

Migration Networks and Location Decisions: Evidence from US Mass Migration[†]

By BRYAN A. STUART AND EVAN J. TAYLOR*

This paper studies how birth town migration networks affected long-run location decisions during historical US migration episodes. We develop a new method to estimate the strength of migration networks for each receiving and sending location. Our estimates imply that when one randomly chosen African American moved from a Southern birth town to a destination county, then 1.9 additional Black migrants made the same move on average. For White migrants from the Great Plains, the average is only 0.4. Networks were particularly important in connecting Black migrants with attractive employment opportunities and played a larger role in less costly moves. (JEL J15, J61, N32, N92, R23, Z13)

Theoretical and empirical research emphasizes the role of expected real wages, amenities, and moving costs in individuals' location decisions (Sjaastad 1962, Greenwood 1997, Kennan and Walker 2011). While theory suggests that social networks might matter as well (Carrington, Detragiache, and Vishwanath 1996), estimating the importance of this factor has proven difficult because of a lack of suitable datasets and research designs. For example, it is well known that immigrants from the same country tend to live in the same place, but this fact does not distinguish between the role of social networks and numerous common factors, such as moving costs, human capital, and language. Evidence on the effects of social networks on location decisions would inform our understanding of past and future migration episodes, the equilibration of local labor markets, and the impacts of policies that affect migration incentives. Furthermore, social networks might continue to attract migrants to their chosen destination for many years, thus limiting adjustments as economic conditions change and ultimately contributing to spatial mismatch.

*Stuart: George Washington University and IZA (email: bastuart@gwu.edu); Taylor: University of Arizona (email: evantaylor@e-mail.arizona.edu). Alexandre Mas was coeditor for this article. For helpful comments and suggestions, we thank anonymous referees, Martha Bailey, Dan Black, John Bound, Leah Boustan, Charlie Brown, John DiNardo, Joseph Ferrie, Paul Rhode, Seth Richards-Shubik, Seth Sanders, Jeff Smith, Lowell Taylor, and numerous seminar and conference participants. Thanks to Seth Sanders and Jim Vaupel for facilitating access to the Duke SSA/Medicare data, and Maggie Levenstein for help accessing the 1940 census data. During work on this project, Stuart was supported in part by an NICHD training grant (T32 HD007339) and an NICHD center grant (R24 HD041028) to the Population Studies Center at the University of Michigan. This paper was previously circulated under the title "Social Interactions and Location Decisions: Evidence from US Mass Migration."

[†]Go to <https://doi.org/10.1257/app.20180294> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

This paper provides new evidence on the effects of social networks on location decisions. We focus on the mass migrations in the mid-twentieth century of African Americans from the US South and Whites from the Great Plains. We proxy for social networks using birth towns, which are particularly relevant in this setting, and we use administrative data that measure town of birth and county of residence at old age for most of the US population born from 1916 to 1936. Our setting and data provide a unique opportunity to study the long-run effects of migration networks. We use detailed geographic information to distinguish the effect of birth town migration networks from other determinants of location decisions, such as moving costs determined by geography or railroad lines. For example, we observe that 51 percent of African American migrants born from 1916 to 1936 in Pigeon Creek, Alabama, moved to Niagara County, New York, while less than 6 percent of Black migrants from nearby towns moved to the same county. This comparison underlies our research design, which asks whether individuals born in the same town are more likely to live in the same destination in old age than individuals born in nearby towns.

We combine this transparent research design with a new method of characterizing birth town migration networks. Our new parameter, which we call the network index, allows us to estimate the effect of migration networks on location decisions for each receiving and sending location and then relate these estimates to locations' economic characteristics. We show that existing methods may mischaracterize the effect of migration networks in our setting. In particular, the influential approach of Bayer, Ross, and Topa (2008) could estimate strong effects for popular destinations even if true effects are relatively weak, and as a result misstate the overall effect of networks. This arises because their model measures network strength using the covariance of decisions, which can be large only because the probability of choosing an option is large. Our method does not suffer from this problem. Under straightforward and partly testable assumptions, the network index identifies the effect of birth town migration networks and maps directly to structural network models. Throughout, we estimate how migration networks affect where individuals move, conditional on migrating.

We find that migration networks strongly influenced the location decisions of Southern Black migrants. Our estimates imply that when one randomly chosen African American moves from a birth town to a destination county, then 1.9 additional Black migrants make the same move on average. Migration networks drew African Americans to destinations with a higher share of 1910 employment in manufacturing, a particularly attractive sector for Black workers in our sample cohorts. This evidence highlights an important role for migration networks in providing job referrals or information about employment opportunities. Moreover, networks are stronger in destinations with a smaller Black-White residual wage gap in 1940, raising the possibility that networks helped migrants find destinations with less labor market discrimination. We also find that networks drew Black migrants to destinations that were closer and more connected by railroads, pointing to the importance of access to information and low moving costs in the functioning of these networks.

We estimate weaker effects of migration networks on the location decisions of Whites. For migrants from the Great Plains, our results imply that when one

randomly chosen migrant moves from a birth town to a destination county, then 0.4 additional White migrants make the same move on average. Results for Southern White migrants are similarly small. Furthermore, migration networks among Whites are less sensitive to employment opportunities and moving costs. There are many possible explanations for the different effects of networks on Black and White migrants. Given the myriad unobserved differences between these groups, this paper does not attempt to explain the Black-White gap. However, one explanation supported by historical context and our results is that Black migrants relied more heavily on their networks to counteract discrimination in labor and housing markets and a lack of financial resources.

To further study the role of migration networks, we map the network index to a structural model that generalizes Glaeser, Sacerdote, and Scheinkman (1996). We estimate that 34 percent of Southern Black migrants and 13 percent of Great Plains White migrants chose their long-run destination because of migration networks. In the absence of networks, Chicago would have 29 percent fewer Southern Black migrants, and Los Angeles, Detroit, Philadelphia, and Baltimore would have 11 to 25 percent fewer Black migrants. Eliminating migration networks would reduce the number of Great Plains White migrants in several places in California, including Los Angeles, Bakersfield, and Fresno. While our model does not account for all possible general equilibrium effects, the direction of these effects is not clear: reducing migration from a town to a county could make that destination more attractive, because of higher wages or lower housing costs, or less attractive, because of fewer individuals with a similar background. Still, the model suggests that migration networks had important effects on the spatial distribution of the US population.

We use the structural model to examine whether migrants would live in destinations with better economic opportunities in the absence of networks, as could occur if networks contributed to spatial mismatch. In the absence of migration networks, Southern Black migrants would live in counties with a slightly smaller African American population share, unemployment rate, and poverty rate, while Great Plains White migrants would live in counties that are nearly identical. Migration networks have little effect on destination characteristics because migrants who did not follow their network moved to similar destinations.

Our research design identifies network effects as large propensities of individuals from the same birth town to move to the same destination, above and beyond the propensity of individuals from nearby towns. Potential threats to this approach include factors—besides the migration network—that especially induce migrants from a single town to move to a particular destination. For example, this could arise if migrants from Pigeon Creek, Alabama, had especially strong preferences for Niagara County (over 1,000 miles away) or human capital especially suited for the Niagara labor market, compared to other nearby towns in Alabama. Qualitative accounts suggest that such threats are unlikely to be important. Furthermore, several pieces of evidence support the validity of our empirical strategy. The research design implies that destination-level network index estimates should not change when controlling for birth town characteristics, because geographic proximity controls for the relevant determinants of location decisions. Reassuringly, our estimates are essentially unchanged when adding several covariates. We also estimate strong network

effects in certain locations, like Rock County, Wisconsin, for which qualitative work supports our findings (Bell 1933, Rubin 1960, Wilkerson 2010).

This paper makes three contributions. First, we develop a new method of characterizing migration networks. Our approach integrates previous work by Glaeser, Sacerdote, and Scheinkman (1996) and Bayer, Ross, and Topa (2008), contains desirable theoretical and statistical properties, and can be used to study networks in other settings and for outcomes besides migration. Second, we provide new evidence on the importance of birth town migration networks and the types of individuals and economic conditions for which networks are most important.¹ Previous work shows that individuals tend to migrate to the same place, often broadly defined, as other individuals from the same town or country but does not isolate the role of social networks in the decision of where to move (Bartel 1989; Bauer, Epstein, and Gang 2005; Beine, Docquier, and Özden 2011; Giulletti, Wahba, and Zenou 2018; Spitzer 2016).² Third, our results inform landmark migration episodes that have drawn interest from economists for a century (Scroggs 1917; Smith and Welch 1989; Margo 1990; Carrington, Detragiache, and Vishwanath 1996; Collins 1997; Boustan 2009, 2010, 2017; Hornbeck 2012; Hornbeck and Naidu 2014; Black et al. 2015; Collins and Wanamaker 2014, 2015; Johnson and Taylor 2019; Long and Siu 2018). Our results complement the small number of interesting but unrepresentative historical accounts suggesting that networks were important in these migration episodes (Jamieson 1942, Rubin 1960, Gottlieb 1987, Gregory 1989).

Our paper also complements recent work by Chay and Munshi (2015). They find that, above a threshold, migrants born in counties with higher population density tend to move to fewer locations, as measured by a Herfindahl-Hirschman Index, and show that this nonlinear relationship accords with a network formation model with fixed costs of participation. We also find some evidence that networks were stronger in denser sending communities. We differ in our research design, empirical methodology, study of White migrants, examination of how network effects vary across destinations, and use of a structural model to examine counterfactuals.

I. Historical Background on Mass Migration Episodes

The Great Migration saw nearly 6 million African Americans leave the South from 1910 to 1970 (Kreps, Slater, and Plotkin 1979). Although migration was concentrated in certain destinations, like Chicago, Detroit, and New York, other cities also experienced dramatic changes. For example, Chicago's Black population share increased from 2 to 32 percent from 1910 to 1970, while Racine, Wisconsin, experienced an increase from 0.3 to 10.5 percent (Gibson and Jung 2005). Migration out of the South increased from 1910 to 1930, slowed during the Great Depression, and then resumed forcefully from 1940 to 1970.

¹This complements research on the effects of social networks on labor market outcomes (e.g., Topa 2001; Munshi 2003; Ioannides and Loury 2004; Bayer, Ross, and Topa 2008; Hellerstein, McInerney, and Neumark 2011; Beaman 2012; Burks et al. 2015; Schmutte 2015; Heath 2018). These papers do not focus on the formation of social networks, which in some cases, like Munshi (2003), arise from location decisions.

²One exception is Chen, Jin, and Yue (2010), who study the impact of peer migration on temporary location decisions in China. However, they lack detailed geographic information on where individuals move.

Several factors contributed to the exodus of African Americans from the South. World War I, which simultaneously increased labor demand among Northern manufacturers and decreased labor supply from European immigrants, helped spark the Great Migration, although many underlying causes existed long before the war (Scroggs 1917, Scott 1920, Gottlieb 1987, Marks 1989, Jackson 1991, Collins 1997, Gregory 2005). These causes include a less developed Southern economy, the decline in agricultural labor demand due to the boll weevil's destruction of cotton crops (Scott 1920; Marks 1989, 1991; Lange, Olmstead, and Rhode 2009), widespread labor market discrimination (Marks 1991), and racial violence and unequal treatment under Jim Crow laws (Tolnay and Beck 1991).

Migrants tended to follow paths established by railroad lines. For example, Mississippi-born migrants predominantly moved to Illinois and other midwestern states, and South Carolina-born migrants predominantly moved to New York and Pennsylvania (Scott 1920; Carrington, Detragiache, and Vishwanath 1996; Collins 1997; Boustan 2010; Black et al. 2015). Labor agents, offering paid transportation, employment, and housing, directed some of the earliest migrants, but historical accounts suggest that their role diminished sharply after the 1920s and most individuals paid for the expensive train fares themselves (Gottlieb 1987, Grossman 1989).³ African American newspapers from the largest destinations circulated throughout the South, providing information on life in the North (Gottlieb 1987, Grossman 1989).⁴

A small number of historical accounts suggest a role for migration networks in location decisions. Social networks, consisting primarily of family, friends, and church members, sometimes provided valuable job references or shelter (Scott 1920, Rubin 1960, Gottlieb 1987). For example, Rubin (1960) finds that migrants from Houston, Mississippi, had close friends or family at two-thirds of all initial destinations.⁵ These accounts emphasize interactions between individuals from the same birth town, which motivates our focus on birth town migration networks.

The experience of John McCord captures many important features of early Black migrants' location decisions.⁶ Born in Pontotoc, Mississippi, nineteen-year-old McCord traveled in search of higher wages in 1912 to Savannah, Illinois, where a fellow Pontotoc-native connected him with a job. McCord moved to Beloit, Wisconsin, in 1914 after hearing of employment opportunities and quickly began working as a janitor at the manufacturer Fairbanks Morse and Company. After two years in Beloit, McCord spoke to his manager about returning home for a vacation. The manager asked McCord to recruit workers during the trip, and McCord returned with 18 unmarried men, all of whom were soon hired. Thus began a persistent flow of African Americans from Pontotoc to Beloit: among individuals born from 1916

³In 1918, train fare from New Orleans to Chicago cost \$22 per person, when Southern farmers' daily wages typically were less than \$1 and wages at Southern factories were less than \$2.50 (Henri 1975).

⁴The *Chicago Defender*, perhaps the most prominent African American newspaper of the time, was read in 1,542 Southern towns and cities in 1919 (Grossman 1989).

⁵Rubin (1960) studied individuals from Houston, Mississippi, because so many migrants from Houston moved to Beloit, Wisconsin. While interesting, this sample is clearly not representative.

⁶The following paragraph draws on Bell (1933). See also Knowles (2010).

to 1936, 14 percent of migrants from Pontotoc lived in Beloit's county in old age (Table 2, discussed in Section IIIA).

Migration out of the Great Plains has received less academic attention than the Great Migration, but nonetheless represents a landmark reshuffling of the US population. Considerable out-migration from the Great Plains started around 1930 (Johnson and Rathge 2006). Explanations for the out-migration include the decline in agricultural prices due to the Great Depression, a drop in agricultural productivity due to drought, and the mechanization of agriculture (Gregory 1989, Curtis White 2008, Hurt 2011, Hornbeck 2012). Some historical work points to an important role for migration networks (Jamieson 1942, Gregory 1989). For example, Jamieson (1942) finds that almost half of migrants to Marysville, California, had friends or family living there.

The mass migrations out of the South and Great Plains share several features. In both episodes, millions of people made long-distance moves in search of better economic and social opportunities. Both episodes occurred around the same time and saw a similar share of the population undertake long-distance moves, as we describe in Sections IIA and IIIA. In addition, both African American and White migrants experienced discrimination in many destinations, although African Americans faced far more severe discrimination and had less wealth (Gregory 2005).

II. Estimating the Effects of Migration Networks on Location Decisions

A. Data on Location Decisions

To measure location decisions, we use the Duke University SSA/Medicare data, which covers over 70 million individuals who received Medicare Part B from 1976 to 2001 (Duke Social Security–Medicare Part B Matched Data, 2002). The data contain race, sex, date of birth, date of death (if deceased), and the zip code of residence at old age (death or 2001, whichever is earlier). In addition, the data include a 12-character string with self-reported birth town information from the SSA NUMIDENT file, which is matched to places, as described in Black et al. (2015). We use the data to measure long-run migration flows from birth town to destination county for individuals born from 1916 to 1936.⁷ These cohorts lie at the center of both episodes and have among the highest out-migration rates (online Appendix Figure A.1). As seen in Figure 1, which we construct using repeated cross sections of decennial census data, the vast majority of Southern Black migrants and Great Plains White migrants born from 1916 to 1936 migrated between 1940 and 1960. Most of these migrants were 15–35 years old when they moved (online Appendix Figure A.2). To improve the reliability of our estimates, we restrict the sample to birth towns with at least ten migrants and, separately for each birth state, combine all destination counties with less than ten migrants.

Figure 2 displays the states we include in the South and Great Plains. For migration out of the South, we study individuals born in Alabama, Florida, Georgia,

⁷Our sample begins with the 1916 cohort because coverage rates are low for prior years (Black et al. 2015) and ends with 1936 because that is the last cohort available in the data.

