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

Dongkyu Yang*

September 30, 2025

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 in the South that lost workers, increasing capital per worker and output in both agriculture and manufacturing. Labor was further reallocated from agriculture to manufacturing and local services. I then develop a dynamic spatial general equilibrium model that incorporates 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 the adjustments to South-to-North migration.

Keywords: Factor abundance, Internal Migration, Labor Scarcity, Structural change.
JEL: N92, O11, R12, R23.

*KDI School of Public Policy and Management. E-mail: dongkyuyang@kdischool.ac.kr. I sincerely thank my dissertation committee members, Richard Mansfield, Sergey Nigai, Taylor Jaworski, and Wolfgang Keller, for their guidance and unwavering support. I received valuable comments and suggestions from Erhan Artuc, Hoyt Bleakley, Maggie Chen, Justin Cook, José-Antonio Espín-Sánchez, Richard Hornbeck, Maggie Jones, Yeonha Jung, Jeanne Lafortune, Brian Marein, Kris Mitchener, Paul Rhode, Martin Rotemberg, and Zachary Ward. All errors are my own.

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 (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), 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 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. I then quantitatively assess the mechanisms behind the empirical findings by developing a computational general equilibrium model featuring migration and the Heckscher–Ohlin force in trade.

First, through a simple model, I interpret the Great Migration as an economy-wide change in the relative abundance of capital and labor. The U.S. can be abstractly thought of as consisting of 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. As labor becomes relatively scarcer, agriculture more flexibly substitutes the now more expensive labor with capital. The increase in capital usage in agriculture can also induce technical change biased toward capital, further releasing agricultural workers. They are absorbed by local non-agriculture. Hence, the out-migration alone can stimulate labor reallocation across sectors. The

[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.

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; Oberfield and Raval, 2021).

The internal trade mechanism could have also operated 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; Wright, 1986). Hence, the decrease in labor endowment would have led to a relative contraction in agriculture but an increase in non-agriculture production. Such a quasi-Rybczynski effect further allows non-agricultural capital accumulation through an accompanying non-agricultural expansion.

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, between counties that experienced different levels of net out-migration.[2]

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 migration matrix as shares and Northern migration pull factors as shifts. The aim here is to capitalize on 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 Northern migration exposure between Southern counties. In essence, the instrument measures how much each Southern county is connected to Northern destinations in terms of pre-period migration shares that happened to have different levels of pull factors during the Second Great Migration. I also include state-by-year and county fixed effects so that the estimation relies on variation in relative change between counties within the same state. In the first stage regression on migration response, one standard deviation greater exposure to Northern pull factors induces 3.6% to 4.8% higher out-of-county out-migration rates between 1940 and 1970.

The main analysis presents reduced-form estimates from regressions of outcomes on Northern migration exposure. I find that stronger Northern migration pull factors, through their effect on relative labor scarcity, contributed to structural change, capital accumulation, and technol-

[2] 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.

ogy adoption in the South in the late 20th century. 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 was a key determinant of the agricultural mechanization (Manuelli and Seshadri, 2014).

However, the adjustment to out-migration did not end in agriculture. Non-agriculture results show that a 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%. Their sales and payroll also increased.

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 human capital in the South during this period, often pointed out as a source of the North-South convergence (Caselli and Coleman II, 2001).

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 constant elasticity of substitution (CES) production. 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 experienced a loss of 0.4%. A contribution analysis using welfare effects shows that the factor substitution played the major role, accounting for 70% of the total adjustment. The trade adjustment played an important supplementary role. 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 economic 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.

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 output mix adjustments (Lafortune et al., 2015), directed technological change (San, 2023), labor/capital substitution (Hornbeck and Naidu, 2014; Clemens et al., 2018; Abramitzky et al., 2023), and human capital investment (Caballero et al., 2023), among others. This paper proposes a new channel through which out-migration can lead to structural change, the origin's re-optimization of factor usage. 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 (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; Boustan 2010; Collins and Wanamaker 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). 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.

The findings in this paper relate to the structural change literature, especially to the studies that focus on the role of physical capital (Barro and Sala-i Martin, 1992; Acemoglu and Guerrieri, 2008; Alvarez-Cuadrado et al., 2017). I add empirical evidence on the mechanism underlying the structural change process. 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 driven by agricultural income growth.

This paper also adds to the literature on the economic development of the American South (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 the 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 in driving structural change.

1 Backgrounds

1.1 Southern Economic Development and the Great Migration

A large literature in economics studies the influence of the Great Migration (Kirby, 1983; Boustan, 2010; Collins and Wanamaker, 2015; Bazzi et al., 2023), while others 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 a different view and investigates how the Great Migration also transformed the Southern economy by focusing on the role of physical capital.[3]

The Great Migration, spanning 1910–1930 and 1940–1970, was one of the largest internal migrations in U.S. 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.[4] This study focuses on how the Second Great Migration (1940-1970) 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.[5] 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. 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, which abolished the immigration quota system and

[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] 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.

[5] International migration was largely shut down with the Immigration Act of 1924, which limited the number of immigrants through a national origins quota. The Bracero program (1942-1964), which allowed short-term Mexican laborers in limited industries, is a notable exception during this period.

significantly increased international migration flows thereafter and until today.

While many Southerners left their origin in pursuit of economic opportunity, the South's economy also matured during this period. By 1940, around 30% of employed Southerners were working in agricultural sectors, compared to 12% in the North, while lagging manufacturing employment by 22% to 33%. However, coinciding with the Second Great Migration, the two regions converged in terms of industry employment share at least by 1990. The non-agricultural wages also caught up with the North.

