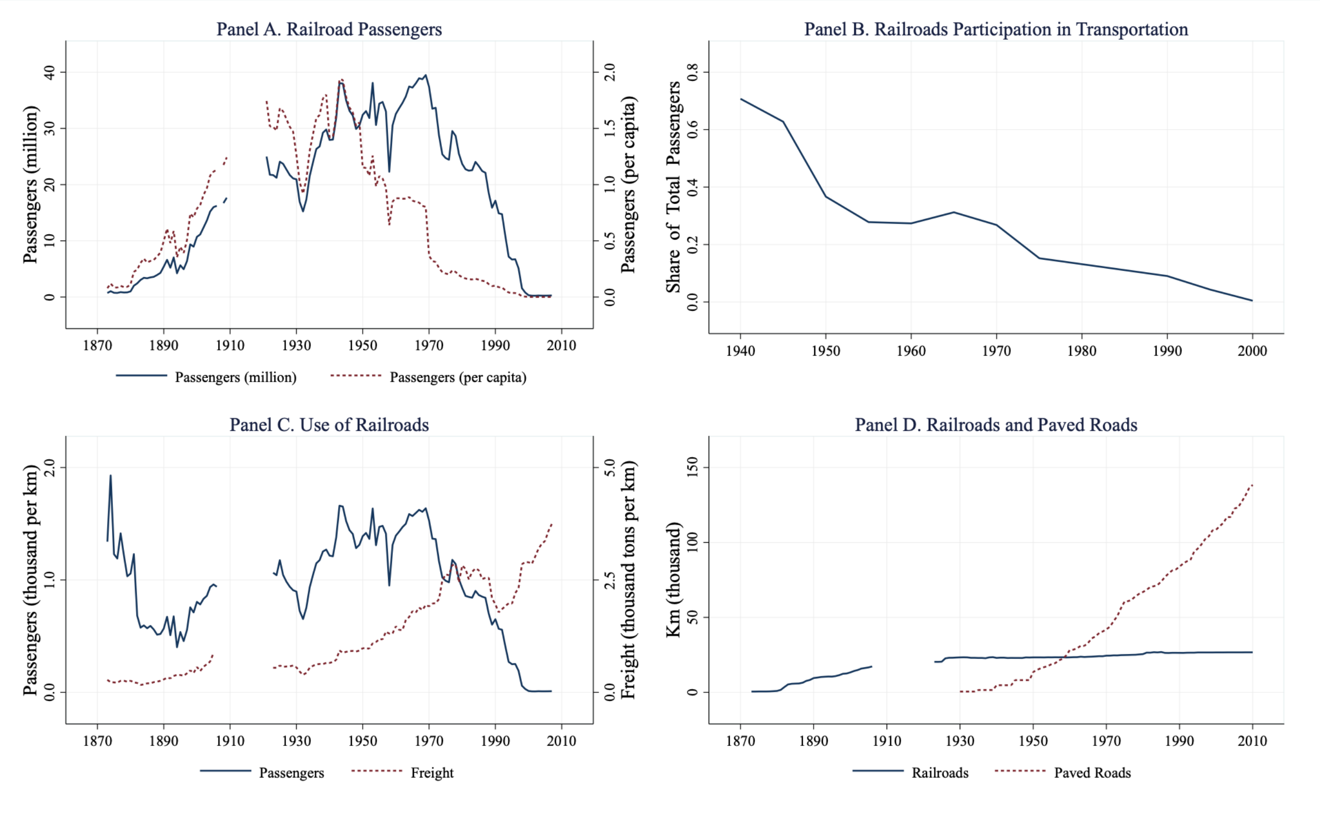
Source: Matrícula Consular de Alta Seguridad (MCAS).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Migration rates (per thousand inhabitants) are calculated based on the immigrants' place of birth. *railroad access* is an indicator variable for observations located within 40 km of the railroad network as it existed by 1910. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index for the period 2006-2012, and the observation's surface area (km^2). *Roads* includes the distance (linear and quadratic) to the nearest paved road. *Land reform* includes the share of land redistributed through outright grants or restitutions and the land granted (km^2) per beneficiary (1916-1976). *Bracero program* is an indicator for municipalities with bracero recruitment centers (1942-1964). *Historical networks* includes an indicator for districts with any migration to the US before 1910 and the mean annual migration rate (1906-1910). All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table 9: Formation of Migration Corridors. District level

	1	2	3	4	5	6
	Share of Migrants by Destination					
	California-Arizona		New Mexico		Texas	
Railroad Network						
Western = 1	0.149***	0.140***	-0.015***	-0.011***	-0.086***	-0.081***
	(0.031)	(0.030)	(0.005)	(0.004)	(0.025)	(0.025)
	[0.035]	[0.034]	[0.005]	[0.004]	[0.031]	[0.030]
Central = 1	-0.064**	-0.051*	0.024***	0.017***	-0.050*	-0.058**
	(0.026)	(0.026)	(0.008)	(0.005)	(0.027)	(0.028)
	[0.026]	[0.024]	[0.008]	[0.005]	[0.027]	[0.027]
Eastern = 1	-0.088***	-0.094***	-0.003	0.001	0.120***	0.119***
	(0.022)	(0.023)	(0.003)	(0.002)	(0.024)	(0.025)
	[0.025]	[0.026]	[0.004]	[0.003]	[0.029]	[0.029]
Ambiguous = 1	0.023	0.020	0.005**	0.007***	-0.048**	-0.049**
	(0.023)	(0.023)	(0.002)	(0.002)	(0.021)	(0.021)
	[0.031]	[0.031]	[0.002]	[0.002]	[0.021]	[0.021]
Observations	386	386	386	386	386	386
R-squared	0.738	0.747	0.606	0.763	0.875	0.876
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Roads	✓	✓	✓	✓	✓	✓
Historical	✓	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓
Predicted destination		✓		✓		✓

Source: Matrícula Consular de Alta Seguridad (MCAS).
Note: The unit of observation is a district. All regressions are estimated by ordinary least squares. Dependent variables are the share of migrants by destination. These are calculated as the number of migrants to each state divided by total migrants to all US destinations. Independent variables are indicators for observations located within 40 km of each railroad network. The omitted category are observations with no railroad access. *Baseline controls* include distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index for the period 2006-2012, and the observation's surface area (km^2). *Roads* includes the distance (linear and quadratic) to the nearest paved road. *Historical* includes the share of land redistributed through outright grants or restitutions and the land granted (km^2) per beneficiary (1916-1976); an indicator for municipalities with bracero recruitment centers (1942-1964); and an indicator for districts with any migration to the US before 1910 and the mean annual migration rate (1906-1910). *Predicted destination* is a categorical variable that classifies each observation into one of three predicted US destination groups according to its nearest border crossing by road: California-Arizona, New Mexico, and Texas. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