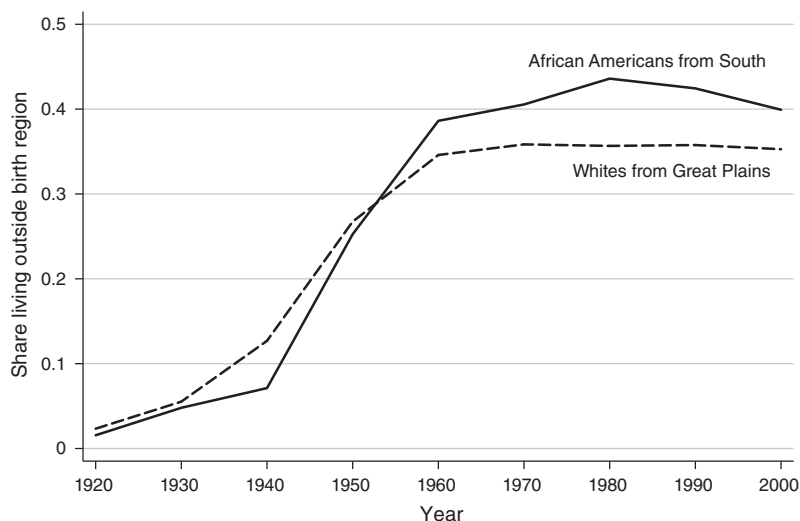


FIGURE 1. SHARE LIVING OUTSIDE BIRTH REGION, 1916–1936 COHORTS, BY YEAR

Notes: The solid line shows the percent of African Americans born from 1916 to 1936 in the seven Southern birth states we analyze (dark grey states in Figure 2, panel A) living outside the South (light and dark grey states) at the time of census enumeration. The dashed line shows the percent of Whites born from 1916 to 1936 from the Great Plains states living outside the Great Plains or border states.

Source: Authors' calculations using Ruggles et al. (2019)

Louisiana, Mississippi, North Carolina, and South Carolina. We define a migrant as someone who moved out of the 11 former Confederate states.⁸ For migration out of the Great Plains, we study individuals born in Kansas, Nebraska, North Dakota, Oklahoma, and South Dakota. We define a migrant as someone who moved out of the Great Plains and a border region, shaded in light grey in Figure 2, panel B.⁹ We make these choices to focus on the long-distance moves that characterize both migration episodes.

Our data capture long-run location decisions, as we only observe individuals' location at birth and old age. We cannot identify return migration: if an individual moved from Mississippi to Wisconsin, then returned to Mississippi at age 60, we do not identify that person as a migrant. It would be interesting to examine short- and medium-run location decisions, but unfortunately the available data do not allow this.¹⁰ Still, the effect of social networks on long-run location decisions is

⁸These include the seven states already listed, plus Arkansas, Tennessee, Texas, and Virginia. The former Confederate states are arguably more culturally, economically, and historically homogeneous during this time than the census definition of the South.

⁹This border region includes Arkansas, Colorado, Iowa, Minnesota, Missouri, Montana, New Mexico, Texas, and Wyoming. We do not focus primarily on Dust Bowl migration. Our Great Plains states did experience soil erosion in the 1930s, but other states also experienced soil erosion (see Hornbeck 2012), and the Southern Great Plains states of Colorado, Kansas, Oklahoma, and Texas are most associated with the Dust Bowl (Long and Siu 2018).

¹⁰To study short-run location decisions, we linked individuals between the 1920 and 1940 complete count censuses, as in Abramitzky, Boustan, and Eriksson (2017). The resulting sample size was too small to generate reliable estimates. For example, of the 334,605 Southern Black migrants in the 1940 census, we were only able to use 18,312 migrants (5.5 percent) to estimate our network index. This low coverage rate is mainly due to our ability to match only 12.5 percent of Southern Black migrants from the 1940 to 1920 census (in line with the match rates for

Panel A. South



Panel B. Great Plains

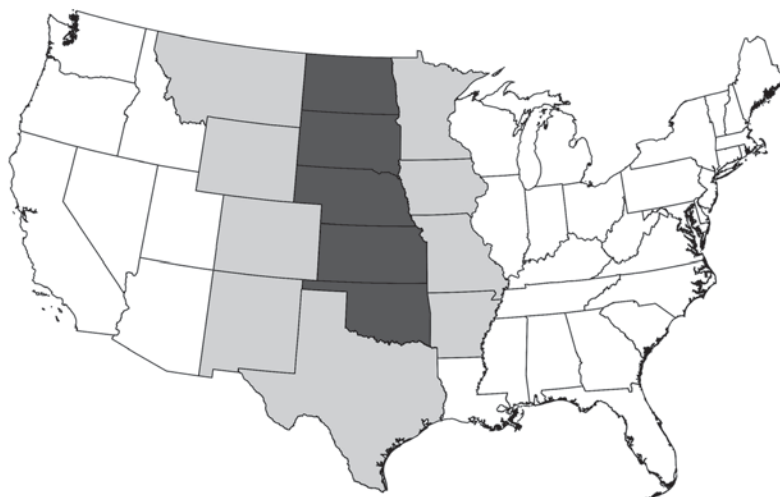


FIGURE 2. GEOGRAPHIC COVERAGE

Notes: For the South, our sample includes migrants born in the seven states in dark grey (Alabama, Georgia, Florida, Louisiana, Mississippi, North Carolina, South Carolina). A migrant is someone who at old age lives outside of the former Confederate states, which are the dark and light grey states. For the Great Plains, our sample includes migrants born in the five states in dark grey (Kansas, Nebraska, North Dakota, Oklahoma, South Dakota). A migrant is someone who at old age lives outside of the Great Plains states and the surrounding border area.

of substantial interest. We also do not observe individuals who die before age 65 or do not enroll in Medicare. We discuss the implications of these measurement issues in online Appendix D.

African Americans in Eriksson, forthcoming). The low coverage rate also stems from our exclusion of birth towns (minor civil divisions in the census) with fewer than ten migrants. The coverage rate for Whites from the Great Plains is also too low (8.4 percent) to generate reliable results.

B. Econometric Model: The Network Index

A natural starting point for an econometric model is the influential approach of Bayer, Ross, and Topa (2008), which leverages detailed geographic data to identify the effects of networks. Using data from the 1990 census, they estimate whether individuals are more likely to work in the same location when they live on the same census block compared to when they live on different blocks in the same block group (a larger geographic area). They measure the strength of neighborhood job referrals as the additional propensity of neighbors to work together.

Our empirical strategy also uses detailed geographic data to identify the strength of networks. In particular, we aim to distinguish the effect of birth town migration networks from other determinants of location decisions, such as moving costs determined by geography or railroad lines. Following Bayer, Ross, and Topa (2008), one approach is to estimate whether migrants from the same birth town are more likely to live in the same destination than migrants from different nearby towns. Mapping their model to our setting yields

$$(1) \quad D_{i,j(i),k} D_{i',j(i'),k} = \alpha_{g,k} + \sum_{j \in g} \beta_{j,k} \mathbf{1}[j(i) = j(i') = j] + \epsilon_{i,i',k},$$

where $D_{i,j,k} = 1$ if migrant i moves from birth town j to destination county k and $D_{i,j,k} = 0$ if migrant i moves elsewhere, $j(i)$ is the birth town of migrant i , and both i and i' live in birth town group g . As described in Section IIC, we define birth town groups in two ways: counties and square grids independent of county borders. The fixed effect $\alpha_{g,k}$ equals the average propensity of migrants from birth town group g to collocate in destination k , and $\beta_{j,k}$ equals the additional propensity of migrants from the same birth town j to collocate in k . Equation (1) allows location decision determinants to vary arbitrarily at the birth town group-by-destination level through $\alpha_{g,k}$ (e.g., because of differences in migration costs due to railroads or highways). The parameter of interest, $\beta_{j,k}$, is identified from within birth town group comparisons. This equation slightly generalizes the main specification in Bayer, Ross, and Topa (2008) by allowing parameters to depend on destination, k .¹¹

The parameters governing networks in this setting are the probability of moving to a destination and the covariance of location decisions among migrants from the same town. We denote the probability that a migrant born in town j chooses destination k as $P_{j,k} \equiv E[D_{i,j,k}]$. This ex ante probability reflects individuals' preferences, resources, and the expected return to migration, but does not depend on other individuals' realized location decisions. The average covariance of location decisions for two migrants from the same town is $C_{j,k} \equiv \sum_{i \neq i' \in j} \text{cov}[D_{i,j,k}, D_{i',j,k}] / (N_j(N_j - 1))$.¹² The number of people who move from j to k is $N_{j,k} \equiv \sum_{i \in j} D_{i,j,k}$, and the number of migrants from birth town j is $N_j \equiv \sum_k N_{j,k}$.

¹¹ In their initial specification, $\alpha_{g,k}$ does not vary by k , and $\beta_{j,k}$ does not vary by j or k . In other specifications, they allow the slope coefficient to depend on observed characteristics of the pair (i, i') .

¹² Glaeser, Sacerdote, and Scheinkman (1996) use the "excess variance" of decisions to infer the presence of social networks. This approach is very closely related, as a positive covariance of decisions increases the variance (see also Graham 2008). Brock and Durlauf (2001) and Blume et al. (2011) provide comprehensive discussions of the related topic of estimating social interactions.

To better understand the reduced form in equation (1), we map the parameters of the generalized Bayer, Ross, and Topa (2008) model, $(\alpha_{g,k}, \beta_{j,k})$, into parameters governing social networks, $(P_{j,k}, C_{j,k})$. Doing so requires two assumptions. The most important assumption is that $P_{j,k}$ is constant across birth towns in the same group.

ASSUMPTION 1: $P_{j,k} = P_{j',k}$ for different birth towns in the same birth town group, $j \neq j' \in g$.

Assumption 1 formalizes the idea that there are no ex ante differences across nearby birth towns in the value of moving to each destination. This assumption is consistent with the presence of pull and push factors, as long as these factors do not vary across birth town-destination pairs. For example, this assumes away the possibility that migrants from Pigeon Creek, Alabama, had preferences or human capital particularly suited for Niagara Falls, New York, relative to migrants from a nearby town, such as Oaky Streak, which is six miles away. Assumption 1 attributes large differences in realized moving propensities across nearby towns to migration networks.

We do not restrict the probability of moving from birth town group g to destination k , $P_{g,k}$, so pull and push factors can vary arbitrarily across birth town group-destination pairs. For example, allowing $P_{g,k}$ to vary flexibly across birth town groups accords with the fact that some Great Plains migrants chose specific destinations in California to pick cotton (Gregory 1989). Assumption 1 covers the probability of choosing a destination, conditional on migrating, which is the focus of our paper; it does not restrict out-migration probabilities.

Assumption 1 is plausible in our setting. Preferences for destination features, such as wages or climate, and information about destinations likely did not vary sharply across nearby birth towns. Furthermore, individuals tended to work in different industries after migrating (online Appendix Table A.1), suggesting a negligible role for human capital specific to a destination county that differed across nearby birth towns. Conditional on migrating, the cost of moving to a given destination likely did not vary sharply across nearby towns.¹³

Section IIIB describes evidence that supports the validity of Assumption 1. Most importantly, we show that using birth town covariates to explain moving probabilities does not affect network index estimates. This implies that geographic proximity adequately controls for the relevant determinants of location decisions, as embedded in Assumption 1. In addition, our results are similar for individuals born from 1916–1925 and 1926–1936; the latter group was much less likely to serve in World War II, which suggests that our results are not driven by networks formed in the military.

The second assumption is that migrants' location decisions are not influenced by migrants from other birth towns.

¹³ Assumption 1 is not violated if the cost of moving to all destinations varied sharply across birth towns (e.g., because of proximity to a railroad), as we focus on where people move, conditional on migrating.

ASSUMPTION 2: $\text{cov}[D_{i,j,k}, D_{i',j',k}] = 0$ for migrants from different birth towns, $j \neq j'$.

Assumption 2 allows us to map the parameters of the extended Bayer, Ross, and Topa (2008) model, $(\alpha_{g,k}, \beta_{j,k})$, into the parameters governing social networks, $(P_{j,k}, C_{j,k})$. Migration networks that extend across nearby towns, which violate Assumption 2, would lead us to underestimate the effect of birth town migration networks. Section IIIB describes evidence that supports the validity of this assumption.

Under Assumptions 1 and 2, the slope coefficient in equation (1) equals the covariance of location decisions from birth town j to destination k : $\beta_{j,k} = C_{j,k}$.¹⁴ In addition, the fixed effect in equation (1) equals the squared moving probability: $\alpha_{g,k} = P_{g,k}^2$. This analysis demonstrates that the Bayer, Ross, and Topa (2008) model uses the covariance of decisions to measure the effect of networks.

The Bayer, Ross, and Topa (2008) model could mischaracterize network effects when the moving probability varies across destinations, because the covariance depends on the moving probability. To see this, let $\mu_{j,k} \equiv E[D_{i,j,k} | D_{i',j',k} = 1]$ be the probability that a migrant moves from birth town j to destination k , conditional on a randomly chosen migrant from j making the same move. Slightly manipulating the definition of the covariance of location decisions yields

$$(2) \quad C_{j,k} = P_{g,k}(\mu_{j,k} - P_{g,k}).$$

Equation (2) shows that variation in $C_{j,k}$ arises from two sources: the probability of moving to a destination, $P_{g,k}$, and the “marginal network effect,” $\mu_{j,k} - P_{g,k}$. For example, $C_{j,k}$ could be large for a popular destination like New York because $P_{g,k}$ is large, even if $\mu_{j,k} - P_{g,k}$ is small. For less popular destinations, $\mu_{j,k} - P_{g,k}$ could be large, but $C_{j,k}$ will be small if $P_{g,k}$ is sufficiently small. Because $P_{g,k}$ varies tremendously in our setting, the covariance of location decisions or any aggregation of the covariance is not an attractive measure of the effect of networks.

To measure the effect of birth town migration networks, we propose an intuitive network index that equals the expected increase in the number of people from birth town j that move to destination county k when an arbitrarily chosen person i makes the same move,

$$(3) \quad \Delta_{j,k} \equiv E[N_{-i,j,k} | D_{i,j,k} = 1] - E[N_{-i,j,k} | D_{i,j,k} = 0],$$

¹⁴ Proof:

$$\begin{aligned} \beta_{j,k} &= E[D_{i,j(i),k} D_{i',j(i'),k} | j(i) = j(i') = j] - E[D_{i,j(i),k} D_{i',j(i'),k} | j(i) \neq j(i')] \\ &= E[D_{i,j(i),k} D_{i',j(i'),k} | j(i) = j(i') = j] - (E[D_{i,j,k}])^2 \\ &= \text{cov}[D_{i,j,k} D_{i',j',k}] = C_{j,k}. \end{aligned}$$

The first line follows directly from equation (1). The second line follows from Assumptions 1 and 2. The third line follows from the definition of covariance.

where $N_{-i,j,k}$ is the number of people who move from j to k , excluding person i . A positive value of $\Delta_{j,k}$ indicates that the network increases the number of people who move from j to k , while $\Delta_{j,k} = 0$ indicates no effect of the network.

The network index, $\Delta_{j,k}$, possesses several attractive properties. The network index permits meaningful comparisons, in intuitive units, of effects across heterogeneous receiving and sending locations. The network index also requires minimal assumptions about the specific behaviors that lead to network effects. For example, correlated location decisions could arise because individuals value living near their friends and family or because networks provide information about job opportunities. The network index also is consistent with multiple structural models and can be mapped directly to them. For example, suppose that all migrants in town j form coalitions of size s , all members of a coalition move to the same destination, and all coalitions move independently of each other. In this case, the network index for each destination k depends only on the structural parameter s : $\Delta_{j,k} = s - 1$ because whenever one person moves to a destination, the other members of the coalition follow. In contrast, the covariance of location decisions depends on the moving probability as well: $C_{j,k} = (s - 1)P_{g,k}(1 - P_{g,k}) / (N_j - 1)$.¹⁵ As another example, we connect the network index to a richer structural model in Section IV. The network index also can be estimated nonparametrically with increasingly available data.