Did out-migration induce regional convergence, or did the process of structural change generate out-migration? One persuasive and pervasive view is that Southern economic changes, such as agricultural development, drove the out-migration of workers into the North and into manufacturing (Grove and Heinicke, 2003; Boustan, 2016). However, the influences may not be unidirectional, as labor scarcity from out-migration could have also spurred endogenous responses in the Southern economy. The out-migration could have incentivized the South to invest more in physical capital in response to labor scarcity, similar to other historical episodes in the United States (Hornbeck and Naidu, 2014; Clemens et al., 2018; Abramitzky et al., 2023).

Online Appendix Section A and Figure A1 illustrate economic changes in the South during and after the Great Migration, focusing on industry employment shares and capital investment by sector. While experiencing catch-up in structural change with the North, the South invested heavily in agricultural machinery and manufacturing capital, surpassing the North in agricultural mechanization by 2000 and in manufacturing capital spending throughout the 20th century. This paper explores how large-scale out-migration (Figure A1, Panel A) contributed to structural change (Panels B and C), emphasizing the role of capital deepening (Panels E and F) as a response to labor scarcity.

1.2 Economic Interpretation on the Impacts of the Great Migration

This section interprets the out-migration and economic changes in the South using a simplified two-period, two-country framework with two industries and two factors of production—labor and capital. I take the Southern out-migration as given and sketch its economic implications. The aim here is to examine how relative labor scarcity can drive structural change from tensions between labor and capital. More details are discussed in Online Appendix Section B. Empirical strategy in Section 2 is designed to replicate "out-migration as given," at least from the perspective of the South, using an SSIV strategy by isolating pull factors of Northern destinations. The dynamic spatial general equilibrium model in Section 4 generalizes the model elements into realistic geography and multiple periods. The empirical analysis checks the common predictions, and the quantitative analysis assesses the potential contribution of each model element based on simulation outcomes.

The production function is assumed to take the constant elasticity of substitution (CES) structure using labor L^s and capital K^s, in each sector s for agriculture and non-agriculture (denoted "a" and "m"):

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

where Z_L^s and Z_K^s represent labor- and capital-augmenting technologies. They are assumed not to exogenously grow but endogenously respond to the changes in factor allocation through a directed technical change process governed by the weak equilibrium bias (Acemoglu, 2002, 2007). I interpret capital as mobile across sectors but a geographically immobile variable factor, such as local structures for production. I assume that both factors are fully employed and perfectly mobile across sectors within the region.

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 the CES elasticity estimates in the literature.[6] In other words, agriculture is assumed to be more flexible in substituting between labor and capital, whereas the two factors are closer to complements in non-agriculture. Because the two sectors combine factors in heterogeneous ways, a common shock can lead to a factor reallocation across sectors in opposite directions.

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 parameter values. Here, I assume that Southern agriculture is labor-intensive, with labor accounting for a larger share of production costs than in non-agriculture. I also assume that the South was labor-abundant before the Great Migration, with a lower capital-to-labor ratio than the North, consistent with the "Old South" method of production (Bateman and Weiss, 1981; Wright, 1986). These assumptions imply that the South had a comparative advantage in labor-intensive agriculture, at least before the out-migration.

As population flows out, labor becomes scarcer and more expensive relative to capital. Because of the different factor substitution patterns, agriculture is more flexible in substituting now scarce labor, releasing labor and absorbing capital. Given the model setting, the shares of labor and capital move in opposite directions from changes in the relative abundance of labor and capital (Alvarez-Cuadrado et al., 2017), implying the labor share in non-agriculture would conversely increase. Hence, a response to labor scarcity across sectors can result in structural change, reallocating labor from agriculture to non-agriculture.

[6] 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. 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.

The factor cost share influences economic allocation through internal trade between the South and the North. Given agriculture's higher labor cost share, a decline in the South's labor endowment would raise agricultural production costs relatively more, thereby weakening the region's comparative advantage in that sector. This shift results in a relative contraction of agriculture and an expansion of non-agricultural production (Romalis, 2004).

Non-unitary elasticities can also generate weak equilibrium biases in technological development (Acemoglu, 2007). Since labor and capital are substitutes in agriculture, a relative increase in capital creates a greater incentive to use the more abundant capital more efficiently, via technology development or learning-by-doing, for instance. Such capital-augmenting technical change could also have reinforced the reallocation of labor from agriculture to non-agriculture.

2 Empirical Strategy

2.1 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 linked individuals between 1910 through 1940 using the Census Tree approach (Buckles et al., 2023).[7] I also report results based on shares constructed from the 1940 Census, which asks for residence 5 years ago. I use this information to construct an alternative migration matrix between 1935 and 1940.[8]

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.

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 services. The CBP contains information related to employment and economic activity for detailed industry codes covering all counties in the United States. The number of net mi-

[7]The Census Tree capitalizes on manual matches created by individuals conducting 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.

[8]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.

grants 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).

The main sample is 1,096 counties in the South between 1940 and 2010. The definition of the South mainly follows the Census definition,[9] 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. I restrict the sample to balanced counties for agriculture and manufacturing dependent variables, except for the number of combines and manufacturing capital spending that are started to be collected in 1950 and have less coverage.

