Time to Accumulate: The Great Migration and the Rise of the American South

Dongkyu Yang*

(Job Market Paper)

[Latest version here] [Online Appendix here]

November 27, 2024

Abstract

The idea that labor scarcity can induce economic development has long been hypothesized, but the evidence remains limited. This paper examines how the Second Great Migration (1940–1970) spurred structural change in the American South between 1970 and 2010. Empirical results using shift-share instruments show that out-migration incentivized capital investment and capital-augmenting technical change, increasing capital per worker and output in both agriculture and manufacturing, at least until 2010. Labor was reallocated from agriculture to non-agriculture. I then develop a dynamic spatial equilibrium model that allows for substitution between factors of production, factor-biased technical change, and factor abundance-based trade to characterize this process. The quantitative analysis indicates that labor-capital substitution played a major role in adjustments to South-to-North migration.

Keywords: Factor abundance, Internal Migration, Labor Scarcity, Structural change.
JEL: N32, O11, R12, R23.

*Department of Economics, University of Colorado - Boulder. E-mail: Dongkyu.Yang@colorado.edu. Website: www.dongkyuyang.com.

I would like to thank my dissertation committee members, Richard Mansfield, Sergey Nigai, Taylor Jaworski, and Wolfgang Keller, for their guidance and support. I received valuable comments and suggestions from Erhan Artuc, Hoyt Bleakley, Maggie Chen, Justin Cook, José-Antonio Espín-Sánchez, Maggie Jones, Yeonha Jung, Jeanne Lafortune, Brian Marein, Kris Mitchener, Paul Rhode, Martin Rotemberg, and Zachary Ward. All errors are my own. Please feel free to reach out to me with any comments.

Introduction

Economies undergo structural change from agriculture to non-agriculture as a crucial part of the development process. A large literature has pointed out various sources of structural change (Baumol, 1967; Caselli and Coleman II, 2001; Acemoglu and Guerrieri, 2008; Boppart, 2014). Such an economic transformation is often accompanied by large internal migration, with its impacts felt on the origins and destinations (Bryan and Morten, 2019; Derenoncourt, 2022; Lagakos et al., 2023). While the role of economic change as a driver of migration is evident (Lewis, 1954; Harris and Todaro, 1970; Gollin et al., 2002), migration as a source of structural change has received relatively less attention.[1] In this paper, I propose migration-induced labor scarcity and the following capital accumulation as a source of structural change by analyzing one of the largest labor reallocation episodes in the United States history: the Great Migration.

During the era of the Great Migration (1910-1970), millions of Black and White migrants left the American South ("the South"). This historical episode is often divided into the first (1910-1930) and the second phase (1940-1970), with the migration flows during the latter being much larger (Gregory, 2005). The second period is also characterized by rapid industrialization and structural change in the South. By 1940, the agricultural share of employment in the South (30%) was almost three times higher than the rest of the country (referred to as "the North" for simplicity). However, the two economies converged by the 1980s in terms of industry employment share. Here, I focus on the second wave and investigate its long-running economic impacts.

I show that the relative labor scarcity from out-migration spurred capital investment and structural change in the South, with the heterogeneous labor-capital substitution pattern by industry playing a key role. First, by adopting a shift-share instrumental variable (SSIV) design, this paper examines how differential exposures to Northern migration pull factors between 1940 and 1970 induced economic changes in the South until 2010, while controlling for the overall levels of migration push factors in the Southern origin.[2] I then quantitatively assess the mechanisms behind the empirical findings by developing a computational model featuring migration and the Heckscher–Ohlin force in trade, building on dynamic spatial equilibrium frameworks.

Economics literature has long studied the influence of the Great Migration (Kirby, 1983; Grossman, 1989; Boustan, 2010; Collins and Wanamaker, 2014, 2015; Bazzi et al., 2023), while others

[1] Compared to the existing studies on labor mobility and structural change, I focus on regional out-migration and resulting labor scarcity at the origin, rather than industry switching or out-migration from a specific industry.

[2] The recent literature examining the impacts of the Great Migration often constructs instruments for Northern inflows based on Southern push factors. I apply the same strategy in the opposite direction in a reduced-form.

have investigated why the American South lagged in economic development and why it later caught up with the North (Whatley, 1985; Wright, 1986; Bleakley, 2007; Depew et al., 2013). On the one hand, the maturing of the Southern economy has been pointed out as a contributor to the Great Migration (Day, 1967; Grove and Heinicke, 2003; Boustan, 2016). This paper, on the other hand, takes an alternative view and investigates how the Great Migration also transformed the Southern economy by focusing on the role of physical capital.[3]

I first lay out a simple theoretical framework motivated by macroeconomics and trade literature. The model features two regions–the South and the North, two industries–agriculture and non-agriculture, and two factors of production–labor and capital. Agriculture is assumed to be more flexible in substituting labor and capital and to be more labor-intensive, relative to non-agriculture.[4]

I interpret the Great Migration as an economy-wide change in the capital-to-labor ratio. As labor becomes relatively more scarce, agriculture substitutes now more expensive labor with capital. The increase in capital usage in agriculture also induces technical change biased toward capital,[5] further releasing agricultural workers. They are absorbed by local non-agriculture. Hence, the out-migration alone can stimulate labor reallocation across sectors. The above changes are driven by the assumption that the elasticity of substitution between labor and capital ("σ") is higher than one in agriculture but less than one in non-agriculture, consistent with the estimates based on the 20th-century United States setting (Herrendorf et al., 2015; Caunedo and Keller, 2024).

Following labor reallocation, non-agriculture is also incentivized to invest in capital due to the complementarity between labor and capital. Nonetheless, capital accumulation may not materialize if the size of the industry is constrained by local demand. The trade mechanism operates through a distinct channel: relative differences in factor intensity, measured as the factor cost share. At least in the early stage of the migration, the Southern economy can be characterized as labor-abundant compared to the North, and Southern agriculture as labor intensive (Bateman and Weiss, 1981; Whatley, 1985; Wright, 1986). Hence, the decrease in labor endowment would have led to a relative contraction of agriculture but an increase in non-agriculture production. Such a quasi-

[3]Contemporary (Raper, 1946) and later studies (Peterson and Kislev, 1986; Holley, 2000; Hornbeck and Naidu, 2014) point out the potential role of labor scarcity on agricultural mechanization in the South. Instead, I focus on its role in non-agricultural development and structural change.

[4]By adopting a constant elasticity of substitution (CES) production function, I model factor substitutability using the elasticity of substitution parameter. Factor intensity is modeled as the factor cost share, determined in equilibrium by factor prices and the CES production parameters.

[5]In agriculture, with $\sigma > 1$, an increase in the relative use of a factor induces technical change biased in favor of that factor (weak equilibrium bias, Acemoglu 2007). Intuitively, when labor and capital are easily substitutable, the economy has a greater incentive to use whichever factor that becomes more abundant more efficiently because doing so yields a higher marginal value product.

Rybczynski effect further allows non-agricultural capital accumulation through an accompanying expansion of non-agricultural production.

How much are these predictions relevant to the actual development of the Southern economy? To answer this question, I examine the economic changes in the South between 1940 and 2010, using a county-level decadal panel constructed from the Census of Agriculture (Haines et al., 2018), Population Census (Ruggles et al., 2024a,b), Economic Census (Census Bureau, 2013), and County and City Data Book (Haines et al., 2010). I estimate year-specific changes in economic outcomes after 1970, relative to their levels in 1940 and 1950,[6] between counties that experienced different levels of net out-migration.[7]

The major identification challenges are reverse causality–where the underlying forces in the South could have affected the regional out-migration–and omitted variable bias resulting from other accompanying changes in the South. To limit such concerns, the baseline strategy combines two sources of variation in predetermined shares (migration matrices between 1910 and 1940, separately for Blacks and Whites) and Northern migration pull factors as shifts (OLS in-migration prediction by race between 1940 and 1970). The aim here is to isolate variation in out-migration that is explained by what happened in the Northern destinations (migration "pull"), rather than what happened in the Southern origins (migration "push").

The constructed instrument captures the level of migration exposure to the North between Southern counties, proxying migration-driven changes in the capital-to-labor ratio. In essence, it measures how much each Southern county is connected to Northern destinations in terms of 1910-1940 migration shares that happened to have different levels of pull factors between 1940 and 1970. I primarily report the reduced-form estimates by directly regressing the outcomes by the Northern migration exposure. I also document the second-stage estimates using the county-level net out-migration rate as the endogenous variable.[8] In the first stage, one standard deviation greater exposure to Northern pull factors induces 3.6% to 4.8% higher out-of-county out-migration between 1940 and 1970.

Nonetheless, even with the SSIV strategy described above, other changes in the South may confound the results. To address this remaining concern, I include state-by-year and county fixed

[6] The baseline strategy drops the 1960 outcomes from the analysis because they may contain the influences of the migration flows in earlier periods (1940-1960). I include the 1950 values to avoid using a single year (1940) as the base. Robustness check documents the results by dropping 1950 or including 1960 outcomes in the base.

[7] To clarify the directions of the migration flows, I explicitly use the terminology in-migration and out-migration. I also use the terminologies "Blacks" and "Whites" for Americans with mainly African and European heritages.

[8] The net out-migration rate is defined to be a negative value of the out-of-county net migration rate during the 30-year period (1940-1970).

effects so that the estimation relies on variation in relative change between counties within the same state. To further account for the influence of initial differences within the same state, I include time-interacted values of the (1) time-invariant county characteristics, (2) pre-period out-migration rates (between 1910 and 1940), (3) predicted out-migration rates by race between 1940 and 1970, indexing overall levels of Southern migration push factors, (4) 1940 agriculture variables that condition the initial agricultural push factors, and (5) changes in trade exposure.

To summarize, the estimation compares the relative changes between counties within the same state with similar levels of Southern push factors and similar pre-migration characteristics, assuming that these counties would have changed the same in the absence of the differential exposure to Northern migration pull factors. I conduct robustness checks by using alternative shares, alternative shifts, and an alternative standard error, among others. I also report the results by controlling for the weighted average of the other counties' exposure as controls using either migration shares within the South or inverse squared distance as the weight.

The baseline results show that relative labor scarcity from the Second Great Migration contributed to structural change, capital accumulation, and technology adoption in the South at least until 2010. Southern counties that were more exposed to Northern pull factors released more agricultural labor and used less farmland but adopted more tractors. Farm outputs were relatively less affected. The above changes are consistent with agricultural mechanization in response to shrinking labor supply from a natural disaster (Hornbeck and Naidu, 2014) or from abrupt changes in migration policy (Clemens et al., 2018; Abramitzky et al., 2023) during similar periods in the United States. Such findings suggest that out-migration could have contributed to the rapid diffusion of tractors in the post-war South (Olmstead and Rhode, 2001), as the relative cost of labor and capital were a key determinant of the agricultural mechanization (Manuelli and Seshadri, 2014).

However, the adjustment to the out-migration did not end in agriculture. Non-agriculture results show that one standard deviation increase in the exposure to Northern pull factors raised manufacturing employment by 8.2% between 1970 and 2010, relative to its level in 1940 and 1950. There was an accompanying increase in manufacturing capital spending by 13.6%. Such changes, in turn, raised manufacturing value added and payroll by 8.6% and 15.4%, respectively. A similar development occurred in the local retail and wholesale sectors, with their employment increasing by 18.6% and 7.5%. The sales and payroll also increased in these sectors.

Year-specific estimates suggest that the increases in physical capital continued to grow or maintained at least until 2010 in agriculture and manufacturing. Such patterns can be rationalized by capital-biased technical change, with more efficient capital usage further incentivizing capital investment. The accumulation of physical capital could have complemented the overall improvements in education and human capital in the South during this period, often pointed out as a source of the North-South convergence (Caselli and Coleman II, 2001). However, at the same time, the relative labor scarcity and physical capital accumulation might have reduced the incentive for human capital investment if the out-migration raised wages regardless. The educational outcomes show that overall levels of education did not experience meaningful relative improvement or even relatively decreased in the more migration-exposed counties, suggesting that out-migration provided an additional, alternative channel on how the South caught up with the North.

In the final part of the paper, I construct a quantitative model featuring trade and migration by capitalizing on recent advancements in dynamic spatial equilibrium frameworks. (Eaton and Kortum, 2002; Artuc et al., 2010; Caliendo et al., 2019; Kleinman et al., 2023; Fan et al., 2023). The model generalizes the simple framework into multiple periods and realistic geography. It considers two sets of industries, agriculture and non-agriculture, where the latter is further divided into tradable and non-tradable sectors. All industries use two factors of production, labor and capital, with CES production structures. However, they are assumed to have different values of factor substitutability and share parameters. The structural parameters are either externally calibrated (CES production function), estimated (demand parameters), or internally calibrated using the estimated changes in agriculture and manufacturing employment (productivity parameters). The model quantification compares the baseline economy with a counterfactual scenario that prohibits migration from the South to the North during the Second Great Migration period.

The counterfactual results show that the South-to-North migration between 1940 and 1970 increased the United States' consumption welfare by 0.6% per capita by 1970, with the South experiencing a gain of 3.2%, whereas the North a loss of 0.4%. A decomposition using the welfare effects suggests that factor substitution channel played the major role, accounting for 70% of the total adjustment. The trade adjustment and directed technical change played important supplementary roles. Computationally, the adjustments to the South-to-North migration reduced the agricultural employment share by 2 percentage points by 2010 in the South, suggesting that the economics adjustments to the migration could have contributed to around 7% of the total decreases during this period. Instead, the model also predicts an increase in capital allocated to agriculture.

Related literature. This paper extends several dimensions in the economics literature. First, it extends our understanding of the impacts of out-migration and, specifically, of the Great Migration. In terms of migration, recent literature identifies the influences of out-migration on origin through labor market upgrading (Akram et al., 2018), output mix adjustments (Lafortune et al., 2015), directed technological change (Andersson et al., 2022; San, 2023), labor/capital substitution (Hornbeck and Naidu, 2014; Clemens et al., 2018; Abramitzky et al., 2023), human capital investment (Theoharides, 2018; Caballero et al., 2023), and trade integration (Egger et al., 2024). Relatedly, this paper proposes a new channel on how out-migration can lead to structural change through the origin's re-optimization of its factor usages. The findings in this paper support promoting rural out-migration as a policy tool for correcting spatial misallocation of labor and capital that is still prevalent across the world (Banerjee and Duflo, 2005; Hsieh and Klenow, 2009; Adamopoulos and Restuccia, 2014; Gollin et al., 2014).

The economic history literature has long studied the impacts of the Great Migration on migrants themselves or receiving regions (Kirby 1983; Collins 1997; Grossman 1989; Boustan 2010; Collins and Wanamaker 2014, 2015; Stuart and Taylor 2021; Derenoncourt 2022; Bazzi et al. 2023; see Collins (2021) for a review), while recent work also pays attention to the impacts on the Southern origin's political economy (Feigenbaum et al., 2020) and racial inequality and segregation (Clay et al., 2020; Chapel and Hung, 2024; Montrose, 2024). Among them, this project is closest to Hornbeck and Naidu (2014), who study the influences of the Great Mississippi Flood of 1927 on subsequent Black out-migration and agricultural mechanization until 1970. This paper, on the other hand, focuses on structural change and non-agricultural development in the post-war South.

Findings in this paper add to the structural change literature (Baumol, 1967; Caselli and Coleman II, 2001; Boppart, 2014; Porzio et al., 2022), especially studies that focus on the role of physical capital (Barro and Sala-i Martin, 1992; Acemoglu and Guerrieri, 2008; Alvarez-Cuadrado et al., 2017; Alonso-Carrera and Raurich, 2018; Caunedo and Keller, 2024). I add empirical evidence on the mechanism underlying the structural change process (Michaels et al., 2012; Fajgelbaum and Redding, 2018; Bustos et al., 2020; Dinkelman et al., Forthcoming) by documenting that labor scarcity could facilitate a "big push" out of labor-intensive equilibrium through capital accumulation. Among them, this paper is closest to Bustos et al. (2020), who show that capital accumulation from agricultural productivity improvement led to structural change in Brazil. The major difference is that I emphasize the role of capital demand from the factor substitution channel, while Bustos et al. (2020) highlight the role of capital supply from a positive agricultural income shock.

This paper also contributes to the literature examining why the American South lagged behind in economic development and how it later caught up with the rest of the United States (Whatley, 1985; Wright, 1986; Caselli and Coleman II, 2001; Grove and Heinicke, 2003; Bleakley, 2007; Depew et al., 2013; Jung, 2020). The results in this paper support the hypothesis that the abundance of labor and the lack of physical capital hampered economic advancement in the American South

(Bateman and Weiss, 1981). This paper is closest to Caselli and Coleman II (2001), who study the role of structural change in North-South convergence through a quantitative model. However, while they emphasize the importance of education and human capital, this paper stresses the role of physical capital on structural change.

Finally, the quantitative framework developed in this paper contributes to a rapidly growing dynamic spatial equilibrium literature (Eaton and Kortum, 2002; Artuc et al., 2010; Caliendo et al., 2019; Kleinman et al., 2023; Fan et al., 2023; Eckert and Peters, 2023). Compared to the existing quantitative framework that features the Heckscher–Ohlin force (Chor, 2010; Caron et al., 2014; Burstein and Vogel, 2017), I incorporate recent innovations in modeling migration (Artuc et al., 2010; Caliendo et al., 2019), capital investment (Kleinman et al., 2023), and structural change (Fan et al., 2023) with realistic geography. The quantitative model can be extended to include multiple factors of production and multiple industries that are distinguished by different factor substitutability and factor intensity.

1 Historical Background

The Great Migration, roughly dated between 1910 and 1970, was one of the largest internal migration episodes in United States history. During this period, approximately six million Blacks left the South in the pursuit of economic and educational opportunities and escaping oppressive systems symbolized by Jim Crow. Moreover, Southern-born Whites moved to the North for better living conditions and economic prospects. This White Migration even exceeded in the total number.[9] The Great Migration is often divided into the first (1910-1930) and the second flows (1940-1970), with the latter being larger in numbers. This study focuses on how the Second Great Migration shaped the economic outcomes in the South, with 1940 and 1970 as the start and end periods.

The start of the second wave coincided with the end of the Great Depression and the beginning of the Second World War, where increased labor demand in the North during the war boom and mobilization was unmet by international migration.[10] Instead, the Southern-born population started to migrate into the Northern and Western cities, leaving behind dire economic conditions at home. The migration flow continued after the war and remained at high levels until the 1960s.

[9] Although it is hard to know the exact number of migrants, Gregory (2005) calculates that more than 27 million southerners left the South either permanently or temporarily over the course of the 20th-century.

[10] International migration was largely shut down with the Immigration Act of 1924, which limited the number of immigrants allowed entry into the United States through a national origins quota. The Bracero program (1942-1964) is a notable exception during this period.

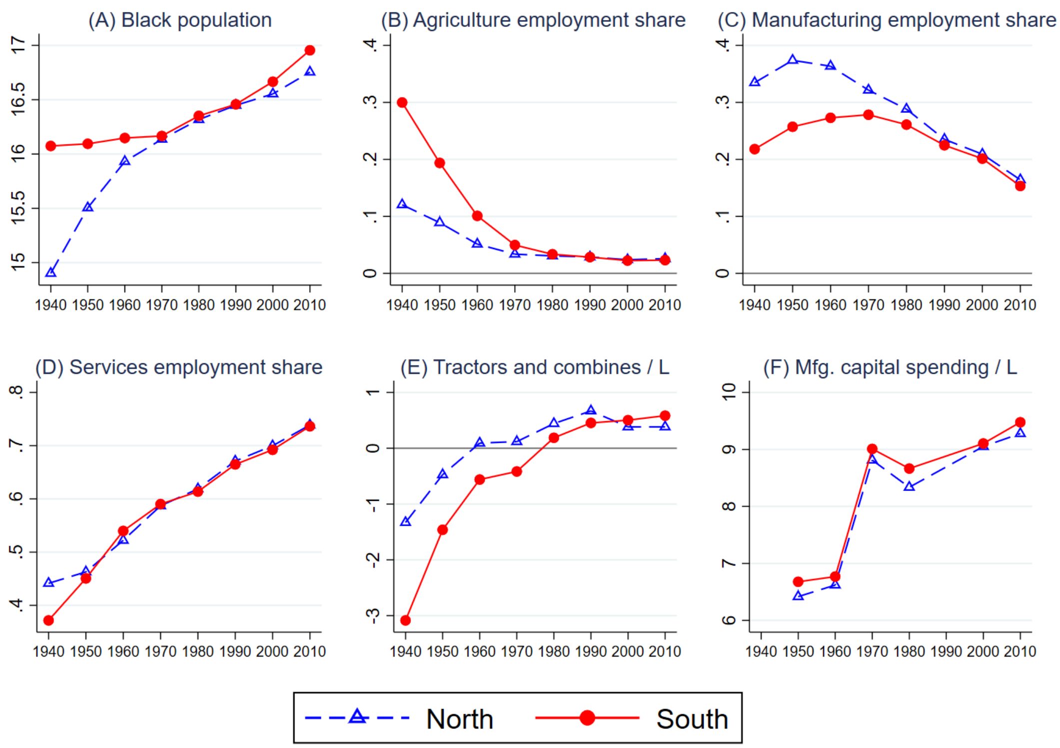

Figure 1: Great Migration and economic changes in the South.