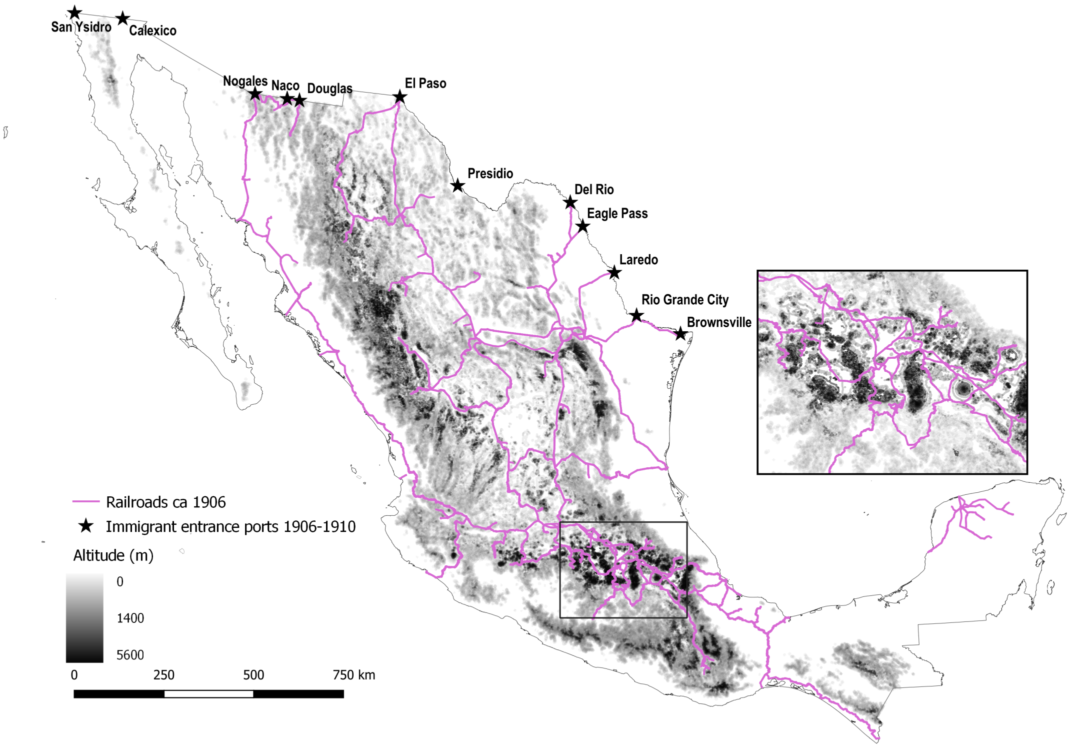
Figure 1: Rise and Fall of Railroad Passenger Transportation



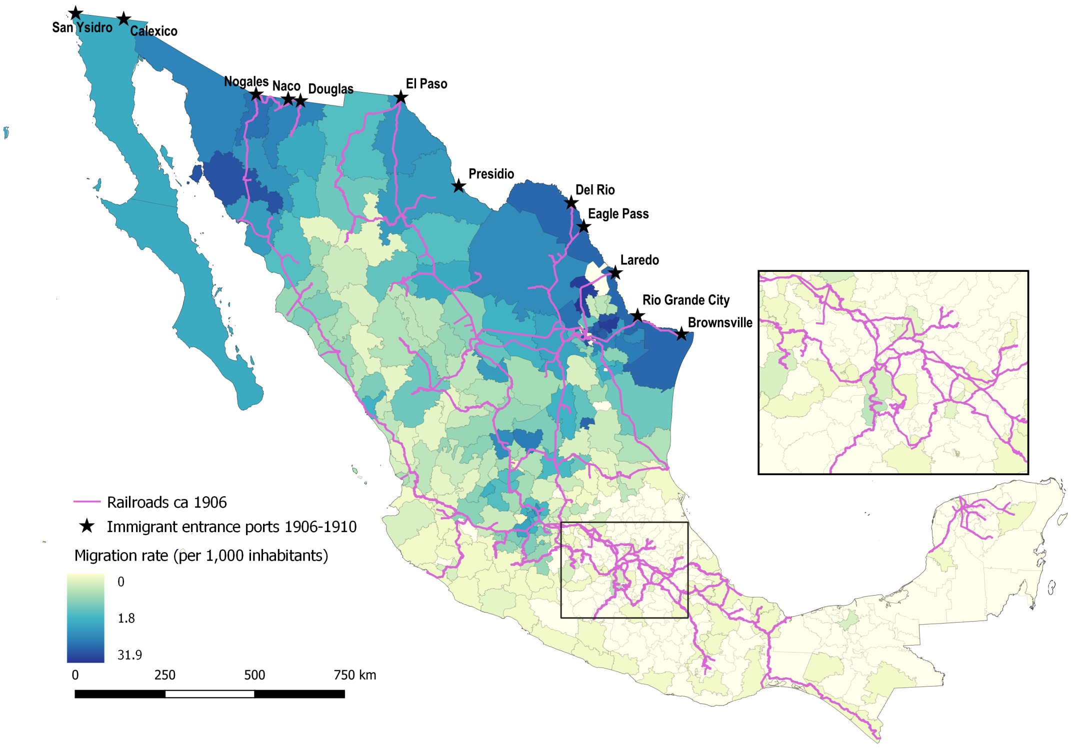
Source: INEGI (1950–2000) and INEGI (2010).
Note: Passengers per capita from 1870 to 1950 are calculated using census data. Passengers per capita from 1951 to 2010 are calculated using population data from CONAPO (2023) (Panel A). Transportation includes medium and long-distance passenger transportation by bus, railroad, and airplane (Panel B). Paved roads consist of free-toll roads and tolled highways (Panel D).