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 borders across different years are adjusted to the 1990 boundaries (Eckert et al., 2020). Given the long 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.

2.2 Estimating equation

The empirical analysis aims to examine how Southern regions that faced different levels of out-migration during the Great Migration period experienced distinct economic changes after the migration. A straightforward empirical implementation would estimate year-specific differences between counties by the levels of net out-migration rates between 1940 and 1970:

$$Y_{c,t} = \beta_t \left(Net\_Out\_Mig^{1940-1970}\right)_c + \alpha_{s,t} + \alpha_c + \gamma X_{c,t} + \gamma_t X_c + \varepsilon_{c,t}. \tag{2}$$

The main regressor summarizes the out-of-county out-migration at the county level during the Second Great Migration. The main identification challenge here 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:

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

by replacing the main regressor to "$\left(\widehat{Northern\_Exposure}^{1940-1970}\right)_c$", 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

[9] 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 Columbia.

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. Conversely, I calculate the migration exposure in the Southern origins using Northern pull factors.

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 changes in outcome, relative to each county's base value, between counties that experienced different levels of Northern pull factors within the same state in the same year.

The omitted base years are 1940 and 1950, and I drop the 1960 value from the regression. Hence, the β_t captures the changes in outcome $Y_{c,t}$ in each year relative to its levels in 1940 and 1950. Compared to using only 1940 values as a base, using both decadal years would minimize the influence of a single base year. The 1960 value is omitted as it may be affected by earlier migration flows (1940–1960).

Baseline estimation is not weighted, and the estimates report the average outcome per county. The results are similar using the 1940 population as the weight. All standard errors are clustered at the county level to take into account serial correlation within a county. I also report Conley standard errors that allow for spatial correlation.

2.2.1 Shift-share instrumental variable strategy

The measure of Northern pull factors combines the predetermined migration as instruments and predicted in-migration as weights (Goldsmith-Pinkham et al., 2020). In the zero-stage regression of in-migration prediction, as in Equation (4), I use OLS regression with variables selected to represent demographic, economic, social, and political environments in the North:[10]

$$(in\_mig\_rate)^r_{d,t} = f(pull\_factors_{d,t-10}) + \varepsilon_{d,t}, \tag{4}$$

where I separately predict Black and White in-migration (Table A1) to take into account heterogeneous migration patterns by race (Collins and Wanamaker, 2015). I also report the results with alternative shifts using actual in-migration rates as in Card (2001) and predicted rates using a machine-learning technique.

Then, I allocate the predicted number of migrants back to Southern counties using pre-

[10]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

period migration share matrices,[11] $\omega_{do}^{r,1910-1940}$, between 1910 and 1940 as in Equation (5):

$$(\widehat{out\_migrants}^{1940-1970})_o = \sum_{t=1950}^{1970} \sum_d \sum_r \left((\widehat{in\_migrants})_{d,t}^{r} \cdot \omega_{do}^{r,1910-1940} \right), \tag{5}$$

where o and d index origin and destination, and r stands for race $\in \{Black, White\}$. I restrict the shares to only include migration flows between the North and the South. Hence, each Southern county is assigned 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.

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 (6). Finally, I standardize the predicted out-migration rate and denote the resulting measure as the Northern pull factor exposure. I also simply refer to it as the Northern exposure or migration exposure.

$$(\widehat{out\_mig\_rate}^{1940-1970})_o = \frac{(\widehat{out\_migrants}^{1940-1970})_o}{(population^{1940})_o}, \tag{6}$$

$$standardize\left((\widehat{out\_mig\_rate}^{1940-1970})_o \right) \equiv \widehat{Northern\_Exposure}^{1940-1970}{}_o. \tag{7}$$

The constructed instrument quantifies the levels of migration linkages of a Southern county to Northern destinations that are predicted to receive different levels of migration between 1940 and 1970.

As an example, while Mississippi and Louisiana are adjacent, counties 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 Los Angeles is predicted to receive relatively more migrants than Chicago, the Northern exposure of Louisiana counties will be higher. Even within the state, counties with stronger pre-period linkages to Los Angeles are assigned greater exposure.

2.2.2 Control variables

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-

[11] 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 quasi-randomly selected Black Southerner who moved to a Northern destination county, 1.9 additional Black migrants made the same move.

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. To do so, I include time-interacted values of (1) time-invariant county characteristics, (2) 1940 agriculture variables, (4) pre-period migration rates, (4) predicted out-migration rates between 1940 and 1970, and (5) international trade exposure.

First, I include the 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 was 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,[12] total farm acres, and acres harvested in cotton, tobacco, corn, and hay, respectively.[13] 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). The predicted rates are constructed using only Southern variables. 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 (4) and Table A1).

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.[14] 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 at

[12]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).

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

[14]Note that the county fixed effect removes the fixed level of trade exposure while the state-by-year fixed effects 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 influence counties in the same state differently.

least 4 million Mexican workers between 1942 and 1964. 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.

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.

2.3 Discussion of the identification strategy

Given the constructed Northern pull factors, 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.

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 can also potentially take into account the 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 and interpretation follow this view.

Alternatively, the identification of SSIV can rely on exogenous shifts (Borusyak et al., 2022). This strategy capitalizes on the idiosyncrasies of the shifts. Here, I use a large set of Northern pull factors as the shifts, which can plausibly be assumed to be unrelated to the changes in Southern economic outcomes. Note that state-by-year fixed effects remove any 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 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. Note that the baseline estimation controls for 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).