In online Appendix A, we show how to express the network index as

$$(4) \quad \Delta_{j,k} = \frac{(\mu_{j,k} - P_{g,k})(N_j - 1)}{1 - P_{g,k}} = \frac{C_{j,k}(N_j - 1)}{P_{g,k} - P_{g,k}^2}.$$

The network index transforms the covariance of location decisions in two ways. First, the covariance is multiplied by $N_j - 1$, which is the number of migrants potentially affected by the location decision of an arbitrarily chosen migrant. Second, the covariance is divided by $P_{g,k} - P_{g,k}^2 = P_{g,k}(1 - P_{g,k})$. This reflects two offsetting forces. On the one hand, the marginal network effect, $\mu_{j,k} - P_{g,k}$, is divided by $1 - P_{g,k}$. The higher the moving probability, the less “room” there is for $\mu_{j,k}$ to exceed $P_{g,k}$, and the denominator amplifies the marginal network effect for these destinations. On the other hand, note that the final expression stems from equation (2), which shows that $\mu_{j,k} - P_{g,k} = C_{j,k} / P_{g,k}$. The intuition for this expression is discussed above: the covariance is deflated for destinations with high moving probabilities, which increase the covariance holding the marginal network effect constant. In our setting, the relevant moving probabilities are well below one-half, where equation (4) assigns less weight to destinations with higher-moving probabilities.

Several other features of equation (4) are noteworthy. The network index depends on the parameters governing social networks, $(P_{g,k}, C_{j,k})$. The network index increases in the marginal network effect, $\mu_{j,k} - P_{g,k}$. If migrants move independently of each other, then $\mu_{j,k} - P_{g,k} = \Delta_{j,k} = 0$. Finally, the network index

¹⁵To see this, note that $\mu_{j,k} \equiv E[D_{i,j,k} | D_{i',j,k} = 1]$ equals 1 if i and i' are in the same coalition, which happens with probability $(s - 1) / (N_j - 1)$, and $\mu_{j,k} = P_{g,k}$ if i and i' are in a different coalition. Simplifying equation (2) then yields the result.

does not necessarily increase in the number of migrants from birth town j , N_j , as the marginal network effect might decrease in N_j .¹⁶

The network index captures actions that generate a positive correlation of location decisions among migrants from the same birth town, relative to what would be predicted by the decisions of migrants from nearby towns. While social networks might affect location decisions in other ways, the network index does not measure them. For example, if social networks affected whether individuals migrated, but not where they moved, then the network index would equal zero. Relatedly, the network index is an average over all migrants, so it could vary with the set of migrants if individuals differ in how much they influence and are influenced by others.¹⁷

The network index equals the expected increase in the number of people that move from j to k when a randomly chosen person makes the same move. This does not necessarily equal the expected increase in the number of people that move from j to k because a randomly chosen person makes the same move. The relationship between these two parameters depends on the underlying structural model. For example, in the coalition model described above—where all migrants in town j form coalitions of size s , all members of a coalition move to the same destination, and all coalitions move independently of each other—these two parameters are identical and equal to $s - 1$. Alternatively, if each coalition has one leader, and all other members of the coalition follow the leader, then the network index equals $s - 1$, but the expected increase in the number of people that move from j to k because a randomly chosen person makes the same move is $(s - 1)/s$. This distinction arises because the network index relies on weak assumptions about the underlying structural model. The weakness of these assumptions and the ability to map the network index directly to several structural models are valuable features of our approach.

C. Estimating the Network Index

As suggested by equation (4), estimating the network index is straightforward. We first define birth town groups, and then nonparametrically estimate the underlying parameters $P_{g,k}$, $P_{g,k}^2$, and $C_{j,k}$.

We define birth town groups in two ways. Our preferred approach balances the inclusion of very close towns, for which Assumption 1 likely holds, with the inclusion of towns that are farther away and lead to a more precise estimate of $P_{g,k}$. We divide each birth state into a grid of squares with sides x^* miles long and choose x^* separately for each state using leave-one-out cross validation. This technique is regularly used for bandwidth selection of matching estimators (e.g., Black and Smith 2004), and it chooses the grid size that minimizes the mean squared error of the observed migration propensities and “out-of-sample” forecasts from other towns in

¹⁶In addition, $-1 \leq \Delta_{j,k} \leq N_j - 1$. At the upper bound, all migrants from j move to the same location, while at the lower bound, migrants displace each other one-for-one.

¹⁷Our approach allows migration networks to influence out-migration, but we do not separately examine this channel.

the same birth town group.¹⁸ Given x^* , the location of the grid is determined by a single latitude-longitude reference point. Network index estimates are very similar across four different reference points, so we average estimates across them.¹⁹

An alternative definition of a birth town group is a county. If the value of choosing a destination varied sharply across county borders in the sending region, then this definition would be appropriate. However, differences across counties, such as local government policies and elected officials, do not necessarily imply that counties are better birth town groups, as what matters is whether these differences affect the ex ante probability of choosing a destination, conditional on migrating. An advantage of cross validation is that it facilitates comparisons across birth states, which differ in average county size for many reasons not related to migration incentives. We emphasize results based on cross validation in the main text and include results based on counties in the online Appendix.²⁰

We estimate the probability of moving from birth town group g to destination county k as the total number of people who move from g to k divided by the total number of migrants in g ,

$$(5) \quad \hat{P}_{g,k} = \frac{\sum_{j \in g} N_{j,k}}{\sum_{j \in g} N_j}$$

We estimate the squared moving probability and covariance of location decisions using the closed-form solution implied by equation (1),²¹

$$(6) \quad \hat{P}_{g,k}^2 = \frac{\sum_{j \in g} \sum_{j' \neq j \in g} N_{j,k} N_{j',k}}{\sum_{j \in g} \sum_{j' \neq j \in g} N_j N_{j'}}$$

$$(7) \quad \hat{C}_{j,k} = \frac{N_{j,k}(N_{j,k} - 1)}{N_j(N_j - 1)} - \hat{P}_{g,k}^2$$

The final component of the network index is the number of migrants from birth town j , N_j .

¹⁸That is,

$$x^* = \arg \min_x \sum_j \sum_k (N_{j,k}/N_j - \hat{P}_{g(x),-j,k})^2,$$

where $\hat{P}_{g(x),-j,k} = \sum_{j' \neq j \in g(x)} N_{j',k} / \sum_{j' \neq j \in g(x)} N_{j'}$ is the average moving propensity from the birth town group of size x , excluding moves from town j . If there is only one town within a group g , then we define $\hat{P}_{g(x),-j,k}$ to be the statewide moving propensity. We search over even integers for convenience. Online Appendix Table A.2 reports the values of x^* chosen by cross validation.

¹⁹To construct reference points, we use the mean latitude in a state and the mean latitude plus one-third of x^* , scaled in appropriate units. We use analogous reference points for longitude.

²⁰Online Appendix Figures A.3 and A.4 describe the number of birth towns per group when groups are defined using cross validation for Southern Black migrants and Great Plains White migrants. The median number of towns per group is 15 for African Americans and 39 for Whites from the Great Plains. Online Appendix Figures A.5 and A.6 describe the number of towns per county. All groups used in estimation have at least two towns, because we cannot estimate $C_{j,k}$ or $P_{g,k}^2$ without multiple towns in the same group.

²¹Equation (6) yields an unbiased estimate of $P_{g,k}^2$ under Assumptions 1 and 2. In contrast, simply squaring $\hat{P}_{g,k}$ would result in a biased estimate.

Given $(\hat{P}_{g,k}, \hat{P}_{g,k}^2, \hat{C}_{j,k}, N_j)$, we can estimate the network index, $\Delta_{j,k}$, using equation (4). However, each estimate $\hat{\Delta}_{j,k}$ depends largely on a single birth town observation. To conduct inference, increase the reliability of our estimates, and decrease the number of parameters reported, we aggregate network index estimates across all birth towns in each state,

$$(8) \quad \hat{\Delta}_k = \sum_j \left(\frac{\hat{P}_{g(j),k} - \hat{P}_{g(j),k}^2}{\sum_{j'} \hat{P}_{g(j'),k} - \hat{P}_{g(j'),k}^2} \right) \hat{\Delta}_{j,k}$$

where $g(j)$ is the group of town j . The weights in equation (8) arise from a generalized method of moments (GMM) estimator in which $\Delta_{j,k}$ is assumed to not vary across birth towns within a state and each birth town group receives equal weight.²² The weights equal the ex ante variance of a migrant’s location decision (i.e., $\text{var}[D_{i,j,k}] = P_{g,k}(1 - P_{g,k})$ because $E[D_{i,j,k}] = P_{g,k}$). For each destination, birth town groups with a moving probability closer to one-half receive greater weight, as these towns provide more information about the influence of social networks. Intuitively, if a group’s migrants are nearly certain to move or not move to a destination, then this group is less valuable for identification. Furthermore, the destination-level network index estimate, $\hat{\Delta}_k$, is robust to small estimates of $P_{g,k}$, which can blow up estimates of $\Delta_{j,k}$. We also construct birth county-level network index estimates by aggregating across destinations and towns within birth county c ,

$$(9) \quad \hat{\Delta}_c = \sum_k \sum_{j \in c} \left(\frac{\hat{P}_{g(j),k} - \hat{P}_{g(j),k}^2}{\sum_{k'} \sum_{j' \in c} \hat{P}_{g(j'),k'} - \hat{P}_{g(j'),k'}^2} \right) \hat{\Delta}_{j,k}$$

Birth county-level network index estimates have similar conceptual and statistical properties as destination-level network index estimates.

To facilitate exposition, we have described estimation of the network index in terms of four distinct components, $(\hat{P}_{g,k}, \hat{P}_{g,k}^2, \hat{C}_{j,k}, N_j)$. However, network index estimates depend only on observed population flows, and equation (8) forms the basis of an exactly identified GMM estimator. To estimate the variance of $\hat{\Delta}_k$, we treat the birth town group as the unit of observation and use a GMM variance estimator. This is akin to calculating heteroskedastic robust standard errors clustered by birth town group.²³ Online Appendix B contains details.

D. An Extension to Assess the Validity of Our Empirical Strategy

The key threat to our empirical strategy is that the ex ante value of moving to a destination differs across nearby birth towns in the same group. If, contrary to this threat, Assumption 1 were true, then geographic proximity would adequately

²² See online Appendix B for details.

²³ Treating birth town groups as the units of observation has no impact on the point estimate, $\hat{\Delta}_k$. We estimate clustered standard errors because the estimates $\hat{P}_{g,k}$ and $\hat{P}_{g,k}^2$ are common to all birth towns within g .

control for the relevant determinants of location decisions, and using birth town covariates to explain moving probabilities would not affect network index estimates.

We assess this threat by allowing moving probabilities to depend on birth town covariates,

$$(10) \quad P_{j,k} = \rho_{g,k} + X_j \pi_k,$$

where $\rho_{g,k}$ is a birth town group-destination fixed effect, and X_j is a vector of birth town covariates whose effect on the moving probability can differ across destinations. We consider two sets of variables for X_j . First, we use the Duke SSA/Medicare data and the railroad information used in Black et al. (2015) to construct indicators for being along a railroad and having above-median Black population share and town size (based on the 1916–1936 cohorts in the Duke data). Second, we use the complete count 1910 census data to construct indicators for whether towns have an above-median value of the following variables: percent age 0–17, percent literate (age 10–39), percent homeowner, percent farmer/farm laborer (age 18–39), percent interstate migrant (age 18–60), and percent immigrant.²⁴ We match the aggregated 1910 census data to the Duke SSA/Medicare data using place names.²⁵ These variables capture potentially relevant determinants of location decisions. For example, migrants born in towns that are larger or have higher literacy rates might have more human capital or information, and these resources might make certain destinations more attractive, causing our network index estimates to reflect variables correlated with birth town size instead of migration networks.

To implement this extension, we construct alternative network index estimates using an alternative moving probability estimate, $\tilde{P}_{j,k}$, equal to the fitted value from the OLS regression

$$(11) \quad \frac{N_{j,k}}{N_j} = \rho_{g,k} + X_j \pi_k + e_{j,k}.$$

We use fitted values from a separate OLS regression, implied by equation (10), to form an alternative squared moving probability estimate, $\tilde{P}_{j,k}^2$.²⁶ We estimate all

²⁴We construct medians separately for each birth state, and all variables are race-specific except for percent immigrant.

²⁵The census data that have been processed by IPUMS do not contain individuals' town of residence, but do include information on minor civil division (MCD), which is a subcounty administrative unit. We are not aware of a crosswalk from 1910 MCDs to town FIPS codes. However, the census includes MCD titles, and these often include town names. Consequently, we are able to match the Duke and census data using a match on town and MCD names within the county. We achieved a match for 61 percent of the Great Plains towns for Whites and 58 percent of Southern towns for African Americans. Large towns are more likely to be matched, and so we have a census match for 87 percent of both Great Plains White migrants and Southern Black migrants. For towns that do not have a match in the census data, we calculate covariates by using the *unmatched* MCDs that are in the same county. Although town name is not available in the data cleaned by IPUMS, the Census Bureau did collect information on town of residence, and one could create a crosswalk from the "cleaned" to "uncleaned" datasets (both of which are available through the NBER); we thank an anonymous referee for bringing the possibility of this crosswalk to our attention.

²⁶We estimate $\tilde{P}_{j,k}^2$ using fitted values from the OLS regression

$$\frac{N_{j,k}}{N_j} \frac{N_{j',k}}{N_{j'}} = \rho_{g(j),k} \rho_{g(j'),k} + X_j \pi_k \rho_{g(j'),k} + X_{j'} \pi_k \rho_{g(j),k} + (X_j \pi_k)(X_{j'} \pi_k) + e'_{j,j',k}$$

for different birth towns, $j \neq j'$.

TABLE 1—LOCATION AT OLD AGE, 1916–1936 COHORTS

Birth state	People (1)	Percent living in location		
		Outside birth region (2)	In birth region	
			Birth state (3)	Other state (4)
<i>Panel A. Black individuals from South</i>				
Alabama	209,128	47.2	39.5	13.3
Florida	79,237	26.1	67.1	6.8
Georgia	218,357	36.3	44.2	19.5
Louisiana	179,445	32.4	52.7	14.9
Mississippi	218,759	56.1	28.9	15.0
North Carolina	200,999	40.2	49.7	10.1
South Carolina	163,650	43.4	41.9	14.7
Total	1,269,575	41.8	44.0	14.1
<i>Panel B. White individuals from Great Plains</i>				
Kansas	462,490	30.4	43.3	26.3
Nebraska	374,265	36.0	42.0	22.0
North Dakota	210,199	44.1	31.8	24.1
Oklahoma	635,621	31.8	41.6	26.6
South Dakota	196,266	40.4	35.4	24.2
Total	1,878,841	34.6	40.3	25.1
<i>Panel C. White individuals from South</i>				
Alabama	469,698	9.8	62.1	28.1
Florida	231,071	12.7	68.5	18.8
Georgia	454,286	7.4	65.5	27.1
Louisiana	384,601	8.7	71.1	20.2
Mississippi	275,147	11.0	57.0	32.0
North Carolina	588,674	8.5	71.6	19.8
South Carolina	238,697	6.6	70.6	22.8
Total	2,642,174	9.0	66.9	24.0

Notes: Column 1 contains the number of people from the 1916–1936 birth cohorts observed in the Duke SSA/Medicare data. Columns 2–4 display the share of individuals living in each location at old age (2001 or date of death, if earlier). Figure 2 displays birth regions.

Source: Authors' calculations using Duke SSA/Medicare data

equations separately for each birth state. Similarity between the baseline and alternative network index estimates would provide support for our empirical strategy.²⁷

III. Results: The Effects of Migration Networks on Location Decisions

A. Network Index Estimates

Table 1 provides an overview of the long-run population flows that we use to estimate the effects of migration networks. Our data contain 1.3 million African Americans born in the South from 1916 to 1936, 1.9 million Whites born in the Great Plains, and 2.6 million Whites born in the South. In old age, 42 percent of African Americans born in the South and 35 percent of Whites born in the Great

²⁷ An alternative way of assessing the validity of Assumption 1 is testing whether the parameter vector $\pi_k = 0$ in equation (11). We prefer to test the difference in network index estimates because this approach allows us to consider the substantive significance of any differences.