Figure 2: Railroads and Migration to the United States

Panel A. Railway Network and Ruggedness

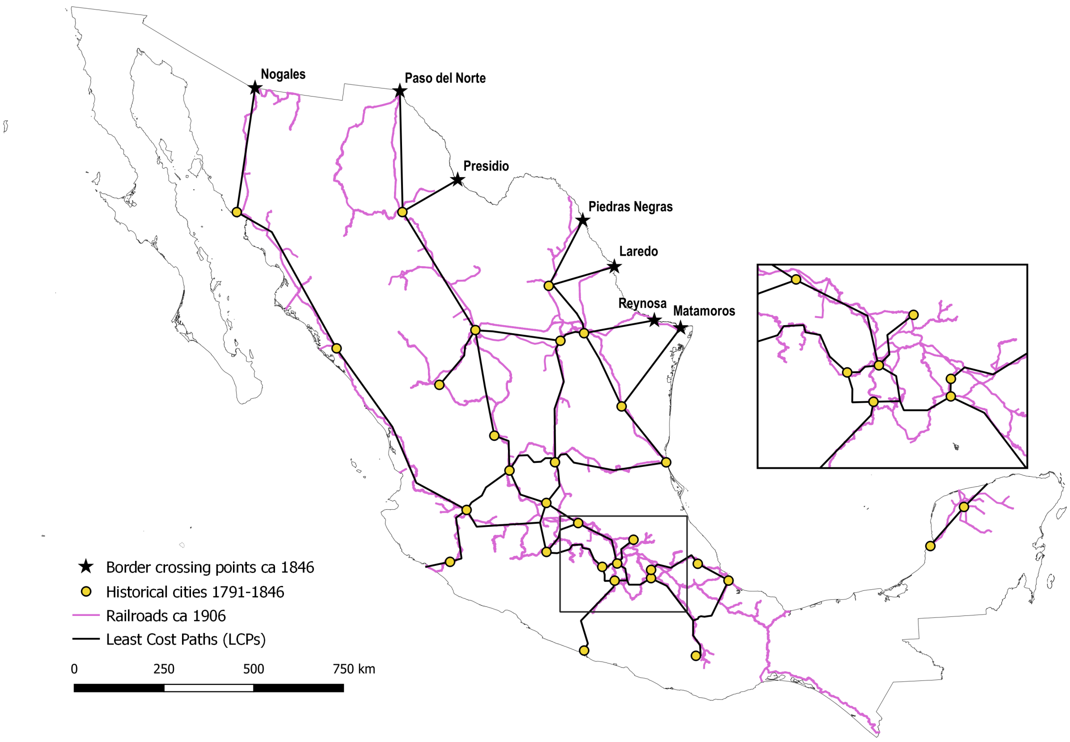


Panel B. Migration Rates by District



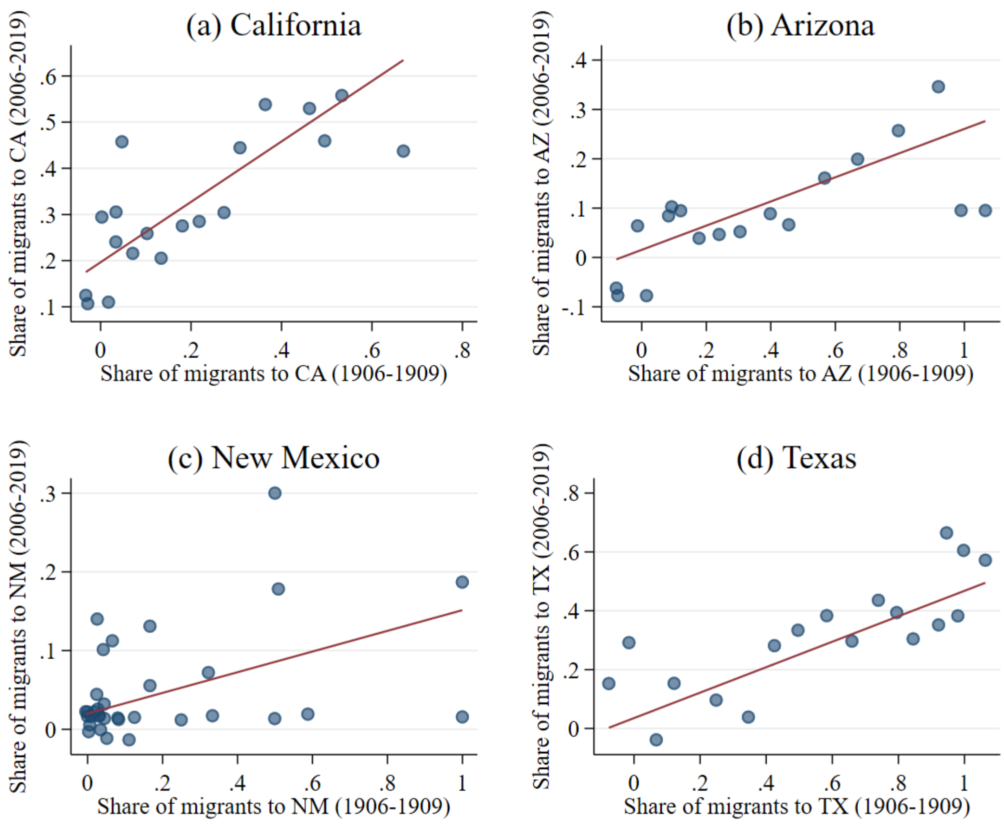
Source: Mexican Border Crossing Records (MBCRs), DGE (1918), and FNM (1914).
Note: The figure shows the terrain ruggedness of Mexico and the railroad network ca 1906 (Panel A). The polygons display migration rates per 1000 inhabitants by district (Panel B).

Figure 3: Least Cost Paths



Source: Castro Aranda (2010), Disturnell (1847), and FNM (1914).
Note: The figure shows the location of target destinations (historical cities and historical border crossing points) used to construct the least cost paths.

Figure 4: Persistence in Destination Choices



Source: Mexican Border Crossing Records (MBCRs) and Matriculas Consulares (MCAS).
Note: Each panel shows a binscatter plot of the relationship between historical and contemporary migrant destination choices. All plots control for region fixed effects to account for geographic proximity to the border. The sample includes municipalities with positive migration in both periods. Destination shares are calculated as the number of migrants to each state divided by total migrants to all US destinations.

Online Appendix

June 11, 2026

A. First stage and reduced form

Table A.1: Railroad Access and International Migration. Immigrants' Place of Birth

	1	2	3	4	5
Panel A. OLS					
railroad access = 1	0.406**	0.403**	0.409*	0.422**	0.389**
	(0.200)	(0.198)	(0.221)	(0.213)	(0.189)
Observations	386	386	386	386	386
R-squared	0.360	0.367	0.377	0.377	0.378
Panel B. Reduced Form					
distance to least cost path (km)	-0.003**	-0.002**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	386	386	386	386	386
R-squared	0.358	0.364	0.376	0.376	0.377
Panel C. IV					
railroad access = 1	0.721**	0.586**	0.731**	0.748**	0.696**
	(0.293)	(0.268)	(0.322)	(0.318)	(0.294)
Observations	386	386	386	386	386
Panel D. First Stage	Dependent variable: railroad access = 1				
distance to least cost path (km)	-0.0038***	-0.0039***	-0.0040***	-0.0039***	-0.0037***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. The dependent variable in Panels A-C is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the interaction term between latitude and longitude. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* is the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* is the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table A.2: Railroad Access and International Migration. Immigrants' Place of Last Residence

	1	2	3	4	5
Panel A. OLS					
railroad access = 1	0.655**	0.654**	0.690**	0.687**	0.621**
	(0.283)	(0.284)	(0.318)	(0.312)	(0.278)
Observations	386	386	386	386	386
R-squared	0.282	0.283	0.297	0.297	0.299
Panel B. Reduced Form					
distance to least cost path (km)	-0.006***	-0.006**	-0.006**	-0.006**	-0.006**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Observations	386	386	386	386	386
R-squared	0.282	0.283	0.298	0.298	0.300
Panel C. IV					
railroad access = 1	1.445***	1.408**	1.605***	1.613***	1.540**
	(0.548)	(0.615)	(0.620)	(0.620)	(0.666)
Observations	386	386	386	386	386
Panel D. First Stage	Dependent variable: railroad access = 1				
distance to least cost path (km)	-0.0038***	-0.0039***	-0.0040***	-0.0039***	-0.0037***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. The dependent variable in Panels A-C is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of last residence. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