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 constructed instrument captures the component of out-migration driven by Northern pull factors, which may differ from variations in a randomized setting. Still, the LATE here is not necessarily a limitation, since the estimation is based primarily on variation driven by migrants who responded to external incentives. The findings here could be more applicable to a setting where the government can incentivize people to move out of labor-abundant regions.

2.3.1 First-stage results

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 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.

Panel A, Columns 4 and 5, and Panel B explore the bounds of the migration response by accounting for non-linear effects and by considering potential biases from 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.

Panel B reports the results with a migration-linkage corrected estimation suggested by Borusyak et al. (2023).[15] While they account for within-South migration linkages, the overall results remain similar.

[15] 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.

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 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 accounting for within-South migration linkages. Column 1 reports the raw relationship, and Columns 2 through 3 add state fixed effects and control variables listed in Section 2.2.2. 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. Higher-order terms are not statistically different from zero at the 10% level and are not reported. The Kleibergen–Paap robust F-statistics are reported in Panel A, and the adjusted R-squared is reported in Panel B. Stars represent: $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

3 Empirical Evidence

This section examines how the Southern out-migration 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. I conclude this section by introducing robustness checks and statistical tests that assess the validity of the estimation strategy.

3.1 Agriculture

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 (2) on Panel A and SSIV reduced-form estimates using Equation (3) 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 for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted base 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 2.2.2. 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. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

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 experienced 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, would have 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

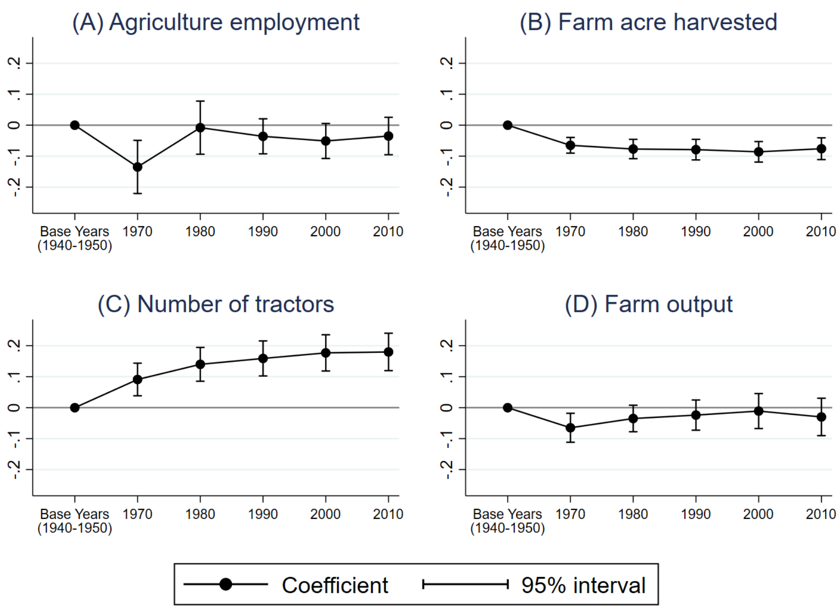

Figure 1: Time trends in agricultural estimates.

Note: The figure presents the SSIV reduced-form estimates on agricultural outcomes using Equation (3), 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 base years of 1940 and 1950. Robust standard errors clustered by county.

Northern counties that had higher levels of migration pull factors.[16] 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 tend 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 (3). For Panel D, I report R-squared and the Kleibergen-Paap robust F[17] from the first-stage regression.

Table 2, Panel D, Columns 1 to 3, show that a one standard deviation increase in Northern migration exposure reduced agricultural employment by 4.2%, the number of farms by 2.3%,

[16]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.

[17]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).

and farm acres harvested by 8.2%[18,19] 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).[20] 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 have experienced only negligible changes.

Figure 1 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 (3). 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 the lower level at least until 2010 (Panel B).

By contrast, the number of tractors increased by 9.1% in 1970, continued to grow to 17.7% in 2000, and remained at 18.0% in 2010 (Panel C). The observed pattern of continued increases in tractor usage can be rationalized by the directed technical change (Acemoglu, 2002, 2007). As agriculture uses less labor and more capital, Southern agriculture becomes better at using capital, which in turn incentivizes further capital investment.[21] In Panel D, farm outputs initially decreased but recovered with accompanying increases in tractor adoption.

Agricultural economics literature documents that low labor costs delay mechanization, while labor scarcity can induce the adoption of labor-saving technologies and capital (see Gallardo and Sauer (2018) for a review). In the 20th-century United States setting, shrinking labor supply from a natural disaster (Hornbeck and Naidu, 2014) or from abrupt changes in migration policy (Clemens et al., 2018; San, 2023; Abramitzky et al., 2023) indeed facilitated the adoption of labor-saving capital and technologies in agriculture. The out-migration induced by pull factors in this paper exhibits a similar pattern. This labor scarcity may also have encouraged

[18] 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.

[19] Note an important limitation of the estimating strategy: Equation (3) 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.

[20] 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.

[21] The diffusion of tractors in the wider U.S. exhibited a similar pattern. Tractor adoption began with narrow applications that directly substituted for labor, but its use subsequently expanded with generalization to broader purposes (Gross, 2018).

related changes in non-agriculture.