TABLE 2—EXTREME EXAMPLES OF CORRELATED LOCATION DECISIONS, SOUTHERN BLACK MIGRANTS AND GREAT PLAINS WHITE MIGRANTS

Birth town (1)	Largest city in destination county (2)	Total birth town migrants (3)	Town- destination flow (4)	Destination share of birth town migrants (5)	Destination share of birth state migrants (6)	SD under independent binomial moves (7)	Moving probability estimate (8)	Network index estimate (9)
<i>Panel A. Southern Black migrants</i>								
Pigeon Creek, AL	Niagara Falls, NY	85	43	50.6%	0.5%	64.5	4.5%	8.5
Marion, AL	Fort Wayne, IN	1,311	200	15.3%	0.7%	63.7	3.8%	8.8
Greeleyville, SC	Troy, NY	215	34	15.8%	0.1%	62.2	1.7%	15.2
Athens, AL	Rockford, IL	649	64	9.9%	0.2%	61.0	2.0%	5.6
Pontotoc, MS	Janesville, WI	456	62	13.6%	0.2%	59.4	3.3%	6.5
New Albany, MS	Racine, WI	599	97	16.2%	0.4%	58.7	4.9%	11.4
West, MS	Freeport, IL	336	35	10.4%	0.1%	56.9	0.8%	6.2
Gatesville, NC	New Haven, CT	176	88	50.0%	1.6%	51.8	8.1%	7.1
Statham, GA	Hamilton, OH	75	22	29.3%	0.3%	50.0	3.0%	4.4
Cochran, GA	Paterson, NJ	259	62	23.9%	0.6%	49.4	4.1%	6.3
<i>Panel B. Great Plains White migrants</i>								
Krebs, OK	Akron, OH	210	32	15.2%	0.1%	82.6	0.2%	8.1
Haven, KS	Elkhart, IN	144	22	15.3%	0.1%	51.1	0.4%	7.1
McIntosh, SD	Rupert, ID	299	20	6.7%	0.1%	50.9	0.6%	4.5
Hull, ND	Bellingham, WA	55	24	43.6%	0.5%	44.6	1.9%	4.2
Lindsay, NE	Moline, IL	226	29	12.8%	0.2%	41.5	0.4%	5.1
Corsica, SD	Holland, MI	253	26	10.3%	0.2%	39.6	0.3%	6.4
Corsica, SD	Grand Rapids, MI	253	34	13.4%	0.3%	37.2	0.5%	6.1
Montezuma, KS	Merced, CA	144	21	14.6%	0.3%	32.7	0.9%	2.7
Hillsboro, KS	Fresno, CA	407	65	16.0%	0.9%	32.0	1.2%	2.3
Henderson, NE	Fresno, CA	146	32	21.9%	0.7%	31.1	0.9%	2.3

Notes: Each panel contains the most extreme examples of correlated location decisions, as determined by column 7. Column 7 equals the difference, in standard deviations, of the actual moving propensity (column 5) relative to the prediction with independent moves following a binomial distribution governed by the statewide moving propensity (column 6). Column 8 equals the estimated probability of moving from town j to county k using observed location decisions from nearby towns, where the birth town group is defined by cross validation. Column 9 equals the destination-level network index estimate for the relevant birth state. When choosing these examples, we restrict attention to town-destination pairs with at least 20 migrants.

Source: Authors' calculations using Duke SSA/Medicare data

Plains lived outside their birth region, while only 9 percent of Whites born in the South lived elsewhere.²⁸ We focus on Southern-born African Americans and Great Plains-born Whites, and leave results for Southern-born Whites for the online Appendix. On average, there are 142 Southern Black migrants and 181 Great Plains White migrants per birth town (online Appendix Table A.3).²⁹

We begin with some examples that illustrate how we identify the effects of birth town migration networks. Table 2 shows the birth town to destination county migration flows that would be most unlikely in the absence of such networks. Panel A shows that 10–50 percent of African American migrants from these birth towns lived in the same destination county in old age, far exceeding the 0.1–1.6 percent of migrants from each birth state that lived in the same county. The observed moving

²⁸Census data show that return migration was quite low among Southern-born African Americans and much higher among Southern-born Whites (Gregory 2005).

²⁹Online Appendix Tables A.4–A.6 draw on matched census data to describe individuals who did and did not migrate between 1920–1930 and 1930–1940. Relative to White migrants, Black migrants were less likely to attend school, be literate, live in owner-occupied housing, and live in a city. There is mixed evidence on whether migrants became more or less positively selected over time. For related analyses, see Collins and Wanamaker (2014, 2015) and Boustan (2017).

TABLE 3—AVERAGE NETWORK INDEX ESTIMATES, BY BIRTH STATE

Birth state	Number of migrants (1)	Unweighted average (2)	Weighted average (3)
<i>Panel A. Southern Black migrants</i>			
Alabama	96,269	0.770 (0.049)	1.888 (0.195)
Florida	19,158	0.536 (0.052)	0.813 (0.117)
Georgia	77,038	0.735 (0.048)	1.657 (0.177)
Louisiana	55,974	0.462 (0.039)	1.723 (0.478)
Mississippi	120,454	0.901 (0.050)	2.303 (0.313)
North Carolina	78,420	0.566 (0.039)	1.539 (0.130)
South Carolina	69,399	0.874 (0.054)	2.618 (0.301)
All states	516,712	0.736 (0.020)	1.938 (0.110)
<i>Panel B. Great Plains White migrants</i>			
Kansas	139,374	0.128 (0.007)	0.255 (0.024)
Nebraska	134,011	0.141 (0.008)	0.361 (0.082)
North Dakota	92,205	0.174 (0.012)	0.464 (0.036)
Oklahoma	200,392	0.112 (0.008)	0.453 (0.036)
South Dakota	78,541	0.163 (0.009)	0.350 (0.026)
All states	644,523	0.137 (0.004)	0.380 (0.022)

Notes: Column 2 is an unweighted average of destination-level network index estimates, $\hat{\Delta}_k$. Column 3 is a weighted average, where the weights are the number of people who move from each state to destination k . Birth town groups are defined by cross validation. Standard errors are in parentheses.

Source: Authors' calculations using Duke SSA/Medicare data

propensities are 49–65 standard deviations larger than what would be expected if migrants moved independently of each other according to the statewide moving propensities. The estimated moving probabilities, $\hat{P}_{g,k}$, exceed the statewide moving propensities, suggesting a meaningful role for local conditions in location decisions. Most importantly, the observed moving propensities are much larger than the estimated moving probabilities, consistent with positive covariance and network index estimates. The results in panel B for Great Plains White migrants are similar.

To summarize the effects of migration networks for all location decisions in our data, Table 3 reports averages of destination-level network index estimates, $\hat{\Delta}_k$. For African Americans, unweighted averages vary from 0.46 (Louisiana) to 0.90 (Mississippi). Averages weighted by the number of migrants in each destination

vary from 0.81 (Florida) to 2.62 (South Carolina) and are larger because migration networks have stronger effects in destinations with more migrants. We prefer the weighted average as a summary measure because it better reflects the experience of a randomly chosen migrant and depends less on our decision to combine destination counties with fewer than ten migrants. Across all states, the migrant-weighted average of destination-level network index estimates is 1.94; this means that when one randomly chosen African American moves from a birth town to a destination county, then 1.94 additional Black migrants from the same birth town make the same move on average. Panel B contains results for White moves out of the Great Plains. The weighted average for Whites is 0.38, only one-fifth the size of the Black average.³⁰ African American migrants relied on birth town migration networks more heavily in making their long-run location decisions.

We provide a more complete picture in Figure 3, which plots the distribution of destination-level network index estimates.³¹ Across the board, network index estimates for African Americans are larger than those for Whites. Migration networks have particularly strong effects for some destinations, especially for Black migrants, and relatively weak effects for most destinations. In Section IIID, we examine whether destinations' economic characteristics can explain this heterogeneity.

To examine the effects of migration networks even more closely, Figure 4 plots the spatial distribution of destination-level network index estimates for Mississippi-born African Americans. We estimate strong network effects for several destinations: 23 counties have a network index estimate greater than 3, and 58 counties have a network index estimate between 1 and 3. These counties lie in the Midwest and, to a lesser degree, the Northeast. The figure also shows that African Americans moved to a relatively small number of destination counties, consistent with limited opportunities, information, or interest in moving to many places in the United States.³² We estimate a strong network effect ($\hat{\Delta}_k > 3$) in Rock County, Wisconsin, consistent with historical accounts of African Americans who moved from Mississippi to Beloit, which is located there (Bell 1933, Rubin 1960, Wilkerson 2010).

Figure 5 maps the destination-level network index estimates for Whites from North Dakota. We find little evidence of strong network effects, although one exception is San Joaquin county ($\hat{\Delta}_k > 3$), an area described memorably in *The Grapes of Wrath* (Steinbeck 1939).³³ Unlike Black migrants, White migrants moved to a large number of destinations throughout the United States. The difference between the number of destinations chosen by Mississippi Black migrants and North Dakota White migrants is striking, especially because our data contain almost 30,000 more migrants from Mississippi. Although some factors, like discrimination, that led African Americans to move to a smaller number of destinations could also explain their greater reliance on migration networks, the limited number of destinations

³⁰ Online Appendix Table A.7 shows that average network index estimates for Southern Whites are somewhat smaller than for Whites from the Great Plains.

³¹ Online Appendix Figure A.7 displays the associated *t*-statistic distributions, and online Appendix Figures A.8 and A.9 display analogous results for Whites from the South. A destination county can appear multiple times in these figures because we estimate destination-level network indices separately for each birth state.

³² In Figure 4, the counties in white received less than ten migrants.

³³ In *The Grapes of Wrath*, the Joad family travels from Oklahoma to the San Joaquin Valley. Gregory (1989) notes that the (fictional) Joads were poorer than many migrants from the Great Plains.

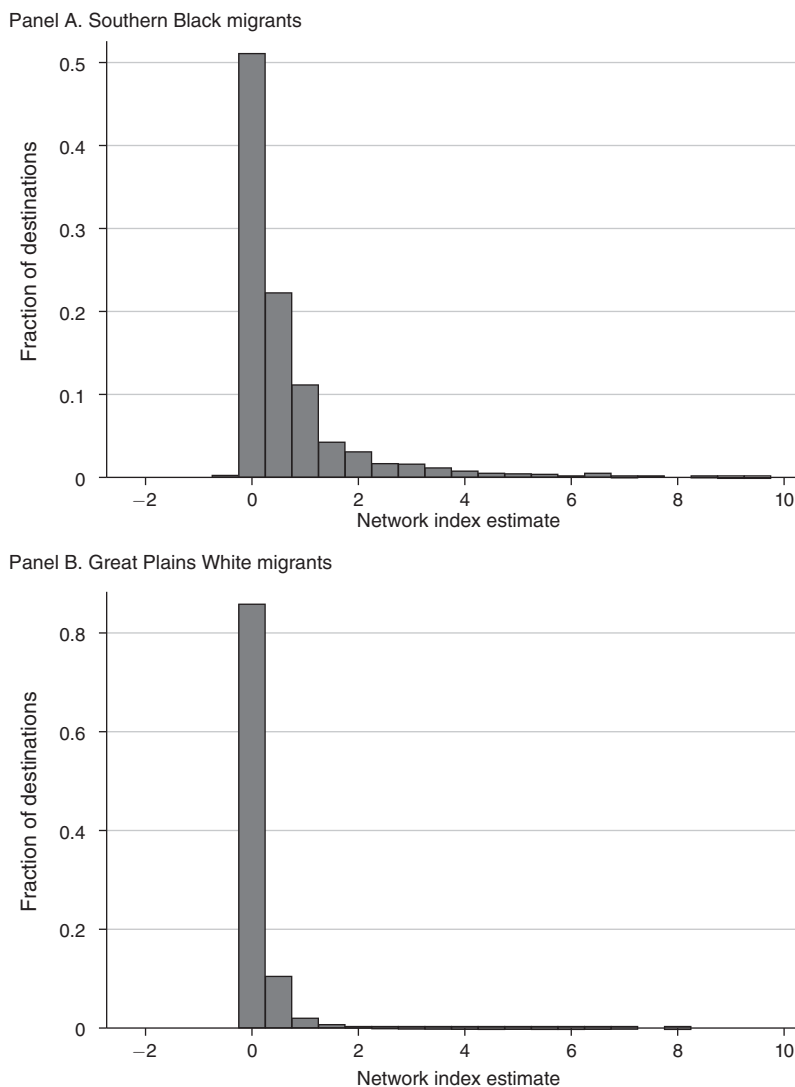


FIGURE 3. DISTRIBUTION OF DESTINATION-LEVEL NETWORK INDEX ESTIMATES

Notes: Bin width is 1/2. Birth town groups are defined by cross validation. Panel A omits the estimate $\hat{\Delta}_k = 11.4$ from Mississippi to Racine County, Wisconsin, $\hat{\Delta}_k = 15.2$ from South Carolina to Rensselaer County, New York, and $\hat{\Delta}_k = 18.1$ from Florida to St. Joseph County, Indiana.

Source: Authors' calculations using Duke SSA/Medicare data

chosen by African Americans does not mechanically generate stronger network effects, because we identify these effects using the location choices of nearby migrants.³⁴ Online Appendix Figures A.10 and A.11, for South Carolina-born Black migrants and Kansas-born White migrants, show similar patterns.

³⁴ Factors that limit the destinations chosen by a group, like discrimination, will tend to increase the probability of moving to a destination, but as discussed above, a higher moving probability does not mechanically increase the network index.

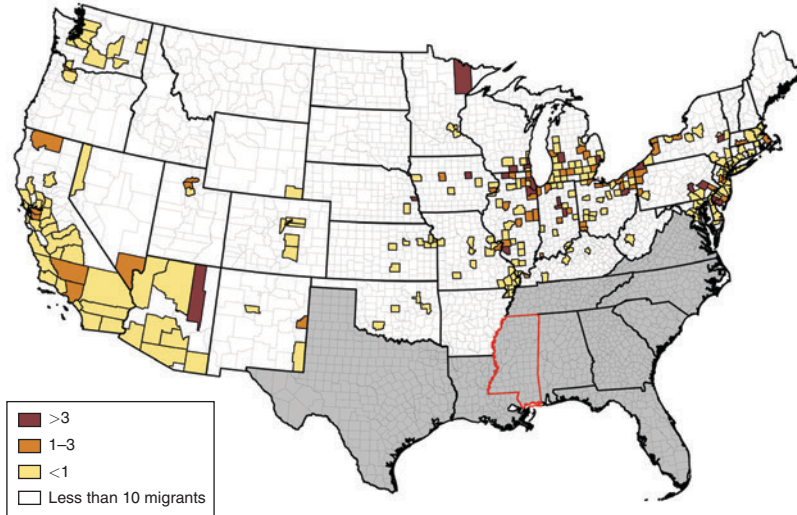


FIGURE 4. SPATIAL DISTRIBUTION OF DESTINATION-LEVEL NETWORK INDEX ESTIMATES, MS-BORN BLACK MIGRANTS

Notes: Figure displays destination-level network index estimates, $\hat{\Delta}_k$, across US counties for Mississippi-born Black migrants. The South is shaded in grey, with Mississippi outlined in red. Destinations to which less than ten migrants moved are in white. Among all African American estimates, $\hat{\Delta}_k = 3$ corresponds to the ninety-fifth percentile, and $\hat{\Delta}_k = 1$ corresponds to the eightieth percentile.

Source: Authors' calculations using Duke SSA/Medicare data

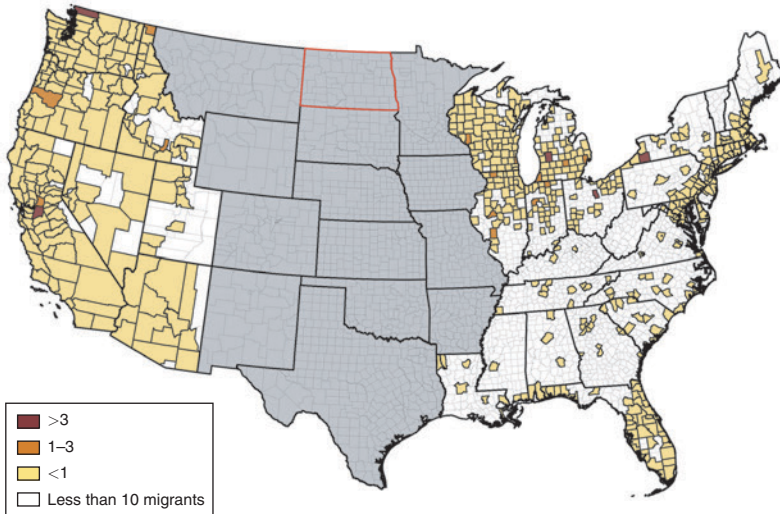


FIGURE 5. SPATIAL DISTRIBUTION OF DESTINATION-LEVEL NETWORK INDEX ESTIMATES, ND-BORN WHITE MIGRANTS

Notes: See note to Figure 4. Among all Great Plains White estimates, $\hat{\Delta}_k = 3$ is greater than the ninety-ninth percentile, and $\hat{\Delta}_k = 1$ corresponds to the ninety-eighth percentile.