B. First stage heterogeneity

Table B.1: IV Estimates with First Stage Heterogeneity

	1	2	3	4	5
Panel A. Place of Birth					
railroad access = 1	0.648**	0.591**	0.575**	0.575*	0.520
	(0.274)	(0.272)	(0.288)	(0.294)	(0.324)
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	22.07	25.43	25.78	25.38	21.03
Panel B. Last Residence					
railroad access = 1	1.237**	1.248**	1.298**	1.299**	1.188*
	(0.509)	(0.579)	(0.571)	(0.584)	(0.667)
Observations	386	386	386	386	386
Kleibergen-Paap F-statistic	22.07	25.43	25.78	25.38	21.03
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface Area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path interacted with region indicators as the instruments for railroad access. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

C. Robustness checks

Table C.1: Railroad Access and International Migration. Alternative Controls

	1	2	3	4	5	6
Panel A. Place of Birth						
railroad access = 1	0.696**	0.827***	0.912***	0.899***	0.893***	0.857**
	(0.294)	(0.321)	(0.315)	(0.322)	(0.325)	(0.345)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	78.88	70.16	59.98	60.85	60.80	59.79
Panel B. Place of Last Residence						
railroad access = 1	1.540**	1.499**	1.711**	1.646**	1.654**	1.716**
	(0.666)	(0.674)	(0.729)	(0.691)	(0.694)	(0.687)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	78.88	70.16	59.98	60.85	60.80	59.79
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Demographic structure		✓	✓	✓	✓	✓
Labor market			✓	✓	✓	✓
Living standards				✓	✓	✓
Landholding					✓	✓
Information						✓
Region FE	✓	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. The dependent variable is the mean annual migration rate per thousand inhabitants calculated based on immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes all controls from Table 3: distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, mean altitude, the latitude-longitude interaction term, land suitability indices for Mexico's main staple crops, the Palmer Drought Severity Index averaged over 1906–1910, and district surface area in km^2. *Demographic structure* includes the share of prime-age population, the share of agricultural peons, the sex ratio, and the share of indigenous-language speakers. *Labor market* includes the unemployment rate and the US-Mexico wage gap. *Living standards* includes urban density, the illiteracy rate, and the share of huts. *Landholding* includes the number of large estates. *Information* includes the share of US-born residents and the number of telegraph offices. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.2: IV Estimates with Conley Standard Errors

	1	2	3	4
	Robust SE	SC 100 km	SC 200 km	SC 300 km
Panel A. Place of Birth				
railroad access = 1	0.696**	0.696**	0.696*	0.696**
	(0.294)	(0.297)	(0.382)	(0.318)
Observations	386	386	386	386
Panel B. Place of Last Residence				
distance to railroad (km)	1.540**	1.540**	1.540**	1.540***
	(0.666)	(0.630)	(0.603)	(0.502)
Observations	386	386	386	386
Controls				
Baseline	✓	✓	✓	✓
Region FE	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). All continuous control variables are standardized. Robust standard errors in column 1. Standard errors corrected for spatial correlation in columns 2-4 (Conley, 1999). We use different distance cutoffs (kilometers) beyond which the correlation of the error term between districts is assumed to be zero. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.3: IV Estimates Controlling for Market Access. Quadratic Specification

	1	2	3	4	5	6
Panel A. Place of Birth						
railroad access = 1	0.721**	0.586**	0.731**	0.748**	0.696**	0.700**
	(0.293)	(0.268)	(0.322)	(0.318)	(0.294)	(0.294)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88	79.26
Panel B. Place of Last Residence						
railroad access = 1	1.445***	1.408**	1.605***	1.613***	1.540**	1.514**
	(0.548)	(0.615)	(0.620)	(0.620)	(0.666)	(0.668)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88	79.26
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓	✓
Agricultural suitability			✓	✓	✓	✓
Weather shocks				✓	✓	✓
Surface area					✓	✓
Market access						✓
Region FE	✓	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. *Market access* is estimated as $MA_o \approx \sum_{d \neq o} \tau_{od}^{-\theta} N_d$ following Donaldson and Hornbeck (2016), with θ = 5.03—the median value of trade elasticity in the literature (Head and Mayer, 2014). This measure captures the economic opportunities available to each district by summing the population of all other districts, weighted by bilateral trade costs $\tau_{od}^{-\theta}$, where districts with lower trade costs and larger populations contribute more to the market access of district o. We control for market access and its square to allow for a non-linear relationship between migration and domestic market access. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.4: IV Estimates Controlling for Market Access. Non-Parametric Specification

	1	2	3	4	5	6
Panel A. Place of Birth						
railroad access = 1	0.721**	0.586**	0.731**	0.748**	0.696**	0.540*
	(0.293)	(0.268)	(0.322)	(0.318)	(0.294)	(0.321)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88	80.32
Panel B. Place of Last Residence						
railroad access = 1	1.445***	1.408**	1.605***	1.613***	1.540**	1.442**
	(0.548)	(0.615)	(0.620)	(0.620)	(0.666)	(0.685)
Observations	386	386	386	386	386	386
Kleibergen-Paap F-statistic	89.83	100.3	96.47	93.62	78.88	80.32
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓	✓
Agricultural suitability			✓	✓	✓	✓
Weather shocks				✓	✓	✓
Surface area					✓	✓
Market access						✓
Region FE	✓	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).

Note: The unit of observation is a district. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth (Panel A) and last residence (Panel B). *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. *Market access* is estimated as $MA_o \approx \sum_{d \neq o} \tau_{od}^{-\theta} N_d$ following Donaldson and Hornbeck (2016), with $\theta = 5.03$—the median value of trade elasticity in the literature (Head and Mayer, 2014). This measure captures the economic opportunities available to each district by summing the population of all other districts, weighted by bilateral trade costs $\tau_{od}^{-\theta}$, where districts with lower trade costs and larger populations contribute more to the market access of district o. We calculate market access deciles and include them as fixed effects to flexibly control for domestic market access. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.5: IV Estimates Excluding Outliers. Immigrants' Place of Birth

	1	2	3	4	5
Panel A. Excluding districts with historical cities					
railroad access = 1	0.770**	0.619**	0.746**	0.758**	0.775**
	(0.301)	(0.277)	(0.321)	(0.319)	(0.311)
Observations	356	356	356	356	356
Kleibergen-Paap F-statistic	80.50	89.93	85.52	82.80	71.71
Panel B. Excluding districts within 20 km from target destinations					
railroad access = 1	0.745**	0.577**	0.720**	0.734**	0.674**
	(0.297)	(0.267)	(0.323)	(0.322)	(0.301)
Observations	360	360	360	360	360
Kleibergen-Paap F-statistic	86.34	96.09	90.26	87.84	73.13
Panel C. Excluding districts at the US border					
railroad access = 1	0.938***	0.867***	1.085***	1.089***	1.072***
	(0.327)	(0.317)	(0.413)	(0.411)	(0.385)
Observations	375	375	375	375	375
Kleibergen-Paap F	70.73	73.54	71.22	70.18	64.19
Panel D. Excluding districts from Mexico City area					
railroad access = 1	0.722**	0.580**	0.735**	0.754**	0.702**
	(0.293)	(0.268)	(0.324)	(0.319)	(0.295)
Observations	377	377	377	377	377
Kleibergen-Paap F	89.43	99.82	95.91	92.65	78.11
Panel E. Excluding districts with high urban density					
railroad access = 1	0.731**	0.581**	0.660**	0.671**	0.625**
	(0.306)	(0.278)	(0.325)	(0.323)	(0.314)
Observations	348	348	348	348	348
Kleibergen-Paap F	78.89	88.17	82.11	80.16	66.64
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface Area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.6: IV Estimates Excluding Outliers. Immigrants' Place of Last Residence