3.2 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 after the out-migration. These contrasting patterns suggest that the role of manufacturing development as a migration push factor and the role of out-migration in shaping manufacturing development 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

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 (2) on Panel A and SSIV reduced-form estimates using Equation (3) 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 for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted base 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 2.2.2. 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. Stars represent: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

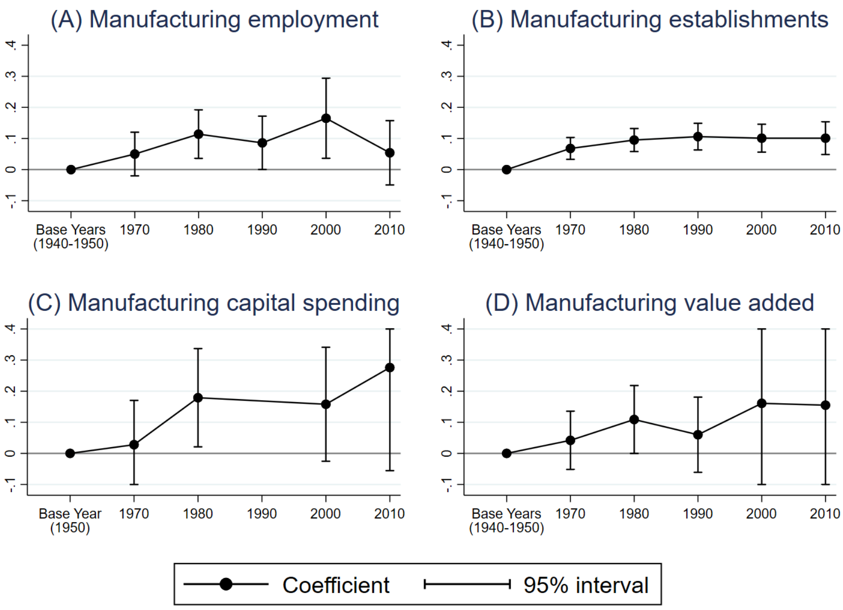

Figure 2: Time trends in manufacturing estimates.

Note: The figure presents the SSIV reduced-form estimates on manufacturing outcomes using Equation (3), 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 base years of 1940 and 1950. Robust standard errors clustered by county.

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.

Panels C and D add state-by-year and county fixed effects and control variables. The baseline estimates in Panel D reveal that Northern migration exposure modestly increased manufacturing employment and the levels of manufacturing development. The pull factor-induced out-migration would have raised 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 relative magnitudes of the estimates suggest that the payroll per worker increased as well.

Figure 2 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. 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,

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 (3) 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 for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted base 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. Stars represent: $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Figure 3: Time trends in wholesale and retail estimates.

Note: The figure presents the SSIV reduced-form estimates on wholesale and retail outcomes using Equation (3), 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 base years of 1940 and 1950. Robust standard errors clustered by county.

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 establishments and sales also grew after 1970. Figure 3 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.

3.3 Structural change

This subsection examines how differential exposure to Northern pull factors influenced overall industry composition. Table 5 reports the baseline results using the employment share in each industry as the dependent variable. Columns 1 and 2 suggest that out-migration contributed to structural change out of agriculture and into manufacturing. One standard deviation increase in

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

	(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

Note: The table reports SSIV reduced-form estimates for industry employment shares, using Equation (3), with county-year as the unit of observation. All results include state-by-year and county fixed effects and control variables described in Section 2.2.2. Each column reports the changes in the indicated outcome variable for the years 1970 to 2010 by the Northern migration exposure, relative to the omitted base years between 1940 and 1960, using periods when information is available. I supplement the main analysis using CBP for 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: $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

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 services tended to experience negative changes (Column 5). One possible explanation is that producer services, which tend to be less capital-intensive than other industries, would have been less likely to be affected by economic reallocation driven by changes in the regional relative labor and capital abundance.

3.4 Robustness and validity check

Here, I introduce the second-stage estimates, statistical tests on the instrument, and additional robustness checks documented in the Online Appendix. First, Table A2 documents the second-stage estimates for the main outcomes in Tables 2, 3, 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.

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 A3, I estimate Equation (3) on pre-period outcomes for the variables that have pre-period values.[22] The test examines whether pre-period changes in the main outcomes

[22] The number of combines and manufacturing capital spending are available after 1950, and retail variables are available after 1940. Instead, I use manufacturing intermediate goods spending and manufacturing revenue. They

(the values in 1920 and 1930 compared to 1940) are systemically correlated with the instrument. Overall, they do not show a clear pattern and are statistically not different from zeros, suggesting that consequential pretrends that drove the main results were less likely to exist.

A couple of exceptions are agricultural employment and farm output. Table A3, 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. These controls account for any changes after 1940 that may have arisen due to differences in the pre-period outcome.

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 A2. 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.

In addition to the statistical tests discussed in the main text, I introduce various placebo tests in terms of random share (Table A6) and random shift (Table A7), and falsification tests in terms of outcomes (Table A8). For the placebo share, I randomly reallocate the migration cell when constructing the migration exposure in the shift-share design. For the shift, I assign a random number of immigrants in the North and reallocate them back to the South. These placebo experiments lead to statistically insignificant estimates, with their signs and magnitudes not systematically related to the main estimates.

Table A8 presents falsification tests using public employment and payroll as dependent variables. The simple framework highlights the private industries' adjustments to the change in relative factor abundance. On the contrary, the public sector would not have been much affected by such a change. Indeed, the public sector did not respond in terms of employment and annual payroll (Columns 1 to 3). Local government employment may have even slightly decreased (Column 2). The results also suggest that the empirical results on non-agricultural

are not used in the main analysis as they are not available or are only sparsely available during the main study periods.

development are not likely to be driven by government policies.