Note: The figure presents time trends in the selected variables for the North and the South. The values in Panels A, E, and F are logged. The definition of the region is described in Footnote 11, where "the North" encompasses both Northern and Westerns states. Panel A, the number of the Black population, is calculated using Haines et al. (2010). Panels B to D, the industry employment shares, are calculated from 1% and 5% samples of the Population Census and American Community Survey (Ruggles et al., 2024b). The industry definition is based on the 1950 Census Bureau industrial classification system. Agriculture consists of agriculture, forestry, and fishing but excludes mining. Panel E, the number of tractors and combines per agriculture employment, is calculated from Haines et al. (2018). Panel F, the manufacturing capital per manufacturing employment, is calculated from Haines et al. (2010).

Two important changes in the U.S. could have contributed to the slowing flows after the 1960s. The Civil Rights Act of 1964 prohibited discrimination in public places and made living conditions in the South less harsh for Blacks. Another important change was the Immigration and Nationality Act of 1965, also known as the Hart-Celler Act. The act was enacted in 1967 and abolished the quota system, significantly increasing international migration flows thereafter and until today.

Figure 1, Panel A, presents the log Black population in the South (red line) and the North (blue dashed line).[11] The steep convergence between 1940 and 1970 gives a hint of the sizes of migration flows between the two regions, where the share of Blacks living in the South decreased

[11]The definition of the South mainly follows the Census definition. It contains contain former Confederate states (South Carolina, Mississippi, Florida, Alabama, Georgia, Louisiana, Texas, Virginia, Arkansas, Tennessee, and North Carolina), Delaware, Kentucky, Maryland, Oklahoma, and West Virginia, but excludes the District of Colombia.

from 76% to 51% during the 30-year period. There was accompanying out-migration of Whites: by 1940, 11% of Southern-born Whites lived outside the South, while the share increased to 20% by 1970 (Bazzi et al., 2023). However, the flow of migrants plateaued and reversed afterward; many gradually returned, and new migrants entered the South. Panel A hints that the share of Blacks in the South again increased after 1990. This new migration flow is often dubbed the New Great Migration.

While many Southerners left their origin in pursuit of economic opportunity, the South's economy also matured during the same period. The next three Panels plot time trends on employment share in three major industries by region. Panels A through D highlight that the Southern economy rapidly caught up with the rest of the United States while simultaneously experiencing large out-migration. By 1940, around 30% of employed Southerners were working in agriculture, forestry, and fishing, compared to 12% in the North. However, coinciding with the Second Great Migration, the two regions converged in terms of industry employment share at least by 1990.

Did out-migration induce regional convergence, or did the process of structural change generate out-migration? One persuasive and pervasive explanation is that agricultural development and mechanization drove the out-migration of workers into the North and into manufacturing (Grove and Heinicke, 2003; Boustan, 2016). It is natural that the agricultural development would have simultaneously decreased employment and led to out migrations. However, the influences may not be uni-directional, as labor scarcity from out-migration could have also incentivized endogenous responses in the Southern economy.

Panels E and F present proxies for the changes in the capital-to-labor ratio in both agriculture and manufacturing. I measure agricultural mechanization using the number of tractors and combines[12] divided by the number of agricultural workers. For manufacturing, I use the reported manufacturing capital spending.[13] By 1940, Southern agriculture was much less mechanized and relied on labor-intensive practices, often dubbed as the "Old South" method of production (Wright, 1986). However, Panel E suggests that the South eventually caught up with the North in terms of mechanization.[14] Two channels explain the agricultural convergence: relative increases in the

[12]Census of Agriculture started to collect the number of tractors in 1925, but combines in 1950. Hence, the 1940 value only includes the number of tractors.

[13]Census of Manufactures did not collect manufacturing capital stock during the early 20th-century but started to report manufacturing capital spending in 1947 with a few exceptions (e.g. 1992). The value for 1990 is missing.

[14]Historical account of the spatial diffusion of tractors accords with the trends in Panel E. By 1920, the absolute majority of American farms depend on horses and mules for power. Then initially, tractors were adopted in the Wheat Belt states (North Dakota, South Dakota, and Kansas) during the 1920s and diffused to the Corn Belt states (Iowa, Illinois, and Nebraska) during the 1930s (Gross, 2018). In the South, tractors was rapidly employed in the post-World

number of tractors and combines (numerator) and relative decreases in employment (denominator). Although the decreases in employment would have contributed to the increase in the trend, the South may have also invested more in mechanization, as in other historical episodes in the United States (Hornbeck and Naidu, 2014; Clemens et al., 2018). Empirical analysis examines both channels.

Furthermore, Panel E shows that the Southern manufacturing invested more in capital per worker relative to the North. If the initial capital-to-labor ratio in the South had been low by 1950, the higher spending per worker in the South would have resulted in relative convergence in the manufacturing capital per worker. Note that if workers simply reallocated from agriculture to manufacturing in the South, capital per worker would have likely to be lower in the South. Instead, relatively higher capital spending suggests that the economy might have responded to labor scarcity by raising capital investment. This paper examines the relationship between large out-migration (Panel A) and regional convergence (Panels B and C). I especially focus on the potential role of capital deepening (Panels E and F) as a response to labor scarcity from out-migration.

2 A Model of Migration and Structural Change

This section outlines a simplified model of a two-period, two-country small open economy framework with two industries and two factors of production—labor and capital. I start with a static, closed-economy version of the model and subsequently add an additional period and another region to study dynamic and open-economy implications. The closed-economy model highlights the importance of factor substitutability and factor-augmenting technical change, whereas open-economy predictions emphasize the role of the Heckscher-Ohlin force in trade specialization through relative factor abundance and factor intensity. The aim is to study the major mechanisms that can induce structural change in a model with two factors of production. The interpretation introduced here can be applied to another setting where the initial economy contains a sufficient share of the workers in labor-intensive agriculture.

I proceed by interpreting well-established theoretical results in relation to the out-migration and subsequent economic development. Specifically, I take the out-migration in the first period as given (the Great Migration) and analyze the implication of resulting relative labor scarcity on the economy and structural change.[15] Subsection 2.4 summarizes the core predictions and discusses

War 2 periods (Olmstead and Rhode, 2001).

[15] Empirical analysis is designed to replicate "out-migration as given," at least in the perspective of the South, using

the endogenous migration in the second period. The dynamic spatial general equilibrium model in Section 5 generalizes the model elements into realistic geography and multiple periods. I implicitly assume a constant consumption share with Cobb-Douglas utility in this section. I allow non-homothetic preferences for the quantitative model (Section 5).

2.1 Agriculture versus non-agriculture

The production function for both agriculture and non-agriculture (denoted "a" and "m") is assumed to take the constant elasticity of substitution (CES) structure using labor L^s and capital K^s in each sector s (Arrow et al., 1961; David and van de Klundert, 1965):[16]

$$Y^s = \left(\rho^s (Z_L^s L^s)^{\frac{\sigma^s - 1}{\sigma^s}} + (1 - \rho^s)(Z_K^s K^s)^{\frac{\sigma^s - 1}{\sigma^s}} \right)^{\frac{\sigma^s}{\sigma^s - 1}}, \tag{1}$$

with labor- and capital-augmenting technologies, Z_L^s and Z_K^s. They are assumed to not exogenously grow but are endogenously affected by the changes in factor allocation. In other words, the production function abstracts from Hicks-neutral and the exogenous components of the factor-augmenting technologies, as they are not needed for the core predictions.[17,18] For simplicity, I assume that technology cannot adapt in the first period when the shock occurs but can endogenously adjust in the second period through the directed technical change process (Acemoglu, 2002, 2007).

The CES production function allows for flexible factor usage with a restriction that the elasticity of substitution between labor and capital, σ, is constant. The value of σ is assumed to be greater than one for agriculture but less than one for non-agriculture by following CES elasticity estimates

shift-share instrumental variable (SSIV) strategy by isolating pull factors of Northern destinations.

[16] Appendix Section 1 adds land as another factor of production in agriculture. Land and capital, respectively, would represent geographically immobile fixed and variable factors of production, while labor can be viewed as a geographically mobile variable factor. I abstract from land in the main model as the addition does not make much difference in terms of core predictions. Still, I outline an additional implication for strongly labor-saving economic development (Acemoglu, 2010) that the introduction of land can generate.

[17] In the real economy, technological advances would consist of both Hicks-neutral and factor-augmenting (non-neutral) components, the latter of which could be either exogenous or endogenous. With Cobb-Douglas demand, the growth of Hicks neutral technology does not affect any predictions of the model, as long as its growth rate is the same across industries. The differences in technical growth between industries, on the other hand, lead to the classic Baumol (1967) effect. See Duernecker et al. (2023) for a recent theoretical treatment. The same growth rate of factor-augmenting technologies within an industry is equivalent to Hicks-neutral technology. However, differences in growth rates can act as another source of structural change. See Alvarez-Cuadrado et al. (2017) for related results.

[18] Empirical analysis aims to be consistent with such an abstraction through the parallel trend assumption. In other words, the strategy assumes that regions with different levels of out-migration changed the same in terms of Hicks-neutral technology and the exogenous components of factor-augmenting technologies. Section 3 illustrates the approach. The quantitative model in Section 5 allows Hicks-neutral technology and the exogenous components of the factor-augmenting technologies. Still, the baseline quantification focuses on the endogenous component.

in the literature. For instance, Herrendorf et al. (2015) estimate the value of σ for agriculture (1.58), manufacturing (0.80), and service (0.75), using the U.S. macro data for 1947-2010.[19] Oberfield and Raval (2021), by focusing on manufacturing, estimate the elasticity parameter at plant-level (0.3-0.5) and macro-level (0.5-0.7) using the U.S. Census of Manufactures for 1972-2007. Using global panel datasets, Boppart et al. (2023) estimate elasticity in contemporary agriculture to be 1.90. Hence, agriculture can be regarded as a flexible sector, while non-agriculture is relatively inflexible in factor usage.[20]

The factor intensity is defined as the cost share of each factor. The share parameter ρ measures the production weight and influences the relative importance of labor in production. As the elasticity of substitution becomes unity, the share parameters become the Cobb-Douglas exponents and solely determine the factor intensity. With the CES production, the cost share is determined in equilibrium by factor prices and the elasticity and share parameter.[21]

2.2 Closed-economy force: Factor substitutability

I start with a closed economy implication for the South, which highlights the role of factor substitutability. Due to the differences in flexibility in combining labor and capital, the two factors reallocate across industries in the opposite direction from the common shock. The non-unitary elasticities also give rise to weak equilibrium biases in technological development.

As the population flows out, labor becomes more scarce and expensive relative to capital. Here, I document how the share of capital allocated to the agriculture, $\kappa = K^a/(K^a+K^m)$, and labor share in agriculture, $\lambda = L^a/(L^a+L^m)$, responds to the change in regional capital-to-labor ratio, $k = (K^a+K^m)/(L^a+L^m)$. Note that the initial industry of the migrants is irrelevant to the changes in k. I assume that both factors are fully employed and perfectly mobile across sectors within the region. I interpret capital as mobile across sectors but a geographically immobile variable factor, such as local structures for production.

Predictions 1 and 2 reinterpret Alvarez-Cuadrado et al. (2017) and amend their results to examine the changes from out-migration and endogenous technology adoption. In order to obtain

[19]More recently, Caunedo and Keller (2024) estimate the value of σ for agriculture (1.23), manufacturing (0.84), and service (0.74) between 1948-2020 using the methodology of Herrendorf et al. (2015).

[20]Since Arrow et al. (1961), the CES elasticity estimates on the U.S. have tended to report a value significantly less than one for the aggregate economy or non-agriculture (see Chirinko (2008) and León-Ledesma et al. (2010) for a review). Notable exceptions include the estimated value of 1.25 in Karabarbounis and Neiman (2014) and the range [1.3, 1.6] in Piketty (2014), who investigate the sources of the labor share decline.

[21]Specifically, the labor cost share ξ is given as $\xi = \rho^\sigma w^{(1-\sigma)}/(\rho^\sigma w^{(1-\sigma)} + (1-\rho)^\sigma r^{(1-\sigma)})$, with the wage rate w and the rental rate of capital r.

analytical results, I use $\sigma_A > 1$ but set $\sigma_M = 1$. The same result can be obtained with $\sigma_A = 1$ and $\sigma_M < 1$. The results would only be strengthened with $\sigma_A > 1$ and $\sigma_M < 1$.

Prediction 1 (Static response)**.** *Assume that the elasticity of substitution between labor and capital for the flexible sector (agriculture) is greater than one, $\sigma_A > 1$, while the elasticity of the inflexible sector is equal to one, $\sigma_M = 1$. As the economy-wide capital-to-labor ratio, k, increases, the fraction of capital allocated to the more flexible sector (agriculture) increases, while the fraction of labor decreases. In particular,*

$$\frac{\partial \kappa}{\partial k} = \frac{(1-\sigma)}{\sigma G(\kappa) k} > 0$$
$$\frac{\partial \lambda}{\partial k} = \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{\lambda(\kappa)}{\kappa}\right)^2 \frac{\sigma - 1}{\sigma G(\kappa) k} < 0.$$

where $G(\kappa) \equiv \left[\frac{1}{\sigma(1-\lambda(\kappa))} + \frac{1}{\lambda(\kappa)}\right]\left(\frac{\lambda(\kappa)}{\kappa}\right)\left(\frac{\alpha}{1-\alpha}\right) + \left[\frac{1}{\kappa} + \frac{1}{\sigma(1-\kappa)}\right].$

Proof. See Online Appendix A. □

Prediction 1 focuses on factor reallocation channel. It clarifies how out-migration leads to an increase in capital allocated to agriculture while also inducing structural change out of agriculture. As labor becomes scarcer, the flexible sector substitutes now more expensive labor with capital, releasing labor and absorbing capital. Due to $G(\kappa)$, labor and capital shares always take the opposite direction from the changes in the capital-to-labor ratio.

Now, I posit a second period that allows endogenous technology adoption and capital investment. I use the prime notation ($'$) to denote the second period. The model abstracts from capital depreciation. First, I introduce the following remark:

Remark 1 (Remark on Prediction 1)**.** *Assuming that the South optimizes its levels of technology, the direction of technical change would exhibit weak equilibrium bias as follows (Acemoglu, 2007):*

$$\frac{d(Z_K^{s\prime}/Z_L^{s\prime})^{\frac{\sigma^s-1}{\sigma^s}}}{d(K^s/L^s)} > 0, \tag{2}$$

where the term Z_K^s/Z_L^s represents the relative level of capital- to labor-augmenting technology.

In other words, an increase in the sectoral capital-to-labor ratio induces technological change biased toward capital or labor depending on the value of σ. Remark 1 is equivalent to imposing additional assumptions on the technology environment as in Acemoglu (2007).[22] The CES function meets the required assumption for the production side. The technical changes can be thought of as generated by learning-by-doing (Arrow, 1962) or directed R&D efforts (Kennedy, 1964). Examples of learning by doing are Southern farmers becoming more proficient at operating farm machinery and local mechanics gaining expertise in repairing them. Directed R&D efforts include enhancing farm machines to better suit Southern crops.

Two competing forces can influence the direction of technical change: price effects that are biased toward scarce factors and market size effects that benefit abundant factors. Given that labor and capital are gross substitutes in agriculture, increases in capital usage raise the relative profitability of capital-augmenting technology with $\sigma > 1$. Hence, it becomes more profitable for the economy to focus on using capital more efficiently. On the contrary, the price effects dominate in non-agriculture as two factors are gross complements. Note that Prediction 1 anticipates an increase in labor allocated to non-agriculture. Because both the value of elasticity and factor allocation take the opposite direction for non-agriculture, weak equilibrium bias would again favor capital in non-agriculture. Such technical changes further raise the economy-wide capital-to-labor ratio, leading to the following Prediction for the second period:

Prediction 2 (Dynamic response)**.** *Assume that the elasticity of substitution between labor and capital for the flexible sector (agriculture) is greater than one, $\sigma_A > 1$, while the elasticity of the inflexible sector is equal to one, $\sigma_M = 1$. With technology in both sectors exhibiting weak equilibrium bias, the fraction of capital allocated to the more flexible sector (agriculture) increases while the fraction of labor decreases.*

To sum up, the increase in the capital-to-labor ratio from the out-migration dynamically incentivizes the adoption of capital-augmenting technology, further raising the capital-to-labor ratio, k, in the second period. It again leads to the reallocation of labor from agriculture to non-agriculture. Hence, the simple framework features the out-migration as a source of "big push" as an endogenous outcome.

2.3 Open-economy force: Factor intensity

I introduce another region, the North, to investigate the open economy implications. Factor substitutability, the focus of the closed-economy predictions, does not yield direct implications for the

[22] Acemoglu (2007) lays out a menu of different assumptions that are unrelated to, but needed in addition to, the above production structure that can lead to equilibrium bias of technology. For instance, there could be a technologist monopolist that supplies technologies to good producers through supplying intermediate goods.

open economy because trade depends on relative comparison to the North. However, a potential tension between labor and capital from factor intensity can give rise to the Heckscher-Ohlin force. The key idea is that relative factor abundance can determine the regional comparative advantage, and hence, the pattern of trade and production.

I introduce additional assumptions regarding factor abundance and factor intensity: the South is labor-abundant, characterized by a lower capital-to-labor ratio compared to the North (Figure 1, Panel E), and the Southern agricultural production is intensive in labor, with labor accounting for a larger share of production costs (Bateman and Weiss, 1981; Wright, 1986). In other words, labor is relatively cheaper in the South, giving it a comparative advantage in agriculture. I also assume that the relative factor intensity is not reversed from the factor reallocation. The resulting Prediction 3 states the quasi-Rybczynski effect (Romalis, 2004):[23]

Prediction 3 (Open economy)**.** *At constant relative goods prices, the decrease in the labor endowment in the South in the first period leads to a relative contraction of the agricultural sector and a relative increase in non-agriculture production. Non-agriculture shares of labor and capital increase.*

The related proof and discussion are in Online Appendix Section A. The results follow from the changed pattern of comparative advantage. As the South lost labor, its comparative advantage in the labor-intensive sector decreased, incentivizing the economy to reallocate resources toward the non-agricultural sector, which had relatively gained a comparative advantage.

The open-economy forces predict a decrease in labor share in agriculture, as in the closed-economy forces. However, the open-economy forces relatively expand non-agriculture while shrinking agriculture, which has different implications for capital allocation. I summarize the common and competing predictions from these different perspectives in the next subsection. In the second period, economy-wide capital accumulation would lead to a further expansion of non-agriculture with a relative decrease in agriculture production.

2.4 Discussion

Summary of the predictions. Closed- and open-economy results rely on related but distinct assumptions. The closed-economy perspective focuses on the differences in labor-capital substi-

[23]When a factor endowment increases, the Rybczynski theorem (Rybczynski, 1955) predicts a more than a proportional expansion of a sector which uses that factor more intensively through a magnification effect (Jones, 1965). However, his sharp prediction relies on factor price equalization, which may not hold in practice. Instead, I state a weaker version that requires a weaker set of assumptions.

tutability and factor reallocation across sectors, whereas the open-economy view relies on the differences in factor intensity between the industries and the changes in comparative advantages.

First, there are common and non-competing predictions. In the first period, both types of models predict the reallocation of labor from agriculture to non-agriculture. The trade channel raises non-agricultural production. Although the closed-economy model assumes that the industry share is fixed by the consumption share, there is no force from the closed economy that works against the relative expansion of the non-agricultural sector. In the second period, the direction of technical change predicts a relative improvement in capital efficiency, which could result in capital accumulation and further expansion of the non-agriculture sector.

However, there are notable differences in terms of agricultural capital and production. In the first period, the increase in the economy-wide capital-to-labor ratio induces capital adoption in agriculture, while trade effects lessen it. Which effects dominate depends on the relative strength of the closed-economy force (factor substitutability) and the quasi-Rybczynski effect (factor intensity). However, note that even if the agricultural share of capital in the economy decreases, the sectoral capital-to-labor ratio in agriculture would still increase with its elasticity of substitution greater than one. It will result in capital-biased technical change and capital accumulation, which may, in turn, increase agricultural production in the long run.

If agricultural output decreases, it favors the Heckscher–Ohlin channels, and one could expect accompanying decreases in agricultural capital. On the other hand, increases in agricultural capital are more consistent with closed-economy prediction, and hence, agricultural output is likely not to experience much change or even increase. In either case, with capital-biased technical change, the agricultural capital stock would progressively increase as time passes.

The empirical analysis first checks the common and non-contradictory predictions. To be consistent with both frameworks, labor should reallocate from agriculture to non-agriculture, and non-agriculture production and capital should increase. On the other hand, the changes in agricultural capital and production are left as empirical questions. I then use quantitative analysis to assess the potential contribution of each component based on the changes in model outcomes.

Endogenous migration in the second period. The model takes the first period Southern out-migration as given, but it can allow endogenous migration in the second period. In the view of the standard migration settings (e.g. Rosen-Roback framework), the wage increases raise the value of living in the South in the second period, holding migration costs and amenities constant. Thus, the out-migration in the first period itself would decrease gross migration flow from the South to the North and increase the flow in the opposite direction. Still, the net effect is ambiguous and depends on the exact value of living in each region. However, if structural change and capital accumulation sufficiently raise the Southern wage, the net migration flow could be reversed in the second period. In addition to the role of return migrants, this is another channel that the first period out-migration could incentives the second period in-migration.