Source: Authors' calculations using Duke SSA/Medicare data

B. Support for Empirical Strategy, Additional Results, and Robustness

To assess the validity of Assumption 1, we examine whether network index estimates change when using birth town covariates to explain moving probabilities, as discussed in Section IID. Columns 1–3 of Table 4 report weighted averages of destination-level network index estimates without covariates (our baseline, in column 1) and with them. In particular, column 2 includes covariates from the Duke/SSA data and column 3 adds covariates from the 1910 census. The different sets of estimates are very similar. When pooling all states together, the estimates are 1.94, 1.92, and 1.88 for Southern Black migrants and 0.38, 0.36, and 0.32 for Great Plains White migrants. Moreover, the destination-level network index estimates with and without covariates are highly correlated: the linear (rank) correlation between the estimates in columns 1 and 2 is 0.914 (0.992) for African Americans from the South and 0.939 (0.988) for Whites from the Great Plains. The column 1 and 3 correlation is 0.896 (0.952) for Black migrants and 0.943 (0.945) for White migrants. On net, we view this evidence as indicating that geographic proximity adequately controls for the relevant determinants of location decisions, supporting the validity of Assumption 1.

Violations of Assumption 2 will generally lead us to underestimate the strength of birth town networks. We can relax this assumption by allowing for cross-town interactions between migrants. We first implement this by adding to equation (1) an indicator for whether towns in the same group are within 10 miles of each other. We allow the coefficient on this and subsequent indicators to vary by destination.³⁵ With this additional variable, the modified assumption is that there are no interactions across towns more than 10 miles away from each other. The results are in column 4 of Table 4. Column 5 adds an indicator for whether towns are within 20 miles of each other and below the statewide median in population to allow for the possibility that cross-town interactions are larger in smaller towns. Column 6 further includes an indicator for whether towns lie along the same railroad. The different estimates are quite similar to each other, which implies that violations of Assumption 2 are of little importance.

Table 5 shows that our results are not driven by migration from the largest birth towns or migration to the largest destinations and, relatedly, that there is limited heterogeneity in network index estimates on these dimensions. Birth town size could be correlated with unobserved determinants of migration networks, such as the level of social and human capital or information about destinations. Still, it is not clear beforehand whether networks will vary with the size of receiving or sending locations. For reference, column 1 of Table 5 reports weighted averages of destination-level network index estimates when including all birth towns and destinations. In column 2, we exclude birth towns with at least 20,000 residents in 1920 when estimating each destination-level network index.³⁶ Column 3

³⁵We continue to use equations (5) and (6) to estimate $P_{g,k}$ and $P_{g,k}^2$.

³⁶The excluded birth towns are Birmingham, Mobile, and Montgomery, Alabama; Jacksonville, Miami, Pensacola, and Tampa, Florida; Atlanta, Augusta, Columbus, Macon, and Savannah, Georgia; Baton Rouge, New Orleans, and Shreveport, Louisiana; Jackson and Meridian, Mississippi; Asheville, Charlotte, Durham, Raleigh, Wilmington, and Winston-Salem, North Carolina; Charleston, Greenville, and Spartanburg, South Carolina; Hutchinson, Kansas City, Topeka, and Wichita, Kansas; Lincoln and Omaha, Nebraska; Fargo, North Dakota; Muskogee, Oklahoma City, and Tulsa, Oklahoma; and Sioux Falls, South Dakota.

TABLE 4—AVERAGE NETWORK INDEX ESTIMATES, ROBUSTNESS TO IDENTIFYING ASSUMPTIONS

Birth state	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Southern Black migrants</i>						
Alabama	1.888 (0.195)	1.852 (0.189)	1.648 (0.246)	2.120 (0.198)	2.121 (0.198)	2.178 (0.212)
Florida	0.813 (0.117)	0.742 (0.119)	0.737 (0.132)	0.829 (0.116)	0.829 (0.117)	0.923 (0.125)
Georgia	1.657 (0.177)	1.689 (0.175)	1.903 (0.191)	1.768 (0.179)	1.788 (0.179)	1.731 (0.192)
Louisiana	1.723 (0.478)	1.651 (0.474)	1.648 (0.715)	1.756 (0.455)	1.761 (0.452)	1.709 (0.451)
Mississippi	2.303 (0.313)	2.295 (0.306)	2.098 (0.402)	2.364 (0.312)	2.362 (0.312)	2.119 (0.307)
North Carolina	1.539 (0.130)	1.482 (0.127)	1.367 (0.133)	1.606 (0.130)	1.606 (0.130)	1.751 (0.135)
South Carolina	2.618 (0.301)	2.636 (0.304)	2.830 (0.354)	2.860 (0.297)	2.871 (0.297)	2.683 (0.315)
All states	1.938 (0.110)	1.917 (0.108)	1.873 (0.143)	2.059 (0.109)	2.063 (0.108)	2.003 (0.110)
<i>Panel B. Great Plains White migrants</i>						
Kansas	0.255 (0.024)	0.233 (0.024)	0.204 (0.021)	0.257 (0.024)	0.257 (0.024)	0.320 (0.028)
Nebraska	0.361 (0.082)	0.349 (0.082)	0.309 (0.074)	0.361 (0.082)	0.361 (0.082)	0.455 (0.105)
North Dakota	0.464 (0.036)	0.445 (0.035)	0.412 (0.037)	0.475 (0.036)	0.475 (0.036)	0.480 (0.038)
Oklahoma	0.453 (0.036)	0.439 (0.036)	0.383 (0.034)	0.460 (0.036)	0.460 (0.036)	0.455 (0.038)
South Dakota	0.350 (0.026)	0.331 (0.026)	0.305 (0.026)	0.354 (0.026)	0.354 (0.026)	0.366 (0.027)
All states	0.380 (0.022)	0.363 (0.022)	0.323 (0.020)	0.385 (0.022)	0.385 (0.022)	0.418 (0.026)
Assumption 1: Allowing moving probability to depend on covariates from						
Duke SSA		X	X			
1910 US census			X			
Assumption 2: Allowing for social interactions among towns that are						
≤10 miles				X	X	X
≤20 miles and small					X	X
Same railroad						X

Notes: All columns contain weighted averages of destination-level network index estimates, $\hat{\Delta}_k$, where the weights are the number of people who move from each state to destination k . Column 1 is our baseline specification. Column 2 allows the moving probability to depend on indicators for being along a railroad and having above-median Black population share and town size (based on the 1916–1936 cohorts in the Duke data). Column 3 additionally controls for indicators for whether towns have an above-median value of the following variables: percent age 0–17, percent literate (age 10–39), percent homeowner, percent farmer (age 18–39, combining farmer and farm laborer categories), percent interstate migrant (age 18–60), and percent immigrant. We construct medians separately for each birth state, and all variables are race-specific except for percent immigrant. Column 4 allows for social interactions between towns in the same group that are within 10 miles of each other, by adding an indicator for this to equation (1) when estimating the covariance of location decisions. Column 5 additionally allows for interactions between same-group towns that are within 20 miles and both below the statewide median in population. Column 6 additionally allows for interactions between same-group towns on the same railroad line. Birth town groups are defined by cross validation. Standard errors are in parentheses.

Sources: Authors' calculations using Duke SSA/Medicare data; Black et al. (2015) data; and Minnesota Population Center and Ancestry.com (2013)

TABLE 5—AVERAGE NETWORK INDEX ESTIMATES, BY SIZE OF BIRTH TOWN AND DESTINATION

Exclude largest birth towns:	No	Yes	No	Yes
Exclude largest destinations:	No	No	Yes	Yes
Birth state	(1)	(2)	(3)	(4)
<i>Panel A. Southern Black migrants</i>				
Alabama	1.888 (0.195)	1.780 (0.148)	2.056 (0.285)	2.185 (0.267)
Florida	0.813 (0.117)	0.604 (0.059)	1.323 (0.229)	1.212 (0.209)
Georgia	1.657 (0.177)	1.460 (0.091)	1.696 (0.170)	1.768 (0.132)
Louisiana	1.723 (0.478)	1.116 (0.093)	0.971 (0.182)	0.965 (0.176)
Mississippi	2.303 (0.313)	2.298 (0.304)	2.085 (0.210)	2.032 (0.205)
North Carolina	1.539 (0.130)	1.445 (0.125)	0.743 (0.064)	0.694 (0.059)
South Carolina	2.618 (0.301)	2.565 (0.283)	1.784 (0.241)	1.745 (0.234)
All states	1.938 (0.110)	1.792 (0.089)	1.755 (0.108)	1.782 (0.101)
<i>Panel B. Great Plains White migrants</i>				
Kansas	0.255 (0.024)	0.220 (0.019)	0.243 (0.021)	0.228 (0.019)
Nebraska	0.361 (0.082)	0.252 (0.014)	0.265 (0.019)	0.252 (0.017)
North Dakota	0.464 (0.036)	0.464 (0.036)	0.527 (0.046)	0.529 (0.046)
Oklahoma	0.453 (0.036)	0.396 (0.029)	0.450 (0.040)	0.428 (0.038)
South Dakota	0.350 (0.026)	0.340 (0.026)	0.387 (0.034)	0.381 (0.033)
All states	0.380 (0.022)	0.331 (0.012)	0.374 (0.016)	0.361 (0.015)

Notes: All columns contain weighted averages of destination-level network index estimates, $\hat{\Delta}_k$, where the weights are the number of people who move from each state to destination k . Column 1 includes all birth towns and destinations. Column 2 excludes birth towns with 1920 population greater than 20,000 when estimating each $\hat{\Delta}_k$. Column 3 excludes all destination counties which intersect in 2000 with the ten largest non-South CMSAs as of 1950: New York, Chicago, Los Angeles, Philadelphia, Boston, Detroit, Washington, DC, San Francisco, Pittsburgh, and St. Louis, in addition to counties which received fewer than ten migrants. Column 4 excludes large birth towns and large destinations. Birth town groups are defined by cross validation. Standard errors are in parentheses.

Source: Authors' calculations using Duke SSA/Medicare data

excludes destination counties that intersect with the ten largest non-Southern consolidated metropolitan statistical areas (CMSAs) as of 1950, in addition to counties that received less than ten migrants.³⁷ We exclude both large birth towns and large

³⁷The ten CMSAs are New York; Chicago; Los Angeles; Philadelphia; Boston; Detroit; Washington, DC; San Francisco; Pittsburgh; and St. Louis. The first nine of these are also the largest non-Great Plains (and border region) CMSAs.

destinations in column 4. The average network index estimates are similar across all four specifications for both Southern African Americans and Great Plains Whites.³⁸

To further understand the nature of migration networks, we examine whether the location decisions of Black migrants influenced White migrants from the same Southern birth town, and vice versa. While African Americans and Whites could have shared information about opportunities in the North, segregation in the Jim Crow South makes cross-race interactions unlikely. Online Appendix C describes how we estimate the effects of cross-race migration networks, and online Appendix Table A.9 displays little evidence of such effects. In addition, there is little correlation between destination-level network index estimates for African Americans and Whites from the South: the linear (rank) correlation is 0.076 (0.149). This also implies that our network index estimates do not simply reflect unobserved characteristics of certain Southern towns.

Online Appendix Table A.10 shows that results are similar when we define birth town groups using counties. For Southern Black migrants, the linear (rank) correlation between the destination-level network index estimates using cross validation and counties is 0.858 (0.904). For Whites from the Great Plains, the linear (rank) correlation is 0.965 (0.891). Online Appendix Table A.11 displays results where, instead of choosing the grid size by cross validation, we use grid sizes of 50, 100, and 200 miles. Network estimates increase somewhat with the grid size.³⁹ Most importantly, network index estimates for African Americans exceed those of Whites from the Great Plains by a similar magnitude for all grid sizes. This implies that our results are not driven by Whites having more dispersed migration networks.

While the Duke SSA/Medicare data include most individuals born from 1916 to 1936, coverage rates are not perfect. Online Appendix D discusses the consequences of this measurement error in detail. We believe that imperfect coverage most likely leads us to understate the importance of migration networks.

C. The Role of Family Migration

The network index might capture the effect of family members from the same birth town on migrants' location decisions. While family migration is not a threat to our empirical strategy, it would be interesting to know the extent to which migration networks reflect within-family connections. Unfortunately, we do not observe family relationships and so we cannot study this question directly. However, we can examine whether our results stem entirely from the migration of male-female couples. If this were true, we would estimate negligible network indices when using male-only or female-only samples. Online Appendix Table A.13 shows that network index estimates are similar in magnitude among men and women, implying that our

³⁸ Online Appendix Table A.8 reports similar results for Southern-born Whites.

³⁹ This could arise because violations of Assumption 1 are more likely or violations of Assumption 2 are less consequential with larger birth town groups. The results in Table 4 suggest that violations of Assumption 1 are more likely. Given the tradeoffs, we prefer to choose the grid size using cross validation.

results do not simply reflect the migration of couples.⁴⁰ Our sample likely contains very few sets of parents and children, since we only include individuals born from 1916 to 1936.

A related question is whether differences in family size explain the Black-White network effect gap. As a first step, we use the 1940 census to measure the average within-household family size for individuals born from 1916 to 1936. African Americans from the South had families that were 17 percent larger than Whites from the Great Plains (6.16 versus 5.25). This difference is too small to explain our finding that average network index estimates are 410 percent larger among African Americans. To construct an upper bound on extended family size, we use the 100 percent sample of the 1940 census to count the average number of individuals in a county born from 1916 to 1936 with the same last name (Minnesota Population Center and Ancestry.com 2013). We find that Southern Black family networks likely were no more than 270 percent larger than those for Great Plains Whites (54.5 versus 14.7). This upper bound is sizable, but still less than the 410 percent difference in network effects. Online Appendix E contains a more formal discussion. We conclude that differences in family size might explain some, but not all, of the difference in network effects between Black and White migrants.⁴¹

D. Migration Networks and Economic Characteristics of Receiving and Sending Locations

To better understand why birth town migration networks affected location decisions, we relate network index estimates to economic characteristics of receiving and sending locations.

We first consider the characteristics of receiving locations. Employment opportunities were one of the most important considerations, and relatively high wages made manufacturing jobs particularly attractive. In the presence of imperfect information among migrants about employment opportunities, networks might have directed their members to destinations with more manufacturing employment. This is the story of John McCord, told in Section I. Because individuals living in the South and Great Plains had more information about the largest destinations, the imperfect information channel suggests a stronger relationship between network effects and manufacturing employment intensity in smaller destinations. In contrast, if information about employment opportunities was widespread, then network effects might not be stronger in destinations with more manufacturing. Similar patterns could arise if workers relied on their networks for job referrals.⁴² Destinations with more agriculture employment also might have been attractive

⁴⁰The similarity between men and women is not surprising given the relative sex balance among migrants in this period (Gregory 2005). The sizable effects among women only also indicate that our results are not driven by individuals serving together during World War II. Further evidence of this comes from the similarity of the results for individuals born from 1916–1925 and 1926–1936 (online Appendix Table A.13).

⁴¹Conditional on family size, Black and White migrants could have differed in the extent to which they tended to follow other family members. We do not have data that let us examine this possibility.

⁴²There is a large literature on social networks and employment opportunities. Recent examples include Topa (2001); Munshi (2003); Ioannides and Loury (2004); Bayer, Ross, and Topa (2008); Hellerstein, McInerney, and Neumark (2011); Beaman (2012); Burks et al. (2015); Schmutte (2015); and Heath (2018).

because migrants had experience in this sector. Pecuniary moving costs, which were largely determined by distance and railroads, represented another key consideration. Lower moving costs could have fostered networks by facilitating the transmission of information. On the other hand, migrants might have been willing to pay high moving costs only if they received information or benefits from a network.