Online Appendix Section D documents the main results from alternative approaches. They include (1) using the Northern exposure based on the 1935-1940 migration matrix, (2) by adopting alternative approaches for in-migration prediction (random forest algorithm and actual number of in-migrants), (3) by limiting the sample to former confederate states, and (4) by adding a weighted average of other Southern counties' migration exposures. While the exact results are influenced by empirical choices, there appear to be no systematic patterns threatening the validity of the estimation strategy.

4 Quantitative Strategy

4.1 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). I extend the aforementioned frameworks by introducing the role of factor substitution and factor intensity in driving structural change and economic allocation, thereby capturing how shifts in the relative abundance of labor and capital shape economic outcomes.

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 can be thought of as consisting of manufacturing and (tradable) production services. The nontradable sector can be regarded as consumer services that are locally provided. For simplicity, I also respectively refer to them as manufacturing and local services.

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 migration costs.[23] In addition, the general equilibrium model al-

[23]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 to quantify labor market impacts of the Great Migration. He focuses on different substitutability

lows non-homothetic preferences on the demand side by introducing the non-homothetic Price-Independent Generalized Linear (PIGL) class in utility. It accounts for demand-side structural change, while factor reallocation captures that in the supply side. [24]

I make two simplifications for tractability. First, to obtain analytical expressions for trade and migration shares in 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.

4.2 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:

$$U(C_{i,t}) = \log\Big(C(e_{i,t}, P_{i,t})\Big). \tag{8}$$

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. The indirect utility of consumption for an individual with expenditure e facing local price of P_i takes the form:

$$C(e, P_i) = \frac{1}{\varepsilon}\left(\frac{e}{(P_i^a)^{\phi^a}(P_i^m)^{\phi^m}(P_i^l)^{\phi^l}}\right)^{\varepsilon} - \sum_{s\in\{a,m,l\}} \nu^s \ln P_i^s, \tag{9}$$

over sectoral value-added CES aggregates of all regional varieties. 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 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:

$$\varphi^s(P_i, e) = \phi^s + \nu^s\left(\frac{e}{P_i}\right)^{-\varepsilon}. \tag{10}$$

An individual's consumption share depends on the price index in region i and her income. For instance, the share on necessity declines as workers' real income rises. Workers do not have

across different labor groups by race and age, while I emphasize the tensions between labor and capital.

[24] 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. The demand side income effects here only account for the second-order income effects due to the Great Migration.

access to the investment technology and live hand-to-mouth. Regional aggregate demand is derived by summing up individual demand in each location.

Intratemporal labor supply. In each period, factors of production are mobile across sectors within the region. To rationalize the observed difference in factor prices, I introduce Roy-type machinery by assuming 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, $P(a_i^s \leq a) = \exp(-(a/A_i^{w,s})^{-\zeta^w})$. 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:

$$s_i^{w,s} = (A_i^{w,s} w_i^s/\bar{w}_i)^{\zeta^w} \text{ where } \bar{w}_i = \left((A_i^{w,a} w_i^a)^{\zeta^w} + (A_i^{w,m} w_i^m)^{\zeta^w} + (A_i^{w,l} w_i^l)^{\zeta^w}\right)^{1/\zeta^w}, \tag{11}$$

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:

$$L_i^s = \Gamma_{\zeta^w} A_i^{w,s} (A_i^{w,s} w_i^s/\bar{w}_i)^{\zeta^w - 1} L_i \equiv A_i^{w,s} \tilde{L}_i^s, \tag{12}$$

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:

$$\mathbb{V}_{i,t} = U(C_{i,t}) + \max_{\{n\}} \left\{\beta \mathbb{E}[\mathbb{V}_{n,t+1}] - \kappa_{ni,t} + \eta u_{n,t}\right\}, \tag{13}$$

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 share is:

$$\mathbb{M}_{in,t} = \frac{\exp\left(\beta \mathbb{E}_t(v_{n,t+1} - \kappa_{ni})/\eta\right)}{\sum_{j=1}^N \exp\left(\beta \mathbb{E}_t(v_{j,t+1} - \kappa_{nj,t})/\eta\right)}, \tag{14}$$

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

4.2.1 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:

$$v_{i,t}^{k} = \mathbb{E}_t \sum_{t=0}^{\infty} \beta^t \frac{(C_{i,t}^{k})^{1-1/\psi}}{1-1/\psi} \text{ subject to } \bar{r}_{i,t} K_{i,t} = P_{i,t}\left(C_{i,t}^{k} + K_{i,t+1} - (1-\delta)K_{i,t}\right), \tag{15}$$

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 $\psi = 1$, the utility function 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 CES consumption 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 an assumption that the capitalists have surpassed the income threshold to reach the asymptotes.

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):

$$\varsigma_{i,t}^{-1} = 1 + \mathbb{E}_t \beta^{\psi}([R_{i,t+1}^{\frac{\psi-1}{\psi}} \varsigma_{i,t+1}^{-\frac{1}{\psi}}])^{\psi}, \tag{16}$$

The consumption share is solved recursively using Equation (16) 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:

$$A_i^{k,s} = \bar{A}_i^{k,s} \times F^s(\cdot). \tag{17}$$

The endogenous component incorporates the dynamic weak equilibrium bias in a reduced-form way. 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. Alternatively, it can be thought of as generated by regional technology developers in shadow as in Acemoglu (2007). For simulation analysis, the values of $F^s(\cdot)$ are calibrated using the estimated changes in agriculture and manufacturing employment with a one standard deviation increase in the Northern exposure. In contrast, the exogenous component $\bar{A}$ cancels out during model calculation.