	1	2	3	4	5
Panel A. Excluding districts with historical cities					
railroad access = 1	1.470***	1.452**	1.614**	1.614**	1.643**
	(0.566)	(0.650)	(0.648)	(0.649)	(0.684)
Observations	356	356	356	356	356
Kleibergen-Paap F-statistic	80.50	89.93	85.52	82.80	71.71
Panel B. Excluding districts within 20 km from target destinations					
railroad access = 1	1.464***	1.395**	1.603**	1.614**	1.519**
	(0.548)	(0.620)	(0.629)	(0.631)	(0.683)
Observations	360	360	360	360	360
Kleibergen-Paap F-statistic	86.34	96.09	90.26	87.84	73.13
Panel C. Excluding districts at the US border					
railroad access = 1	1.072***	0.814**	1.141**	1.136**	1.090**
	(0.363)	(0.333)	(0.462)	(0.458)	(0.444)
Observations	375	375	375	375	375
Kleibergen-Paap F	70.73	73.54	71.22	70.18	64.19
Panel D. Excluding districts from Mexico City area					
railroad access = 1	1.446***	1.408**	1.610***	1.617***	1.543**
	(0.548)	(0.617)	(0.622)	(0.621)	(0.668)
Observations	377	377	377	377	377
Kleibergen-Paap F	89.43	99.82	95.91	92.65	78.11
Panel E. Excluding districts with high urban density					
railroad access = 1	1.468***	1.448**	1.556**	1.558**	1.514**
	(0.569)	(0.638)	(0.638)	(0.639)	(0.700)
Observations	348	348	348	348	348
Kleibergen-Paap F	78.89	88.17	82.11	80.16	66.64
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface Area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. The dependent variable is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of last residence. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.7: Rail Line Orientation and South-North Mobility

	1	2	3
	Internal migration from states located to the south		
	All rail lines	South-North rail lines	East-West rail lines
railroad access = 1	46.702***	55.840***	18.264
	(17.822)	(17.856)	(29.843)
Observations	386	322	194
Kleibergen-Paap F-statistic	48.82	48.01	19.02
Controls			
Baseline	✓	✓	✓
Region FE	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Column 1 reports estimates of Table 4 (Panel C) for reference. Column 2 excludes districts with large shares of east-west oriented rail lines. Column 3 excludes districts with large shares of north-south oriented rail lines. The dependent variable is the rate of internal migrants (from states located to the south) per thousand inhabitants. *railroad access* is an indicator variable for districts within 40 km from the nearest rail line. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index averaged over 1906-1910, and the observation's surface area (km^2). All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

D. Continuous treatment

Table D.1: Continuous Treatment. Immigrants' Place of Birth

	1	2	3	4	5
Panel A. OLS					
distance to railroad (km)	-0.000	-0.001	-0.002	-0.002	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Observations	386	386	386	386	386
R-squared	0.355	0.363	0.374	0.374	0.375
Panel B. Reduced Form					
distance to least cost path (km)	-0.003**	-0.002**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	386	386	386	386	386
R-squared	0.358	0.364	0.376	0.376	0.377
Panel C. IV					
distance to railroad (km)	-0.008**	-0.006**	-0.008**	-0.008**	-0.008**
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	386	386	386	386	386
Panel D. First Stage	Dependent variable: distance to railroad (km)				
distance to least cost path (km)	0.348***	0.383***	0.361***	0.350***	0.322***
	(0.058)	(0.053)	(0.056)	(0.054)	(0.051)
Observations	386	386	386	386	386
Kleibergen-Paap F	35.53	53.10	42.22	41.59	39.48
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. The dependent variable in Panels A-C is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of birth. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table D.2: Continuous Treatment. Immigrants' Place of Last Residence

	1	2	3	4	5
Panel A. OLS					
distance to railroad (km)	-0.004	-0.005*	-0.006*	-0.006*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	386	386	386	386	386
R-squared	0.277	0.278	0.293	0.293	0.295
Panel B. Reduced Form					
distance to least cost path (km)	-0.006***	-0.006**	-0.006**	-0.006**	-0.006**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Observations	386	386	386	386	386
R-squared	0.282	0.283	0.298	0.298	0.300
Panel C. IV					
distance to railroad (km)	-0.016**	-0.014**	-0.018**	-0.018**	-0.018**
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Observations	386	386	386	386	386
Panel D. First Stage	Dependent variable: distance to railroad (km)				
distance to least cost path (km)	0.348***	0.383***	0.361***	0.350***	0.322***
	(0.058)	(0.053)	(0.056)	(0.054)	(0.051)
Observations	386	386	386	386	386
Kleibergen-Paap F	35.53	53.10	42.22	41.59	39.48
Controls					
Baseline	✓	✓	✓	✓	✓
Location fundamentals		✓	✓	✓	✓
Agricultural suitability			✓	✓	✓
Weather shocks				✓	✓
Surface area					✓
Region FE	✓	✓	✓	✓	✓

Source: Mexican Border Crossing Records (MBCRs) and DGE (1918).
Note: The unit of observation is a district. The dependent variable in Panels A-C is the mean annual migration rate per thousand inhabitants. This variable is calculated based on the immigrants' place of last residence. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), a Yucatan Peninsula indicator, and region fixed effects. *Location fundamentals* includes mean altitude and the latitude-longitude interaction term. *Agricultural suitability* includes land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat). *Weather shocks* includes the Palmer Drought Severity Index averaged over 1906-1910. *Surface area* includes the district's surface area in km^2. All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

E. Long-run effects. Municipality-level analysis

Table E.1: Long-Run Effects of Railroad Access. Municipality level

	1	2	3	4	5
railroad access = 1	-3.653***	-3.706***	-3.738***	-3.770***	-3.757***
	(0.919)	(0.926)	(0.934)	(0.933)	(0.934)
	[1.702]	[1.693]	[1.705]	[1.696]	[1.706]
Observations	2,469	2,469	2,469	2,469	2,469
Kleibergen-Paap F-statistic	148.4	144.6	143.9	145.6	144.2
Controls					
Baseline	✓	✓	✓	✓	✓
Roads		✓	✓	✓	✓
Land reform			✓	✓	✓
Bracero program				✓	✓
Historical networks					✓
State FE	✓	✓	✓	✓	✓