To explore these hypotheses, we regress destination-level network index estimates on county covariates. Column 1 of Table 6 shows that network effects among African Americans are larger in destinations with a higher 1910 manufacturing employment share: a one standard deviation increase is associated with an increase in the network index of 0.2 people.⁴³ Column 2 shows that the positive relationship between manufacturing employment and network effects is almost six times larger in smaller destinations.⁴⁴ There is little relationship between network effects and the agriculture employment share. We also find stronger network effects in destinations that were closer to and could be reached by rail directly or with one stop from migrants' birth state. Network effects are stronger in destinations with a smaller Black population share in 1900, suggesting that networks helped migrants find opportunities in new places. There is little relationship between church members per capita in 1916 and network strength. This is not particularly surprising: while existing churches could have served as a substitute or complement for the services provided by migrant networks, historical accounts describe religious leaders from the South directing migration flows and establishing new churches in the North.

One possible concern is that these results do not reflect characteristics of destination counties but instead characteristics of birth states linked to destinations via vertical migration patterns. Column 3 indicates that this concern is unimportant, as adding birth state fixed effects has very little impact. Columns 4–6 present results for White migrants from the Great Plains. For this group, there is little relationship between network effects and the share of employment in manufacturing or agriculture.⁴⁵ Network effects are again stronger in destinations that could be reached more easily by rail and were closer.⁴⁶

In online Appendix Table A.17, we also examine measures of racial wage gaps and residential segregation. We construct a Black-White relative wage for each county, where a higher value indicates less racial discrimination in the labor market (see online Appendix F for details). To study discrimination against White migrants from the Great Plains (which existed, albeit less severely), we construct a similar relative wage for White men who are born in the five Great Plains states or outside the border region shown in Figure 2. To explore residential segregation, we use the measure from Logan and Parman (2017b), which captures the extent to which Black households

⁴³ Online Appendix Table A.14 contains summary statistics. Online Appendix Figure A.12 plots the bivariate relationship between network index estimates and 1910 manufacturing employment share, showing the considerable variation in manufacturing employment share across destinations.

⁴⁴ Small destination counties are those that do not intersect with the ten largest non-South CMSAs in 1950 (New York; Chicago; Los Angeles; Philadelphia; Boston; Detroit; Washington, DC; San Francisco; Pittsburgh; and St. Louis).

⁴⁵ For destinations that intersect with the largest CMSAs, networks are actually weaker in destinations with more manufacturing.

⁴⁶ Results are similar when using counties to define birth town groups (online Appendix Table A.15). Results for Southern Whites are in online Appendix Table A.16.

TABLE 6—NETWORK INDEX ESTIMATES AND DESTINATION COUNTY CHARACTERISTICS

	Dependent variable: Destination-level network index estimate					
	Southern Black migrants			Great Plains White migrants		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing employment share, 1910	1.775 (0.528)	0.414 (0.664)	0.371 (0.672)	-0.086 (0.101)	-0.279 (0.136)	-0.280 (0.136)
Manufacturing employment share by small destination indicator		2.418 (0.984)	2.462 (0.976)		0.268 (0.178)	0.275 (0.177)
Agriculture employment share, 1910	0.070 (0.287)	-0.391 (0.447)	-0.451 (0.450)	0.093 (0.045)	0.192 (0.125)	0.188 (0.125)
Agriculture employment share by small destination indicator		0.681 (0.477)	0.606 (0.490)		-0.104 (0.125)	-0.098 (0.124)
Small destination indicator		-0.494 (0.262)	-0.496 (0.262)		0.053 (0.065)	0.050 (0.065)
Log distance from birth state	-0.436 (0.064)	-0.403 (0.066)	-0.401 (0.063)	0.062 (0.035)	0.076 (0.036)	0.067 (0.037)
Direct railroad connection from birth state	0.305 (0.112)	0.307 (0.112)	0.293 (0.128)	0.203 (0.041)	0.201 (0.042)	0.191 (0.043)
One-stop railroad connection from birth state	0.221 (0.076)	0.211 (0.074)	0.166 (0.078)	0.079 (0.017)	0.073 (0.017)	0.072 (0.016)
Log population, 1910	0.061 (0.052)	0.063 (0.056)	0.069 (0.057)	0.023 (0.010)	0.034 (0.010)	0.034 (0.010)
Percent African American, 1910	-2.081 (0.358)	-1.915 (0.354)	-1.837 (0.349)	-0.226 (0.036)	-0.248 (0.038)	-0.244 (0.038)
Percent rural, 1910	-0.285 (0.197)	-0.268 (0.212)	-0.224 (0.216)	-0.049 (0.044)	-0.039 (0.044)	-0.038 (0.044)
Black/White church members per capita, 1916	-0.321 (0.179)	-0.235 (0.191)	-0.220 (0.193)	-0.111 (0.039)	-0.095 (0.039)	-0.095 (0.039)
Birth state fixed effects			X			X
R ²	0.096	0.103	0.115	0.033	0.037	0.037
Observations (birth state-destination county pairs)	1,515	1,515	1,515	4,104	4,104	4,104
Destination counties	382	382	382	1,230	1,230	1,230

Notes: The sample contains only counties that received at least ten migrants. Birth town groups are defined by cross validation. We measure distance from the centroid of destination counties to the centroid of birth states. Columns 1–3 include Black church members per capita, and columns 4–6 include White church members per capita. Standard errors, clustered by destination county, are in parentheses.

Sources: Authors' calculations using Duke SSA/Medicare data; Black et al. (2015) data; US Bureau of the Census (1992); Haines and ICPSR (2010); and Ruggles et al. (2019)

had non-Black next-door neighbors in 1940 relative to the expected value under complete integration. The most intriguing result is that Black migration networks were stronger in destinations where African Americans had relatively higher wages in 1940. One possible interpretation is that networks helped Black migrants identify and move to areas where they faced less labor market discrimination.⁴⁷ Online

⁴⁷This is not the only interpretation, as the 1940 measure of racial wage gaps could be affected by migration networks, because of causal effects of networks on labor market outcomes (either directly through job referrals or indirectly through an effect on the number of migrants) or because networks attracted individuals with different levels of unobserved human capital. However, to the extent that networks attracted individuals with less human capital (Stuart and Taylor 2021b; online Appendix Table A.7), this would lead to a lower relative wage.

Appendix Figures A.13 and A.14 display nonlinear relationships based on models with restricted cubic splines.⁴⁸

Overall, the results in Table 6 suggest that Black migration networks responded more than White networks to attractive employment opportunities, especially in smaller destinations, and to moving costs. This is consistent with Black migrants relying more heavily on their networks for information about employment opportunities or job referrals, possibly because they faced greater discrimination in labor and housing markets or had fewer resources. The results in online Appendix Table A.17 provide some support for the conclusion that networks also helped Black migrants avoid the most discriminatory labor markets.

We next consider the relationship between migration networks and characteristics of sending counties. Networks could have been particularly valuable in locating jobs or housing for migrants from poorer communities who had fewer resources to engage in costly search (McKenzie and Rapoport 2007). Better labor market opportunities could have reduced the incentive to invest in social networks. Factors that increased social interactions in origin communities include population density (Chay and Munshi 2015) and church attendance.⁴⁹ We also consider proxies for educational achievement (literacy and school attendance) and, for African Americans, access to Rosenwald schools, which improved educational attainment in this period (Aaronson and Mazumder 2011b). The relationship between education and network effects is theoretically ambiguous, as education could promote social ties while also increasing the relative return to choosing a non-network destination. Other factors we explore include railroad exposure—which could have facilitated the transmission of information through both network and non-network channels—and, for the South, the share of votes that went to Strom Thurmond in the 1948 presidential election—which proxies for the degree of racism.

Table 7 displays results from regressing birth county-level network index estimates on county characteristics.⁵⁰ Columns 1 and 2 contain results for Black moves out of the South. Network effects were stronger in counties with higher Black farm ownership rates but weaker in counties where a higher share of individuals lived in owner-occupied housing.⁵¹ Consequently, there is little evidence for a relationship between wealth/resources and network strength. Networks are weaker in places with a higher share of employment in manufacturing: a one standard deviation increase

⁴⁸For Southern Black migrants, one notable result is the negative, concave relationship with distance. Longer-distance destinations tend to be in the West (especially California), so this result is generally consistent with historical accounts of the Great Migration, which emphasize networks in the Midwest and Northeast. Even more interesting is the positive, concave relationship with the Black-White relative wage. One interpretation is that networks helped migrants avoid especially discriminatory labor markets; as we note above, other interpretations are possible because of potential reverse causality.

⁴⁹Drawing upon the 1906 Census of Religious Bodies, we consider the following to be Black churches: Baptists-National Convention, Colored Primitive Baptists in America, United American Freewill Baptists, Church of the Living God (Christian Workers for Friendship), Free Christian Zion Church of Christ, Union American Methodist Episcopal Church, African Methodist Episcopal Church, African Union Methodist Protestant Church, African Methodist Episcopal Zion Church, Congregational Methodist Church, Colored Methodist Episcopal Church, Reformed Zion Union Apostolic Church, and Colored Cumberland Presbyterian Church. We define White churches to be all others.

⁵⁰Online Appendix Table A.18 contains summary statistics for birth county characteristics.

⁵¹These variables are highly correlated ($\rho = 0.8$), but estimating models that only include one of them does not lead to meaningfully different results.

TABLE 7—NETWORK INDEX ESTIMATES AND BIRTH COUNTY CHARACTERISTICS

	Dependent variable: Birth county-level network index estimate			
	Southern Black migrants		Great Plains White migrants	
	(1)	(2)	(3)	(4)
Percent of Black/White farmers who are owners, 1910	2.306 (2.100)	2.809 (2.279)	0.682 (0.567)	1.001 (0.647)
Percent of Black/White individuals in owner-occupied housing, 1910	-3.350 (2.972)	-3.575 (2.990)	0.579 (0.665)	0.177 (0.735)
Percent of Black/White workers in agriculture, 1910	-0.884 (1.844)	-0.461 (1.888)	-0.247 (0.494)	-0.058 (0.547)
Percent of Black/White workers in manufacturing, 1910	-5.952 (3.284)	-6.031 (3.605)	1.606 (3.085)	1.582 (3.019)
Percent of farm acreage in cotton, 1910	-1.441 (2.951)	-1.003 (3.244)	1.658 (0.899)	2.015 (0.905)
Log Black/White population density, 1910	0.985 (0.555)	0.851 (0.561)	0.228 (0.090)	0.252 (0.095)
Black/White church members per capita, 1916	0.223 (0.739)	0.217 (0.662)	0.400 (0.269)	0.405 (0.280)
Rosenwald school exposure	-0.595 (0.843)	-0.974 (0.878)		
Black/White literacy rate (10+), 1910	2.021 (2.408)	4.073 (3.180)	-8.216 (2.845)	-9.290 (2.952)
Black/White school attendance rate (6–14), 1910	1.170 (1.519)	2.199 (1.554)	-1.637 (0.615)	-1.700 (0.689)
Railroad exposure	-0.310 (0.464)	-0.252 (0.478)	-0.144 (0.073)	-0.184 (0.084)
Percent African American, 1910	-0.273 (1.711)	0.332 (1.734)	-0.214 (0.974)	0.210 (0.939)
Percent rural, 1910	-0.508 (1.736)	-0.675 (1.723)	-0.591 (0.262)	-0.675 (0.277)
Percent voting for Strom Thurmond, 1948	0.613 (0.524)	-0.298 (0.929)		
Birth state fixed effects		X		X
R^2	0.089	0.100	0.282	0.304
Observations (birth counties)	546	546	383	383

Notes: Columns 1–2 include indicated Black-specific variables, and columns 3–4 include indicated White-specific variables. Railroad exposure is the share of migrants in a county that lived along a railroad. Rosenwald exposure is the average Rosenwald coverage experienced over ages 7–13. Heteroskedastic robust standard errors are in parentheses.

Sources: Authors' calculations using Duke SSA/Medicare data; Aaronson and Mazumder (2011b) data; Black et al. (2015) data; US Bureau of the Census (1992); Haines and ICPSR (2010); ICPSR (1999); and Ruggles et al. (2019)

in the manufacturing employment share is associated with a 0.5 person decrease in the network index. Network strength is positively correlated with Black population density, church members per capita, literacy rates, and school attendance rates, but none of these coefficients are statistically significant. Indeed, a general conclusion is that nearly all of these variables are weakly related to network strength. Results are similar in column 2, where we include birth state fixed effects to address the possibility that our results are driven by destination factors, such as labor demand, that are linked to certain areas of the South through vertical migration patterns.

Columns 3 and 4 present results for White moves out of the Great Plains.⁵² Our explanatory variables explain a higher share of the variance in White network effects. White population density and church members per capita are positively correlated with network strength, although only the relationship with density is distinguishable from zero in both columns. A notable difference from columns 1 and 2 is that White literacy rates and school attendance are negatively correlated with network strength. Literacy and school attendance rates were much higher in the origin counties of White migrants (see the nonlinear estimates in online Appendix Figures A.15 and A.16). One possible explanation is that the general relationship between human capital and network strength is inverse-U-shaped. Another possibility is that only Whites with relatively little human capital relied on their social networks to obtain employment, while African Americans with higher human capital relied on their networks to overcome the more severe discrimination they faced.

IV. A Structural Model of Migration Networks and Location Decisions

As discussed above, the network index is consistent with and can be mapped to multiple structural models. In this section, we map the network index to one such model, in which migration networks arise because some individuals follow other migrants to a destination. Our model shares this basic structure with Glaeser, Sacerdote, and Scheinkman (1996), but we extend previous work by modeling the interdependence between various destinations—as is necessary in a multinomial choice problem—and allowing for more than two types of agents. The additional structure in the model allows us to examine counterfactual location decisions in the absence of migration networks.

A. Model

Migrants from birth town j are indexed on a circle by $i \in \{1, \dots, N_j\}$, where N_j is the total number of migrants from j . For migrant i , destination k belongs to one of three preference groups: high (H_i), medium (M_i), or low (L_i). The high preference group contains a single destination. In the absence of migration networks, the destination in H_i is most preferred, and destinations in M_i are preferred over those in L_i .⁵³ A migrant never moves to a destination in L_i . A migrant chooses a destination in M_i if and only if their neighbor, $i - 1$, chooses the same destination. A migrant chooses a destination in H_i if their neighbor chooses the same destination or their neighbor selects a destination in L_i .

Migrants from the same birth town differ in their preferences over destinations. The probability that destination k is in the high preference group for a migrant from town j is $h_{j,k} \equiv \Pr[k \in H_i | i \in j]$, and the probability that destination k is in the medium preference group is $m_{j,k} \equiv \Pr[k \in M_i | i \in j]$. These probabilities arise

⁵²Columns 3 and 4 exclude Rosenwald school exposure because these schools existed primarily in the South. We also exclude the Strom Thurmond vote share because he received a negligible number of votes in these states.

⁵³The assumption that H_i is a nonempty singleton ensures that migrant i has a well-defined location decision in the absence of networks. We could allow H_i to contain many destinations and specify a decision rule among the elements of H_i . This extension would complicate the model without adding any new insights.

from expected utility maximization problems solved by migrants. We do not need to specify migrants' utility functions, but expected wages and transportation costs are among the relevant factors. We also do not need to specify why some migrants choose the same destination as their neighbor. For example, neighbors might provide information about employment opportunities, or migrants might value living near friends and family. As with the network index, this model considers how networks affect where individuals move, conditional on migrating.

The share of migrants from birth town j living in destination k that chose their destination because of the network equals $m_{j,k}$.⁵⁴ Hence, the number of migrants who chose destination k because of the network is $N_k^{network} \equiv \sum_j N_{j,k} m_{j,k}$, where $N_{j,k}$ is the number of migrants who moved from j to k . In the absence of networks, where $m_{j,k} = 0$, migrants move to the destination in H_j . As a result, in the counterfactual where networks do not influence location decisions, the probability of moving from j to k is $h_{j,k}$, and the number of migrants in destination k is $N_k^{cf} \equiv \sum_j N_j h_{j,k}$.

Appendix A describes how we estimate the structural parameters, $m_{j,k}$ and $h_{j,k}$, using estimates of moving probabilities, $P_{j,k}$, and network indices, $\Delta_{j,k}$. While the structural parameters are jointly identified, estimates of $m_{j,k}$ tend to reflect the network index. Estimation depends on Assumptions 1 and 2, plus the additional structure imposed by the model of local social interactions.