4.3 Production.

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

$$Y_i^s = z_i^s \left(\rho_i^s (A_i^{w,s} \tilde{L}_i^s)^{\frac{\sigma^s - 1}{\sigma^s}} + (1 - \rho_i^s)(A_i^{k,s} \tilde{K}_i^s)^{\frac{\sigma^s - 1}{\sigma^s}} \right)^{\frac{\sigma^s}{\sigma^s - 1}}. \tag{18}$$

The production side is given similarly to Section 1.2, with the addition of Hicks-neutral technology z_i^s. The Hicks-neutral technology, z_i^s,is a realization of the Frechet distribution with local fundamental Z_i^s and the shape parameter θ^s,[25] as in the standard Eaton and Kortum (2002) setting. The local Hicks-neutral 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 factor cost shares.

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

$$p_{n,t}^s = \min_{1 \le i \le N} \left\{ \frac{x_{i,t}^s \tau_{ni,t}^s}{z_{i,t}^s} \right\}, \tag{19}$$

where the term inside Equation (19) 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:

$$x_{i,t}^s = \min \left(w_{i,t}^s A_{i,t}^{w,s} \tilde{L}_{i,t}^s + r_{i,t}^s A_{i,t}^{k,s} \tilde{K}_{i,t}^s \right) = \left((\rho_i^s)^{\sigma^s} (w_{i,t}^s)^{1-\sigma^s} + (1 - \rho_i^s)^{\sigma} (r_{i,t}^s)^{1-\sigma^s} \right)^{\frac{1}{1-\sigma^s}}. \tag{20}$$

The changes in factor allocation and resulting changes in factor prices affect the regional comparative advantage through Equation (20).[26]

[25] 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.

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

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

$$\mathbb{S}^s_{ni,t} = \frac{Z^s_{i,t}\left(x^s_{i,t}\tau^s_{ni,t}\right)^{-\theta^s}}{\sum_{j=1}^N Z^s_{j,t}\left(x^s_{j,t}\tau^s_{nj,t}\right)^{-\theta^s}}, \tag{21}$$

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.

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 $Y^k_{n,t} = \sum_{i=1}^N \mathbb{S}^k_{ni,t} X^k_{n,t}$.

4.4 Taking the model to the data

Table 6 summarizes the model parameters. First, the consumption side and migration parameters follow Yang (2024), 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 for 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 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 3. First, I use the CES production function estimates in Herrendorf et al. (2015), who uses U.S. macro data between 1947 and 2010. As theirs, the model here adopts the value-added production 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. These 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.[27] Still, the direct values on location-industry-specific cost shares are required for simulation. Here, I adjust the estimates on ρ^s by regional proxies for the capital-

[27] Compared 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.

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		
EoS 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

to-labor ratio. I use variation in the number of tractors and combines per agricultural worker as a proxy for the 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.

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-Mead 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 1, Panels A and D), manufacturing (Figure 2, Panels A and D), and retail (Figure 3, 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.

4.5 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 C). 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.

5 Quantitative Results

In this section, I first outline counterfactual outcomes in terms of welfare effects and introduce contribution analysis. I focus on the changes in the workers' consumption welfare (Equation 8), which is henceforth simply referred to as welfare.[28] 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.

5.1 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 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.[29]

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

[29] 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.

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: Row A shows the consumption welfare effect by each scenario, and Row B decomposes this effect into a contribution from each model element. 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 shares in Row B are calculated by comparing the welfare effects of the full model with those of constrained versions in which each model component is turned off one by one.

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 across regions, 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%. I then run a constrained model separately for each channel by turning off one model component at a time. Columns 2 to 4 present the welfare effect in each scenario. Finally, in Row B, 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, Column 2, shows that factor substitution takes into account the lion's share of the response to the South-to-North migration flows, accounting for 69.8% of the adjustments. The trade adjustment and directed technical change played supplementary roles, each contributing to 12.8% and 20.4%. The total value equals 103.1%, where the excess of the 3.1% is generated by interaction effects. Given the model parameters, the trade adjustment played a supplementary role in the adjustments.

5.2 Distribution of the shock

Figure 4 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 predicted South-to-North migration rate for the Southern states, the shock used in the counterfactual analysis. It serves as a proxy for the magnitude of the Second Great Migration. 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 the modeled Great Migration. Panel B shows the changes in aggregate capital-to-labor

