I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12709

Migration Networks and Location Decisions: Evidence from U.S. Mass Migration

Bryan A. Stuart
Evan J. Taylor

OCTOBER 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12709

Migration Networks and Location Decisions: Evidence from U.S. Mass Migration

Bryan A. Stuart
George Washington University and IZA

Evan J. Taylor
University of Arizona

OCTOBER 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

Migration Networks and Location Decisions: Evidence from U.S. Mass Migration*

This paper studies how birth town migration networks affected long-run location decisions during historical U.S. 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 Classification: J61, N32, O15, R23, Z13

Keywords: migration networks, location decisions, social interactions, Great Migration

Corresponding author:
Bryan A. Stuart
George Washington University
Department of Economics
2115 G Street NW
Washington, DC 20052
USA

E-mail: bastuart@gwu.edu

---

* For helpful comments and suggestions, we thank 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 U.S. Mass Migration."

1 Introduction

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 data sets 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.

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 U.S. 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 U.S. population born from 1916–1936. 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 1916–1936 in Pigeon Creek, Alabama moved to Niagara County, New York, while less than six 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 U.S. 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 that 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), has 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 Ozden, 2011; Giuletti, Wahba and Zenou, 2014; 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,

---

[1] This 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, 2016). These papers do not focus on the formation of social networks, which in some cases, like Munshi (2003), arise from location decisions.

[2] One 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.

2017; Hornbeck, 2012; Hornbeck and Naidu, 2014; Black et al., 2015; Collins and Wanamaker, 2014, 2015; Johnson and Taylor, 2016; 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.

2 Historical Background on Mass Migration Episodes

The Great Migration saw nearly six million African Americans leave the South from 1910 to 1970 (Census, 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 two to 32 percent from 1910–1970, 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.

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

---

[3] In 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).

[4] The *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).

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

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

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 1916–1936, 14 percent of migrants from Pontotoc lived in Beloit's county in old age (Table 2, discussed in Section 4.1).

Migration out of the Great Plains has received less academic attention than the Great Migration, but nonetheless represents a landmark reshuffling of the U.S. 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 3.1 and 4.1. 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).

3 Estimating the Effects of Migration Networks on Location Decisions

3.1 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 1976–2001. 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 1916–1936.[7] These cohorts lie at the center of both episodes and have among the highest out-migration rates (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 and Great Plains white migrants born from 1916–1936 migrated between 1940 and 1960. Most of these migrants were 15–35 years old when they moved (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, 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 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

---

[7] Our 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.

[8] These 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.

[9] This 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).

[10] To 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

is 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 Appendix D.

3.2 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

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

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 3.3, 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 co-locate in destination k, and $\beta_{j,k}$ equals the additional

---

(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 African Americans in Eriksson (2016)). The low coverage rate also stems from our exclusion of birth towns (Minor Civil Divisions in the Census) with fewer than 10 migrants. The coverage rate for whites from the Great Plains is also too low (8.4 percent) to generate reliable results.

propensity of migrants from the same birth town j to co-locate 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 $P_{j,k} \equiv \mathbb{E}[D_{i,j,k}]$. 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 $N_j \equiv \sum_k N_{j,k}$.

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

---

[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.

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 (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 4.2 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 1916–25 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:

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

---

[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.

effect of birth town migration networks. Section 4.2 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 $\mu_{j,k} \equiv \mathbb{E}[D_{i,j,k}|D_{i',j,k} = 1]$ 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

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

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

---

[14] Proof:

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

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.

county k when an arbitrarily chosen person i makes the same move,

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

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 and can be mapped directly to multiple structural models. 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 5. The network index also can be estimated nonparametrically with increasingly available data.

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

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

The network index transforms the covariance of location decisions in two ways. First, the covari-

---

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

ance 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 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

---

[16] In 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.

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

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.

3.3 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 further 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 the same birth town group.[18] Given x^*, the location of the grid is determined by a single latitude-longitude

---

[18] That is,

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

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. Appendix Table A.2 reports the values of x^* chosen by cross validation.

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 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,

$$\widehat{P_{g,k}} = \frac{\sum_{j \in g} N_{j,k}}{\sum_{j \in g} N_j}. \tag{5}$$

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

$$\widehat{P^2_{g,k}} = \frac{\sum_{j \in g} \sum_{j' \neq j \in g} N_{j,k} N_{j',k}}{\sum_{j \in g} \sum_{j' \neq j \in g} N_j N_{j'}} \tag{6}$$

$$\widehat{C_{j,k}} = \frac{N_{j,k}(N_{j,k}-1)}{N_j(N_j-1)} - \widehat{P^2_{g,k}}. \tag{7}$$

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

---

[19] To 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.

[20] 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 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. 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 $P^2_{g,k}$ without multiple towns in the same group.

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

Given $(\widehat{P_{g,k}}, \widehat{P^2_{g,k}}, \widehat{C_{j,k}}, N_j)$, we can estimate the network index, $\Delta_{j,k}$, using equation (4). However, each estimate $\widehat{\Delta_{j,k}}$ 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,

$$\widehat{\Delta_k} = \sum_j \left(\frac{\widehat{P_{g(j),k}} - \widehat{P^2_{g(j),k}}}{\sum_{j'} \widehat{P_{g(j'),k}} - \widehat{P^2_{g(j'),k}}} \right) \widehat{\Delta_{j,k}}, \tag{8}$$

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 $\mathbb{E}[D_{i,j,k}] = P_{g,k}$). 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. Furthermore, the destination-level network index estimate, $\widehat{\Delta_k}$, 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,

$$\widehat{\Delta_c} = \sum_k \sum_{j \in c} \left(\frac{\widehat{P_{g(j),k}} - \widehat{P^2_{g(j),k}}}{\sum_{k'} \sum_{j' \in c} \widehat{P_{g(j'),k'}} - \widehat{P^2_{g(j'),k'}}} \right) \widehat{\Delta_{j,k}}. \tag{9}$$

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, $(\widehat{P_{g,k}}, \widehat{P^2_{g,k}}, \widehat{C_{j,k}}, N_j)$. 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 $\widehat{\Delta_k}$, we treat the birth town group as the unit of observation and use a

---

[22] See Appendix B for details.