Source: Matrícula Consular de Alta Seguridad (MCAS).
Note: The unit of observation is a municipality. All regressions are estimated by instrumental variables with distance to the nearest least cost path as the instrument for railroad access. Migration rates (per thousand inhabitants) are calculated based on the immigrants' place of birth. *railroad access* is an indicator variable for observations located within 40 km of the railroad network as it existed by 1910. *Baseline* includes distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index for the period 2006-2012, and the observation's surface area (km^2). *Roads* includes the distance (linear and quadratic) to the nearest paved road. *Land reform* includes the share of land redistributed through outright grants or restitutions and the land granted (km^2) per beneficiary (1916-1976). *Bracero program* is an indicator for municipalities with bracero recruitment centers (1942-1964). *Historical networks* includes an indicator for districts with any migration to the US before 1910 and the mean annual migration rate (1906-1910). All continuous control variables are standardized. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

Table E.2: Formation of Migration Corridors. Municipality level

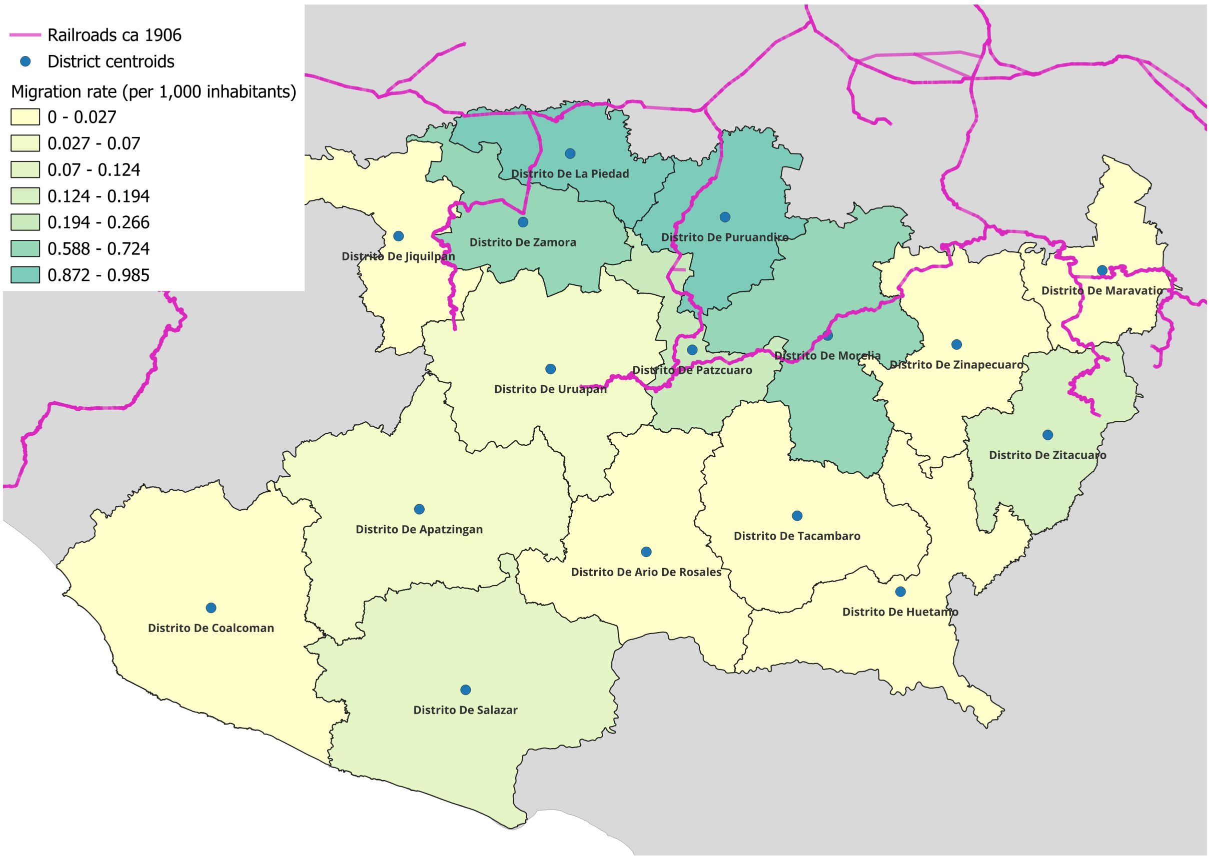
	1	2	3	4	5	6
	Share of Migrants by Destination					
	California-Arizona		New Mexico		Texas	
Railroad Network						
Western = 1	0.095***	0.095***	-0.005***	-0.004***	-0.023**	-0.023**
	(0.019)	(0.019)	(0.001)	(0.001)	(0.011)	(0.011)
	[0.039]	[0.039]	[0.002]	[0.002]	[0.017]	[0.018]
Central = 1	-0.046**	-0.045**	0.017***	0.015***	-0.012	-0.009
	(0.019)	(0.019)	(0.005)	(0.005)	(0.019)	(0.019)
	[0.026]	[0.027]	[0.006]	[0.006]	[0.026]	[0.025]
Eastern = 1	-0.072***	-0.072***	-0.003	-0.003	0.100***	0.101***
	(0.015)	(0.015)	(0.003)	(0.003)	(0.017)	(0.017)
	[0.026]	[0.026]	[0.003]	[0.003]	[0.025]	[0.025]
Ambiguous = 1	0.039***	0.039***	0.002**	0.001**	-0.013**	-0.012*
	(0.012)	(0.012)	(0.001)	(0.001)	(0.006)	(0.006)
	[0.029]	[0.029]	[0.001]	[0.001]	[0.012]	[0.011]
Observations	2,469	2,469	2,469	2,469	2,469	2,469
R-squared	0.481	0.482	0.769	0.788	0.811	0.812
Controls						
Baseline	✓	✓	✓	✓	✓	✓
Roads	✓	✓	✓	✓	✓	✓
Historical	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Predicted destination		✓		✓		✓

Source: Matrícula Consular de Alta Seguridad (MCAS).

Note: The unit of observation is a municipality. All regressions are estimated by ordinary least squares. Dependent variables are the share of migrants by destination. These are calculated as the number of migrants to each state divided by total migrants to all US destinations. Independent variables are indicators for observations located within 40 km of each railroad network. The omitted category are observations with no railroad access. *Baseline controls* include distance to LCP target destinations, distance to the US border, distance to the nearest historical seaport (all linear and quadratic), an indicator for observations in the Yucatan Peninsula, mean altitude, the interaction between latitude and longitude, land suitability indices for Mexico's main staple crops (beans, carrot, coffee, maize, potato, tomato, and wheat), the Palmer Drought Severity Index for the period 2006-2012, and the observation's surface area (km^2). *Roads* includes the distance (linear and quadratic) to the nearest paved road. *Historical* includes the share of land redistributed through outright grants or restitutions and the land granted (km^2) per beneficiary (1916-1976); an indicator for municipalities with bracero recruitment centers (1942-1964); and an indicator for districts with any migration to the US before 1910 and the mean annual migration rate (1906-1910). *Predicted destination* is a categorical variable that classifies each observation into one of three predicted US destination groups according to its nearest border crossing by road: California-Arizona, New Mexico, and Texas. Robust standard errors in parentheses. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level. Conley standard errors in brackets using a 100 km cutoff.