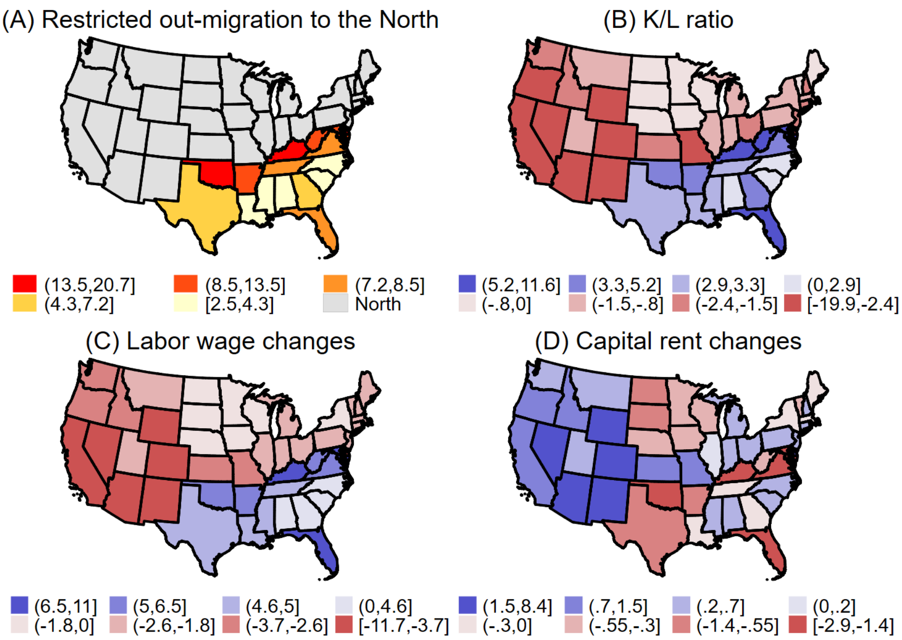

Figure 4: Maps of the Great Migration and outcome variables.

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 (3). Red means higher levels of predicted out-migration relative to yellow. The predicted out-migration from 1940 to 1970 is used as a shock, serving as a proxy for the Second Great Migration. Panels B to D map the simulation outcomes 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.

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. Panel C documents wage effects that mirror the distribution of changes in the capital-to-labor ratio, whereas Panel D shows the opposite pattern for the rental rate of capital. Regions that accumulated more capital relative to labor tended to experience higher wages but lower capital rental rates.

Figure 5 plots the changes in economic distribution between 1940 and 2010. 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 corresponds to 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

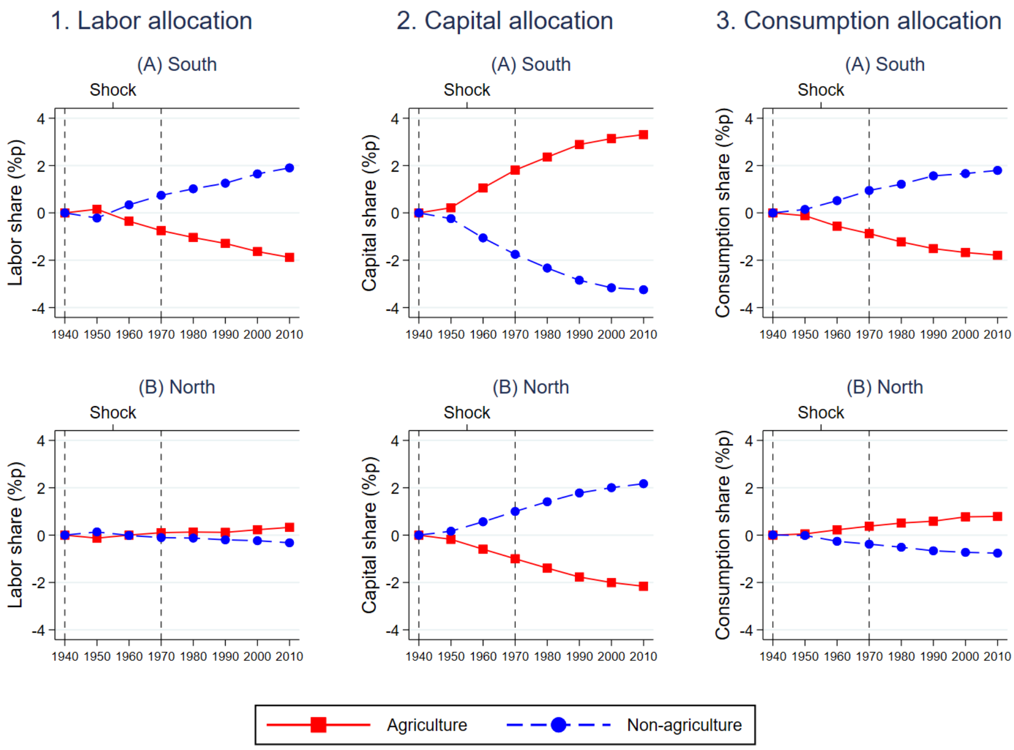

Figure 5: Simulated Changes in Economic Allocation by Region, 1940-2010.

Note: The figure illustrates the percentage-point changes in sectoral allocation of labor, capital, and consumption between 1940 and 2010, by region. The red straight line represents agriculture, and the blue dashed line indicates non-agriculture.

South between 1940 and 2010. Hence, in the model's view, the Great Migration played a supplementary yet important role in reallocating labor out of agriculture, complementing broader changes in the South such as technological progress and improvements in human capital.

Panel 2 shows opposite capital allocation patterns across sectors, indicating that Southern agriculture mechanized in response to shifts in the capital-to-labor ratio (Panel 2, Row A). Panel 3 shows changes in consumption allocation due to the Second Great Migration, reflecting second-order income effects driven by shifts in the capital-to-labor ratio. Row A suggests that wage increases and resulting consumption adjustment also supported the expansion of non-agricultural sectors in the South.

6 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 structural change. In response to 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-augmenting 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), 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 Platt 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.

_ **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.

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

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

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.

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

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

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.

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.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, "Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition," 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.

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.

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., "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**, "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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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].

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.

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.