3 Empirical Strategy

3.1 Data

First, I use the complete-count Census between 1910 and 1940 (Ruggles et al., 2024a,b) to generate county-level variables and to construct county-to-county level transition matrices for the shift-share design. I use two sets of migration matrices, separately constructed for Blacks and Whites. The baseline migration share uses matched individuals between 1910 through 1940 using the Census Tree approach (Buckles et al., 2023). The Census Tree capitalizes on manual matches created by people doing research on their own family histories using FamilySearch.org. Buckles et al. (2023) then extend these linkages using both traditional and machine learning matching strategies. The datasets provide the largest matches among publicly available methods and also provide links for women and Blacks. I also document the robustness in terms of the share using the 1940 Census, which asks for state and country of residence 5 years ago ("MIGCOUNTY"). I use this information to construct an alternative migration matrix between 1935 and 1940.[24]

Secondly, I use county-level datasets: "*Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*" (Haines et al., 2010), henceforth HDES, and agriculture Census compiled in Haines et al. (2018). HDES contains county and state-level information on agriculture, manufacturing, retail, and wholesale activity and various county characteristics. "*United States Agriculture Census, 1840 - 2012*" (Haines et al., 2018), henceforth Agriculture Census provides detailed county-level agricultural information, such as farm output and acres harvested.

I supplement the datasets from Census Bureau (2013) for non-agricultural outcomes after 2002 and the County Business Patterns (Eckert et al., 2022; Census Bureau, 2023), henceforth CBP, for sub-sectors in services. The CBP contains information related to employment and economic activity for detailed industry codes covering all counties in the United States. Given the long

[24] For the migration between counties in the contiguous U.S., there are 9,622,404 ($= 3102 \times 3102$) possible combinations, including stayers. The Census Tree approach between 1910 and 1940 generates 1,748,472 (18.2%) non-zero migration cells among all possible flows, while the 1940 Census approach between 1935 and 1940 generates 830,892 (8.6%) non-zero cells.

study periods, I match the aggregate values from the above datasets using time-series data from the "*Historical Statistics of the United States*" (Carter et al., 2006) whenever possible.

The main sample is 1,096 counties in the South between 1940 and 2010. The definition of the South mainly follows the Census definition. The sample states contain former Confederate states (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, Virginia), Delaware, Kentucky, Maryland, Oklahoma, and West Virginia, but excludes the District of Colombia. I denote counties in these states as "the South," while I use the terminology "the North" to denote all counties in the contiguous U.S. outside the South. Outliers in terms of the top and bottom 1% for both the 1940-1970 out-migration rate (endogenous variable) and Northern migration exposure (excluded instrument) are excluded from the analysis to avoid distortion caused by extreme values. I restrict the sample to balanced counties for agriculture and manufacturing dependent variables, except for the number of combines and manufacturing capital spending that started to be collected in 1950 and have less coverage. I also report the results by limiting the sample to balanced counties for all main outcome variables or by limiting the sample to former Confederate states. Note that even within the balanced counties, the number of observations can differ between variables as some variables are not reported in certain years.

The values from different datasets are linked to the closest decadal year. For instance, the Census of Agriculture is taken every five years and was conducted in 1997 and 2002. For 2000 values, I take an average of the two nearby values. The CBP, on the other hand, is conducted annually and its 2000 values are matched to the year 2000. The county border across different years is adjusted to 1990 boundaries (Eckert et al., 2020). Further details on data are described in the Online Appendix Section B.

3.2 Estimating equation

Empirical analysis examines the relationship between out-migration between 1940 and 1970 with economic changes in the South after 1970, relative to their values in 1940 and 1950. The identification relies on the parallel-trend assumption: counties with different levels of out-migration would have changed the same. While this assumption is unlikely to hold, I present the modified assumption after introducing the shift-share design and control variables.

Conceptually, the baseline specification would estimate year-specific differences between counties with different levels of net out-migration rates between 1940 and 1970, "$\left(Net\_Out\_Mig^{1940-1970}\right)_c$"

$$Y_{c,t} = \beta_t \left(Net\_Out\_Mig^{1940-1970}\right)_c + \alpha_{s,t} + \alpha_c + \gamma X_{c,t} + \gamma_t X_c + \varepsilon_{c,t}. \tag{3}$$

The main regressor summarizes the out-of-county out-migration at the county level during the Second Great Migration period. It is calculated as the negative value of the number of net migrants between 1940 and 1970, divided by the 1940 population. The number of net migrants is calculated from decadal net migration rate estimates (Gardner and Cohen, 1992; Voss et al., 2005; Fuguitt et al., 2010; Winkler et al., 2013; Bowles et al., 2016). I estimate a pooled version of Equation (3) as a difference-in-differences estimator: it compares the outcomes before and after the Second Great Migration between counties that experienced different levels of out-migration. For dependent variables, I focus on the changes in employment and measures of outputs and capital in agriculture, manufacturing, and wholesale and retail.

The omitted base years are 1940 and 1950. Hence, the β_t captures the changes in outcome $Y_{c,t}$ relative to its levels in 1940 and 1950.[25] Compared to using only 1940 values as an initial condition, using both decadal years would minimize the influences of potential changes related to World War II during the 1940s (Jaworski and Yang, 2024). I drop the 1960 outcomes from the analysis because they may partially contain the influences of the migration flows in earlier periods (1940-1960). Online Appendix documents the results by either excluding 1950 or incorporating 1960 outcomes.

I include state-by-year fixed effects ($\alpha_{s,t}$) to account for national- and state-level trends and county fixed effects (α_c) to remove county-level time-invariant unobservable factors that may confound the results. Hence, the identifying variation uses the differences in the changes in outcome, relative to each county's base value, between counties that experienced different levels of out-migration rates within the same state in the same year.

3.2.1 Shift-share instrument design

The main identification challenge is that the dependent variables could have reversely affected the out-migration rate in Southern county c. To ameliorate the concern, I limit the attention to the component explained by Northern pull factors:

[25] The observations include values in 1940, 1950, 1970, 1980, 1990, 2000, and 2010. A set of estimates $\{\hat{\beta}_{1970}, \cdots, \hat{\beta}_{2010}\}$ captures relative effects in 1970, $\cdots$, 2010, compared to its average value between 1940 and 1950.

$$Y_{c,t} = \beta_t \left(Northern\_\widehat{Exposure}^{1940-1970}\right)_c + \alpha_{s,t} + \alpha_c + \gamma X_{c,t} + \gamma_t X_c + \varepsilon_{c,t}, \tag{4}$$

by replacing the main regressor to "$\left(Northern\_\widehat{Exposure}^{1940-1970}\right)_c$", a standardized measure that summarizes the Northern migration pull factors experienced by Southern county c between 1940 and 1970. This measure utilizes a shift-share instrumental variable (SSIV) strategy that combines two sources of variation through fixed, predetermined migrant networks ("share") with pull factors of receiving cities ("shifts"). In the Great Migration setting, Boustan (2010), Derenoncourt (2022), and Bazzi et al. (2023) construct predicted shifts based on push factors of Southern origins to instrument the number of in-migrants to Northern destinations. I instead calculate migration exposure in the Southern origins using predicted shifts based on Northern pull factors, by leveraging that the Great Migration was driven by a combination of push and pull factors (Collins, 1997).

In the zero-stage regression of in-migration prediction in Equation (5), I use OLS regression with latitude, longitude, log values of the total, Black, and White population, urbanization, median income, median rents, total housing units, 1940 values of the share of foreigners, Black and White, 1940 values of employment share and median occupational score, and Republican vote share of presidential elections between 1940 and 1972, selected to represent demographic, economic, social, and political environments in the North:

$$(in\_mig\_rate)^r_{d,t} = f(pull\_factors_{d,t-10}) + \varepsilon_{d,t}, \tag{5}$$

where I separately predict Black and White in-migration (Table A1). I also report the results with alternative shifts using actual in-migration rates as in Card (2001) and predicted rates using random forest algorithm, an ensemble machine-learning technique based on the decision tree method.

Then, I allocate the predicted number of migrants back to Southern counties using pre-period migration share matrices,[26] $\omega^{r,1910-1940}_{do}$, between 1910 and 1940 as in Equation (6):

$$(out\_\widehat{migrants}^{1940-1970})_o = \sum_{t=1950}^{1970} \sum_d \sum_r \left((in\_\widehat{migrants})^r_{d,t} \cdot \omega^{r,1910-1940}_{do} \right), \tag{6}$$

where o and d index origin and destination, and r stands for race $\in \{Black, White\}$. The shifts and shares are constructed separately for the Black and White migration to take into account hetero-

[26] The practice of using pre-period migration shares capitalizes on empirical regularity dubbed the "chain migration." During the Great Migration era, Stuart and Taylor (2021) document that for every one randomly selected Black Southerner who moved to a Northern destination county, 1.9 additional Black migrants made the same move.

geneous migration patterns by race (Collins and Wanamaker, 2015). I restrict the shares to only include migration flows between the North and the South. Hence, each Southern county is assigned the same one unit of the total Northern linkage. The instrument measures how much of this linkage is allocated to Northern counties that experienced relatively higher levels of in-migration between 1940 and 1970. The baseline shares use matched individuals using the Census Tree approach (Buckles et al., 2023; Ruggles et al., 2024a). I also report the results using the 1935-1940 migration matrices constructed from direct data on 1935 locations in the 1940 Census.

Dividing the allocated out migrants by the initial 1940 population yields the predicted out-migration rate between 1940 and 1970 in Southern origin county o as in Equation (7). Finally, I standardize the predicted out-migration rate and denote the resulting measure as the Northern pull factor exposure. For simplicity, I also refer to it as Northern exposure or migration exposure.

$$(\widehat{out\_mig\_rate}^{1940-1970})_o = \frac{(\widehat{out\_migrants}^{1940-1970})_o}{(population^{1940})_o}, \tag{7}$$

$$standardize\Big((\widehat{out\_mig\_rate}^{1940-1970})_o\Big) \equiv \widehat{Northern\_Exposure}^{1940-1970}{}_o. \tag{8}$$

The constructed Northern exposure quantifies the levels of migration linkages of a Southern county to Northern destinations that are predicted to receive different levels of migration during the 1940 and 1970 periods. For instance, while Mississippi and Louisiana are adjacent, counties

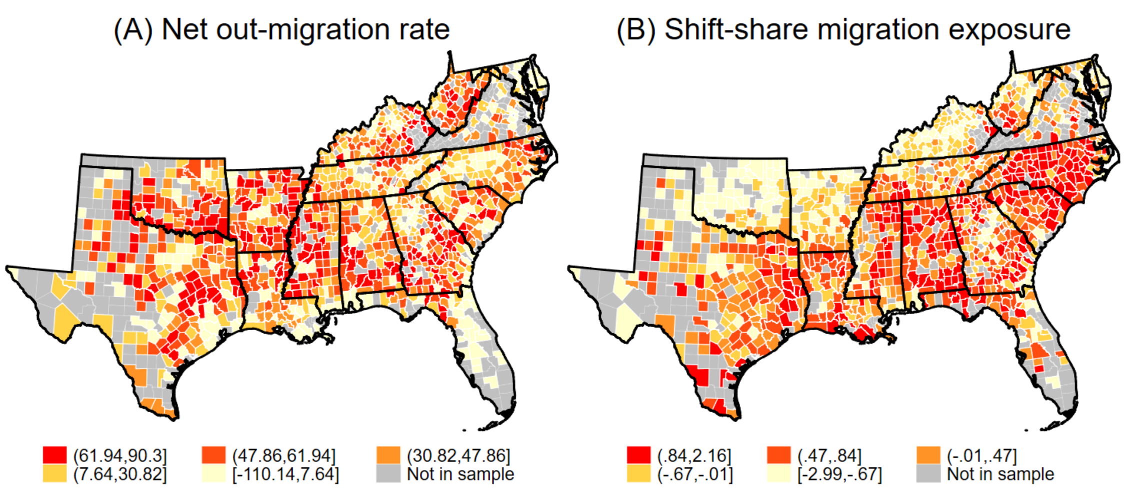

Figure 2: Map of the out-migration rate and SSIV predicted migration exposure (1940-1970).

Note: The figure presents the geographical distribution of the endogenous variable, the actual out-migration rate between 1940 and 1970, on Panel A, and the excluded instrument, the predicted Northern pull factors-based migration exposure, on Panel B. Both variables are not residualized. The instrument is constructed by the SSIV design described in Section 3.2.1. Red indicates greater levels of net out-migration and Northern migration exposure, relative to yellow.

in Mississippi tend to have higher migration linkages to Chicago. Mississippian migrants often traveled on the Illinois Central Railroad, which passed through the state and ended in Chicago. On the contrary, many Louisianans migrated to Los Angeles, linked through the Sunset Limited Train. The SSIV design measures the strength of these linkages through the pre-period (1910-1940) migration matrices, which encompass transportation linkages, the degree of migration enclaves, or any fundamentals that could have influenced the migration between 1940 and 1970. If L.A. is predicted to receive relatively more migrants than Chicago in the zero stage prediction, the Northern exposure in Louisianan counties ("parishes") would be higher. Even within the state, counties with higher pre-period linkages to Los Angeles would be assigned greater exposure.

Figure 2 plots the map of the non-residualized values of the endogenous regressor (actual out-migration rate) on Panel A and the excluded instrument (predicted Northern migration exposure) on Panel B. Red indicates greater levels of net out-migration and migration exposure, relative to orange and yellow. Sample selection criteria are described in Section 3.1.

3.2.2 Control variables

As a set of time-varying controls, $X_{c,t}$, I use the log population and contemporaneous net migration rate. They play important roles in taking into account the changes in county sizes and the potential role of migration after 1970 that may be correlated with the migration flows. However, one might be concerned about the presence of contemporaneous variables as they would have been affected by the main explanatory variable itself. Hence, I also document the results without them.

Even within the same state, counties with different initial conditions before 1940 may not have changed the same after the Great Migration period. Hence, I include extensive sets of time-interacted variables to ensure that the estimation relies on comparisons between counties in the same state with similar pre-migration characteristics and with similar levels of Southern push factors. Specifically, I include time-interacted values of (1) time-invariant county characteristics, (2) pre-period migration rates, (3) predicted out-migration rates that summarize the Southern push factors, (4) 1940 agriculture variables capturing agricultural push factors, and (5) trade exposure.

First, I include log land area, longitude, latitude, and 1940 values of log population. They adjust for time-varying effects of initial county sizes and basic suitability for agriculture. Next, I include time-interacted values of agriculture variables, given that agriculture practices in the United States continued to be developed during the 20th-century. For instance, during this period, cotton production rapidly mechanized, while tobacco production continually declined, where both of them had been traditional cash crops in the South (Whatley, 1985; Holley, 2000; Jung, 2020). To minimize the role of such "agricultural push factors," I use 1940 values of the share of sharecroppers,[27] total farm acres, and acres harvested in cotton, tobacco, corn, and hay, respectively.[28] I also include the shares of farms in five different farm-size bins, as initial farm size may have influenced the adoption of tractors and combines or other agricultural practices (Grove and Heinicke, 2003; Manuelli and Seshadri, 2014).

I add an extensive list of decadal migration rates for pre-period out-migration (for three 1910-1920, 1920-1930, and 1930-1940 flows) and predicted rates during the treatment period (for 1940-1950, 1950-1960, and 1960-1970 flows, by race). They are aimed to capture any regional fundamentals that are expected to drive residents out of the county. In other words, the control variables summarize the "Southern push factors," whereas the excluded instrument is constructed to proxy for the "Northern pull factors." The pre-period out-migration rates are calculated from the Census Tree approach (Buckles et al., 2023; Ruggles et al., 2024a). For predicted out-migration between 1940 and 1970, I use the value predicted by Derenoncourt (2022) for Blacks and the zero-stage predicted value for Whites (Equation (5) and Table A1). I also report the results by dropping the predicted out-migration rates.

A control for the trade exposure uses the Japanese import penetration measure from Batistich and Bond (2023), as the Japan shock is the most relevant to the main study period.[29] Regarding international migration, first note that the contemporaneous net migration rate includes both internal and international migration. Furthermore, between the Johnson-Reed Act of 1924 and the Immigration and Nationality Act of 1965, U.S. international migration was largely restricted with a quota system. An important exception during the Great Migration period was the Bracero program, a government-sponsored program that temporarily received Mexican workers for farm and railroad between the years 1942 and 1964. At least 4.2 million Mexicans entered the US through the program. The best available data on the direct measure of the program is state-level Bracero exposure digitized by Clemens et al. (2018). The state-by-year fixed effects capture relevant variation.

Baseline estimation is not weighted, and the estimates report the average outcome per county.

[27] In 1940, around 19% of agriculture workers in the South were sharecroppers. The share was as high as 37% in the Deep South (Alabama, Georgia, Louisiana, Mississippi, and South Carolina), especially pronounced in the Mississippi Delta. Such a prevalence may have influenced subsequent agricultural development (Day, 1967; Ferleger, 1993).

[28] Corn production took up the largest land in terms of acre harvested. Acre in hay is included as a proxy for livestock production.

[29] Note that the county fixed effect removes the fixed level of trade exposure while the state-by-year fixed effect remove overall state-level changes in exposure. Hence, the primary role of a control variable is to take into account the influences of the changes in trade exposure that may influences counties in the same state differently.

The results are similar using the 1940 population as the weight, which represents the average outcomes per person in the initial period. All standard errors are clustered at the county level to take into account serial correlation within a county. As a robustness check, I report Conley standard errors that allow for spatial correlation.

3.3 Assessing strategy validity

3.3.1 Discussion of the identification strategy

Given the constructed SSIV measure of the Northern pull factors and the set of fixed effects and control variables, the baseline strategy assumes that Southern counties would have changed the same after 1970 in the absence of the differential exposure to Northern pull factors between 1940 and 1970, when compared to other counties in the same state with similar levels of Southern push factors and with similar pre-migration characteristics. Recent literature shows that the consistency of an SSIV estimator can rely on either share (Goldsmith-Pinkham et al., 2020) or shift (Borusyak et al., 2022). Here, I stress the identification in terms of the share, but the empirical strategy is designed to also take into account the potential role of shift exogeneity.

Goldsmith-Pinkham et al. (2020) show that SSIV is numerically equivalent to a GMM estimator with the shares as a large set of instruments and a weight matrix constructed from shifts. Shares are allowed to be correlated with the levels of outcomes since the strategy asks whether differential exposure to common shocks leads to differential changes in the outcome. Here, this condition requires migration linkages before 1940 (predetermined migration share) to be orthogonal to the changes in outcomes after 1970, conditional on observables. Note that county fixed effects isolate variation in changes and remove any time-invariant county characteristics that could have influenced the levels of migration linkages before 1940. The share strategy can be viewed as a DiD-IV that requires a parallel trend assumption, which is central to this paper's strategy. The statistical tests (Section 3.3.3 and Online Appendix Section D) and interpretation follow this view.

Alternatively, the identification of SSIV can rely on exogenous shifts (Borusyak et al., 2022). This strategy is closer to a standard IV and capitalizes on the idiosyncrasies of the shifts. Here, I use a large set of Northern pull factors as the shifts, which can plausibly assumed to be unrelated to the changes in Southern economic outcomes. Note that state-by-year fixed effects remove national- and state-level trends and, hence, any common shocks to the destinations and origins.

In applying SSIV on migration, Jaeger et al. (2018) cautions the potentially confounding influences of serially correlated migration. While migration can induce both short—and long-term changes, the flows of migrants themselves tend to be correlated, making it hard to distinguish between longer-term adjustments and the influences of lagged migration. To limit such a concern, I restrict the attention only to longer-term changes induced by the 1940-1970 flow, which is unique in its size and breadth compared to the migration flows before or after (Section 1). Note that the baseline estimation controls for both the pre-period and contemporaneous net migration rates.

Relatedly, the presence of spatially correlated migration exposure may bias the results. I report the robustness exercises by adding weighted average of the other Southern counties' migration exposure as additional time-interacted controls. The weight is either proportional to the squared inverse distance for adjacent counties or proportional to pre-period (1910-1940) migration linkages within the Southern counties, by following Borusyak et al. (2023).

The exclusion restriction for the two-stage least squares assumes that the Northern pull factors influenced the Southern economic outcomes through out-of-county out-migration. Although I cannot distinguish out-of-South migration from out-of-county, within-South migration, the construction of the instrument would likely induce out-of-South migration after accounting for state-wide trends. This is supported by the patterns that including proxies for potentially correlated migration exposure as controls do not significantly affect the migration estimates, as discussed in the next subsection. Nonetheless, I mainly present the reduced-form estimates to clearly highlight the identifying variation of the instrument.

3.3.2 First-stage results

Figure 3 shows the first-stage regression in Panel A, while Panels B and C separately report the results by race. Overall, there exist strong correlations between actual out-migration rates and constructed migration exposure. However, the relationship tends to be weaker for Black out-migration. This could be due to the limitation of the linking approach in general, as it is harder to link historically similar Black names. An alternative approach using 1935-1940 migration shares exhibits a relatively stronger relationship for Blacks but overall weaker relationships when combined with Whites (Online Appendix Figure OA1). The baseline estimation combines both Black and White out-migration, as in Panel A, while the robustness check also separately reports the results by race, as in Panels B and C. Both are similar to the baseline outcomes in direction but less precise.

Table 1 reports the first-stage results using the out-of-county net out-migration rate as the dependent variable. In Panel A, Column 1 shows the basic relationship between the dependent variable and the excluded instrument. Columns 2 and 3 each add state fixed effects and control

Table 1: First-stage results on migration response.