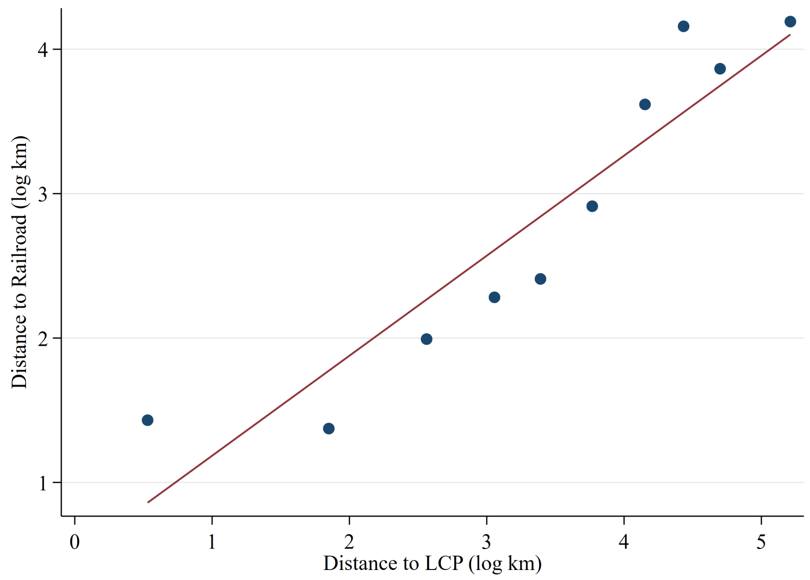
F. Additional figures

Figure F.1: Railroad Access and Migration to the United States. State of Michoacan



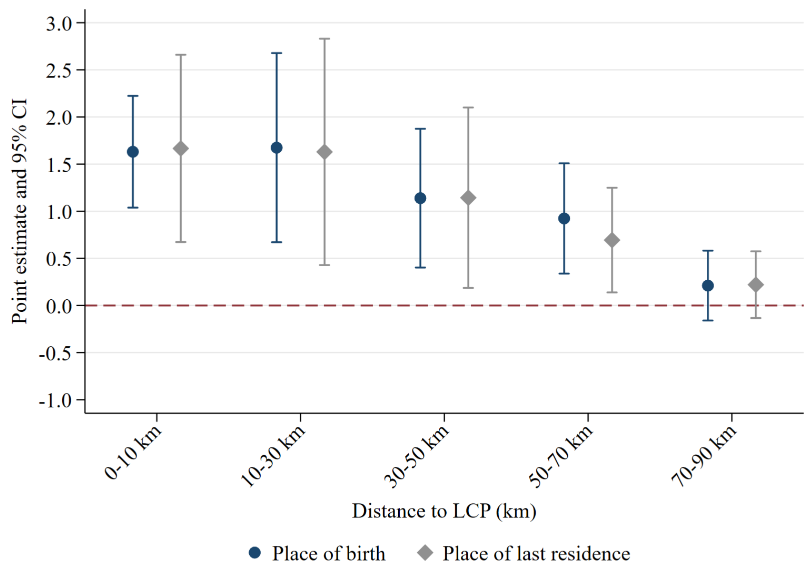
Source: Mexican Border Crossing Records (MBCRs), DGE (1918), and FNM (1914).
Note: The figure shows the railroad network ca 1906 and migration rates per 1000 inhabitants by district.

Figure F.2: Distance to Railroads and Distance to Least Cost Paths



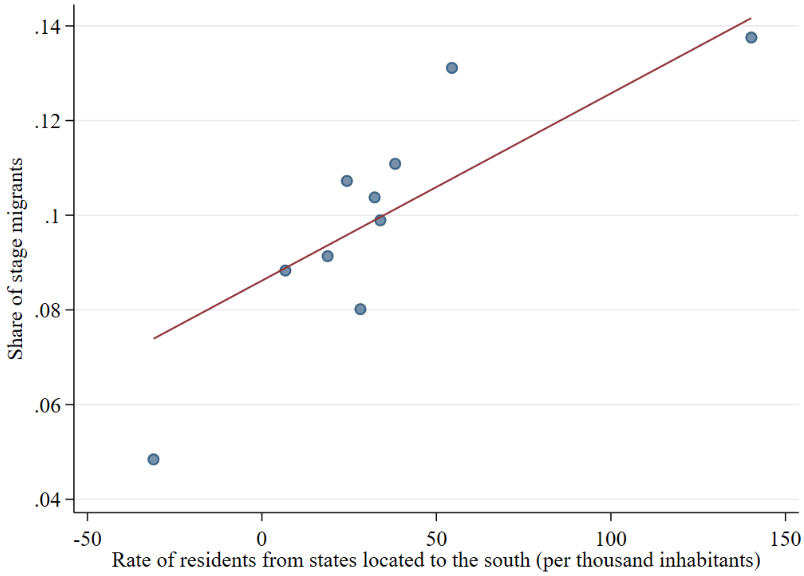
Source: Authors' estimates.
Note: The figure shows a binscatter plot—conditional on region fixed effects—of the relationship between distance to the railroad network and distance to the least cost paths.