B. Results

Table 8 reports estimates of the percent of migrants who chose their destination because of migration networks, calculated as migrant-weighted averages of $100 \times (\hat{N}_k^{network}/N_k)$. On average, we estimate that 34 percent of Southern Black migrants chose their long-run location because of their birth town migration network. There is considerable variation across destination regions. For example, of Mississippi-born migrants, 17 percent of Northeast-bound, 40 percent of Midwest-bound, and 23 percent of West-bound migrants chose their location because of their migration network.⁵⁵ On average, 13 percent of Great Plains White migrants chose their long-run location because of their migration network.

Table 9 illustrates the effects of migration networks for selected destinations. We report the actual number of migrants and the number of migrants in a counterfactual without migration networks, for counties with the largest increases and decreases in migration. In the absence of migration networks, we estimate that Cook County, Illinois (home of Chicago) would experience a 29 percent decline in Southern Black

⁵⁴The share of migrants from birth town j living in destination k that chose their destination because of the network is $\Pr[k \in M_j | D_{i,j,k} = 1]$. By Bayes' theorem, this equals

$$\Pr[k \in M_j | D_{i,j,k} = 1] = \frac{\Pr[D_{i,j,k} = 1 | k \in M_j] \Pr[k \in M_j]}{\Pr[D_{i,j,k} = 1]} = \frac{P_{j,k} m_{j,k}}{P_{j,k}} = m_{j,k},$$

because $\Pr[D_{i,j,k} = 1 | k \in M_j] = \Pr[D_{i-1,j,k} = 1] = P_{j,k}$. This relationship between the distribution of preferences among all migrants and among migrants in each destination results in part from the assumption that individuals' social network (i.e., their neighbor) is independent of their preferences.

⁵⁵This regional variation is also apparent in estimates of the network index (online Appendix Tables A.20 and A.21).

TABLE 8—ESTIMATED PERCENT OF MIGRANTS WHO CHOSE THEIR DESTINATION BECAUSE OF MIGRATION NETWORK

Birth state	Destination region				
	All (1)	Northeast (2)	Midwest (3)	West (4)	South (5)
<i>Panel A. Southern Black Migrants</i>					
Alabama	34.3	24.5	40.1	22.5	—
Florida	22.8	24.3	23.4	13.5	—
Georgia	32.9	32.3	36.5	17.0	—
Louisiana	35.0	20.3	29.9	38.7	—
Mississippi	36.0	17.4	39.8	23.3	—
North Carolina	32.2	34.5	21.1	8.3	—
South Carolina	36.8	39.2	26.4	11.0	—
All states	34.2	32.9	37.4	28.5	—
<i>Panel B. Great Plains White Migrants</i>					
Kansas	9.1	3.1	10.3	10.5	3.4
Nebraska	12.7	4.6	11.6	14.5	4.2
North Dakota	13.9	5.2	10.0	15.7	4.9
Oklahoma	14.5	4.6	8.5	17.3	5.2
South Dakota	12.0	3.7	11.1	13.6	4.0
All states	12.6	4.1	10.3	14.7	4.4

Notes: Table contains migrant-weighted average estimates of $100 \times (N_k^{network} / N_k)$, the percent of migrants who chose their destination because of their birth town migration network. See the text for details.

Source: Authors' calculations using Duke SSA/Medicare data

migrants. Los Angeles, Detroit, Philadelphia, and Baltimore also would have considerably fewer migrants, experiencing declines from 11 to 25 percent. The largest increases in migration would be to Queens, New York; Prince George's County, Maryland (near Washington); and Oakland County, Michigan (near Detroit). In the absence of migration networks, there would be considerably fewer Great Plains White migrants in several California counties: those containing Los Angeles, Bakersfield, Fresno, and Stockton would experience declines of 20 to 28 percent. These results show that migration networks account for a sizable portion of the migration to prominent destinations, and consequently that migration networks had important effects on the distribution of population across the United States.⁵⁶

Since migration networks clearly affected where individuals moved, a natural question is whether these networks led migrants to live in areas with worse economic opportunities, as could happen if networks limited later migratory responses to economic shocks. To study this, we examine how characteristics of migrants' location would change in a counterfactual without migration networks.⁵⁷ In Table 10, column 1 of panel A shows that the average Southern Black migrant lived in a county where the unemployment rate was 7.5 percent in 2000. In the no-network counterfactual, this falls to 7.3 percent. Hence, for the 34.5 percent of migrants who would move in the counterfactual, the mean unemployment rate falls by 0.7 percentage

⁵⁶Online Appendix Table A.22 reports counterfactual migration flows from birth state to destination region in the absence of migration networks. The results show that migration networks were important determinants of vertical migration patterns, one of the most widely known features of the Great Migration.

⁵⁷A different question, which we do not answer, is whether migration networks had a causal effect on migrants' labor market outcomes.

TABLE 9—COUNTIES WITH THE FIVE LARGEST INCREASES AND DECREASES IN MIGRATION IN A COUNTERFACTUAL WITHOUT MIGRATION NETWORKS

Destination county (1)	Largest city in destination county (2)	Actual number of migrants (3)	Counterfactual number of migrants (4)	Difference (5)	Percent difference (6)
<i>Panel A. Southern Black Migrants</i>					
Queens, NY	New York	12,507	15,148	2,641	21.1
Prince George's, MD	Bowie	7,241	8,959	1,718	23.7
Oakland, MI	Farmington Hills	3,570	4,774	1,204	33.7
Sacramento, CA	Sacramento	2,939	4,128	1,189	40.5
Alameda, CA	Oakland	8,041	9,002	961	11.9
Baltimore City, MD	Baltimore	12,520	9,381	-3,139	-25.1
Philadelphia, PA	Philadelphia	25,007	21,408	-3,599	-14.4
Wayne, MI	Detroit	42,818	38,200	-4,618	-10.8
Los Angeles, CA	Los Angeles	31,703	25,534	-6,169	-19.5
Cook, IL	Chicago	59,915	42,638	-17,277	-28.8
<i>Panel B. Great Plains White Migrants</i>					
Maricopa, AZ	Phoenix	28,967	29,398	431	1.5
San Bernardino, CA	San Bernardino	13,037	13,453	416	3.2
Pima, AZ	Tucson	8,000	8,383	383	4.8
Mohave, AZ	Lake Havasu City	3,825	4,181	356	9.3
Clark, NV	Las Vegas	9,408	9,755	347	3.7
San Diego, CA	San Diego	19,960	18,739	-1,221	-6.1
San Joaquin, CA	Stockton	7,207	5,653	-1,554	-21.6
Fresno, CA	Fresno	8,329	5,968	-2,361	-28.4
Kern, CA	Bakersfield	10,546	8,134	-2,412	-22.9
Los Angeles, CA	Los Angeles	38,552	30,769	-7,783	-20.2

Notes: Column 3 contains N_k , the actual number of migrants in destination k . Column 4 contains estimates of N_k^{df} , the number of migrants who would have chosen destination county k in the absence of migration networks. Column 6 is equal to column 5 divided by column 3. See the text for details.

Source: Authors' calculations using Duke SSA/Medicare data

points. The poverty rate (a measure of economic disadvantage) and the Black population share (a measure of segregation) in the average migrant's destination county also fall modestly in the no-network counterfactual. Results are similar when we examine county characteristics as of 1980.⁵⁸ For Great Plains White migrants, panel B generally shows even smaller changes in destination characteristics. In sum, these results suggest that migration networks had little or no effect on the characteristics of migrants' chosen destination. This is largely because migrants who did not follow their birth town migration network moved to similar places.

One important caveat is that our model does not account for all possible general equilibrium effects. However, the direction of these effects is not clear: reducing migration from a town to a county could make that destination more attractive, because of higher wages or lower housing costs, or less attractive, because of fewer individuals with the same race and background. Our model also does not account for the possibility that destination characteristics, such as the unemployment rate, could change in the counterfactual. Addressing these issues would require a model with labor demand, housing supply, and endogenous amenities, which is beyond the scope of this paper.

⁵⁸Results also are similar when we consider a counterfactual in which Southern Black migration networks are as strong as those of Great Plains White migrants.

TABLE 10—CHARACTERISTICS OF COUNTIES WHERE MIGRANTS LIVE UNDER REALIZED AND COUNTERFACTUAL MIGRATION NETWORKS

County characteristic	Realized networks mean (1)	No network counterfactual		
		Mean (2)	Mean change, all migrants (3)	Mean change, switchers (4)
<i>Panel A. Southern Black migrants</i>				
Unemployment rate, 2000	7.50	7.27	-0.23	-0.67
Poverty rate, 2000	14.79	14.28	-0.51	-1.48
Percent Black, 2000	25.17	23.70	-1.47	-4.26
Unemployment rate, 1980	8.12	7.95	-0.17	-0.49
Poverty rate, 1980	13.64	13.15	-0.49	-1.42
Percent Black, 1980	22.68	21.13	-1.55	-4.49
<i>Panel B. Great Plains White migrants</i>				
Unemployment rate, 2000	6.64	6.53	-0.11	-0.87
Poverty rate, 2000	12.68	12.48	-0.20	-1.58
Percent Black, 2000	6.61	6.70	0.09	0.71
Unemployment rate, 1980	7.34	7.32	-0.02	-0.16
Poverty rate, 1980	11.21	11.17	-0.04	-0.32
Percent Black, 1980	6.02	6.05	0.03	0.24

Notes: Column 1 contains the migrant-weighted average of county characteristics based on the realized location decisions of migrants. Column 2 contains the migrant-weighted average based on the location decisions that would be made in the absence of migration networks. Column 3 is the difference between column 2 and column 1. Column 4 reports the mean change for migrants who would switch their location under the counterfactual, calculated as column 3 divided by the percent of migrants who would switch their location (34.5 percent in panel A and 12.7 percent in panel B). See the text for details.

Sources: Authors' calculations using Duke SSA/Medicare data; Manson et al. (2019)

V. Conclusion

This paper provides new evidence on the effects of birth town migration networks on location decisions. We use administrative data to study over one million long-run location decisions made during two landmark migration episodes by African Americans born in the US South and Whites born in the Great Plains. We formulate a novel network index that characterizes the effect of migration networks for each receiving and sending location. The network index allows us to estimate the overall effect of migration networks and the degree to which network effects are associated with economic characteristics of receiving and sending locations. The network index can be used for other outcomes and settings to provide a deeper understanding of social networks.

We find very strong network effects among Southern Black migrants and weaker effects among Whites. Estimates of our network index imply that when one randomly chosen African American moves from a birth town to a destination county, then 1.9 additional Black migrants make the same move on average. For White migrants from the Great Plains, the average is only 0.4, and results for Southern Whites are similarly small. Interpreted through a novel structural model, our estimates imply that 34 percent of Black migrants chose their long-run destination because of their birth town migration network, while 13 percent of Great Plains White migrants were similarly influenced. In addition, our results suggest that Black migration networks connected migrants to attractive employment opportunities and played a larger role

in less costly moves. While the goal of this paper is not to explain the Black-White gap, one interpretation of our results is that African Americans relied on migration networks more heavily to overcome greater discrimination in labor and housing markets and a relative lack of resources.

These results shed new light on location decisions. In addition to real wages, amenities, and moving costs, as emphasized by previous work, our results suggest that social networks play a major role in where individuals move. Migration networks appear to help mitigate the substantial information frictions in long-distance location decisions. Networks could play an important role in contemporaneous rural-to-urban migrations in developing nations, which resemble the historical migration episodes we study on several dimensions. Our results also suggest that long-run location decisions will shift labor more effectively to areas with a high marginal product if there are pioneer migrants who can facilitate these moves. Policies that seek to direct migration to certain areas should account for such networks.

Our results also have implications for the effects of migration on destinations. Migration networks continued to operate after location decisions were made, and the Great Migration generated considerable variation in the strength of social networks across destinations. In other work, we use this variation to study the relationship between crime and social connectedness in US cities (Stuart and Taylor 2021b). Examining the importance of migration networks in other settings, and studying other effects of migration networks on destinations, is a valuable direction for future work.

APPENDIX A. DETAILS ON THE STRUCTURAL MODEL

This Appendix provides additional details on the structural model introduced in Section IVA.

The probability that a randomly chosen migrant i moves from j to k is

$$(A1) \quad P_{j,k} \equiv \Pr[D_{i,j,k} = 1] = \Pr[D_{i-1,j,k} = 1, k \in H_i] + \Pr[D_{i-1,j,k} = 1, k \in M_i] \\ + \sum_{k' \neq k} \Pr[D_{i-1,j,k'} = 1, k \in H_i, k' \in L_i].$$

The first term on the right-hand side of equation (A1) is the probability that a migrant's neighbor moves to k , and k is in the migrant's high preference group; in this case, the neighbor's decision reinforces the migrant's desire to move to k . The second term is the probability that a migrant moves to k only because their neighbor moved there. The third term is the probability that a migrant moves to k because it is in the high preference group and the neighbor's chosen destination is in the low preference group. Using the parameters defined in Section IVA, we rewrite equation (A1) as

$$(A2) \quad P_{j,k} = P_{j,k} h_{j,k} + P_{j,k} m_{j,k} + \sum_{k' \neq k} P_{j,k'} h_{j,k} \left(\frac{1 - h_{j,k'} - m_{j,k'}}{1 - h_{j,k'}} \right).$$

To facilitate estimation, we introduce an auxiliary parameter. The probability that destination k is in the medium preference group, conditional on not being in the high preference group, is $\nu_{j,k} \equiv \Pr[k \in M_i | k \notin H_i, i \in j]$. The conditional probability definition for $\nu_{j,k}$ implies that $\nu_{j,k} = m_{j,k} / (1 - h_{j,k})$. Using $\nu_{j,k}$ allows us to simplify equation (A2) to

$$(A3) \quad P_{j,k} = P_{j,k} \nu_{j,k} + \sum_{k'=1}^K P_{j,k'} (1 - \nu_{j,k'}) h_{j,k}.$$

We next connect the structural model to the network index. The model implies that the average covariance of location decisions, $C_{j,k}$, equals

$$(A4) \quad C_{j,k} = \frac{2P_{j,k}(1 - P_{j,k}) \sum_{a=1}^{N_j-1} (N_j - a) \left(\frac{\rho_{j,k} - P_{j,k}}{1 - P_{j,k}} \right)^a}{N_j(N_j - 1)},$$

where $\rho_{j,k} \equiv \Pr[D_{i,j,k} = 1 | D_{i-1,j,k} = 1, i \in j] = h_{j,k} + m_{j,k}$ is the probability that migrant i moves to destination k given that their neighbor moves there.⁵⁹

We continue to maintain Assumption 1, so that the probability of moving from j to k is the same for all birth towns in the same birth town group g . In the structural model, Assumption 1 holds because we assume that $m_{j,k}$ and $h_{j,k}$ are equal for all birth towns in the same group. This implies that $\rho_{j,k}$ is also constant across birth towns in the same group. The justification for this assumption is the same as previously discussed.

Imposing this assumption, substituting equation (A4) into the expression for the network index in equation (4), simplifying, and taking the limit as $N_j \rightarrow \infty$ yields

$$(A5) \quad \Delta_{g,k} = \frac{2(\rho_{g,k} - P_{g,k})}{1 - \rho_{g,k}},$$

where $\Delta_{g,k}$ is the birth town group-destination network index. Equation (A5) can be rearranged to show that

$$(A6) \quad \rho_{g,k} = \frac{2P_{g,k} + \Delta_{g,k}}{2 + \Delta_{g,k}}.$$

We use equation (A6) to estimate $\rho_{g,k}$ with our estimates of $P_{g,k}$ and $\Delta_{g,k}$.

Equation (A3), plus the facts that $\nu_{g,k} = m_{g,k} / (1 - h_{g,k})$ and $\rho_{g,k} = h_{g,k} + m_{g,k}$, implies that

$$(A7) \quad \rho_{g,k} = \nu_{g,k} + \frac{P_{g,k}(1 - \nu_{g,k})^2}{\sum_{k'=1}^K P_{g,k'}(1 - \nu_{g,k'})}.$$

⁵⁹Equation (A4) follows from the fact that the covariance of location decisions for individuals i and $i + n$ is $\text{cov}[D_{i,j,k}, D_{i+n,j,k}] = P_{j,k}(1 - P_{j,k})((\rho_{j,k} - P_{j,k}) / (1 - P_{j,k}))^n$.