Panel A. First-stage regression					
	Net out-of-county out-migration rate (1940-1970)				
	(1)	(2)	(3)	(4)	(5)
Migration exposure (SSIV, 1std)	4.579***	7.098***	3.726***	3.642***	4.841***
Clustered s.e. (county)	(1.270)	(1.538)	(0.777)	(0.842)	(1.036)
State fixed effect	No	Yes	Yes	Yes	Yes
Baseline controls	No	No	Yes	Yes	Yes
Adjacent counties' exposure control	No	No	No	Yes	No
Squared and cubic terms control	No	No	No	No	Yes
First-stage F	13.00	25.15	26.59	18.51	25.26
Counties	1,096	1,096	1,096	1,096	1,096
Panel B. Migration linkage-corrected regression (Borusyak et al., 2023)					
	Net out-of-county out-migration rate (1940-1970)				
	(1)	(2)	(3)	(4)	(5)
Linkage-corrected migration exposure (SSIV, 1std)	5.053	7.625**	3.777***	3.663***	4.582***
Robust s.e.	(3.704)	(3.289)	(0.663)	(0.636)	(0.874)
State fixed effect	No	Yes	Yes	Yes	Yes
Baseline controls	No	No	Yes	Yes	Yes
Adjacent counties' exposure control	No	No	No	Yes	No
Squared and cubic terms control	No	No	No	No	Yes
R-squared	0.003	0.150	0.969	0.970	0.969
Counties	1,096	1,096	1,096	1,096	1,096

Note: The table reports the first-stage estimation results using the out-of-county net out-migration rate as the dependent variable and the Northern migration exposure as the excluded instrument. The unit of observation is county. Panel A presents the standard first-stage regression, while Panel B corresponds to a modified approach from Borusyak et al. (2023) by constructing a modified exposure measure that takes into account the within-South correlation of the migration exposure. Column 1 reports the raw relationship, and Columns 2 through 3 add state fixed effect and control variables listed in Section 3.2.2. Column 4 includes a weighted average of adjacent counties' exposure as a control, with squared inverse distance as the weight. Column 5 adds squared and cubic values of the migration exposure to remove non-linear effects. Higher-order terms are not statistically different from zero at the 10% level and are not reported. The Kleibergen-Paap robust F-statistics for Panel A and adjusted R-squared for Panel B are reported. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

variables. The baseline result in Panel A, Column 3 shows that one standard deviation greater exposure to Northern pull factors induces 3.7% more out-of-county out-migration for the Southern counties. The second-stage results in Table A3 scale the reduced-form estimates using the baseline estimates as in Panel A, Column 3.

I introduce additional specifications that can be interpreted as alternative first-stage regressions. Panel A, Columns 4 and 5, and Panel B explore the bounds for the migration estimate by taking into account non-linear effects or considering potential biases through correlated migration exposure. First, Column 4 adds a weighted average of adjacent counties' exposure to control for spatially correlated migration exposure, with squared inverse distance as the weight. The inclusion reduces the estimated migration response by 0.084 percentage points. Alternatively, in Column 5, I add squared and cubic exposure terms to remove nonlinear effects. Although they are not statistically significant at the 10% level, their inclusion increases the linear estimate to 4.8%, suggesting that the migration response around the mean could have been larger. In Panel B, I report the results with a migration-linkage corrected estimation suggested by Borusyak et al. (2023).[30] While they account for within-South migration linkages, the overall results remain similar.

3.3.3 Pretrend tests

While directly testing the validity of an instrumental strategy is, in general, not feasible, I document the robustness and limitations of the baseline strategy through commonly used statistical tests. For Table A2, I estimate Equation (4) on pre-period outcomes for the variables that have pre-period

[30] Borusyak et al. (2023) point out that the traditional migration regression, as in Panel A, could underestimate the true migration response if the correlated shocks are not properly taken into account. By following their method, I create an alternative measure that takes into account a weighted average migration exposure to other Southern counties, where the weight is given by pre-period (1910-1940) within-South migration shares.

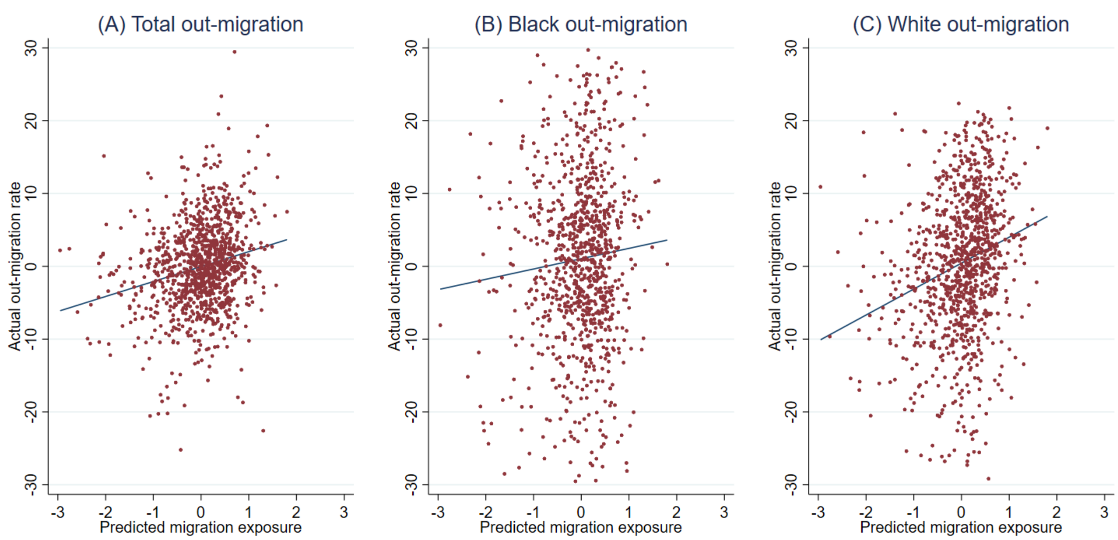

Figure 3: Scatter plot of the residualized first-stage (1940-1970).

Note: The figure presents first-stage regression results. The y-axis plots the net out-migration rates between 1940 and 1970, and the x-axis is the Northern migration exposure between 1940 and 1970, constructed by the SSIV strategy described in Section 3.2.1. The net migration rates are from Gardner and Cohen (1992) and Bowles et al. (2016). Both the left-hand- and right-hand-side variables are residualized by the set of control variables described in Section 3.2.2 and state fixed effects. Counties with x-axis values less than -30% or greater than 30% are excluded from the figure for visibility (3, 101, 10 counties for Panels A, B, and C).

information.[31] The test examines whether pre-period changes in the main outcomes (the values in 1920 and 1930 compared to 1940) are systemically correlated with the instrument.

Panels A to C report the outcomes on agriculture, manufacturing, and wholesale variables. Overall, they do not show a clear pattern and are statistically not different from zeros, suggesting that consequential pretend that drove the main results was less likely to exist.

A couple of exceptions are agricultural employment and farm output. Panel A, Column 1, suggests that agricultural employment in 1940 was lower than in 1920 and 1930 for the counties that were more exposed to Northern pull factors between 1940 and 1970. This could have been driven by a correlation between the changes in agricultural employment in the pre-period and migration patterns during the pre-period. A similar relationship might exist for the changes in farm output (Column 5). To account for potential confounding factors from such relationships, I add time-interacted values of the 1930 outcome as additional controls for agricultural outcomes, except for the number of combines that started to be recorded in 1950. These controls account for any changes after 1940 that may have arisen due to differences in the pre-period outcome. Overidentification tests, as well as placebo tests and falsification exercises, are discussed after presenting the main results.

4 Empirical Evidence

This section examines how the migration flows between 1940 and 1970 shaped the subsequent economic development in the South after 1970. First, I study whether relative labor scarcity from out-migration led to agricultural mechanization. I then document novel findings on the relationships between out-migration and subsequent developments in manufacturing and services. The observed changes in the economy are discussed in terms of regional structural change. Unless mentioned otherwise, the dependent variables are logged values and have semi-elasticity interpretation. For simplicity, I refer to log points as percentage changes.

4.1 Agriculture

Table 2 reports the estimation results on agricultural variables. They compare the outcomes before (1940 and 1950) and after (1970 to 2010) the Second Great Migration between counties that expe-

[31] The number of combines and manufacturing capital spending are available after 1950, and retail variables are available after 1940. Instead, I add manufacturing intermediate goods spending and manufacturing revenue that are not available or only sparsely available during the main study periods.

Table 2: Estimation results for agricultural variables (OLS and reduced-form).

	(1) Agriculture employment	(2) Number of farms	(3) Acres in farmland	(4) Number of tractors	(5) Number of combines	(6) Farm output	(7) Farm value per acre
(A) Out-migration rate (OLS, 1%)	-0.020***	-0.009***	-0.001	0.007***	0.000	0.008***	0.012***
Clustered s.e. (county)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Fixed effects	No	No	No	No	No	No	No
Controls	No	No	No	No	No	No	No
(B) Migration exposure (SSIV, 1std)	-0.003	-0.039	-0.146***	-0.013	-0.056	0.042	0.127***
Clustered s.e. (county)	(0.036)	(0.027)	(0.030)	(0.029)	(0.050)	(0.044)	(0.032)
Fixed effects	No	No	No	No	No	No	No
Controls	No	No	No	No	No	No	No
(C) Migration exposure (SSIV, 1std)	-0.180***	-0.088***	-0.074***	0.192***	0.247***	-0.046	0.010
Clustered s.e. (county)	(0.030)	(0.021)	(0.018)	(0.029)	(0.065)	(0.031)	(0.016)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No	No
(D) Migration exposure (SSIV, 1std)	-0.042*	-0.023*	-0.082***	0.150***	0.043	-0.033	0.020
Clustered s.e. (county)	(0.024)	(0.012)	(0.015)	(0.028)	(0.062)	(0.024)	(0.013)
Conley s.e. (250km)	[0.021]	[0.012]	[0.011]	[0.023]	[0.038]	[0.021]	[0.010]
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.457	0.626	0.353	0.447	0.087	0.968	0.640
First-stage F	26.59	24.62	24.36	25.40	18.48	14.64	26.76
Counties	1,096	1,096	1,096	1,090	1,058	1,090	1,090

Note: The table reports OLS estimates using Equation (3) on Panel A and SSIV reduced-form estimates using Equation (4) on Panels B to D, with county-year as the unit of observation. All dependent variables are logged values and have semi-elasticity interpretation. Each column reports the changes in the indicated outcome variable in logs for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted years of 1940 and 1950, except for the number of combines with the omitted year of 1950. Panels A and B do not include any fixed effects or control variables. Panels C and D add state-by-year and county fixed effects and control variables described in Section 3.2.2. Robust standard errors are clustered by county and reported in parentheses. Panel D also reports Conley (1999) standard errors with 250 km (155 miles) as a cutoff in square brackets and the first-stage Kleibergen-Paap robust F-statistics. The corresponding second-stage estimates are reported in Table A3. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

rienced different levels of out-migration. Panels A and B, respectively, present the OLS and SSIV reduced-form results without any fixed effects or control variables. They demonstrate the raw relationships between the explanatory variable and outcome variables. Panel A shows that a higher out-migration rate is associated with less agricultural employment and fewer farms (Columns 1 and 2), but with higher numbers of tractors and combines (Columns 4 and 5), the measure of agricultural mechanization. The negative association between agricultural development and out-migration suggests that agricultural push factors, such as mechanization, played an important role in driving regional out-migration (Kirby, 1983; Boustan, 2010, 2016).

The OLS captures any correlation between the dependent variables and the out-migration rate, while the SSIV focuses on the relationship explained by the differential exposure to the Northern

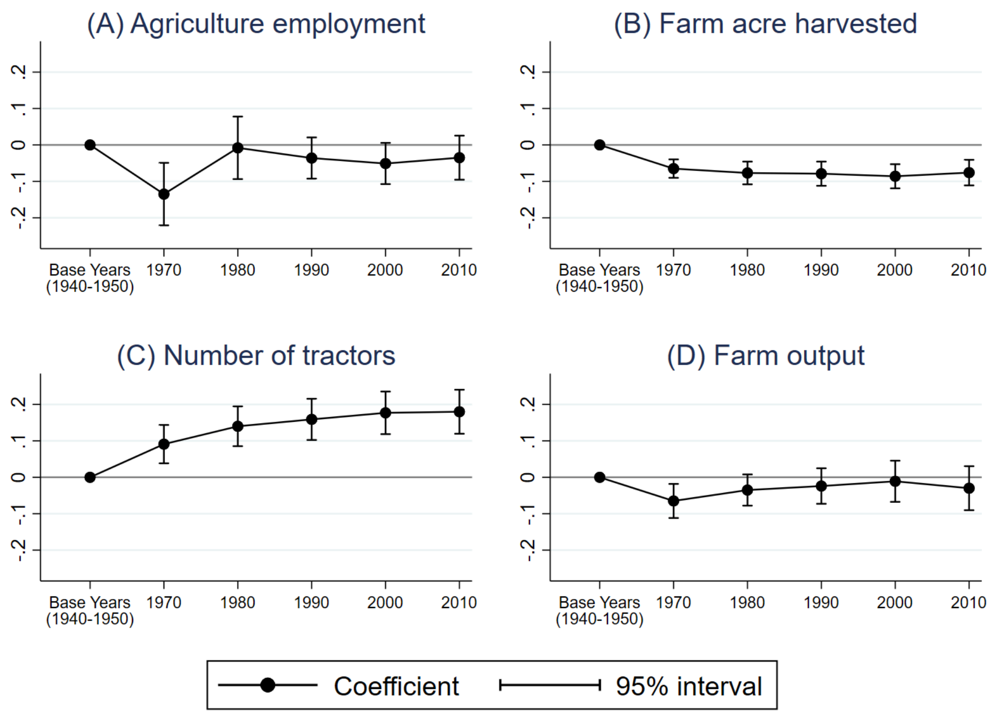

Figure 4: Time trends in agricultural estimates.

Note: The figure presents the SSIV reduced-form estimates on agricultural outcomes using Equation (4), along with the 95% confidence interval. The reported variables are agricultural employment (Panel A), acres in farmland (Panel B), the number of tractors (Panel C), and total farm output (Panel D). They correspond to the year-specific version of Table 2, Panel D, by including the full set of fixed effects and control variables. The coefficients estimate the changes in the indicated outcome variable in each year for one standard deviation greater exposure to Northern pull factors, relative to the omitted years of 1940 and 1950. Robust standard errors clustered by county.

counties that had higher levels of migration pull factors.[32] The reduced-from estimates in Panel B measure how one standard deviation greater exposure to Northern pull factors is associated with the changes in outcomes. At least for agriculture, they tends to display similar patterns as in OLS.

Panel C estimates relative changes in outcomes between counties within the same state in the same year by adding state-by-year and county fixed effects, where the reported estimate is an average across states and years. Panel D further adds control variables so that the comparison relies on counties with similar levels of Southern push factors and similar pre-migration characteristics, where the parallel trend assumption is more likely to hold. It is estimated by the aggregated version of Equation (4). For Panel D, I report R-squared and the Kleibergen-Paap robust F[33] from the first-

[32]This interpretation follows the share view of Goldsmith-Pinkham et al. (2020). The strategy compares the changes in outcomes between counties with different predetermined migration shares before the Second Great Migration, weighted by the levels of pull factors from Northern counties between 1940 and 1970.

[33]With one endogenous variable and one instrument, the value of robust F is equal to the value of efficient F developed by Montiel Olea and Pflueger (2013).

stage regression.

Table 2, Panel D, Columns 1 to 3, show that one standard deviation increase in Northern migration exposure reduced agricultural employment by 4.2%, the number of farms by 2.3%, and farm acres harvested by 8.2%[34,35] when compared to other counties in the same state with similar characteristics. However, as the county-wide capital-to-labor ratio increased, agriculture may have substituted labor with capital, proxied by the number of tractors and combines. Column 4 shows that higher out-migration induced relative adoption of tractors (15.0%). Although not precise, the number of combines could have increased as well (4.3%, Column 4).[36] As a result, the overall level of farm output is not much affected by different degrees of out-migration (Column 6). Similarly, Column 7 reports that the total value of farms, including the value of land, implements, and buildings, tended to experience only negligible changes.

Figure 4 plots the time trend in the estimates for agricultural employment (Panel A), acres in farmland (Panel B), the number of tractors (Panel C), and farm output (Panel D). I use the time-interacted migration exposures using Equation (4). One standard deviation increase in Northern pull factors induced agricultural employment to decrease by 13.5% in 1970, but it recovered to 3.5% by 2010 (Panel A). The acres of farmland decreased by 6.5% in 1970 and stayed at a similar level at least until 2010 (Panel B).

The number of tractors increased steadily until 2000 and maintained at least until the year 2010. It increased by 9.1% in 1970, continued to grow by 17.7% in 2000, and maintained at 18.0% in 2010 (Panel C). The observed pattern of continued increases in tractor usage can be rationalized by the directed technical change (Hicks, 1932; Acemoglu, 2002, 2007). As agriculture uses less labor and more capital, agriculture becomes better at using capital, which in turn incentivizes further capital investment. For instance, the diffusion of tractors could have been initiated with a narrow application that can directly substitute labor but subsequently generalized to broader use (Gross,

[34]The decreased farmland could have been driven by the introduction of tractors. Tractors augment land by freeing up the land previously allocated to feed farm animals. Between 1930 and 1960, acres of cropland used to feed horses and mules decreased from 65 million acres to 5 million acres (Olmstead and Rhode, 2001). Alternatively, a portion of the trend been could have driven by complementarity between labor and land.

[35]Note an important limitation of the estimating strategy: Equation (4) only measures relative effects. Here, the relative changes in agriculture may exaggerate the impact of out-migration because labor scarcity encourages other counties within the same state to specialize more in agriculture. Therefore, the estimates should be understood as indications of relative increases and decreases.

[36]The result on the number of combines is less precisely estimated compared to the number of tractors. This could be due to the differences in their usage. A tractor is essentially a power unit that can move agriculture equipment, and a combine is a combination of tractor and harvesting equipment. While tractors are universally used in all types of farming for plowing, planting, and harvesting, combines are specialized to harvest field crops, which has been less central to Southern agriculture.

2018). In Panel D, farm outputs initially decreased but recovered with accompanying increases in tractor adoption. The pattern suggests that, at least in terms of agriculture, closed-economy forces from factor substitution was stronger than the Heckscher–Ohlin channel from the differences in factor intensity (Section 2.4).

Agricultural economics literature has documented that low labor costs can delay mechanization, while labor scarcity may induce the adoption of labor-saving technologies (see Gallardo and Sauer (2018) for a review). In the 20th United States setting, shrinking labor supply from a natural disaster (Hornbeck and Naidu, 2014) or from an abrupt change in migration policy (Clemens et al., 2018; San, 2023) facilitated the adoption of labor-saving capital and technologies in agriculture. The relative labor shortages from pull factor-induced out-migration exhibit a similar pattern (Table 2, Panel D). Such labor scarcity might have also encouraged related changes in non-agriculture.

4.2 Non-agriculture

Manufacturing. Table 3 reports the manufacturing results. Panels A and B, respectively, report the OLS and SSIV reduced-form estimates without any fixed effects or control variables. Panel A shows that higher levels of regional out-migration are associated with fewer manufacturing workers and lower levels of manufacturing development, represented by manufacturing capital spending, value added, and annual payroll. On the contrary, Panel B suggests that counties more exposed to Northern pull factors had a more developed manufacturing sector. Such contrasting patterns suggest that the role of manufacturing development as migration push factors and the potential role of out-migration on manufacturing have opposite implications.

While agriculture mechanization would have likely pushed workers out of agriculture and out of more agrarian counties (Table 2, Panel A), manufacturing development would have pulled migrants. Hence, Table 3, Panel A shows that counties with more advanced manufacturing tended to experience less out-migration. On the contrary, the influences of pull factor-induced out-migration, as evinced by positive associations in Panel B, could have incentivized labor reallocation into manufacturing and physical capital investment. Such disparities between Panels A and B demonstrate the role of the instrumental variable strategy.

To focus on changes in outcomes related to the relative migration exposure, Panels C and D add state-by-year and county fixed effects and control variables so that estimation relies on comparisons between counties in the same state with similar levels of Southern push factors and pre-migration characteristics. The baseline estimates in Panel D reveal that Northern migration exposure modestly increased manufacturing employment and the levels of manufacturing development.

Recall that pull factor-induced out-migration would raise the regional capital-to-labor ratio. As a response, more flexible agriculture substituted now scarcer labor with capital, releasing workers from agriculture (Table 2, Panel D, Column 1). Some of this labor would have been reallocated to local manufacturing, increasing employment by 8.2% with one standard deviation greater exposure to the Northern pull factors. The number of establishments also increased (Column 2, 9.1%). The relative increases in manufacturing employment incentivized further investment in manufacturing capital due to the labor-capital complementarity; Column 3 finds that relative capital spending increased by 13.6%, more than the relative increase in employment. As a result, manufacturing value added and annual payroll increased by about 8.6% and 15.4% (Columns 4 and 5). The

Table 3: Estimation results for manufacturing variables (OLS and reduced-form).