GMM variance estimator. This is akin to calculating heteroskedastic robust standard errors clustered by birth town group.[23] Appendix B contains details.

3.4 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 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,

$$P_{j,k} = \rho_{g,k} + X_j \pi_k, \tag{10}$$

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

---

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

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

[25] Census data do not contain individuals' town of residence, but do include information on minor civil division (MCD), which is a sub-county 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

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, $\widetilde{P_{j,k}}$, equal to the fitted value from the OLS regression

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

We use fitted values from a separate OLS regression, implied by equation (10), to form an alternative squared moving probability estimate, $\widetilde{P^2_{j,k}}$.[26] We estimate all equations separately for each birth state. Similarity between the baseline and alternative network index estimates would provide support for our empirical strategy.[27]

4 Results: The Effects of Migration Networks on Location Decisions

4.1 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 1916–1936, 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

---

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.

[26] We estimate $\widetilde{P^2_{j,k}}$ 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, $j \neq j'$.

[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.

Great Plains lived outside their birth region, while only nine 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 appendix. On average, there are 142 Southern black migrants and 181 Great Plains white migrants per birth town (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 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, $\widehat{P_{g,k}}$, 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 whites 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, $\widehat{\Delta_k}$. 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 10 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

---

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

[29] 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); Boustan (2017).

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 4.4, 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 U.S.[32] We estimate a strong network effect ($\widehat{\Delta_k} > 3$) 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 ($\widehat{\Delta}_k > 3$), an area described memorably in *The Grapes of Wrath* (Steinbeck, 1939).[33] Unlike black

---

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

[31] Appendix Figure A.7 displays the associated t-statistic distributions, and 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 10 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.

migrants, whites moved to a large number of destinations throughout the U.S. The difference between the number of destinations chosen by Mississippi black and North Dakota white migrants is striking, especially because our data contain almost 30,000 more migrants from Mississippi. While 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 chosen by African Americans does not mechanically generate stronger network effects, because we identify these effects using the location choices of nearby migrants.[34] Appendix Figures A.10 and A.11, for South Carolina-born black and Kansas-born white migrants, show similar patterns.

4.2 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 3.4. 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 whites. 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.

---

[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.

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 of 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 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 10 migrants.[37] We exclude both large birth towns and large destinations in column 4. The average network index estimates are similar across all four specifications for both

---

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

[36] The 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

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

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. Appendix C describes how we estimate the effects of cross-race migration networks, and 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.

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). 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 1916–1936, coverage rates are not perfect. 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.

---

[38] 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.

4.3 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 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. Appendix Table A.13 shows that network index estimates are similar in magnitude among men and women, implying that our 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 1916–1936.

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 1916–1936. 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 1916–1936 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. 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]

---

[40] The 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 (Appendix Table A.13).

[41] Conditional on family size, black and white migrants could have differed in the extent to which they tended to

4.4 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 2. 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 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

---

follow other family members. We do not have data that let us examine this possibility.

[42] There 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); Heath (2016).

[43] Appendix Table A.14 contains summary statistics. 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.

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 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 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 (2017), which captures the extent to which black households had non-black next-door

---

[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, D.C., 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 (Appendix Table A.15). Results for Southern whites are in Appendix Table A.16.

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] 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 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 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 educa-

---

[47] This 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, Forthcoming, Appendix Table A.7), this would lead to a lower relative wage.

[48] For 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.

[49] Drawing 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.

tional achievement (literacy and school attendance) and, for African Americans, access to Rosenwald schools, which improved educational attainment in this period (Aaronson and Mazumder, 2011). 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 SD increase 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

---

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

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

[52] Columns 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.

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 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.

5 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.

5.1 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

---

[53] The assumption that H_i is a non-empty 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.

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 $h_{j,k} \equiv \mathbb{P}[k \in H_i | i \in j]$, and the probability that destination k is in the medium preference group is $m_{j,k} \equiv \mathbb{P}[k \in M_i | i \in j]$. These probabilities arise 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 that chose destination k because of the network is $N_k^{\text{network}} \equiv \sum_j N_{j,k} m_{j,k}$, where $N_{j,k}$ is the number of migrants that moved from j to k. In the absence of networks, where $m_{j,k} = 0$, migrants move to the destination in H_i. 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 $N_k^{\text{cf}} \equiv \sum_j N_j h_{j,k}$.

The Appendix 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 As-

---

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

$$\mathbb{P}[k \in M_i | D_{i,j,k} = 1] = \frac{\mathbb{P}[D_{i,j,k} = 1 | k \in M_i]\, \mathbb{P}[k \in M_i]}{\mathbb{P}[D_{i,j,k} = 1]} = \frac{P_{j,k} m_{j,k}}{P_{j,k}} = m_{j,k}$$

because $\mathbb{P}[D_{i,j,k} = 1 | k \in M_i] = \mathbb{P}[D_{i-1,j,k} = 1] = P_{j,k}$. 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.

sumptions 1 and 2, plus the additional structure imposed by the model of local social interactions.