We use equation (A7) to estimate $\nu_g \equiv (\nu_{g,1}, \dots, \nu_{g,K})$ using our estimates of $(P_{g,1}, \dots, P_{g,K}, \rho_{g,1}, \dots, \rho_{g,K})$. We employ a computationally efficient algorithm that leverages the fact that equation (A7) is a quadratic in $\nu_{g,k}$, conditional on $\sum_{k'=1}^K P_{g,k'}(1 - \nu_{g,k'})$. We initially assume that $\sum_{k'=1}^K P_{g,k'}(1 - \nu_{g,k'}) = \sum_{k'=1}^K P_{g,k'} = 1$, then solve for $\nu_{g,k}$ using the quadratic formula, then construct an updated estimate of $\sum_{k'=1}^K P_{g,k'}(1 - \nu_{g,k'})$, and then solve again for $\nu_{g,k}$ using the quadratic formula. We require that each estimate of $\nu_{g,k}$ lies in $[0, 1]$. This iterated algorithm converges very rapidly in essentially all cases.⁶⁰ Finally, we use equation (A3) to estimate $h_{g,k}$ with our estimates of $\rho_{g,k}$ and $\nu_{g,k}$, and we estimate $m_{g,k}$ using the fact that $m_{g,k} = \rho_{g,k} - h_{g,k}$.

The parameters of the structural model are exactly identified. We jointly identify $m_{j,k}$ and $h_{j,k}$ from estimates of moving probabilities and network indices. Estimates of $m_{j,k}$ tend to reflect the network index: if $m_{j,k} = 0$, then equation (A2) implies that $P_{j,k} = h_{j,k}$ and equation (A5) implies that $\Delta_{j,k} = 0$. Location decisions differ across nearby towns due to exogenous shifters in the location decisions of some migrants. For example, if a migrant moves to destination k for some idiosyncratic reason, then other migrants will tend to follow. This captures the story of John McCord, described in Section I.

REFERENCES

- Aaronson, Daniel, and Bhashkar Mazumder. 2011a. "Data for: The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* 119 (5): 821–88.
- Aaronson, Daniel, and Bhashkar Mazumder. 2011b. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* 119 (5): 821–88.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2017. "Cultural Assimilation during the Age of Mass Migration." NBER Working Paper 22381.
- Bartel, Ann P. 1989. "Where Do the New U.S. Immigrants Live?" *Journal of Labor Economics* 7 (4): 371–91.
- Bauer, Thomas, Gil S. Epstein, and Ira N. Gang. 2005. "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics* 18 (4): 649–62.
- Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116 (6): 1150–96.
- Beaman, Lori A. 2012. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies* 79 (1): 128–61.
- Beine, Michel, Frédéric Docquier, and Çağlar Özden. 2011. "Diasporas." *Journal of Development Economics* 95 (1): 30–41.
- Bell, Velma Fern. 1933. "African Americans—The Negro in Beloit and Madison, Wisconsin." Master's diss. University of Wisconsin.
- Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review* 105 (2): 477–503.
- Black, Dan A., and Jeffrey A. Smith. 2004. "How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics* 121 (1–2): 99–124.

⁶⁰The algorithm converges in all cases for Great Plains White migrants. For Southern Black migrants, there are three birth town groups for which the algorithm does not converge because our estimates of $P_{g,k}$ and $\rho_{g,k}$ do not yield a real solution to the quadratic formula. We set $\hat{\nu}_{g,k}$ equal to zero for any (g, k) observation for which the quadratic formula solution does not exist. The motivation for this is that our estimates of $P_{g,k}$ and $\rho_{g,k}$ in these cases are consistent with negative values of $\nu_{g,k}$, even though this is not a feasible solution. Our results are nearly identical when dropping these cases, which is not surprising because these three birth town groups account for a negligible share of the 223 groups used in our estimation for Southern Black migrants.

- Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides.** 2011. "Chapter 18—Identification of Social Interactions." In *Handbook of Social Economics*, Vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 853–964. Amsterdam: North-Holland.
- Boustan, Leah Platt.** 2009. "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970." *Journal of Economic History* 69 (3): 755–82.
- Boustan, Leah Platt.** 2010. "Was Postwar Suburbanization 'White Flight'? Evidence from the Black Migration." *Quarterly Journal of Economics* 125 (1): 417–43.
- Boustan, Leah Platt.** 2017. *Competition in the Promised Land—Black Migrants in Northern Cities and Labor Markets*. Princeton, NJ: Princeton University Press.
- Brock, William A., and Steven N. Durlauf.** 2001. "Discrete Choice with Social Interactions." *Review of Economic Studies* 68 (2): 235–60.
- Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman.** 2015. "The Value of Hiring through Employee Referrals." *Quarterly Journal of Economics* 130 (2): 805–39.
- Carrington, William J., Enrica Detragiache, and Tara Vishwanath.** 1996. "Migration with Endogenous Moving Costs." *American Economic Review* 86 (4): 909–30.
- Chay, Kenneth, and Kaivan Munshi.** 2015. "Black Networks after Emancipation: Evidence from Reconstruction and the Great Migration." http://www.histecon.magd.cam.ac.uk/km/Chay_Munshi_December2015.pdf.
- Chen, Yuyu, Ginger Zhe Jin, and Yang Yue.** 2010. "Peer Migration in China." NBER Working Paper 15671.
- Collins, William J.** 1997. "When the Tide Turned: Immigration and the Delay of the Great Black Migration." *Journal of Economic History* 57 (3): 607–32.
- Collins, William J., and Marianne H. Wanamaker.** 2014. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics* 6 (1): 220–52.
- Collins, William J., and Marianne H. Wanamaker.** 2015. "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *Journal of Economic History* 75 (4): 947–92.
- Curtis White, Katherine J.** 2008. "Population Change and Farm Dependence: Temporal and Spatial Variation in the U.S. Great Plains, 1900–2000." *Demography* 45 (2): 363–86.
- Duke Social Security–Medicare Part B Matched Data.** 2002. PI: Seth Sanders. Accessed February 12, 2014.
- Eriksson, Katherine.** Forthcoming. "Education and Incarceration in the Jim Crow South: Evidence from Rosenwald Schools." *Journal of Human Resources*.
- Gibson, Campbell, and Kay Jung.** 2005. "Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic Origin, 1970 to 1990, for Large Cities and Other Urban Places in the United States." US Census Bureau Population Division Working Paper 76.
- Giulletti, Corrado, Jackline Wahba, and Yves Zenou.** 2018. "Strong versus Weak Ties in Migration." *European Economic Review* 104: 111–37.
- Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman.** 1996. "Crime and Social Interactions." *Quarterly Journal of Economics* 111 (2): 507–48.
- Gottlieb, Peter.** 1987. *Making Their Own Way: Southern Blacks' Migration to Pittsburgh, 1916–1930*. Urbana: University of Illinois Press.
- Graham, Bryan S.** 2008. "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica* 76 (3): 643–60.
- Gray, Charles P.** 1898. *Gray, Charles P. Gray's New Trunk Railway Map of the United States, Dom. of Canada and Portion of Mexico. Compiled from Latest Official Sources, Showing Carefully Selected List of Cities and Town in Readable Type for Quick Reference*. New York, 1898. Map. <https://www.loc.gov/item/gm71000844/>.
- Greenwood, Michael J.** 1997. "Chapter 12—Internal Migration in Developed Countries." In *Handbook of Population and Family Economics*, Vol. 1B, edited by Mark R. Rosenzweig and Oded Stark, 647–720. Amsterdam: North-Holland.
- Gregory, James N.** 1989. *American Exodus: The Dust Bowl Migration and Okie Culture in California*. New York: Oxford University Press.
- Gregory, James N.** 2005. *The Southern Diaspora—How the Great Migrations of Black and White Southerners Transformed America*. Chapel Hill: University of North Carolina Press.
- Grossman, James R.** 1989. *Land of Hope—Chicago, Black Southerners, and the Great Migration*. Chicago, IL: University of Chicago Press.

- Haines, Michael R., and Inter-university Consortium for Political and Social Research (ICPSR).** 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790–2002 (ICPSR 2896)." ICPSR, Ann Arbor, MI. <https://www.icpsr.umich.edu/web/ICPSR/studies/02896/staff> (accessed August 9, 2019).
- Heath, Rachel.** 2018. "Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories." *Journal of Political Economy* 126 (4): 1691–1746.
- Hellerstein, Judith K., Melissa McInerney, and David Neumark.** 2011. "Neighbors and Coworkers: The Importance of Residential Labor Market Networks." *Journal of Labor Economics* 29 (4): 659–95.
- Henri, Florette.** 1975. *Black Migration: Movement North, 1900–1920*. New York: Doubleday.
- Hornbeck, Richard.** 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review* 102 (4): 1477–1507.
- Hornbeck, Richard, and Suresh Naidu.** 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review* 104 (3): 963–90.
- Hurt, R. Douglas.** 2011. *The Big Empty: The Great Plains in the Twentieth Century*. Tucson: University of Arizona Press.
- Inter-university Consortium for Political and Social Research (ICPSR).** 1999. "United States Historical Election Returns, 1824–1968 (ICPSR 1)." ICPSR, Ann Arbor, MI. <https://www.icpsr.umich.edu/web/ICPSR/studies/1> (accessed August 31, 2019).
- Ioannides, Yannis M., and Linda Datcher Loury.** 2004. "Job Information Networks, Neighborhood Effects, and Inequality." *Journal of Economic Literature* 42 (4): 1056–93.
- Jackson, Blyden.** 1991. "Introduction: A Street of Dreams." In *Black Exodus: The Great Migration from the American South*, edited by Alferdteen Harrison, xi–xx. Jackson: University Press of Mississippi.
- Jamieson, Stuart M.** 1942. "A Settlement of Rural Migrant Families in the Sacramento Valley, California." *Rural Sociology* 7 (1): 49–61.
- Johnson, Janna E., and Evan J. Taylor.** 2019. "The Long Run Health Consequences of Rural-Urban Migration." *Quantitative Economics* 10 (2): 565–606.
- Johnson, Kenneth M., and Richard W. Rathge.** 2006. "Agricultural Dependence and Changing Population in the Great Plains." In *Population Change and Rural Society*, edited by William A. Kandel and David L. Brown, 197–217. New York: Springer.
- Kennan, John, and James R. Walker.** 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1): 211–51.
- Knowles, Lucas W.** 2010. "Beloit, Wisconsin and the Great Migration: The Role of Industry, Individuals, and Family in the Founding of Beloit's Black Community 1914–1955." https://minds.wisconsin.edu/bitstream/handle/1793/44608/Knowles_Lucas.pdf?sequence=2&isAllowed=y.
- Kreps, Juanita M., Courtenay M. Slater, and Manuel D. Plotkin.** 1979. "The Social and Economic Status of the Black Population in the United States, 1790–1978." Washington, DC: US Bureau of the Census.
- Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode.** 2009. "The Impact of the Boll Weevil, 1892–1932." *Journal of Economic History* 69 (3): 685–718.
- Logan, Trevon D., and John M. Parman.** 2017a. "Data for: The National Rise in Residential Segregation." *Journal of Economic History* 77 (1): 127–70. <https://jmparman.people.wm.edu/dataandcode.html> (accessed September 1, 2019).
- Logan, Trevon D., and John M. Parman.** 2017b. "The National Rise in Residential Segregation." *Journal of Economic History* 77 (1): 127–70.
- Long, Jason, and Henry E. Siu.** 2018. "Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants." *Journal of Economic History* 78 (4): 1001–33.
- Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles.** 2019. "IPUMS National Historical Geographic Information System: Version 14.0 [Database]." IPUMS, Minneapolis, MN. <https://ipums.org/projects/ipums-nhgis/d050.v14.0> (accessed April 6, 2018).
- Margo, Robert A.** 1990. *Race and Schooling in the South, 1880–1950—An Economic History*. Chicago, IL: University of Chicago Press.
- Marks, Carole.** 1989. *Farewell, We're Good and Gone: The Great Black Migration*. Bloomington: Indiana University Press.
- Marks, Carole.** 1991. "The Social and Economic Life of Southern Blacks during the Migrations." In *Black Exodus: The Great Migration from the American South*, edited by Alferdteen Harrison, 36–50. Jackson: University Press of Mississippi.

- McKenzie, David, and Hillel Rapoport.** 2007. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico." *Journal of Development Economics* 84 (1): 1–24.
- Minnesota Population Center, and Ancestry.com.** 2013. "IPUMS Restricted Complete Count Data: Version 1.0 [Machine-readable database]." NSF Census Research Network Michigan Node, University of Minnesota, MN. <https://ebp-projects.isr.umich.edu/NCRN/histcensus.html> (accessed March 18, 2020).
- Missouri Census Data Center.** 2000. "Geocorr 2000: Geographic Correspondence Engine." MCDC Data Applications, Missouri Census Data Center, St. Louis, MO. <http://mcdc.missouri.edu/applications/geocorr2000.html> (accessed March 17, 2020).
- Munshi, Kaivan.** 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118 (2): 549–99.
- Rubin, Morton.** 1960. "Migration Patterns of Negroes from a Rural Northeastern Mississippi Community." *Social Forces* 39 (1): 59–66.
- Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek.** 2019. "IPUMS USA: Version 9.0 [dataset]." IPUMS, University of Minnesota, Minneapolis, MN. <https://ipums.org/projects/ipums-usa/d010.v9.0> (accessed September 5, 2017).
- Schmutte, Ian M.** 2015. "Job Referral Networks and the Determination of Earnings in Local Labor Markets." *Journal of Labor Economics* 33 (1): 1–32.
- Scott, Emmett J.** 1920. *Negro Migration During the War*. New York: Oxford University Press.
- Scroggs, William O.** 1917. "Interstate Migration of Negro Population." *Journal of Political Economy* 25 (10): 1034–43.
- Sjaastad, Larry A.** 1962. "The Costs and Returns of Human Migration." *Journal of Political Economy* 70 (5): 80–93.
- Smith, James P., and Finis R. Welch.** 1989. "Black Economic Progress after Myrdal." *Journal of Economic Literature* 27 (2): 519–64.
- Spitzer, Yannay.** 2016. "Pogroms, Networks, and Migration—The Jewish Migration from the Russian Empire to the United States 1881–1914." https://yannayspitzer.files.wordpress.com/2014/11/pogromsnetworksmigration_160727.pdf.
- Steinbeck, John.** 1939. *The Grapes of Wrath*. New York: Viking Press.
- Stuart, Bryan A., and Evan J. Taylor.** 2021a. "Replication data for: Migration Networks and Location Decisions: Evidence from US Mass Migration." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E118301V1>.
- Stuart, Bryan A., and Evan J. Taylor.** 2021b. "The Effect of Social Connectedness on Crime: Evidence from the Great Migration." *Review of Economics and Statistics* 103 (1): 18–33.
- Tolnay, Stewart E., and E.M. Beck.** 1991. "Rethinking the Role of Racial Violence in the Great Migration." In *Black Exodus: The Great Migration from the American South*, edited by Alferdteen Harrison, 20–35. Jackson: University Press of Mississippi.
- Topa, Giorgio.** 2001. "Social Interactions, Local Spillovers and Unemployment." *Review of Economic Studies* 68 (2): 261–95.
- US Board on Geographic Names.** 2017. "State, Territories, Associated Areas of the United States." USGS. <https://www.usgs.gov/core-science-systems/ngp/board-on-geographic-names/download-gnis-data> (accessed August 1, 2017).
- US Bureau of the Census.** 1992. "Censuses of Religious Bodies, 1906–1936." Inter-university Consortium for Political and Social Research (ICPSR), Ann Arbor, MI. <https://www.icpsr.umich.edu/web/ICPSR/studies/8> (accessed August 29, 2019).
- US Department of Housing and Urban Development.** 2010. "USPS Zip Code Crosswalk Files." Office of Policy Development and Research (PD&R), Washington, DC. https://www.huduser.gov/portal/datasets/usps_crosswalk.html (accessed March 16, 2020).
- Wilkerson, Isabel.** 2010. *The Warmth of Other Suns—The Epic Story of America's Great Migration*. New York: Random House.