	(1) Manufacturing employment	(2) Manufacturing establishment	(3) Manufacturing capital spending	(4) Manufacturing value added	(5) Manufacturing annual payroll
(A) Out-migration rate (OLS, 1%)	-0.009***	-0.011***	-0.016***	-0.005***	-0.004**
Clustered s.e. (county)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Fixed effects	No	No	No	No	No
Controls	No	No	No	No	No
(B) Migration exposure (SSIV, 1std)	0.274***	0.192***	0.368***	0.402***	0.387***
Clustered s.e. (county)	(0.048)	(0.034)	(0.084)	(0.081)	(0.072)
Fixed effects	No	No	No	No	No
Controls	No	No	No	No	No
(C) Migration exposure (SSIV, 1std)	0.086**	0.067***	0.214**	0.165**	0.189***
Clustered s.e. (county)	(0.039)	(0.022)	(0.087)	(0.073)	(0.060)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No
(D) Migration exposure (SSIV, 1std)	0.082**	0.091***	0.136	0.086	0.154***
Clustered s.e. (county)	(0.038)	(0.018)	(0.088)	(0.072)	(0.057)
Conley s.e. (250km)	[0.026]	[0.013]	[0.069]	[0.048]	[0.049]
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R-squared	0.157	0.417	0.160	0.144	0.174
First-stage F	23.03	24.33	15.54	21.07	22.23
Counties	1,096	1,096	1,065	1,096	1,096

Note: The table reports OLS estimates using Equation (3) on Panel A and SSIV reduced-form estimates using Equation (4) on Panels B to D, with county-year as the unit of observation. All dependent variables are logged values and have semi-elasticity interpretation. Each column reports the changes in the indicated outcome variable in logs for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted years of 1940 and 1950, except for the manufacturing capital spending with the omitted year of 1950. Panels A and B do not include any fixed effects or control variables. Panels C and D add state-by-year and county fixed effects and control variables described in Section 3.2.2. Robust standard errors are clustered by county and reported in parentheses. Panel D also reports Conley (1999) standard errors with 250 km (155 miles) as a cutoff in square brackets and the first-stage Kleibergen-Paap robust F-statistics. The corresponding second-stage estimates are reported in Table A3. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

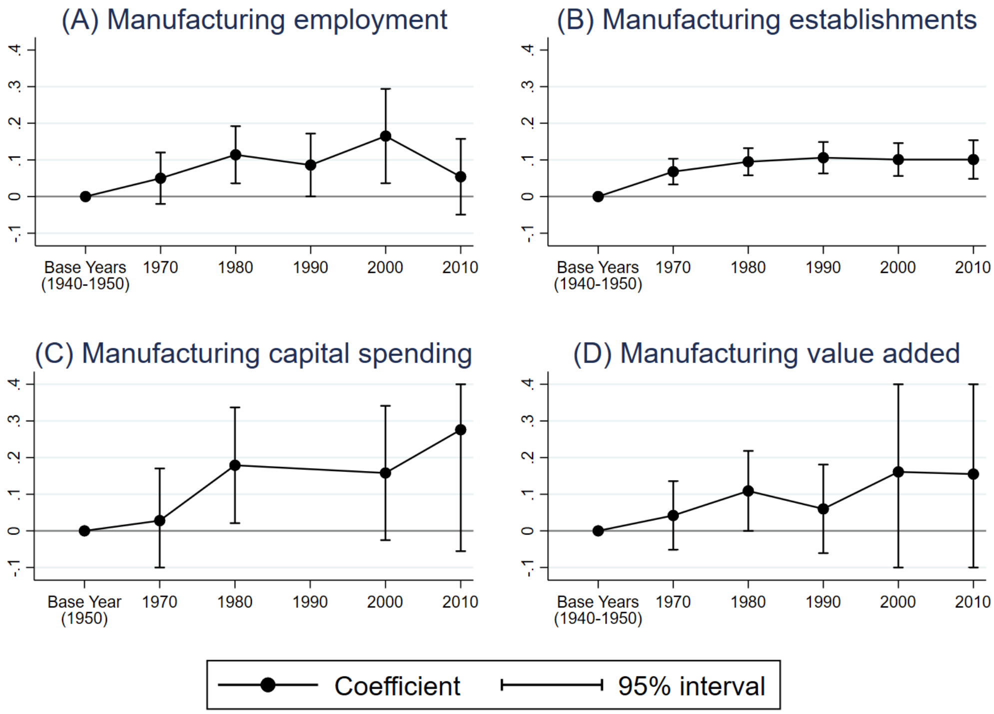

Figure 5: Time trends in manufacturing estimates.

Note: The figure presents the SSIV reduced-form estimates on manufacturing outcomes using Equation (4), along with the 95% confidence interval. The confidence intervals are truncated above at 0.4 and below at -0.1 for visibility. The reported variables are manufacturing employment (Panel A), capital spending (Panel B), value added (Panel C), and annual payroll (Panel D). They correspond to the year-specific version Table 3, Panel D, by including the full set of fixed effects and control variables. The coefficients estimate the changes in the indicated outcome variable in each year for one standard deviation greater exposure to Northern pull factors, relative to the omitted years of 1940 and 1950. Robust standard errors clustered by county.

relative magnitudes of the estimates suggest that the payroll per worker increased as well.

Figure 5 presents the changes in manufacturing employment (Panel A), establishments (Panel B), capital spending (Panel C), and value added (Panel D). Manufacturing employment increased by 4.7% in 1970 and tended to remain at the higher levels until 2000. The influences wore off to 4.2% in 2010 and became less precise. Other outcomes demonstrate that overall results reported in Table 3 are maintained during the study period and even grew at least until 2000 or 2010.

The continued growth of manufacturing in the more exposed counties can be interpreted with the directed technical change and Heckscher-Ohlin framework (Section 2). As manufacturing absorbed labor, weak equilibrium bias in non-agriculture would have favored the complementary factor, capital, if the elasticity of substitution is less than one (Acemoglu, 2007). Such a capital-biased technological growth would have further incentivized capital investment. Moreover, the Heckscher-Ohlin force suggests that if manufacturing is more capital-intensive, both labor scarcity and capital investment would lead to an expansion of its production through the relative increase in comparative advantage, as in the quasi-Rybczynski effect of Romalis (2004).

Local nontradable sectors. Table 4 documents wholesale and retail outcomes, which are used as proxies for local nontradable services. As in manufacturing, both retail and wholesale experienced positive growth from out-migration, with wholesale reporting stronger positive changes. For instance, with one standard deviation higher exposure to the Northern pull factors, employment in wholesale increased by 18.6% and retail by 7.5% between 1970 and 2010, relative to their levels in 1940 and 1950. Total sales in each sector grew by 10.2% and 5.4%. The number of establish-

Table 4: SSIV estimation results for wholesale and retail (reduced-form).

Panel A. Wholesale				
	(1) Wholesale employment	(2) Wholesale establishment	(3) Wholesale sales	(4) Wholesale annual payroll
Migration exposure (SSIV, 1std)	0.186***	0.146***	0.102**	0.132***
Clustered s.e. (county)	(0.031)	(0.022)	(0.051)	(0.050)
Conley s.e. (250km)	[0.024]	[0.015]	[0.042]	[0.044]
Fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R-squared	0.278	0.445	0.141	0.130
First-stage F	27.46	24.93	27.57	22.70
Counties	1,083	1,096	1,086	1,086
Panel B. Retail				
	(1) Retail employment	(2) Retail establishment	(3) Retail sales	(4) Retail annual payroll
Migration exposure (SSIV, 1std)	0.075***	0.036***	0.054***	0.080***
Clustered s.e. (county)	(0.016)	(0.009)	(0.013)	(0.018)
Conley s.e. (250km)	[0.011]	[0.007]	[0.012]	[0.013]
Fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R-squared	0.663	0.728	0.681	0.627
First-stage F	25.07	25.06	25.07	25.04
Counties	1,096	1,096	1,096	1,096

Note: The table reports SSIV reduced-form estimates using Equation (4) for wholesale variables in Panel A and retail variables in Panel B, with county-year as the unit of observation. All dependent variables are logged values and have semi-elasticity interpretation. Each column reports the changes in the indicated outcome variable in logs for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted years of 1940 and 1950. Robust standard errors in parentheses are clustered by county, and Conley (1999) standard errors in square brackets use 250 km (155 miles) as a cutoff. Kleibergen-Paap robust F-statistics are reported. The corresponding second-stage estimates are reported in Table A3. Stars represent: $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Figure 6: Time trends in wholesale and retail estimates.

Note: The figure presents the SSIV reduced-form estimates on wholesale and retail outcomes using Equation (4), along with the 95% confidence interval. The estimation includes the full set of fixed effects and control variables. The reported variables are wholesale employment (Panel A), retail employment (Panel B), wholesale sales (Panel C), and retail sales (Panel D). The coefficients estimate the changes in the indicated outcome variable in each year for one standard deviation greater exposure to Northern pull factors, relative to the omitted years of 1940 and 1950. Robust standard errors clustered by county.

ments and sales also grew after 1970. Figure 6 shows the year-specific changes in employment and total sales in wholesale and retail after 1970. As in manufacturing, the increases in local sector outcomes were maintained at least until 2000 or 2010.

Given that retail and wholesale use both factors of production, closed-economy forces would have generated similar changes to manufacturing. However, the local nature of these industries implies that they are less governed by the open-economy force. Instead, the non-tradable sector was subjected to another channel: local spillover effects. The growth of labor payments in manufacturing and the possible increases in agricultural wages would have positively affected the growth of the local tradable industry.[37]

[37] Wholesale is closer to a tradable industry than retail because wholesale involves selling large quantities to other businesses. Retail sells products directly to consumers. Hence, the differences between the changes in the two industries give a hint of the role of trdability.

Table 5: SSIV estimation results for industry share and education (reduced-form).

Panel A. Structural change (employment share)

	(1) Agriculture	(2) Manufacturing	(3) Services	(4) Consumer services	(5) Producer services
Migration exposure (SSIV, 1std)	-0.081***	0.064*	0.005	0.030*	-0.033
Clustered s.e. (county)	(0.025)	(0.033)	(0.016)	(0.017)	(0.032)
Conley s.e. (250km)	[0.019]	[0.021]	[0.011]	[0.013]	[0.027]
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R-squared	0.345	0.127	0.085	0.258	0.133
First-stage F	25.49	21.60	23.57	21.85	24.27
counties	1,096	1,096	1,096	1,096	1,078

Panel B. Education

	(1) Median school year	(2) Share high school	(3) Share college	(4) Employment in education	(5) Education spending
Migration exposure (SSIV, 1std)	-0.009***	0.008	-0.023	-0.013	-0.015**
Clustered s.e. (county)	(0.003)	(0.007)	(0.029)	(0.011)	(0.007)
Conley s.e. (250km)	[0.002]	[0.006]	[0.028]	[0.007]	[0.006]
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R-squared	0.160	0.391	0.087	0.311	0.448
First-stage F	20.61	30.63	40.96	15.48	23.22
counties	1,096	1,096	1,096	1,094	1,096

Note: The table reports SSIV reduced-form estimates for industry employment shares (Panel A) and educational outcomes (Panel B) using Equation (4), with county-year as the unit of observation. All results include state-by-year and county fixed effects and control variables described in Section 3.2.2. Each column reports the changes in the indicated outcome variable in logs for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted years between 1940 and 1960, using periods when information is available. HDES, the main dataset, does not report detailed service employment, and I supplement the analysis using CBP for Panel A, Columns 3 to 5. Consumer services are defined to be the 2017 NAICS classification in 42-45 and 61-72. Producer services include 51-56. Robust standard errors in parentheses are clustered by county, and Conley (1999) standard errors in square brackets use 250 km (155 miles) as a cutoff. The first-stage Kleibergen-Paap robust F-statistics are reported. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

4.3 Structural change

This subsection examines how differential exposure to Norther pull factors influenced overall industry composition and the levels of education. Table 5, Panel A, reports the baseline results using the employment share in each industry as dependent variables. Panel A, Columns 1 and 2 suggest that out-migration contributed to structural change out of agriculture and into manufacturing. One standard deviation increase in Northern pull factors resulted in an 8.1% decrease in agriculture employment share but a 6.4% increase in manufacturing. Column 3 adds services that experienced relatively small changes in employment. However, the decomposition into consumer and producer services shows that the former experienced a relative gain (Column 4), while producer service tended to experience negative changes (Column 5). A potential driver of such a disparity is discussed in Online Appendix Section B. The key idea is that producer services, which tend to be less capital-intensive among the listed industries, could have been less affected by economic reallocation from the changes in the regional capital-to-labor ratio.

If a region with higher out-migration primarily responded with physical capital investment, the resulting capital deepening might have reduced the incentive for human capital accumulation. Panel B reports outcomes on county-level educational attainment to proxy human capital development. Overall results show that counties with greater migration exposure did not increase or even relatively decrease the median school year (Column A) or high school and college graduates share (Columns B and C). Furthermore, they did not invest more in education, both in terms of spending (Column D) and employment (Column E). Note that the baseline model controls for contemporaneous population and net migration rates, and hence the results are unlikely to be driven by the changes in county sizes. These results highlight that counties that experienced higher out-migration primarily responded by physical capital deepening, evinced by relative increases in agriculture and manufacturing capital relative to employment (Tables 2 and 3).

The results in Table 5, Panel B, do not imply that the South failed to raise its overall level of education. On the contrary, the South experienced rapid growth in terms of educational attainment during and after the Great Migration period, on average.[38] Such overall increases in the level of education were pointed out as an important channel that facilitated structural change in the South (Caselli and Coleman II, 2001). However, the counties with different levels of pull factor exposure did not experience relative growth in educational outcomes. Consequently, the role of out-migration, the main focus of this paper, would have likely provided a complementary channel for Southern economic development during the 20th-century.

4.4 Discussion and robustness

Here, I introduce the second-stage estimates, overidentification tests, and additional robustness checks. First, Table A3 documents the second-stage estimates for the main outcomes in Tables 2, 3,

[38] In the South, the share of college graduates for adults older than 25 years old increased from 6.5% in 1940 to 18.6% in 1990. The share of high school graduates also increased from 25.1% in 1950 to 71.4% in 1990. During the same period, college graduates increased from 7.2% to 21.1% and high graduates from 37.4% to 77.3% in the North.

and 4 using the two-stage least squares with the full set of fixed effects and control variables. They scale the reduced form estimates in the main text by the baseline first-stage migration regression as in Table 1, Panel A, Column 3.

Overidentification tests examine the null hypothesis of constant effects between instruments, and the test requires that the number of instruments exceed the number of endogenous regressors. The rejection indicates that the estimated coefficients between the instruments are statistically different. In Table A4, in addition to the baseline 1910-1940 shares, I add the 1935-1940 shares using the information of county of residence five years ago in the 1940 record. I report the values of heteroskedasticity robust Sargan-Hansen J statistics and associated p-values. In general, the two migration exposures from the two shares yield similar results. As exceptions, the retail outcomes tend to differ in magnitudes (Panel C, Columns 5 to 8), compared to the baseline second-stage estimates in Table A3. Among the reported estimates, 14 out of 20 variables do not reject the null of constant effects between instruments at the 5% level. The rejections are mainly driven by less precise estimates from using the 1935-1940 share, further discussed in the Online Appendix.

Online Appendix Section D documents the main results in Tables 2, 3, and 4 from alternative approaches. They include estimating Equation (4) (1) using the Northern exposure based on the 1935-1940 migration matrix, (2) separately by race, (3) by adopting alternative approaches for in-migration prediction (random forest algorithm and actual number of in-migrants), (4) by limiting sample to former confederate states, (5) by limiting sample to balanced counties for all the main variables, (6) using 1940 population-weighted regression, (7) by dropping the time-varying controls (contemporaneous population and net migration rate), (8) by dropping the predicted migration rates between 1940-1970 (predicted Black out-migration rates from Derenoncourt (2022) and White out-migration rates from Table A1), (9) using alternative base years (1940 or 1940-1960), and (10) by adding weighted average of other Southern counties' migration exposures. These controls are aimed at limiting the concerns about potential spatial spillover effects. The Online Appendix also presents falsification tests using the changes in the number of government employees and annual payroll.

Finally, the baseline estimates can be interpreted as the local average treatment effects (LATE). In an ideal randomized setting, one might randomly allocate the number of migrants across counties and randomize who to migrate. However, the Northern migration exposure measure in this paper isolates the component of out-migration induced by Northern pull factors, which could be different from variations in the randomized setting. Still, the LATE here is not necessarily a limitation for policy implications, as the estimation is based primarily on variation driven by migrants who responded to external incentives. In other words, the findings here could be more applicable to a setting where the government can incentivize people to move out of labor-abundant regions.

5 Quantitative Strategy

5.1 Roadmap to the quantitative model.

In this section, I construct a dynamic spatial general equilibrium model with multiple factors of production. The model is based on canonical models of trade and migration (Eaton and Kortum, 2002; Artuc et al., 2010; Caliendo et al., 2019) with capital accumulation (Kleinman et al., 2023) and structural change (Fan et al., 2023; Eckert and Peters, 2023). Here, I extend the aforementioned frameworks by introducing the role of factor substitution and factor intensity in driving structural change and economic allocation. I do so by generalizing the key forces highlighted in Section 2 into multiple periods and realistic geography. Table A5 summarizes the model elements.

The economy consists of a set of discrete locations ($i = 1, ..., N$) and three industries: agriculture, tradable non-agriculture, and local nontradable non-agriculture ($s = a$, m and l). Time is discrete and indexed by t. All sectors use two factors of production, labor and capital, and they are assumed to have CES production structures with non-unitary elasticity of substitution. Agriculture and tradable non-agriculture are subject to forces arising from trade, while the local nontradable sector is instead influenced by local consumption spillover. Tradable non-agriculture consists of manufacturing and (tradable) production services, and it can be regarded as a goods-producing sector. On the contrary, the nontradable sector can be viewed as consumer services that are locally provided.[39] For simplicity, I also refer to them as manufacturing and local services.[40]

Compared to Section 2, I directly model economic agents and general equilibrium conditions. There are two types of infinitely-lived agents: workers and capitalists ($\mathscr{T} = w$ and k), each supplying labor and capital. Capitalists are geographically immobile and own depreciable capital stocks in their location. They make forward-looking decisions over consumption and investment. Capital is freely mobile across sectors within a region but not across regions. Workers do not have access to investment technology and live hand to mouth, but they are geographically mobile, subject to mi-

[39] Although the simple framework focuses on the tradable sectors, the addition of a nontradable sector in the quantitative model helps to position agriculture and manufacturing in the economy, as noted in Eaton and Kortum (2002).

[40] The model can further separate tradable non-agriculture into two types of tasks: physical-capital-intensive and human-capital-intensive tasks, where the former uses both factors while the latter only uses labor (Online Appendix Section A). The human-capital-intensive tasks would represent the portion of the economy not affected (if nontradable) or mechanically negatively affected (if tradable) by labor scarcity.

gration costs.[41] In addition, the general equilibrium model allows non-homothetic preferences on the demand side by introducing the non-homothetic Price-Independent Generalized Linear (PIGL) class in utility (Boppart, 2014; Fan et al., 2023).[42]

I make two simplifications for tractability. First, to obtain analytical expressions for trade and migration with realistic geography, I rely on extreme value distribution assumptions. The resulting expenditure and migration shares take the standard gravity structure. Second, for model calculation, I follow the dynamic exact hat-algebra approach (Caliendo et al., 2019) to eliminate the need to recover counterfactual-invariant fundamentals of the model.

5.2 Preferences and factor supply

Preferences. Workers' welfare is defined as the discounted sum of the infinite path of consumption indirect utility via the log-utility function:

$$U(C_{i,t}) = \log\Big(C(e_{i,t}, P_{i,t})\Big). \tag{9}$$

Workers inelastically supply their labor and earn wages at competitive market rates. Following Boppart (2014) and Fan et al. (2023), individuals' consumption preferences are in the non-homothetic PIGL class. It represents the structural change on the demand side, while the changes in factor allocations capture the structural change on the supply side. The indirect utility of consumption for an individual with expenditure e facing local price of P_i takes the form:

$$C(e, P_i) = \frac{1}{\varepsilon}\left(\frac{e}{(P_i^a)^{\phi^a}(P_i^m)^{\phi^m}(P_i^l)^{\phi^l}}\right)^{\varepsilon} - \sum_{s\in\{a,m,l\}} \nu^s \ln P_i^s, \tag{10}$$

over sectoral value-added CES aggregates of varieties from all regions. I use $P_i \equiv (P_i^a)^{\phi^a}(P_i^m)^{\phi^m}(P_i^l)^{\phi^l}$ as local price index with $\sum_{s\in\{a,m,l\}} \phi^s = 1$. If $\nu^s = 0$ for all sectors and $\varepsilon = 1$, the consumption utility reduces to a Cobb-Douglas utility with consumption share ϕ^s allocated to each sector s. The income elasticity parameter, $\varepsilon \in (0,1)$, is interpreted as the Engel elasticity. The larger the

[41] The baseline model assumes that Black and White workers are perfectly substitutable, but a model extension can consider potentially different productivity by race and imperfect substitutability through an additional layer of CES composite of labor. For a related approach, refer to Takahashi (2023), who adopts a dynamic spatial equilibrium framework as in here. He focuses on different substitutability across different labor groups by race and age, while I focus on the tensions arising from labor and capital.

[42] The demand side is an essential component in quantitatively modeling structural change (Alvarez-Cuadrado et al., 2017), as it allows the aggregate industry shares to vary with the changes in income and relative goods price.

Engel elasticity, the stronger the effect of real income on demand. As incomes grow to infinity, the consumption share on each good converges to $\phi^s \in (0,1)$ as a consumption asymptote.

By applying Roy's identity to the indirect utility function, the consumption share is given as:

$$\varphi^s(P_i, e) = \phi^s + \nu^s \left(\frac{e}{P_i}\right)^{-\varepsilon}. \tag{11}$$

An individual's consumption share depends on the price index in region i and her income. The consumption share on necessity declines as workers' real income rises. Workers do not have access to the investment technology and their labor income equals total expenditure in each period. Regional aggregate demand is derived by summing up individual demand in each location.