5.2 Results

Table 8 reports estimates of the percent of migrants that chose their destination because of migration networks, calculated as migrant-weighted averages of $100 \times (\widehat{N_k^{\text{network}}/N_k})$. 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 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 U.S.[56]

Since migration networks clearly affected where individuals moved, a natural question is whether

---

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

[56] 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.

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 that would move in the counterfactual, the mean unemployment rate falls by 0.7 percentage 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] Panel B, for Great Plains white migrants, 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 that 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.

---

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

[58] Results also are similar when we consider a counterfactual in which Southern black migration networks are as strong as those of Great Plains white migrants.

6 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 U.S. 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 whites 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 U.S. cities (Stuart and Taylor, Forthcoming). 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: Details on the Structural Model

This appendix provides additional details on the structural model introduced in Section 5.1.

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

$$\begin{aligned} P_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1] &= \mathbb{P}[D_{i-1,j,k} = 1, k \in H_i] + \mathbb{P}[D_{i-1,j,k} = 1, k \in M_i] \\ &\quad + \sum_{k' \neq k} \mathbb{P}[D_{i-1,j,k'} = 1, k \in H_i, k' \in L_i]. \end{aligned} \tag{12}$$

The first term on the right hand side of equation (12) 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 5.1, we rewrite equation

(12) as

$$P_{j,k} = P_{j,k}h_{j,k} + P_{j,k}m_{j,k} + \sum_{k' \neq k} P_{j,k'}h_{j,k}\left(\frac{1 - h_{j,k'} - m_{j,k'}}{1 - h_{j,k'}}\right). \tag{13}$$

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 $\nu_{j,k} \equiv \mathbb{P}[k \in M_i | k \notin H_i, i \in j]$. 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 (13) to

$$P_{j,k} = P_{j,k}\nu_{j,k} + \sum_{k'=1}^{K} P_{j,k'}(1 - \nu_{j,k'})h_{j,k}. \tag{14}$$

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

$$C_{j,k} = \frac{2P_{j,k}(1 - P_{j,k})\sum_{a=1}^{N_j - 1}(N_j - a)\left(\frac{\rho_{j,k} - P_{j,k}}{1 - P_{j,k}}\right)^a}{N_j(N_j - 1)}, \tag{15}$$

where $\rho_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1 | D_{i-1,j,k} = 1, i \in j] = h_{j,k} + m_{j,k}$ 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 (15) into the expression for the network index

---

[59] Equation (15) follows from the fact that the covariance of location decisions for individuals i and $i + n$ is Cov$[D_{i,j,k}, D_{i+n,j,k}] = P_{j,k}(1 - P_{j,k})\left(\frac{\rho_{j,k} - P_{j,k}}{1 - P_{j,k}}\right)^n$.

in equation (4), simplifying, and taking the limit as $N_j \to \infty$ yields

$$\Delta_{g,k} = \frac{2(\rho_{g,k} - P_{g,k})}{1 - \rho_{g,k}}, \tag{16}$$

where $\Delta_{g,k}$ is the birth town group-destination network index. Equation (16) can be rearranged to show that

$$\rho_{g,k} = \frac{2P_{g,k} + \Delta_{g,k}}{2 + \Delta_{g,k}}. \tag{17}$$

We use equation (17) to estimate $\rho_{g,k}$ with our estimates of $P_{g,k}$ and $\Delta_{g,k}$.

Equation (14), plus the facts that $\nu_{g,k} = m_{g,k}/(1 - h_{g,k})$ and $\rho_{g,k} = h_{g,k} + m_{g,k}$, imply that

$$\rho_{g,k} = \nu_{g,k} + \frac{P_{g,k}(1 - \nu_{g,k})^2}{\sum_{k'=1}^{K} P_{g,k'}(1 - \nu_{g,k'})}. \tag{18}$$

We use equation (18) to estimate $\nu_g \equiv (\nu_{g,1}, \ldots, \nu_{g,K})$ using our estimates of $(P_{g,1}, \ldots, P_{g,K}, \rho_{g,1}, \ldots, \rho_{g,K})$. We employ a computationally efficient algorithm that leverages the fact that equation (18) 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 (14) 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

---

[60] The 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.

the network index: if $m_{j,k} = 0$, then equation (13) implies that $P_{j,k} = h_{j,k}$ and equation (16) 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 2.

References

Aaronson, Daniel, and Bhashkar Mazumder. 2011. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy*, 119(5): 821–888.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2017. "Cultural Assimilation during the Age of Mass Migration." *NBER Working Paper 22831*.

Bartel, Ann P. 1989. "Where do the New U.S. Immigrants Live?" *Journal of Labor Economics*, 7(4): 371–391.

Bauer, Thomas, Gil S. Epstein, and Ira N. Gang. 2005. "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics*, 18(4): 649–662.

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): 1150–1196.

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): 128–161.

Beine, Michel, Frédéric Docquier, and Çağlar Ozden. 2011. "Diasporas." *Journal of Development Economics*, 95(1): 30–41.

Bell, Velma Fern. 1933. "The Negro in Beloit and Madison, Wisconsin." Master's diss. University of Wisconsin.

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.

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.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. 2011. "Identification of Social Interactions." In *Handbook of Social Economics*. Vol. 1, ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 853–964. Elsevier.

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): 756–783.

Boustan, Leah Platt. 2010. "Was Postwar Suburbanization 'White Flight'? Evidence from the Black Migration." *Quarterly Journal of Economics*, 125(1): 417–443.