Figure F.3: Distance to Least Cost Paths and Migration to the United States



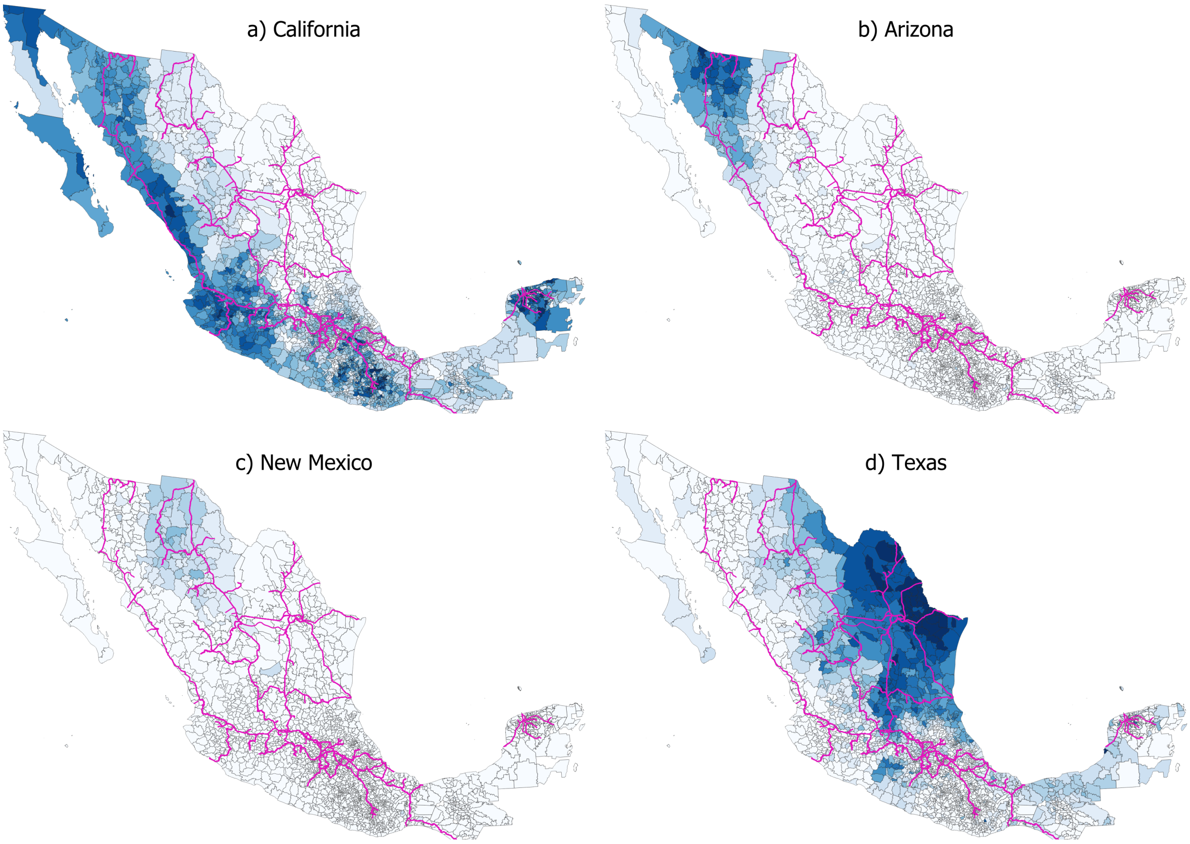
Source: Authors' estimates.
Note: The figure shows estimates of the reduced form relationship between proximity to the least cost paths (LCPs) and migration rates to the United States. Each coefficient represents the difference in migration rates between districts in a given distance bin and districts more than 90 km from the nearest LCP (omitted category). The dependent variable is the mean annual migration rate per thousand inhabitants, calculated based on immigrants' place of birth (navy circles) and place of last residence (grey diamonds). Regressions control for distance to LCP target destinations (linear and quadratic) and district surface area. Capped lines represent 95% confidence intervals based on robust standard errors.

Figure F.4: Internal Mobility and Sequential Migration



Source: Authors' estimates.
Note: The figure shows a binscatter plot—conditional on state fixed effects—of the relationship between internal mobility and stage migration.

Figure F.5: Railroad Access and Segmentation of Destination Choices



Source: MCAS.

Note: The figure shows that historical railroad access induced the geographic segmentation of destination choices.

G. Additional tables

Table G.1: Minimum wage in Mexico by economic activity, 1877–1911.
Cents per day (US dollars)

	All Sectors		**Agriculture**		**Manufactures**		**Mining**	
Year	*Current prices*	*1900 prices*	*Current prices*	*1900 prices*	*Current prices*	*1900 prices*	*Current prices*	*1900 prices*
1877	11	16	11	16	11	16	11	16
1885	11	14	11	13	14	17	13	15
1892	15	14	14	13	16	13	16	15
1898	17	19	15	18	19	25	20	23
1902	18	16	17	16	20	18	23	21
1911	24	15	22	13	29	18	59	36

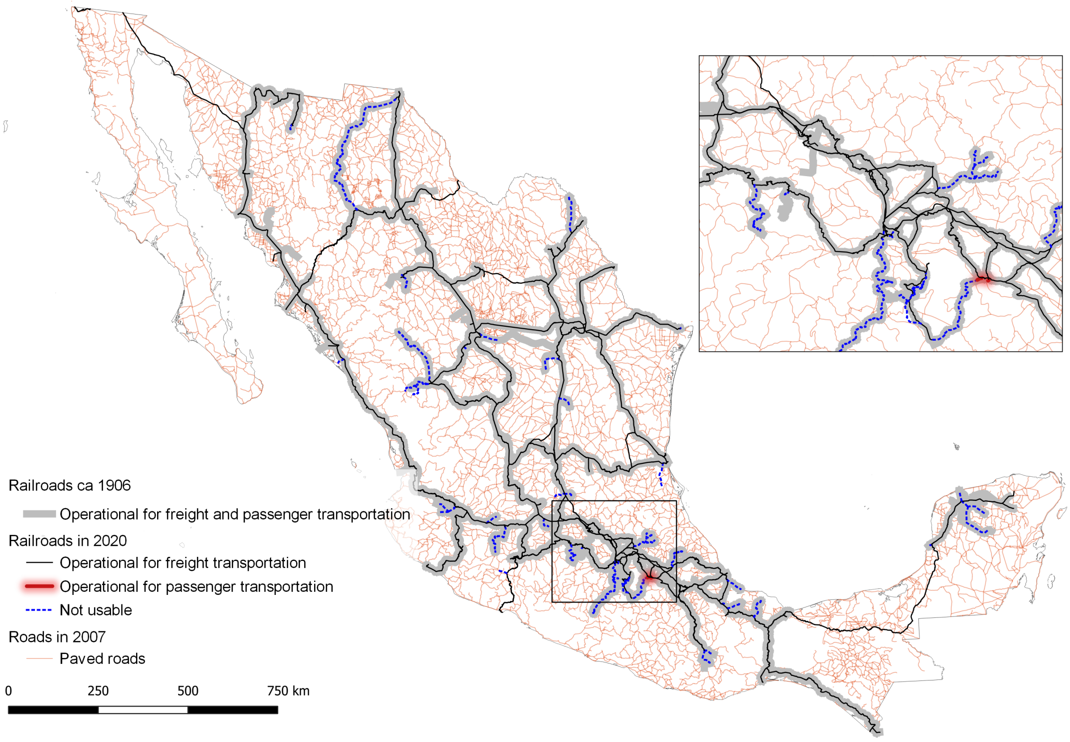
Source: Rosenzweig (1965, p. 447).
Notes: Real wages stagnated during the Porfirian period (1877–1911), especially in agriculture. The exception are wages in the mining sector that grew from 1898. Yet, for most of the Mexican population, wage stagnation translated into large wage differentials between Mexico and the United States.

H. Additional maps

Figure H.1: Emigration Regions in Mexico

Source: Adapted from Durand (2016, p. 28).
Note: Before 1910, the state of Nayarit was called Tepic, and the states of Baja California and Baja California Sur constituted a single territory.

Figure H.2: Railroads in Mexico (1906-2020)



Source: FNM (1914) and Agencia Reguladora del Transporte Ferroviario (2023).
Note: The map shows Mexico's railroad network by use type from 1906 to 2020. Overlapping gray and black lines show railroads that were operational in 1906 and remained operational in 2020. Gray lines alone indicate railroads operational in 1906 that were later removed. Black lines alone represent extensions added to the network after 1906. Overlapping gray and blue dotted lines show railroads from 1906 that were unusable in 2020. Red lines indicate railroads that provided passenger transportation in 2020.