Intratemporal labor supply. In each period, factors of production are freely mobile across sectors within the region. To rationalize the observed difference in factor prices, I introduce Roy-type machinery by modeling that a worker has a different ability in each sector, as in Eckert and Peters (2023). A worker supplies a_i^s efficiency units to sector s that are drawn from a sector-specific Frechet distribution with dispersion parameter ζ^w, $P(a_i^s \le a) = \exp(-(a/A_i^{w,s})^{-\zeta^w})$. The size of $A_i^{w,s}$ represents the fundamental level of region-sector-specific labor-augmenting technology, where the superscript w denotes worker.

Each worker chooses a sector by maximizing her income, and the employment share is:

$$s_i^{w,s} = (A_i^{w,s} w_i^s / \bar{w}_i)^{\zeta^w} \text{ where } \bar{w}_i = \left((A_i^{w,a} w_i^a)^{\zeta^w} + (A_i^{w,m} w_i^m)^{\zeta^w} + (A_i^{w,l} w_i^l)^{\zeta^w}\right)^{1/\zeta^w}, \tag{12}$$

where the dispersion parameter governs the sectoral labor supply. The measure of workers in sector s is then given as $s_i^s L_i$, and the effective unit of labor supplied can be written as:

$$L_i^s = \Gamma_{\zeta^w} A_i^{w,s} (A_i^{w,s} w_i^s / \bar{w}_i)^{\zeta^w - 1} L_i \equiv A_i^{w,s} \tilde{L}_i^s, \tag{13}$$

with the labor-augmenting efficiency term $A_i^{w,s}$. I assume that their values are not affected by the Great Migration. The Gamma function $\Gamma_{\zeta^w} \equiv \Gamma(1 - 1/\zeta^w)$ is an inconsequential scalar term with a restriction $1 - 1/\zeta^w > 0$.

Spatial mobility. Individuals make forward-looking decisions over which region n to live in the next period, taking the expected value of future utilities $\mathbb{V}$ and migration costs κ as given:

$$\mathbb{V}_{i,t} = U(C_{i,t}) + \max_{\{n\}} \left\{ \beta \mathbb{E}[\mathbb{V}_{n,t+1}] - \kappa_{ni,t} + \eta u_{n,t} \right\}, \tag{14}$$

where the idiosyncratic preference shocks, u, follow the Type I extreme value distribution. The parameter η scales the variance of the shock, and $1/\eta$ has an interpretation of migration elasticity. The future expected utility depends on the average wage in each potential destination n, $\bar{w}_{n,t+1}$.

The solution to the above dynamic problem yields the migration share proportional to the migration cost- and elasticity-adjusted utility, compared to that of all other possible destinations. By expressing the expected value of the worker's value function as v, the migration share is:

$$\mathbb{M}_{in,t} = \frac{\exp\left(\beta\mathbb{E}_t(v_{n,t+1} - \kappa_{ni})/\eta\right)}{\sum_{j=1}^{N}\exp\left(\beta\mathbb{E}_t(v_{j,t+1} - \kappa_{nj,t})/\eta\right)}, \tag{15}$$

which, combined with the initial population, yields labor market distributions in the next period.

5.2.1 Capitalists

Capitalists' problem. In each region, geographically immobile capitalists of measure sufficiently close to zero choose their consumption and investment to maximize the expected present value of their consumption utility, subject to the standard budget constraint:

$$v_{i,t}^{k} = \mathbb{E}_t\sum_{t=0}^{\infty}\beta^t\frac{(C_{i,t}^{k})^{1-1/\psi}}{1-1/\psi} \text{ subject to } \bar{r}_{i,t}K_{i,t} = P_{i,t}\left(C_{i,t}^{k} + K_{i,t+1} - (1-\delta)K_{i,t}\right), \tag{16}$$

where the superscript k indexes capitalists. The utility function at the higher level takes the constant intertemporal elasticity of substitution form with parameter $\psi > 0$. With $\psi = 1$, it becomes the log form. At the medium level, I assume that capitalists consume the Cobb-Douglas composite of the three industries, $C_{i,t}^{k} \equiv (C_i^a)^{\phi^a}(C_i^m)^{\phi^m}(C_i^l)^{\phi^l}$, with the CES aggregates of varieties at the lower level. Compared to the workers' problem, capitalists' sectoral consumption shares at the medium level equal the consumption asymptotes. This simplification can be viewed as assuming that the capitalists have surpassed the income threshold to reach the asymptotes. An extension that allows the PIGL preference to capitalists is discussed in the Online Appendix.

The budget constraint states that the capitalists' net nominal income, $\bar{r}_{i,t}K_{i,t}$, is allocated to consumption and investment, where the investment good combines goods from all sectors with the asymptotic consumption share. The regional rental rate of capital $\bar{r}_{i,t}$ is a weighted average of net returns across industries. The parameter δ measures the depreciation rate. The capital is geographically immobile once installed and depreciates gradually at the rate δ. I define the real gross return on capital as: $\bar{R}_{i,t} \equiv 1-\delta+\bar{r}_{i,t}/P_{i,t}$.

In each period t, capitalists consume a fixed share $\varsigma_{i,t}$ of their real gross investment income $\bar{R}_{i,t}K_{i,t}$, as in Kleinman et al. (2023):

$$\varsigma_{i,t}^{-1} = 1+\mathbb{E}_t\beta^{\psi}([R_{i,t+1}^{\frac{\psi-1}{\psi}}\varsigma_{i,t+1}^{-\frac{1}{\psi}}])^{\psi}, \tag{17}$$

The consumption share is solved recursively using Equation (17) with the expected sequences of future returns and the values of the consumption parameters. A special case of log utility ($\psi = 1$) yields a constant consumption rate $\varsigma_{i,t} = 1-\beta$.

Intratemporal capital supply. As in labor, the regional capital stock is allocated across the sectors by assuming the role of capital efficiency drawn from a Frechet distribution with region-sector specific fundamental $A_i^{k,s}$ and dispersion parameter ζ^k. Consequently, the intratemporal capital allocation across sectors is given by the share $s_i^{k,s} = (A_i^{k,s}r_i^s/\bar{r}_i)^{\zeta^k}$. The resulting effective capital for each sector is then given as $K_i^s = \Gamma_{\zeta^k}A_i^{k,s}(r_i^s/\bar{r}_i)^{\zeta^k-1}K_i \equiv A_i^{k,s}\tilde{K}_i^s$.

I assume that capital efficiency consists of exogenous fundamental $\bar{A}$ and endogenous components $F^s(\cdot)$ that depends on regional economic allocations:

$$A_i^{k,s} = \bar{A}_i^{k,s} \times F^s(\cdot). \tag{18}$$

The endogenous component incorporates the dynamic weak equilibrium bias in a reduced form way (Section 2, Remark 1). This term represents how economic allocation, such as relative capital abundance, influences the development of capital-augmenting technologies. For instance, it can be viewed as a function capturing the outcomes of learning-by-doing given the regional factor abundance as in Foster and Rosenzweig (1995). Alternatively, it can be thought of as generated by regional technology developers in shadow as in Acemoglu (2002, 2007). For simulation analysis, the values of $F^s(\cdot)$ are calibrated using the estimated changes in agriculture and manufacturing employment with one standard deviation increase in the Northern exposure. In contrast, the exogenous component $\bar{A}$ cancels out during model calculation.[43]

[43] Similarly, labor efficiency terms can be separated into an exogenous and an endogenous component. I abstract from this distinction for simplicity. The rationale follows from Section 2, Prediction 2, which suggests the relative development of capital-augmenting technical change due to factor reallocation induced by the Great Migration.

5.3 Production.

In each region i, a representative local firm in each sector s uses the following CES technology:

$$Y_i^s = z_i^s \left(\rho_i^s (A_i^{w,s} \tilde{L}_i^s)^{\frac{\sigma^s - 1}{\sigma^s}} + (1 - \rho_i^s)(A_i^{k,s} \tilde{K}_i^s)^{\frac{\sigma^s - 1}{\sigma^s}} \right)^{\frac{\sigma^s}{\sigma^s - 1}}. \tag{19}$$

The production side is given similarly to Section 2, with the addition of Hicks-neutral technology z_i^s. This component is a realization of the Frechet distribution with local fundamental Z_i^s and the shape parameter θ^s,[44] as in the standard Eaton and Kortum (2002) setting. The local fundamental Z_i^s is assumed not to be affected by the changes in factor allocation. The parameter σ^s governs the elasticity of substitution between factors, while ρ_i^s influences the labor cost share.

Trade and Market Clearing. The price of each industry s in importer region n is determined as the minimum unit cost across all regions:

$$p_{n,t}^s = \min_{1 \le i \le N} \left\{ \frac{x_{i,t}^s \tau_{ni,t}^s}{z_{i,t}^s} \right\}, \tag{20}$$

where the term inside Equation (20) is the factory-gate price of one unit of goods multiplied by the iceberg-type trade costs, $\tau_{ni,t}^s$. The unit cost function follows from the CES production structure:

$$x_{i,t}^s = \min \left(w_{i,t}^s A_{i,t}^{w,s} \tilde{L}_{i,t}^s + r_{i,t}^s A_{i,t}^{k,s} \tilde{K}_{i,t}^s \right) = \left((\rho_i^s)^{\sigma^s} (w_{i,t}^s)^{1-\sigma^s} + (1 - \rho_i^s)^{\sigma} (r_{i,t}^s)^{1-\sigma^s} \right)^{\frac{1}{1-\sigma^s}}. \tag{21}$$

The changes in factor allocation and resulting changes in factor prices affect the regional comparative advantage through Equation (21).[45]

With the distributional assumption on the Hicks-neutral productivity term, the bilateral expenditure share is given by:

$$\mathbb{S}_{ni,t}^s = \frac{Z_{i,t}^s \left(x_{i,t}^s \tau_{ni,t}^s \right)^{-\theta^s}}{\sum_{j=1}^N Z_{j,t}^s \left(x_{j,t}^s \tau_{nj,t}^s \right)^{-\theta^s}}, \tag{22}$$

where the denominator can be interpreted as inward market access of region n. It also represents the sectoral price index up to a constant. Finally, the region-industry-level price indices are aggregated by consumption share asymptotes to yield the regional price index.

[44] In each sector, the elasticity of substitution of varieties at the lower-level demand is assumed to be less than $1 + \theta^s$ to have a well-defined sector price index. The exact value of the lower level demand parameter can be ignored as long as this restriction is satisfied.

[45] As labor becomes more expensive, a sector that relies more on labor will relatively lose its comparative advantage and experience a relative decrease production.

Table 6: Parameters for Quantitative Analysis.

Definition	Parameter	Comment
Panel (A) Utility parameters		
Asymptotic consumption share	ϕ = (0.01, 0.33, 0.66)	Moment condition (Yang, 2024)
Preference elasticity	ν = (1.27, -0.27, -1.0)	Moment condition (Yang, 2024)
Engel elasticity	$\eta = 0.39$	Estimation (Yang, 2024)
Migration elasticity	$\chi = 0.84$	Estimation (Yang, 2024)
Discount rate	$\beta = 0.67$	Set to $(0.96)^{10}$
Panel (B) Productivity parameters		
EIS between labor and capital	$\sigma = (1.58, 0.80, 0.75)$	Herrendorf et al. (2015)
Average labor weights in production	$\bar{\rho} = (0.49, 0.71, 0.66)$	Herrendorf et al. (2015)
Factor efficiency distribution	$(\zeta^L, \zeta^K) = (6.9, 6.9)$	Eckert and Peters (2023)
Hicks-neutral productivity distribution	$\theta = (12, 6.5, \infty)$	Nigai (2016)
Capital depreciation rate	$\delta = 0.34$	Hulten and Wykoff (1981)
Changes in Southern capital efficiency	$\dot{\mathbf{F}}^s(\cdot) = (1.168, 1.040, 1.027)$	Internally calibrated

Goods market clearing implies that the regional total expenditure, X_n, equals the regional total value-added output, Y_n. The value-added is distributed to workers as wages and capitalists as rents. Market clearing conditions for the factor market are then given by $Y^k_{n,t} = \sum_{i=1}^{N} \mathbb{S}^k_{ni,t} X^k_{n,t}$.

5.4 Taking the model to the data

Table 6 summarizes the model parameters. First, the consumption side and migration parameters follow Yang (2024),[46] where I estimate the PIGL parameters and migration elasticity in the 20th-century United States setting. The values of the preference elasticities imply that agriculture is a necessity and two non-agriculture sectors are luxuries, with local services having the highest income elasticity. Within non-agriculture, the tradable sector is closer to normal goods. The baseline migration elasticity over a 10-year span is estimated to be $\eta = 0.84$, which suggests a higher migration response compared to 1.88 at the annual frequency (Artuc et al., 2010) and 5.34 at quarterly (Caliendo et al., 2019). I set the discount rate β for the decennial interval to be 0.67, assuming a yearly discount rate of 4%. The intertemporal elasticity of substitution of capitalists is set to 1, and their consumption takes the log form with a fixed saving rate β.

In terms of the dynamic spatial equilibrium setting, relatively distinctive features of the model

[46] In this paper, I study the long-running spillover effects of the American Dust Bowl of the 1930s in terms of welfare and structural change through a dynamic spatial equilibrium model. I use similar settings as here but with production functions using one factor or Cobb-Douglas structure.

are the use of CES production functions and factor-augmenting technologies. For related parameters, I capitalize on estimates in the literature and empirical findings in Section 4. First, I use the CES production function estimates in Herrendorf et al. (2015), who uses U.S. macro data between 1947 and 2010. The model here adopts the same production structure in the value-added form. The study periods align except for 1940-1946, when the required data is unavailable. Using their estimates, the values of σ^s for agriculture and two non-agriculture sectors are set to 1.58, 0.80, and 0.75, respectively. The values suggest that agriculture is flexible, while the non-agriculture sectors are inflexible in factor usage.

Herrendorf et al. (2015) also report estimates for ρ^s as the average factor cost shares during the sample period.[47] Still, the direct values on location-industry-specific cost shares are required for simulation. Here, I adjust the estimates on ρ^s by regional proxy for the capital-to-labor ratio. I use variation in the number of tractors and combines per agricultural worker as a proxy for agricultural capital-to-labor ratio. For non-agriculture, I use the manufacturing capital spending per worker as the proxy. I apply these regional variations to adjust the labor share parameter ρ^s. The resulting weights imply that agriculture tended to be labor-intensive in the South but capital-intensive in the North, consistent with a historical account of regional differences in agricultural practices before the Second Great Migration period.

Factor efficiency dispersion parameters are taken from the labor dispersion parameter from Eckert and Peters (2023), estimated using 1880-1920 U.S. data. They measure the within-factor substitutability across sectors. I calibrate the capital depreciation rate by following Hulten and Wykoff (1981) and trade elasticities by Nigai (2016).

As a final step, I calibrate the capital efficiency parameters for the Southern states between 1940 and 1970 given the values of the other parameters described above. Specifically, I minimize the Euclidean distance between the data and simulated moments with the Nelder and Mead (1965) algorithm, a derivative-free numerical optimization in a multidimensional space. I use twelve data moments: the changes in employment and value-added for agriculture (Figure 4, Panels A and D), manufacturing (Figure 5, Panels A and D), and retail (Figure 6, Panels B and D), each for 1970 and 2010. The calibrated values are $(1.168, 1.040, 1.027)$ for agriculture and non-agriculture (tradable and non-tradable). In simulation, as an example, capital efficiency for Southern agriculture is set to increase by 16.8% during the 30-year span (1940-1970) and stay at the higher level between 1970 and 2010.

[47] Compare to their capital share estimate on agriculture, I use the value that excludes the land, in order to focus on the distinction between labor and physical capital.

5.5 Calculating counterfactuals

For the simulation analysis, I use the migration flows during the Second Great Migration period as a shock. Specifically, I prohibit the migration from the South to the North between 1940 and 1970 and allocate the migrants back to the Southern-origin states as stayers. I allow endogenous migration after 1970. To close the model, I add an additional period that corresponds to the year 2020, where the economy is assumed to reach a stationary equilibrium.

The model calculation adopts the dynamic exact hat algebra approach (Online Appendix Section 3). The method calculates the changes in economic allocation over time, given the shock in time changes. The model calculation in time changes annihilates the need to recover the majority of the time-invariant components as they cancel out during calculation. I first run the quantitative model without the shock to calculate a baseline economy that represents the actual history. I then calculate a counterfactual economy in the absence of the Great Migration, modeled by prohibiting all migration flows from the South to the North between 1940 and 1970. The differences between the two are interpreted as the impacts of the Second Great Migration.

The average wage of workers in the baseline is set to one and used as a numeraire. The total number of workers is 100 and constant. Hence, the labor allocation is determined solely by migration dynamics but not by births and deaths. Unless mentioned otherwise, the reported effects measure the outcomes in 1970.

6 Quantitative Results

In this section, I first outline counterfactual outcomes in terms of welfare effects and introduce contribution analysis. Here, I focus on the changes in the workers' consumption welfare (Equation 9), which is henceforth simply referred to as welfare.[48] Although this section does not directly report the capitalists' welfare, I discuss it in terms of changes in capital rents. Lastly, I show how the Great Migration shaped economic distributions through time and across geography.

6.1 Welfare effects and contribution analysis

The baseline counterfactual analysis shows that the South-to-North migration between 1940 and 1970 increased the United States consumption welfare by 0.66% per capita by 1970. The South experienced a gain of 3.20%, while the North a loss of 0.39%. Table 7 reports the baseline welfare

[48] The consumption welfare effect measures the changes in real income with non-homothetic adjustments.

Table 7: Welfare effect and the contribution of each model element.

	(1) Baseline results	(2) No factor substitution	(3) No trade adjustment	(4) No directed technical change	(5) Without all adjustments
A. Consumption welfare effect	+0.64%	-1.86%	0.18%	-0.09%	-2.94%
B. Contribution of each channel	-	[69.8%]	[12.8%]	[20.4%]	[100%]

Note: This table shows the consumption welfare effect by scenario in Row A and the contribution of each model element in Row B. The baseline analysis quantifies the impact of the Great Migration by restricting the migration from the South to the North between 1940 and 1970. The contribution analysis compares the difference between the welfare effects of the full model and a constrained version by turning off each model component.

effect for the contiguous U.S. and the contribution of each model element in generating welfare. I examine the contribution of factor substitution, trade adjustment, and directed technical change in response to the Second Great Migration. Specifically, I use consumption welfare as the criteria since it is the main outcome of the simulation, summarizing all functions and interactions of the model elements.[49]

Table 7, Row A, reports the welfare effect for the baseline and restricted scenarios. As a benchmark, I run a scenario where all three adjustment mechanisms are held fixed. In other words, I fix the share of labor and capital allocated to each industry, trade share, and the level of factor-augmenting technology to the baseline level in the absence of the migration flows. Column 5 reports that the fully restricted model yields a welfare effect of -2.94% from the Second Great Migration. I then run a constrained model separately for each channel by turning off one model component at a time. The difference in welfare between the baseline model and the one-channel constrained model, divided by the difference in welfare between the baseline model and the fully restricted model, is interpreted as the contribution of each model element.

Row B, Columns 2 to 4, reports the results. As shown in Column 2, factor substitution takes into account the lion's share of the response to the South-to-North migration flows, driving 69.8% of the adjustments. The trade adjustment and directed technical change played supplementary roles, and each contributing to 12.8% and 20.4%. The total equals 103.1%, where the excess of the 3.1% is generated by interaction effects.

Here, the trade adjustment measures how the changes in trade share alleviated the potential welfare loss. Given that the changes in trade share are driven by the changes in factor prices, the

[49] Although this procedure is not a formal decomposition analysis, the exercise provides a useful gauge of which model element is driving the simulation results. For instance, Chor (2010) conducts a similar contribution analysis using how the welfare effects change by turning off each model component to quantify the relative importance of different sources of comparative advantage.

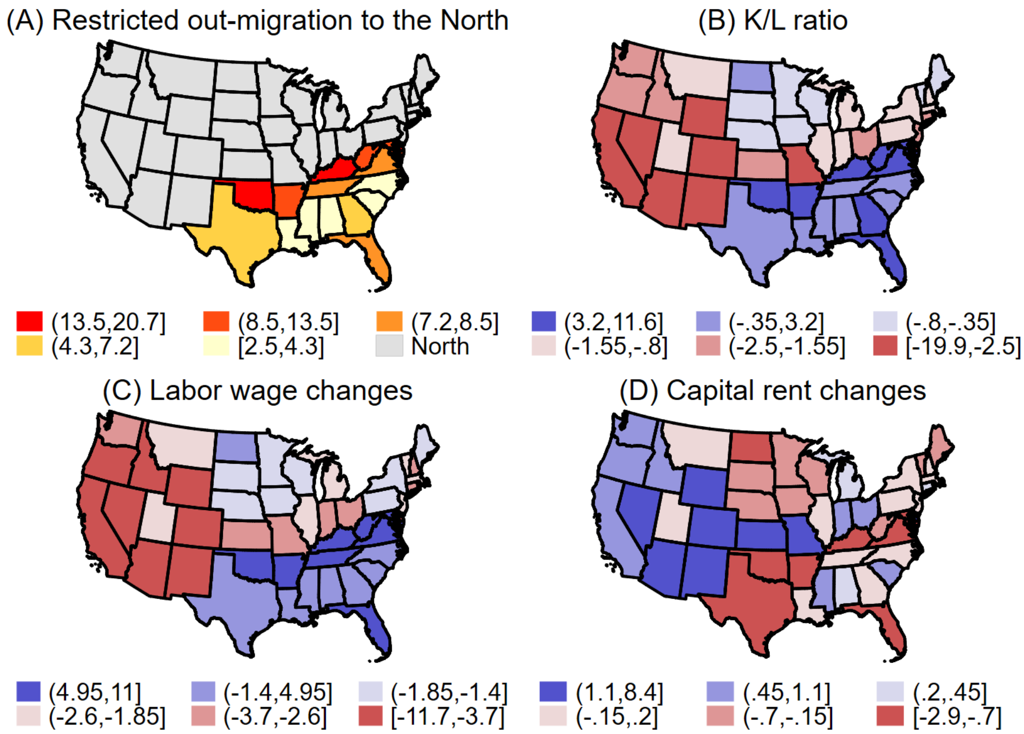

Figure 7: Maps of the Great Migration and outcome variables.