Boustan, Leah Platt. 2017. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets.* Princeton University Press.

Brock, William A., and Steven N. Durlauf. 2001. "Discrete Choice with Social Interactions." *Review of Economic Studies*, 68(2): 235–260.

Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. "The Value of Hiring through Employee Referrals." *Quarterly Journal of Economics*, 130(2): 805–839.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with Endogenous Moving Costs." *American Economic Review*, 86(4): 909–930.

Census, United States Bureau of the. 1979. "The Social and Economic Status of the Black Population in the United States, 1790-1978." *Current Population Reports, Special Studies Series P-23 No. 80.*

Census, United States Bureau of the. 1992. "Censuses of Religious Bodies, 1906-1936. Ann Arbor, MI: ICPSR [distributor]."

Chay, Kenneth, and Kaivan Munshi. 2015. "Black Networks After Emancipation: Evidence from Reconstruction and the Great Migration."

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): 607–632.

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): 947–992.

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): 363–386.

Eriksson, Katherine. 2016. "Education and Incarceration in the Jim Crow South: Evidence from Rosenwald Schools."

Gibson, Campbell, and Kay Jung. 2005. "Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic Origin, 1790 to 1990, For Large Cities and Other Urban Places in the United States." *U.S. Census Bureau Population Division Working Paper No. 76.*

Giuletti, Corrado, Jackline Wahba, and Yves Zenou. 2014. "Strong versus Weak Ties in Migration."

Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics*, 111(2): 507–548.

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): 643–660.

Greenwood, Michael J. 1997. "Internal Migration in Developed Countries." In *Handbook of Population and Family Economics.* Vol. 1, ed. Mark R. Rosenzweig and Oded Stark, 647–720. Elsevier.

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: University of Chicago Press.

Haines, Michael R., and ICPSR. 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. Ann Arbor, MI: ICPSR [distributor]."

Heath, Rachel. 2016. "Why do Firms Hire using Referrals? Evidence from Bangladeshi Garment Factories."

Hellerstein, Judith K., Melissa McInerney, and David Neumark. 2011. "Neighbors and Coworkers: The Importance of Residential Labor Markets." *Journal of Labor Economics*, 29(4): 659–695.

Henri, Florette. 1975. *Black Migration: Movement North, 1900-1920.* New York: Anchor Press/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): 963–990.

Hurt, Douglas R. 2011. *The Big Empty: The Great Plains in the Twentieth Century.* Tucson: University of Arizona Press.

ICPSR. 1999. "United States Historical Election Returns, 1824-1968. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]."

Ioannides, Yannis M., and Linda Datcher Loury. 2004. "Job Information Networks, Neighborhood Effects, and Inequality." *Journal of Economic Literature*, 42(4): 1056–1093.

Jackson, Blyden. 1991. "Introduction: A Street of Dreams." In *Black Exodus: The Great Migration from the American South*. ed. Alferdteen Harrison, xi–xvii. 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. 2016. "The Heterogeneous Long-Run Health Consequences of Rural-Urban Migration."

Johnson, Kenneth M., and Richard W. Rathge. 2006. "Agricultural Dependence and Changing Population in the Great Plains." In *Population Change and Rural Society*. ed. William A. Kandel and David L. Brown, 197–217. Springer.

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica*, 79(1): 211–251.

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."

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. 2017. "The National Rise in Residential Segregation." *The Journal of Economic History*, 77(1): 127–170.

Long, Jason, and Henry E. Siu. 2018. "Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants."

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2019. "IPUMS

National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS."

Margo, Robert A. 1990. *Race and Schooling in the South, 1880-1950: An Economic History.* Chicago: The 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*. ed. 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]."

Minnesota Population Center and Ancestry.com. 2013. "IPUMS Restricted Complete Count Data: Version 1.0 [Machine-readable database]. Minneapolis: University of Minnesota."

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*, 118(2): 549–597.

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]."

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): 1034–1043.

Sjaastad, Larry A. 1962. "The Costs and Returns of Human Migration." *Journal of Political Economy*, 70(5): 80–93.

Smith, James P., and Finis Welch. 1989. "Black Economic Progress After Myrdal." *Journal of Economic Literature*, 27(2): 519–564.

Spitzer, Yannay. 2016. "Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States 1881-1914."

Steinbeck, John. 1939. *The Grapes of Wrath.* New York: The Viking Press.

Stuart, Bryan A., and Evan J. Taylor. Forthcoming. "The Effect of Social Connectedness on Crime: Evidence from the Great Migration." *Review of Economics and Statistics*.

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*. ed. Alferdteen Harrison, 20–35. Jackson: University Press of Mississippi.

Topa, Giorgio. 2001. "Social Interactions, Local Spillovers and Unemployment." *Review of Economic Studies*, 68(2): 261–295.

Wilkerson, Isabel. 2010. *The Warmth of Other Suns: The Epic Story of America's Great Migration.* New York: Random House.