Note: The figure shows the geographic distributions of the shock and outcome variables in the contiguous United States. Panel A plots the predicted out-migration during 1940 and 1970 using Equation (4). Red means higher levels of predicted out-migration relative to yellow. Panels B to D map the simulation outcome evaluated in the year 1970 for the aggregate capital-to-labor ratio, wage of labor, and rental rate of capital. Red indicates a decrease, while blue means an increase, with a darker color representing a larger absolute size. The predicted out-migration between 1940 and 1970 is used as a shock that represents the Second Great Migration.

trade adjustment channel can be interpreted as capturing the quasi-Rybczynski effect. Given the model parameters, the Heckscher–Ohlin forces played a less pivotal role in the adjustments.

6.2 Distribution of the shock

Figure 7 displays the geographical distribution of the shock and model outcomes. The map highlights the relationship between the shock and the main mechanism captured through the model. Panel A shows the South-to-North migration rate for the Southern states, the shock used in the counterfactual analysis. Red means higher levels of out-migration relative to yellow. The Northern states are denoted as grey.

Panels B to D plot the changes in labor wage, capital rents, and capital-to-labor ratio in 1970. Red indicates a decrease, while blue means an increase, with a darker color representing a larger absolute size. The effect size is defined as the percentage point changes in the outcomes due to

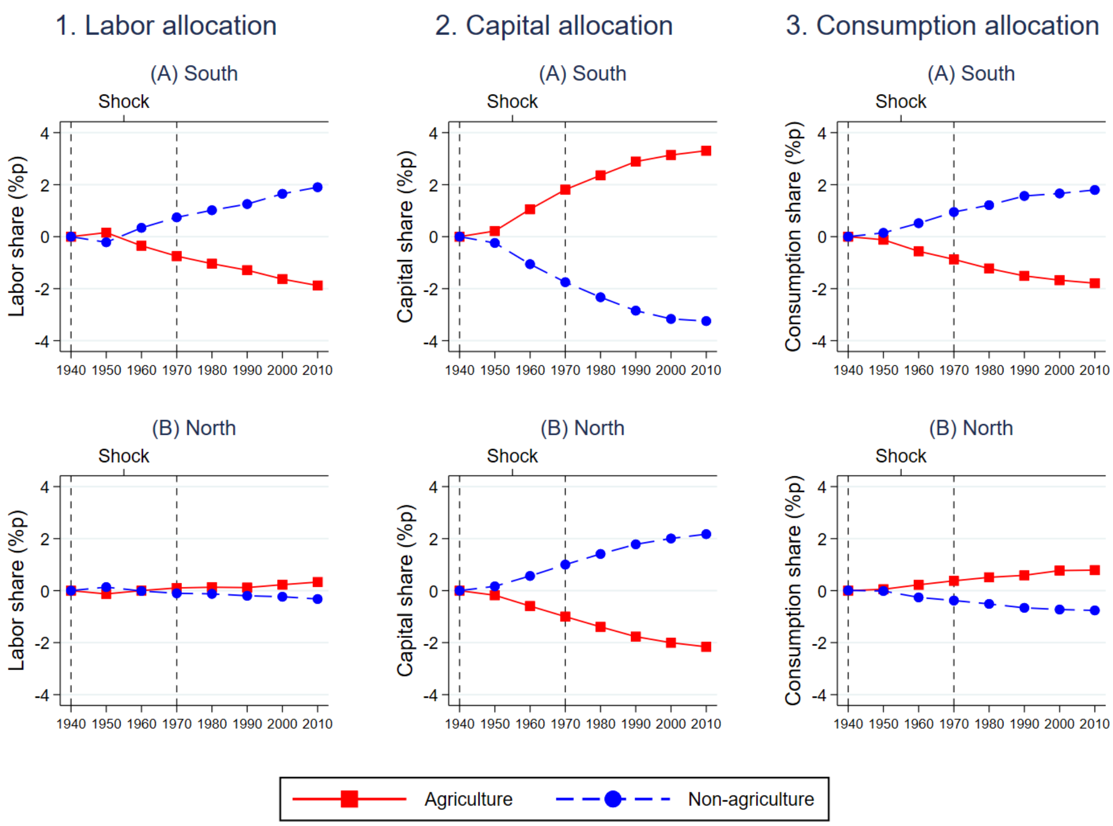

Figure 8: Simulated Changes in Economic Allocation by Region, 1940-2010.

Note: The figure shows the time trends in sectoral allocation of labor, capital, and consumption between 1940 and 2010. The effect sizes are defined as the percentage point changes in outcomes in each region. The red straight line represents agriculture, and the blue dashed line indicates non-agriculture.

the modeled Great Migration. Panel B shows the changes in aggregate capital-to-labor ratio. It contains the mechanical change from the labor decrease from the Great Migration and the endogenous response from capital accumulation. The Southern states increased in capital more relative to labor, while the opposite pattern held for the North, especially for the states that received larger migrants, such as California.

Panels C and D report the impacts of the Great Migration on factor prices. Compared to patterns in Panel B, Panel C documents a similar distribution of wage changes, and Panel D shows the opposite pattern in terms of the rental rate of capital. The region that accumulated more capital relative to labor tended to increase wages but decrease the rental rate of capital.

Figure 8 plots the changes in economic distribution between 1940 and 2010. For Figure 8, I report the results from a model that turns off the factor-augmenting technology channel. The aim here is to evaluate the impacts of the South-to-North migration flows solely based on the existing parameters estimated in the 20th-century United States setting. The Figure shows the changes in the share of labor (Panel 1), capital (Panel 2), and consumption spending (Panel 3) allocated to agriculture (red line) and non-agriculture (blue dashed line), separately for the South in Row A and the North in Row B. The non-agriculture results are aggregated.

Panel 1, Row A, suggests that the Great Migration led to a structural change in labor allocation. The relative labor scarcity incentivized the flexible sector, agriculture, to substitute labor with capital. Concretely, the model calculates that the Great Migration and the following adjustment decreased agricultural employment share by around 2% (Panel 1, Row A). Such a decrease constituted around 7% of the total decrease in agricultural employment during the study period, given that the agricultural employment share decreased from 30% to 2% in the South between 1940 and 2010. Hence, in the view of the model, the Great Migration played a supplementary yet important role in reallocating labor out of agriculture. On the contrary, the share of capital allocated to agriculture increased (Panel 2, Row A).

7 Conclusion

This paper proposes a new perspective on the economic development of the American South during the 20th-century by focusing on the role of labor scarcity in inducing capital accumulation and capital-augmenting technical change. In response to the out-migration, flexible agriculture substituted labor with capital, while the open economy force depressed the size of labor-intensive agriculture. The following labor-capital reallocation induced structural change out of agriculture, expansion of non-agriculture, and capital-biased technical change. These mechanisms highlight a potential channel for inducing structural change by encouraging out-migration in rural areas with a significant share of agricultural workers.

In 2019, around 1.2 billion people worked in the agricultural sector globally, constituting approximately 28% of the employed population, with a significant portion residing in rural areas in low-income countries (Davis et al., 2023). Still to this day, the agricultural sector in developing countries is characterized by low labor productivity compared to the non-agricultural counterpart, due to its ineffective use of labor and land (Gollin et al., 2014; Adamopoulos and Restuccia, 2014; Chen et al., 2023), or relative shortages of capital and intermediate inputs (Gollin and Udry, 2021; Boppart et al., 2023). Hence, policies that can facilitate rural economic development and structural change out of agriculture could yield substantial gains. Further research on the potential costs and benefits of out-migration on the origin's economy in modern contexts could be highly valuable.

References

Abramitzky, Ran, Philipp Ager, Leah Boustan, Elior Cohen, and Casper W. Hansen, "The Effect of Immigration Restrictions on Local Labor Markets: Lessons from the 1920s Border Closure," *American Economic Journal: Applied Economics*, 2023, *15* (1), 164–91.

Acemoglu, Daron, "Directed Technical Change," *The Review of Economic Studies*, 2002, *69* (4), 781–809.

— , "Equilibrium Bias of Technology," *Econometrica*, 2007, *75* (5), 1371–1409.

— , "When Does Labor Scarcity Encourage Innovation?," *Journal of Political Economy*, 2010, *118* (6), 1037 – 1078.

— **and Veronica Guerrieri**, "Capital deepening and nonbalanced economic growth," *Journal of Political Economy*, 2008, *116* (3), 467–498.

Adamopoulos, Tasso and Diego Restuccia, "The Size Distribution of Farms and International Productivity Differences," *American Economic Review*, June 2014, *104* (6), 1667–97.

Akram, Agha Ali, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Effects of Emigration on Rural Labor Markets," *Unpublished Manuscript*, 2018.

Alonso-Carrera, Jaime and Xavier Raurich, "Labor mobility, structural change and economic growth," *Journal of Macroeconomics*, 2018, *56* (April), 292–310.

Alvarez-Cuadrado, Francisco, Ngo Van Long, and Markus Poschke, "Capital-labor substitution, structural change, and growth," *Theoretical Economics*, 2017, *12* (3), 1229–1266.

Andersson, David, Mounir Karadja, and Erik Prawitz, "Mass Migration and Technological Change," *Journal of the European Economic Association*, 05 2022, *20* (5), 1859–1896.

Arrow, Kenneth J., "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 1962, *29* (3), 155–173.

— **, Hollis B. Chenery, B. S. Minhas, and Robert M. Solow**, "Capital-Labor Substitution and Economic Efficiency," *The Review of Economics and Statistics*, 1961, *43* (3), 225–250.

Artuc, Erhan, Shubham Chaudhuri, and John McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, 2010, *100* (3), 1008–1045.

Banerjee, Abhijit and Esther Duflo, "Growth Theory through the Lens of Development Economics," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, 1 ed., Vol. 1, Part A, Elsevier, 2005, chapter 07, pp. 473–552.

Barro, Robert J. and Xavier Sala i Martin, "Convergence," *Journal of Political Economy*, 1992, *100* (2), 223–251.

Bateman, Fred and Thomas Weiss, "A Deplorable Scarcity: The Failure of Industrialization in the Slave Economy.," 1981.

Batistich, Mary Kate and Timothy N Bond, "Stalled Racial Progress and Japanese Trade in the 1970s and 1980s," *The Review of Economic Studies*, 01 2023, *90* (6), 2792–2821.

Baumol, William J., "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *The American Economic Review*, 1967, *57* (3), 415–426.

Bazzi, Samuel, Andreas Ferrara, Martin Fiszbein, Thomas Pearson, and Patrick A Testa, "The Other Great Migration: Southern Whites and the New Right," *The Quarterly Journal of Economics*, 03 2023, *138* (3), 1577–1647.

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

Boppart, Timo, "Structural Change and the Kaldor Facts in a Growth Model With Relative Price Effects and Non-Gorman Preferences," *Econometrica*, 2014, *82* (6), 2167–2196.

_ , Patrick Kiernan, Per Krusell, and Hannes Malmberg, "The Macroeconomics of Intensive Agriculture," *Unpublished Manuscript*, 2023.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

_ , Rafael Dix-Carneiro, and Brian K. Kovak, "Understanding Migration Responses to Local Shocks," June 2023. Unpublished Manuscript.

Boustan, Leah, *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, Princeton: Princeton University Press, 2016.

Boustan, Leah Platt, "Was postwar suburbanization "white flight"'? Evidence from the black migration," *Quarterly Journal of Economics*, 2010, *125*, 417–443.

Bowles, Gladys K., James D. Tarver, Calvin L. Beale, and Everette S. Lee, "Net Migration of the Population by Age, Sex, and Race, 1950-1970," 2016. ICPSR08493-v2.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley Wilbert, "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," 2023. Working Paper.

Burstein, Ariel and Jonathan Vogel, "International Trade, Technology, and the Skill Premium," *Journal of Political Economy*, 2017, *125* (5), 1356–1412.

Bustos, Paula, Gabriel Garber, and Jacopo Ponticelli, "Capital Accumulation and Structural Transformation," *The Quarterly Journal of Economics*, 01 2020, *135* (2), 1037–1094.

Caballero, Maráa Esther, Brian Cadena, and Brian K Kovak, "The International Transmission of Local Economic Shocks Through Migrant Networks," *Journal of International Economics*, 2023, *145.*

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

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 01 2001, *19* (1), 22–64.

Caron, Justin, Thibault Fally, and James R. Markusen, "International Trade Puzzles: A Solution Linking Production and Preferences," *The Quarterly Journal of Economics*, 2014, *129* (3), 1501–1552.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, "Farm Employment: 1910–1999," 2006.

Caselli, Francesco and Wilbur John Coleman II, "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 2001, *109* (3), 584–616.

Caunedo, Julieta and Elisa Keller, "Capital-Embodied Structural Change," *Unpublished Manuscript*, 2024.

Census Bureau, "Economic Census," 2013. https://www.census.gov/data/developers/data-sets/economic-census.2012.html#list-tab-1462300713.

_ , "County Business Patterns," 2023. Accessed: 2023-05-21.

Chapel, Jack and Yi-Ju Hung, "The Great Migration's Impact on Southern Inequality," *Unpublished Manuscript*, 2024.

Chen, Chaoran, Diego Restuccia, and Raul Santaeulàlia-Llopis, "Land Misallocation and Productivity," *American Economic Journal: Macroeconomics*, April 2023, *15* (2), 441–65.

Chirinko, Robert S., "σ: The long and short of it," *Journal of Macroeconomics*, 2008, *30* (2), 671–686. The CES Production Function in the Theory and Empirics of Economic Growth.

Chor, David, "Unpacking sources of comparative advantage: A quantitative approach," *Journal of International Economics*, 2010, *82* (2), 152–167.

Clay, Karen, Ethan J Schmick, and Werner Troesken, "The Boll Weevil's Impact on Racial Income Gaps in the Early Twentieth Century," *NBER Working Paper No. 27101*, 2020.

Clemens, Michael A., Ethan G. Lewis, and Hannah M. Postel, "Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion," *American Economic Review*, June 2018, *108* (6), 1468–87.

Collins, William J., "When the Tide Turned: Immigration and the Delay of the Great Black Migration," *The Journal of Economic History*, 1997, *57* (3), 607–632.

— , "The Great Migration of Black Americans from the US South: A guide and interpretation," *Explorations in Economic History*, 2021, *80*, 1–17.

— **and Marianne H. Wanamaker**, "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, January 2014, *6* (1), 220–52.

— **and** — , "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants," *The Journal of Economic History*, 2015, *75* (4), 947–992.

Conley, T.G., "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

David, Paul A. and Th. van de Klundert, "Biased Efficiency Growth and Capital-Labor Substitution in the U.S., 1899-1960," *The American Economic Review*, 1965, *55* (3), 357–394.

Davis, Benjamin, Esin Mane, Lale Yalçın Gurbuzer, Giulia Caivano, Natalia Piedrahita, Karen Schneider, Nabila Azhar, Mahdi Benali, Naveed Chaudhary, Riad Rivera, Ramanan Ambikapathi, and Paul Winters, "Estimating global and country-level employment in agrifood systems," Technical Report 23-34, Food and Agriculture Organization of the United Nations, Rome 2023.

Day, Richard H., "The Economics of Technological Change and the Demise of the Sharecropper," *American Economic Review*, 1967, *57* (3), 427–449.

Depew, Briggs, Price V. Fishback, and Paul W. Rhode, "New deal or no deal in the Cotton South: The effect of the AAA on the agricultural labor structure," *Explorations in Economic History*, 2013, *50* (4), 466–486.

Derenoncourt, Ellora, "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, 2022, *112* (2), 369–408.

Dinkelman, T, G Kumchulesi, and M Mariotti, "Labor Migration and the Structure of Rural Labor Markets," *Review of Economics and Statistics*, Forthcoming.

Duernecker, Georg, Berthold Herrendorf, and ákos Valentinyi, "Structural Change within the Services Sector and the Future of Cost Disease," *Journal of the European Economic Association*, 05 2023, *22* (1), 428–473.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eckert, Fabian and Michael Peters, "Spatial Structural Change," *Unpublished Manuscript*, 2023.

— **, Andrés Gvirtz, Jack Liang, and Michael Peters**, "A Method to Construct Geographical Crosswalks with an Application to Us Counties since 1790," *Unpublished Manuscript*, 2020.

__ , **Ka leung Lam, Atif Mian, Karsten Muller, Rafael Schwalb, and Amir Sufi**, "The Early County Business Pattern Files: 1946-1974," Technical Report, National Bureau of Economic Research 2022.

Egger, Dennis, Benjamin Faber, Ming Li, and Wei Lin, "Rural-Urban Migration and Market Integration," *Unpublished Manuscript*, 2024.

Fajgelbaum, Pablo and Stephen J. Redding, "Trade Structural Transformation and Development: Evidence from Argentina 1869-1914," *Journal of Political Economy*, 2018, *130* (5), 1249–1318.

Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti, "Growing Like India—the Unequal Effects of Service-Led Growth," *Econometrica*, 2023, *91* (4), 1457–1494.

Feigenbaum, James J, Soumyajit Mazumder, and Cory B Smith, "When Coercive Economies Fail: The Political Economy of the US South After the Boll Weevil," Working Paper 27161, National Bureau of Economic Research 2020.

Ferleger, Louis, "Sharecropping contracts in the late-nineteenth-century South," *Agricultural History*, 1993, *67* (3), 31–46.

Foster, Andrew D. and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Fuguitt, Glenn V., Calvin L. Beale, and Paul R. Voss, "County-Specific Net Migration Estimates, 1980-1990 [United States]," 2010. ICPSR26761.

Gallardo, R Karina and Johannes Sauer, "Adoption of Labor-Saving Technologies in Agriculture," *Annual Review of Resource Economics*, 2018, *10*, 185–206.

Gardner, John and William Cohen, "Demographic Characteristics of the Population of the United States, 1930-1950: County-Level," 1992.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gollin, Douglas and Christopher Udry, "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," *Journal of Political Economy*, 2021, *129* (1), 1–80.

__ , **David Lagakos, and Michael E. Waugh**, "The agricultural productivity gap," *Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

__ , **Stephen Parente, and Richard Rogerson**, "The Role of Agriculture in Development," *American Economic Review*, May 2002, *92* (2), 160–164.

Gregory, James N., *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*, Chapel Hill: University of North Carolina Press, 2005.

Gross, Daniel P., "Scale versus scope in the diffusion of new technology: evidence from the farm tractor," *RAND Journal of Economics*, 2018, *49* (2), 427–452.

Grossman, James R., *Land of Hope: Chicago, Black Southerners, and the Great Migration*, Chicago, IL: University of Chicago Press, 1989.

Grove, Wayne A. and Craig Heinicke, "Better Opportunities or Worse? The Demise of Cotton Harvest Labor, 1949-1964," *The Journal of Economic History*, 2003, *63* (3), 736–767.

Haines, Michael, Price Fishback, and Paul Rhode, "United States Agriculture Data, 1840 - 2012," 2018. Inter-university Consortium for Political and Social Research [distributor].

Haines, Michael R., Inter university Consortium for Political, and Social Research, "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002," 2010. V3.

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *The American Economic Review*, 1970, *60* (1), 126–142.

Herrendorf, Berthold, Christopher Herrington, and ákos Valentinyi, "Sectoral Technology and Structural Transformation," *American Economic Journal: Macroeconomics*, October 2015, *7* (4), 104–33.

Hicks, John R., *The Theory of Wages*, London: Macmillan, 1932.

Holley, Donald, *The Second Great Emancipation: The Mechanical Cotton Picker, Black Migration, and How They Shaped the Modern South*, Fayetteville, AR: University of Arkansas Press, 2000.

Hornbeck, Richard and Suresh Naidu, "When the Levee Breaks: Black Migration and Economic Development in the American South," *American Economic Review*, March 2014, *104* (3), 963–90.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 11 2009, *124* (4), 1403–1448.

Hulten, Charles R. and Frank C. Wykoff, "The Estimation of Economic Depreciation Using Vintage Asset Prices: An Application of the Box-Cox Power Transformation," *Journal of Econometrics*, 1981, *15* (3), 367–396.

Jaeger, David A, Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," NBER Working Paper 24285, National Bureau of Economic Research February 2018.

Jaworski, Taylor and Dongkyu Yang, "Did War Mobilization Cause Regional Development?," *Unpublished Manuscript*, 2024.

Jones, Ronald W., "The Structure of Simple General Equilibrium Models," *Journal of Political Economy*, 1965, *73* (6), 557–572.

Jung, Yeonha, "The long reach of cotton in the US South: Tenant farming, mechanization, and low-skill manufacturing," *Journal of Development Economics*, 2020, *143*, 102432.

Karabarbounis, Loukas and Brent Neiman, "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 2014, *129* (1), 61–103.

Kennedy, Charles, "Induced Bias in Innovation and the Theory of Distribution," *The Economic Journal*, 09 1964, *74* (295), 541–547.

Kirby, Jack Temple, "The Southern Exodus, 1910-1960: A Primer for Historians," *The Journal of Southern History*, Nov 1983, *49* (4), 585–600.