Table 1: Location at Old Age, 1916–1936 Cohorts

		Percent Living in Location		
		Outside Birth	In Birth Region	
Birth State	People (1)	Region (2)	Birth State (3)	Other State (4)
Panel A: Black Individuals from South				
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%
Panel B: White Individuals from Great Plains				
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%
Panel C: White Individuals from South				
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: Duke SSA/Medicare data

Table 2: Extreme Examples of Correlated Location Decisions, Southern Black 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)
Panel A: Southern Black Migrants								
Pigeon Creek, AL	Niagara Falls, NY	85	43	50.6%	0.5%	64.5	4.5%	8.5
Marion, AL	Fort Wayne, IN	1311	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
Panel B: Great Plains White Migrants								
Krebs, OK	Akron, OH	210	32	15.2%	0.1%	82.6	0.3%	7.4
Haven, KS	Elkhart, IN	144	22	15.3%	0.1%	51.1	0.4%	6.9
McIntosh, SD	Rupert, ID	299	20	6.7%	0.1%	50.9	0.6%	4.8
Hull, ND	Bellingham, WA	55	24	43.6%	0.5%	44.6	1.5%	4.3
Lindsay, NE	Moline, IL	226	29	12.8%	0.2%	41.5	0.4%	5.2
Corsica, SD	Holland, MI	253	26	10.3%	0.2%	39.6	0.4%	6.3
Corsica, SD	Grand Rapids, MI	253	34	13.4%	0.3%	37.2	0.7%	6.0
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%	2.2
Henderson, NE	Fresno, CA	146	32	21.9%	0.7%	31.1	0.8%	2.2

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: Duke SSA/Medicare data

Table 3: Average Network Index Estimates, by Birth State

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

Table 4: Average Network Index Estimates, Robustness to Identifying Assumptions

Birth State	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Southern Black Migrants						
Alabama	1.888	1.852	1.648	2.120	2.121	2.178
	(0.195)	(0.189)	(0.246)	(0.198)	(0.198)	(0.212)
Florida	0.813	0.742	0.737	0.829	0.829	0.923
	(0.117)	(0.119)	(0.132)	(0.116)	(0.117)	(0.125)
Georgia	1.657	1.689	1.903	1.768	1.788	1.731
	(0.177)	(0.175)	(0.191)	(0.179)	(0.179)	(0.192)
Louisiana	1.723	1.651	1.648	1.756	1.761	1.709
	(0.478)	(0.474)	(0.715)	(0.455)	(0.452)	(0.451)
Mississippi	2.303	2.295	2.098	2.364	2.362	2.119
	(0.313)	(0.306)	(0.402)	(0.312)	(0.312)	(0.307)
North Carolina	1.539	1.482	1.367	1.606	1.606	1.751
	(0.130)	(0.127)	(0.133)	(0.130)	(0.130)	(0.135)
South Carolina	2.618	2.636	2.830	2.860	2.871	2.683
	(0.301)	(0.304)	(0.354)	(0.297)	(0.297)	(0.315)
All States	1.938	1.917	1.873	2.059	2.063	2.003
	(0.110)	(0.108)	(0.143)	(0.109)	(0.108)	(0.110)
Panel B: Great Plains White Migrants						
Kansas	0.255	0.233	0.204	0.257	0.257	0.320
	(0.024)	(0.024)	(0.021)	(0.024)	(0.024)	(0.028)
Nebraska	0.361	0.349	0.309	0.361	0.361	0.455
	(0.082)	(0.082)	(0.074)	(0.082)	(0.082)	(0.105)
North Dakota	0.464	0.445	0.412	0.475	0.475	0.480
	(0.036)	(0.035)	(0.037)	(0.036)	(0.036)	(0.038)
Oklahoma	0.453	0.439	0.383	0.460	0.460	0.455
	(0.036)	(0.036)	(0.034)	(0.036)	(0.036)	(0.038)
South Dakota	0.350	0.331	0.305	0.354	0.354	0.366
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.027)
All States	0.380	0.363	0.323	0.385	0.385	0.418
	(0.022)	(0.022)	(0.020)	(0.022)	(0.022)	(0.026)
Assumption 1: Allowing moving probability to depend on covariates from						
Duke SSA		x	x			
1910 U.S. Census			x			
Assumption 2: Allowing for social interactions among towns that are						
$\leq$ 10 miles				x	x	x
$\leq$ 20 miles & 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 (combines farmer and farm laborer category, ages 18–39), percent inter-state 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: Duke SSA/Medicare data, Black et al. (2015) data, Minnesota Population Center and Ancestry.com (2013)

Table 5: Average Network Index Estimates, by Size of Birth Town and Destination

Exclude Largest Birth Towns: Exclude Largest Destinations: Birth State	No No (1)	Yes No (2)	No Yes (3)	Yes Yes (4)
Panel A: Southern Black Migrants				
Alabama	1.888	1.784	2.056	2.189
	(0.195)	(0.149)	(0.285)	(0.268)
Florida	0.813	0.607	1.323	1.231
	(0.117)	(0.061)	(0.229)	(0.215)
Georgia	1.657	1.458	1.696	1.772
	(0.177)	(0.092)	(0.170)	(0.133)
Louisiana	1.723	1.106	0.971	0.960
	(0.478)	(0.095)	(0.182)	(0.176)
Mississippi	2.303	2.299	2.085	2.032
	(0.313)	(0.304)	(0.210)	(0.205)
North Carolina	1.539	1.451	0.743	0.687
	(0.130)	(0.126)	(0.064)	(0.059)
South Carolina	2.618	2.556	1.784	1.742
	(0.301)	(0.283)	(0.241)	(0.234)
All States	1.938	1.791	1.755	1.783
	(0.110)	(0.089)	(0.108)	(0.102)
Panel B: Great Plains White Migrants				
Kansas	0.255	0.220	0.243	0.228
	(0.024)	(0.019)	(0.021)	(0.019)
Nebraska	0.361	0.253	0.265	0.253
	(0.082)	(0.014)	(0.019)	(0.017)
North Dakota	0.464	0.464	0.527	0.531
	(0.036)	(0.036)	(0.046)	(0.046)
Oklahoma	0.453	0.395	0.450	0.427
	(0.036)	(0.029)	(0.040)	(0.038)
South Dakota	0.350	0.339	0.387	0.381
	(0.026)	(0.026)	(0.034)	(0.033)
All States	0.380	0.331	0.374	0.361
	(0.022)	(0.012)	(0.016)	(0.016)