Kleinman, Benny, Ernest Liu, and Stephen J. Redding, "Dynamic Spatial General Equilibrium," *Econometrica*, 2023, *91* (2), 385–424.

Lafortune, Jeanne, José Tessada, and Carolina González-Velosa, "More hands, more power? Estimating the impact of immigration on output and technology choices using early 20th century US agriculture," *Journal of International Economics*, 2015, *97* (2), 339–358.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh, "The Welfare Effects of Encouraging Rural-Urban Migration," *Econometrica*, 2023, *91* (3), 803–837.

Lewis, W. Arthur, "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, 1954, *22* (2), 139–191.

León-Ledesma, Miguel A., Peter McAdam, and Alpo Willman, "Identifying the Elasticity of Substitution with Biased Technical Change," *American Economic Review*, September 2010, *100* (4), 1330–57.

Manuelli, Rodolfo E. and Ananth Seshadri, "Frictionless Technology Diffusion: The Case of Tractors," *American Economic Review*, April 2014, *104* (4), 1368–91.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding, "Urbanization and structural transformation," *Quarterly Journal of Economics*, 2012, *127* (2), 535–586.

Montrose, Paige, "Without Deliberate Speed: The Effects of Southern Out-Migration on School Desegregation," *Unpublished Manuscript*, 2024.

Nelder, John A. and Roger Mead, "A simplex method for function minimization," *The Computer Journal*, 1965, *7* (4), 308–313.

Nigai, Sergey, "On Measuring the Welfare Gains from Trade Under Consumer Heterogeneity," *Economic Journal*, June 2016, *126* (593), 1193–1237.

Oberfield, Ezra and Devesh Raval, "Micro Data and Macro Technology," *Econometrica*, 2021, *89* (2), 703–732.

Olea, José Luis Montiel and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business Economic Statistics*, 2013, *31* (3), 358–369.

Olmstead, Alan L. and Paul W. Rhode, “ Reshaping the Landscape: the Impact and Diffusion of the Tractor in American Agriculture, 1910-1960 ,” *The Journal of Economic History*, 2001, *61* (3), 663–698.

Peterson, Willis and Yoav Kislev, “The Cotton Harvester in Retrospect: Labor Displacement or Replacement?,” *The Journal of Economic History*, 1986, *46* (1), 199–216.

Piketty, Thomas, *Capital in the Twenty-First Century*, Cambridge, MA: Belknap Press, 2014.

Porzio, Tommaso, Federico Rossi, and Gabriella Santangelo, “The Human Side of Structural Transformation,” *American Economic Review*, 2022, *112* (8), 2774–2814.

Raper, Arthur, “The Role of Agricultural Technology in Southern Social Change,” *Social Forces*, 1946, *25* (1), 21–30.

Romalis, John, “Factor proportions and the structure of commodity trade,” *American Economic Review*, 2004, *94* (1), 67–97.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren, “IPUMS Ancestry Full Count Data: Version 4.0,” dataset 2024.

_ , Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler, “IPUMS USA: Version 15.0,” 2024. [dataset].

Rybczynski, T. M., “Factor Proportions and Comparative Advantage,” *Economica*, 1955, *22* (88), 336–341.

San, Shmuel, “Labor Supply and Directed Technical Change: Evidence from the Termination of the Bracero Program in 1964,” *American Economic Journal: Applied Economics*, January 2023, *15* (1), 136–63.

Stuart, Bryan A. and Evan J. Taylor, “Migration Networks and Location Decisions: Evidence from US Mass Migration,” *American Economic Journal: Applied Economics*, July 2021, *13* (3), 134–75.

Takahashi, Motoaki, “The Aggregate Effects of the Great Black Migration,” *Unpublished Manuscript*, 2023.

Theoharides, Caroline, “Manila to Malaysia, Quezon to Qatar: International Migration and its Effects on Origin-Country Human Capita,” *Journal of Human Resources*, 2018, *53* (4), 1022–1049.

Voss, Paul R., Scott McNiven, Roger B. Hammer, Kenneth M. Johnson, and Glenn V. Fuguitt, “County-Specific Net Migration by Five-Year Age Groups, Hispanic Origin, Race, and Sex, 1990-2000: [United States],” 2005. ICPSR04171.

Whatley, Warren C, "A History of Mechanization in the Cotton South: The Institutional Hypothesis," *The Quarterly Journal of Economics*, 1985, *100* (4), 1191–1215.

Winkler, Richelle, Kenneth Johnson, Cheng Cheng, Paul Voss, and Katherine J. Curtis, "County-Specific Net Migration by Five-Year Age Groups, Hispanic Origin, Race, and Sex: 2000-2010," 2013. ICPSR34638.

Wright, Gavin, *Old South, New South: Revolutions in the Southern Economy since the Civil War*, New York: Basic Books, 1986.

Yang, Dongkyu, "A General Equilibrium Investigation of the American Dust Bowl," *Unpublished Manuscript*, 2024.

Appendix.

Table A1: Zero-stage in-migration prediction.

	(1)	(2)	(3)	(4)	(5)	(6)
	1940-1950		1950-1960		1960-1970	
	Black	White	Black	White	Black	White
Latitude	-5.550	-0.154	-7.302*	-0.377	-3.029*	-1.355***
	(3.605)	(0.227)	(3.937)	(0.487)	(1.828)	(0.226)
Longitude	-4.769***	0.035	-2.099	-0.412*	-1.936**	-0.248**
	(1.702)	(0.107)	(1.907)	(0.236)	(0.914)	(0.113)
Log population	67.285	3.636	-117.904	21.049	12.575	13.916**
	(73.652)	(4.636)	(106.951)	(13.216)	(49.977)	(6.172)
Log black population	-15.663***	0.101	-63.246***	1.234**	-4.314**	0.490*
	(3.423)	(0.215)	(4.044)	(0.500)	(2.076)	(0.256)
Log white population	-71.899	-3.410	162.889	-29.611**	-11.011	-13.347**
	(74.179)	(4.669)	(108.079)	(13.356)	(50.658)	(6.256)
Urbanization	-0.987***	0.057***	-0.813***	0.251***	-0.307**	-0.029*
	(0.279)	(0.018)	(0.301)	(0.037)	(0.137)	(0.017)
Median income	-3.979	0.016	26.244***	15.626***	14.718***	9.442***
	(4.364)	(0.275)	(8.928)	(1.103)	(4.577)	(0.565)
Log housing units	12.854**	-0.294	14.702**	0.819	2.392	-1.024***
	(5.283)	(0.333)	(5.736)	(0.709)	(2.646)	(0.327)
Median rent	-3.305	-0.064	-7.525	5.714***	-11.123***	5.218***
	(5.549)	(0.349)	(7.098)	(0.877)	(3.430)	(0.424)
1940 Share naturalized	3.413	-0.559***	6.187*	-0.407	5.571***	0.467**
	(3.138)	(0.198)	(3.422)	(0.423)	(1.573)	(0.194)
1940 Share foreigner	-1.768	-0.365	3.385	0.259	-1.126	-0.185
	(3.633)	(0.229)	(3.977)	(0.491)	(1.840)	(0.227)
1940 Employment rate	3.579**	0.874***	1.025	-0.163	1.328	0.469***
	(1.622)	(0.102)	(1.802)	(0.223)	(0.840)	(0.104)
1940 Occupational score	0.036	0.018**	0.227*	0.015	-0.061	-0.020***
	(0.117)	(0.007)	(0.130)	(0.016)	(0.060)	(0.007)
Republican vote share (1944)	-0.055	-0.072***	0.332	0.174***	-0.246	-0.041
	(0.438)	(0.028)	(0.484)	(0.060)	(0.224)	(0.028)
Republican vote share (1948)	-1.028**	0.032	-0.827*	0.012	-0.252	0.087***
	(0.449)	(0.028)	(0.492)	(0.061)	(0.227)	(0.028)
Republican vote share (1952)	0.478	-0.155***	0.052	0.004	0.600*	-0.007
	(0.632)	(0.040)	(0.695)	(0.086)	(0.320)	(0.039)
Republican vote share (1956)	0.101	0.234***	-0.664	-0.135*	-0.233	0.083**
	(0.577)	(0.036)	(0.633)	(0.078)	(0.291)	(0.036)
Republican vote share (1960)	-0.312	0.002	0.494	0.133**	0.090	0.044
	(0.480)	(0.030)	(0.525)	(0.065)	(0.242)	(0.030)
Republican vote share (1964)	0.169	-0.011	-0.533	-0.153***	-0.035	-0.133***
	(0.412)	(0.026)	(0.453)	(0.056)	(0.211)	(0.026)
Republican vote share (1968)	1.429**	0.028	1.636**	0.100	0.631**	0.129***
	(0.600)	(0.038)	(0.660)	(0.082)	(0.305)	(0.038)
Republican vote share (1972)	-0.090	-0.017**	0.071	0.005	-0.048	-0.016*
	(0.134)	(0.008)	(0.147)	(0.018)	(0.068)	(0.008)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observation	3,102	3,102	3,092	3,092	3,080	3,080
R^2	0.152	0.234	0.333	0.261	0.114	0.525

Note: This table reports zero-stage in-migration prediction for the using OLS regression, with county as the unit of observation. The dependent variables, net migration rates by race and by decade, are from Gardner and Cohen (1992) and Bowles et al. (2016). The explanatory variables are listed as variable names. All specifications include state fixed effects. The prediction sample includes all counties in the U.S. Robust standard errors are clustered by county and reported in parentheses. Stars represent: $** \ p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.001$.

Table A2: Exclusion restriction - Pretrend tests.

Panel A. Agriculture

	(1) Employment	(2) Number of farms	(3) Acres in farmland	(4) Number of tractors	(5) Farm output	(6) Farm value per acre
Migration exposure (SSIV, 1std)	0.051***	0.009	0.014	0.029	0.038***	-0.004
Clustered s.e. (county)	(0.013)	(0.011)	(0.018)	(0.021)	(0.012)	(0.013)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.422	0.411	0.201	0.136	0.478	0.298
First-stage F	12.97	12.97	15.99	16.71	12.97	12.97
Counties	1,096	1,096	1,095	1,090	1,096	1,096

Panel B. Manufacturing

	(1) Employment	(2) Number of establishments	(3) Value added	(4) Annual payroll	(5) Intermediate goods	(6) Revenue
Migration exposure (SSIV, 1std)	-0.058	0.022	-0.041	-0.066	-0.007	-0.021
Clustered s.e. (county)	(0.041)	(0.019)	(0.033)	(0.042)	(0.042)	(0.036)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.172	0.204	0.207	0.197	0.195	0.216
First-stage F	14.36	13.05	12.69	12.69	12.69	12.69
Counties	1,035	1,058	994	994	994	994

Panel C. Wholesale

	(1) Wholesale employment	(2) Wholesale establishment	(3) Wholesale sales	(4) Wholesale annual payroll
Migration exposure (SSIV, 1std)	-0.035	-0.001	-0.036	0.003
Clustered s.e. (county)	(0.038)	(0.030)	(0.033)	(0.036)
Fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R-squared	0.133	0.125	0.146	0.183
First-stage F	14.52	14.83	14.77	13.13
Counties	947	1,060	953	943

Note: The table reports estimation results using Equation (4) on pre-period outcomes (1920 and 1930 for agriculture and manufacturing, 1930 for wholesale variables), with county-year as the unit of observation. Panels A to C correspond to the baseline results in Tables 2, 3, and 4 with the full set of fixed effects and control variables. Each column reports the changes in the indicated outcome variable in logs for the years 1920 and 1930, relative to the omitted years of 1940. Robust standard errors are clustered by county and reported in parentheses, and the first-stage Kleibergen-Paap robust F-statistics are reported. Counties with zero values before taking logs for the dependent variable are dropped from the sample. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table A3: SSIV second-stage estimates for main outcomes.

Panel A. Agriculture

	(1) Employment	(2) Number of farms	(3) Acres in farmland	(4) Number of tractors	(5) Number of combines	(6) Farm output	(7) Farm value per acre
Out-migration rate (2SLS, 1%)	-0.011*	-0.006*	-0.023***	0.042***	0.014	-0.004	0.005
Clustered s.e. (County)	(0.007)	(0.004)	(0.006)	(0.010)	(0.020)	(0.004)	(0.003)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F	26.59	24.62	24.36	25.40	18.48	14.64	26.76
Counties	1,096	1,096	1,096	1,090	1,058	1,090	1,090

Panel B. Manufacturing

	(1) Employment	(2) Number of establishment	(3) Capital spending	(4) Value added	(5) Annual payroll
Out-migration rate (2SLS, 1%)	0.022**	0.026***	0.040	0.024	0.043**
Clustered s.e. (County)	(0.011)	(0.007)	(0.028)	(0.021)	(0.017)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
First-stage F	23.03	24.33	15.54	21.07	22.23
Counties	1,096	1,096	1,065	1,096	1,096

Panel C. Wholesale and retail

	(1) Wholesale emp.	(2) Wholesale est.	(3) Wholesale sales	(4) Wholesale payroll	(5) Retail emp.	(6) Retail est.	(7) Retail sales	(8) Retail payroll
Out-migration rate (2SLS, 1%)	0.047***	0.040***	0.026*	0.035**	0.021***	0.010***	0.015***	0.022***
Clustered s.e. (County)	(0.011)	(0.010)	(0.014)	(0.015)	(0.006)	(0.003)	(0.005)	(0.007)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F	27.46	24.93	27.57	22.70	25.07	25.06	25.07	25.04
Counties	1,083	1,096	1,086	1,086	1,096	1,096	1,096	1,096

Note: The table reports the second-stage estimation results, with county-year as the unit of observation. Panels A to C correspond to the baseline results in Tables 2, 3, and 4 with the full set of fixed effects and control variables. The second-stage estimates are calculated by adjusting the reduced form estimates by the first-stage estimates. Robust standard errors are clustered by county and reported in parentheses. Stars represent: $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A4: **Overidentification tests using alternative shares (second-stage estimates).**

Panel A. Agriculture

	(1) Employment	(2) Number of farms	(3) Acres in farmland	(4) Number of tractors	(5) Number of combines	(6) Farm output	(7) Farm value per acre
Out-migration rate (2SLS, 1%)	-0.012*	-0.010**	-0.024***	0.039***	0.019	-0.016**	0.005
	(0.006)	(0.004)	(0.006)	(0.011)	(0.020)	(0.008)	(0.003)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F	13.69	13.16	13.16	13.22	10.93	13.16	13.16
Counties	1,096	1,096	1,096	1,096	1,060	1,096	1,096
Sargan-Hansen J	0.43	0.15	1.74	0.17	1.78	0.03	0.00
Sargan-Hansen p-value	0.51	0.70	0.19	0.68	0.18	0.87	1.00

Panel B. Manufacturing

	(1) Employment	(2) Number of establishment	(3) Capital spending	(4) Value added	(5) Annual payroll
Out-migration rate (2SLS, 1%)	0.020*	0.026***	0.044	0.019	0.039**
	(0.011)	(0.007)	(0.027)	(0.019)	(0.016)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
First-stage F	12.38	12.82	8.00	11.04	11.67
Counties	1,096	1,096	1,065	1,096	1,096
Sargan-Hansen J	1.79	0.03	1.90	4.59	4.20
Sargan-Hansen p-value	0.18	0.87	0.17	0.03	0.04

Panel C. Wholesale and retail

	(1) Wholesale emp.	(2) Wholesale est.	(3) Wholesale sales	(4) Wholesale payroll	(5) Retail emp.	(6) Retail est.	(7) Retail sales	(8) Retail payroll
Out-migration rate (2SLS, 1%)	0.052***	0.042***	0.029**	0.037***	0.019***	0.009***	0.013***	0.020***
	(0.012)	(0.010)	(0.014)	(0.014)	(0.005)	(0.003)	(0.004)	(0.006)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F	14.40	13.25	14.39	12.46	13.16	13.16	13.16	13.15
Counties	1,083	1,096	1,086	1,086	1,096	1,096	1,096	1,096
Sargan-Hansen J	5.89	0.74	1.32	0.60	12.05	2.35	13.59	15.43
Sargan-Hansen p-value	0.02	0.39	0.25	0.44	0.00	0.13	0.00	0.00

Note: The table reports two-stage estimates with two migration exposures (two excluded instruments), separately constructed from (1) 1910-1940 migration share from matched Census (Ruggles et al., 2024a; Buckles et al., 2023) and (2) 1935-1940 migration share from 1940 Census record. The unit of observation is county-year. Panels A to C correspond to the baseline results in Tables 2 to 4 with the full set of fixed effects and control variables. Each column reports the changes in the indicated outcome variable in logs for the years 1970 to 2010 by county-level net out-migration rates, relative to the omitted years of 1940 and 1950. Robust standard errors are clustered by county and reported in parentheses, and the first-stage Kleibergen-Paap robust F-statistics are reported. Heteroskedasticity-robust Sargan-Hansen J statistics under the null of constant effects and related p-values are added. Stars represent: $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A5: Quantitative Environment.

Definition	Parameter
Panel (A) Workers	
Consumption indirect utility	$C(e,P_{i,t}) = \frac{1}{\varepsilon}\left(e/P_{i,t}\right)^{\varepsilon} - \sum_s \nu^s \ln P^s_{i,t}$
Consumption share	$\varphi^s(P_{i,t},e) = \phi^s + \nu^s\left(e/P_{i,t}\right)^{-\varepsilon}$
Intertemporal preferences	$\mathbb{V}^s_{i,t} = U\left(C_{i,t}\right) + \max_{\{n\}}\left\{\beta\mathbb{E}[\mathbb{V}_{n,t+1}] - \kappa_{ni,t} + \eta\varepsilon_{n,t}\right\}$
Average wage	$\bar{w}_{i,t} = \left((A^{w,a}_{i,t}w^a_{i,t})^{\zeta^w} + (A^{w,m}_{i,t}w^m_{i,t})^{\zeta^w} + (A^{w,s}_{i,t}w^r_{i,t})^{\zeta^w}\right)^{1/\zeta^w}$
Sectoral labor allocation	$L^s_{i,t} = \Gamma_{\zeta^w}A^{w,s}_{i,t}\left(A^{w,s}_{i,t}w^s_{i,t}/\bar{w}_{,ti}\right)^{\zeta^w-1}L_{i,t} \equiv A^{w,s}_{i,t}\tilde{L}^s_{i,t}$
Panel (B) Capitalists	
Intertemporal preferences	$v^k_{i,t} = \mathbb{E}_t\sum_{t=0}^{\infty}\beta^t\log(C^k_{i,t}(e,P_{i,t}))$
Budget constraint	$\bar{r}_{i,t}K_{i,t} = P_{i,t}\left(C^k_{i,t} + K_{i,t+1} - (1-\delta)K_{i,t}\right)$
Average nominal net rental rate	$\bar{r}_{i,t} = \left((A^{K,a}_{i,t}r^a_{i,t})^{\zeta^k} + (A^{K,m}_{i,t}r^m_{i,t})^{\zeta^k} + (A^{K,l}_{i,t}r^r_{i,t})^{\zeta^k}\right)^{1/\zeta^k}$
Investment rule	$K_{i,t+1} = \beta\bar{R}_{i,t}K_{i,t}$
Sectoral capital allocation	$K^s_{i,t} = \Gamma_{\zeta^k}A^{k,s}_{i,t}\left(A^{k,s}_{i,t}r^s_{i,t}/\bar{r}_{i,t}\right)^{\zeta^k-1}K_{i,t} \equiv A^{k,s}_{i,t}\tilde{K}^s_{i,t}$
Panel (C) Production technology	
Production function	$Y^s_{i,t} = z^s_{i,t}\left(\rho^s_i(A^{w,s}_{i,t}\tilde{L}^s_{i,t})^{\frac{\sigma^s-1}{\sigma^s}} + (1-\rho^s_i)(A^{k,s}_{i,t}\tilde{K}^s_{i,t})^{\frac{\sigma^s-1}{\sigma^s}}\right)^{\frac{\sigma^s}{\sigma^s-1}}$
Unit cost	$x^s_{i,t} = \left((\rho^s_i)^{\sigma^s}(w^s_{i,t})^{1-\sigma^s} + (1-\rho^s_i)^{\sigma}(r^s_{i,t})^{1-\sigma^s}\right)^{\frac{1}{1-\sigma^s}}$
Panel (D) General equilibrium	
Migration flow	$\mathbb{M}_{in,t} = \frac{\exp\left(\beta\mathbb{E}_t(v_{n,t+1} - \kappa_{ni})/\eta\right)}{\sum_{j=1}^{N}\exp\left(\beta\mathbb{E}_t(v_{j,t+1} - \kappa_{nj,t})/\eta\right)}$
Expenditure share	$\mathbb{S}^s_{ni,t} = \frac{Z^s_{i,t}\left(x^s_{i,t}\tau^s_{ni,t}\right)^{-\theta^s}}{\sum_{j=1}^{N}Z^s_{j,t}\left(x^s_{j,t}\tau^s_{nj,t}\right)^{-\theta^s}}$
Goods market clearing	$Y^s_{i,t} = \sum_{n=1}^{N}\mathbb{S}^s_{ni,t}\left(\sum_{s=1}^{S}\varphi^s_{ns,t}E^s_{n,t} + \varphi^k_{ns,t}E^k_{n,t}\right)$

Note: The table summarizes the main model elements described in Section 5. The subscript i and n index region. For expenditure and migration share, i denotes the exporter and origin, respectively, and n represents the importer and destination. The superscript $s \in \{a,m,l\}$ denotes industry.