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 D.C., San Francisco, Pittsburgh, and St. Louis, in addition to counties which received fewer than 10 migrants. Column 4 excludes large birth towns and large destinations. Birth town groups are defined by cross validation. Standard errors are in parentheses.
Source: Duke SSA/Medicare data

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	0.414	0.371	-0.086	-0.279	-0.280
	(0.528)	(0.664)	(0.672)	(0.101)	(0.136)	(0.136)
Manufacturing employment share by small destination indicator		2.418	2.462		0.268	0.275
		(0.984)	(0.976)		(0.178)	(0.177)
Agriculture employment share, 1910	0.070	-0.391	-0.451	0.093	0.192	0.188
	(0.287)	(0.447)	(0.450)	(0.045)	(0.125)	(0.125)
Agriculture employment share by small destination indicator		0.681	0.606		-0.104	-0.098
		(0.477)	(0.490)		(0.125)	(0.124)
Small destination indicator		-0.494	-0.496		0.053	0.050
		(0.262)	(0.262)		(0.065)	(0.065)
Log distance from birth state	-0.436	-0.403	-0.401	0.062	0.076	0.067
	(0.064)	(0.066)	(0.063)	(0.035)	(0.036)	(0.037)
Direct railroad connection from birth state	0.305	0.307	0.293	0.203	0.201	0.191
	(0.112)	(0.112)	(0.128)	(0.041)	(0.042)	(0.043)
One-stop railroad connection from birth state	0.221	0.211	0.166	0.079	0.073	0.072
	(0.076)	(0.074)	(0.078)	(0.017)	(0.017)	(0.016)
Log population, 1910	0.061	0.063	0.069	0.023	0.034	0.034
	(0.052)	(0.056)	(0.057)	(0.010)	(0.010)	(0.010)
Percent African-American, 1910	-2.081	-1.915	-1.837	-0.226	-0.248	-0.244
	(0.358)	(0.354)	(0.349)	(0.036)	(0.038)	(0.038)
Percent rural, 1910	-0.285	-0.268	-0.224	-0.049	-0.039	-0.038
	(0.197)	(0.212)	(0.216)	(0.044)	(0.044)	(0.044)
Black/white church members per capita, 1916	-0.321	-0.235	-0.220	-0.111	-0.095	-0.095
	(0.179)	(0.191)	(0.193)	(0.039)	(0.039)	(0.039)
Birth state fixed effects			x			x
R-squared	0.096	0.103	0.115	0.033	0.037	0.037
N (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 10 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 capita. Standard errors, clustered by destination county, are in parentheses.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Ruggles et al. (2019)

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	2.809	0.682	1.001
	(2.100)	(2.279)	(0.567)	(0.647)
Percent of black/white individuals in owner-occupied housing, 1910	-3.350	-3.575	0.579	0.177
	(2.972)	(2.990)	(0.665)	(0.735)
Percent of black/white workers in agriculture, 1910	-0.884	-0.461	-0.247	-0.058
	(1.844)	(1.888)	(0.494)	(0.547)
Percent of black/white workers in manufacturing, 1910	-5.952	-6.031	1.606	1.582
	(3.284)	(3.605)	(3.085)	(3.019)
Percent of farm acreage in cotton, 1910	-1.441	-1.003	1.658	2.015
	(2.951)	(3.244)	(0.899)	(0.905)
Log black/white population density, 1910	0.985	0.851	0.228	0.252
	(0.555)	(0.561)	(0.090)	(0.095)
Black/white church members per capita, 1916	0.223	0.217	0.400	0.405
	(0.739)	(0.662)	(0.269)	(0.280)
Rosenwald school exposure	-0.595	-0.974		
	(0.843)	(0.878)		
Black/white literacy rate (10+), 1910	2.021	4.073	-8.216	-9.290
	(2.408)	(3.180)	(2.845)	(2.952)
Black/white school attendance rate (6–14), 1910	1.170	2.199	-1.637	-1.700
	(1.519)	(1.554)	(0.615)	(0.689)
Railroad exposure	-0.310	-0.252	-0.144	-0.184
	(0.464)	(0.478)	(0.073)	(0.084)
Percent African-American, 1910	-0.273	0.332	-0.214	0.210
	(1.711)	(1.734)	(0.974)	(0.939)
Percent rural, 1910	-0.508	-0.675	-0.591	-0.675
	(1.736)	(1.723)	(0.262)	(0.277)
Percent voting for Strom Thurmond, 1948	0.613	-0.298		
	(0.524)	(0.929)		
Destination state fixed effects		x		x
R-squared	0.089	0.100	0.282	0.304
N (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: Duke SSA/Medicare data, Aaronson and Mazumder (2011) data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), ICPSR (1999), Ruggles et al. (2019)

Table 8: Estimated Percent of Migrants That Chose Their Destination Because of Migration Network

	Destination Region				
Birth State	All (1)	Northeast (2)	Midwest (3)	West (4)	South (5)
Panel A: Southern Black Migrants					
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	-
Panel B: Great Plains White Migrants					
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 × $(N_k^{\text{network}}/N_k)$, the percent of migrants that chose their destination because of their birth town migration network. See the text for details.

Source: Duke SSA/Medicare data

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)
Panel A: Southern Black Migrants					
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	12.0
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
Panel B: Great Plains White Migrants					
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	-28.3
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^{cf}, the number of migrants that 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: Duke SSA/Medicare data

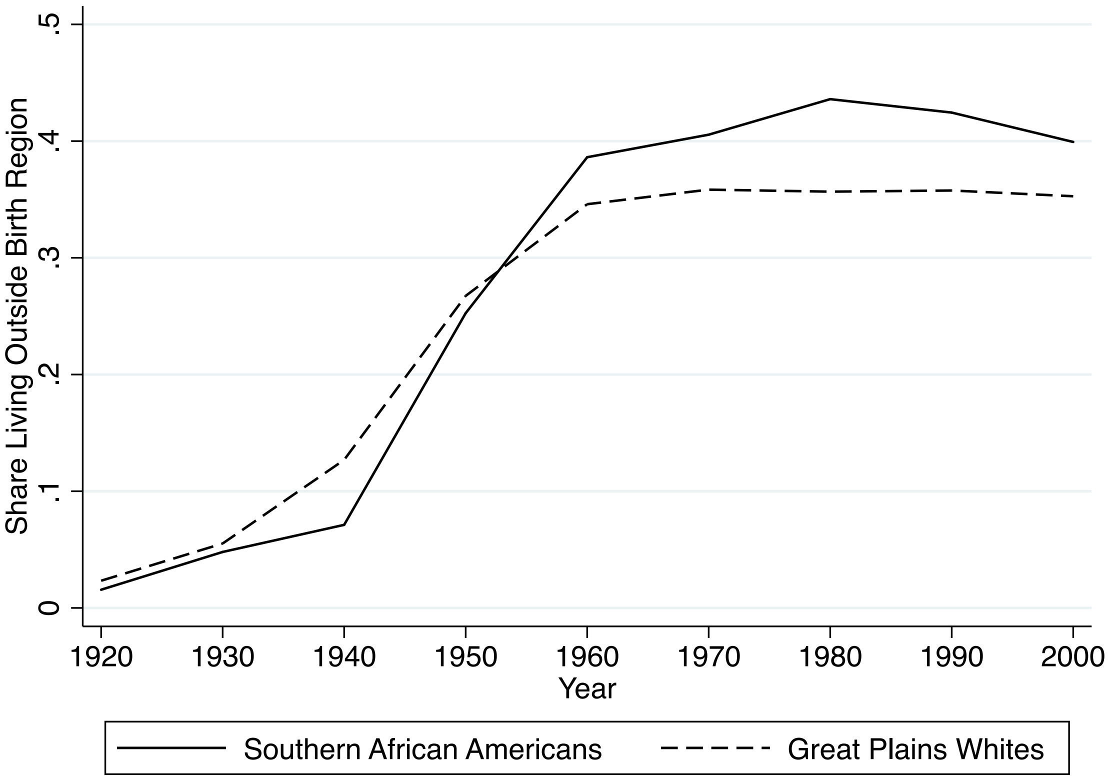
Table 10: Characteristics of Counties where Southern Black Migrants Live Under Realized and Counterfactual Migration Networks

	Realized Networks	No Network Counterfactual		
County Characteristic	Mean (1)	Mean (2)	Mean change, all migrants (3)	Mean change, switchers (4)
Panel A: Southern Black Migrants				
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
Panel B: Great Plains White Migrants				
Unemployment rate, 2000	6.61	6.70	0.09	0.71
Poverty rate, 2000	6.64	6.53	-0.11	-0.87
Percent black, 2000	12.68	12.48	-0.20	-1.58
Unemployment rate, 1980	6.02	6.05	0.03	0.24
Poverty rate, 1980	7.34	7.32	-0.02	-0.16
Percent black, 1980	11.21	11.17	-0.04	-0.32

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 that would switch their location under the counterfactual, calculated as column 3 divided by the percent of migrants that would switch their location (34.5 percent in Panel A and 12.7 percent in Panel B). See the text for details.

Source: Duke SSA/Medicare data, Manson et al. (2019)

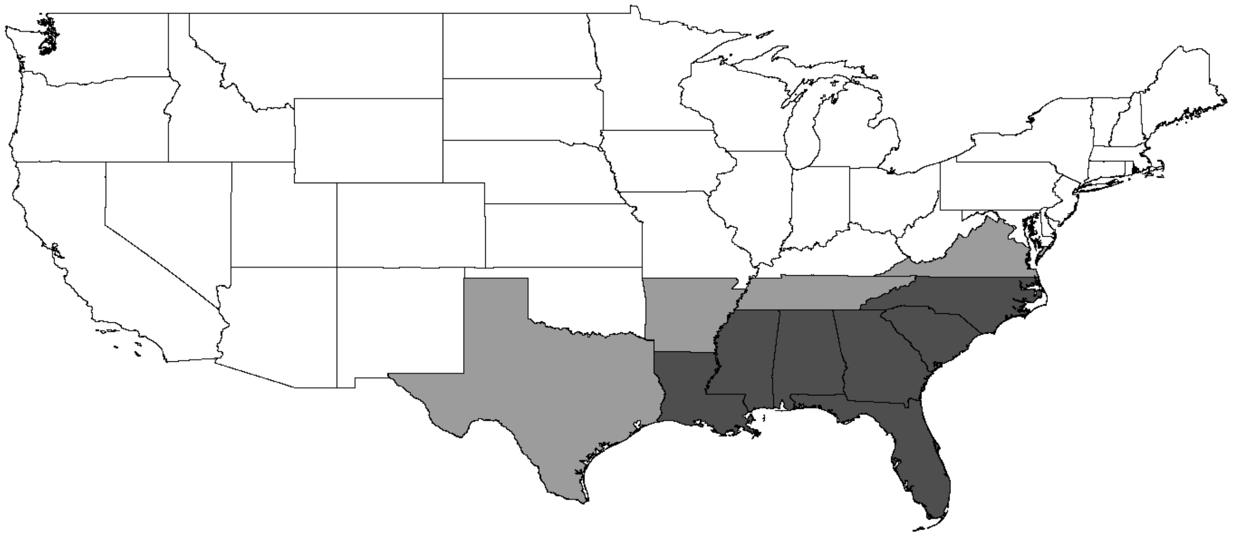
Figure 1: Share Living Outside Birth Region, 1916–1936 Cohorts, by Year



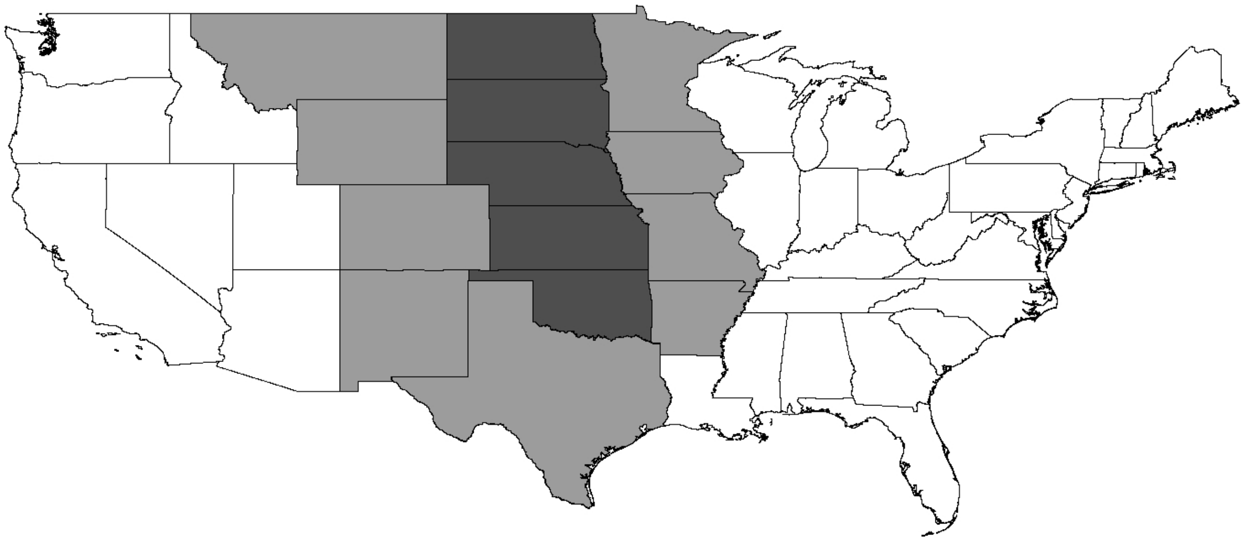
Notes: The solid line shows the percent of African Americans born from 1916–1936 in the seven Southern birth states we analyze (dark grey states in Figure 1a) living outside the South (light and dark grey states) at the time of Census enumeration. The dashed line shows the percent of whites born from 1916–1936 from the Great Plains states living outside the Great Plains or Border States.
Source: Ruggles et al. (2019)

Figure 2: Geographic Coverage

(a) South



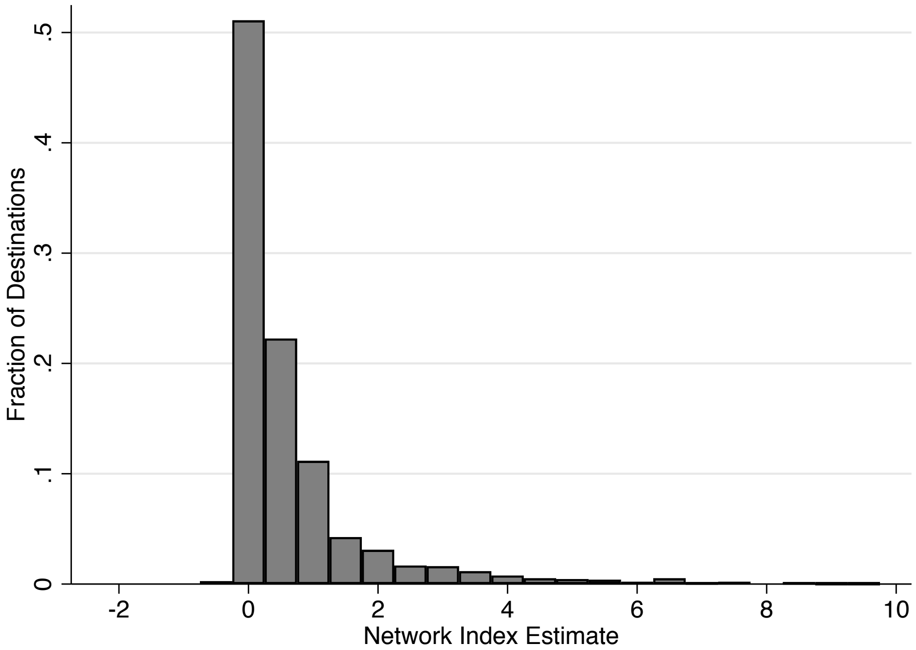
(b) Great Plains



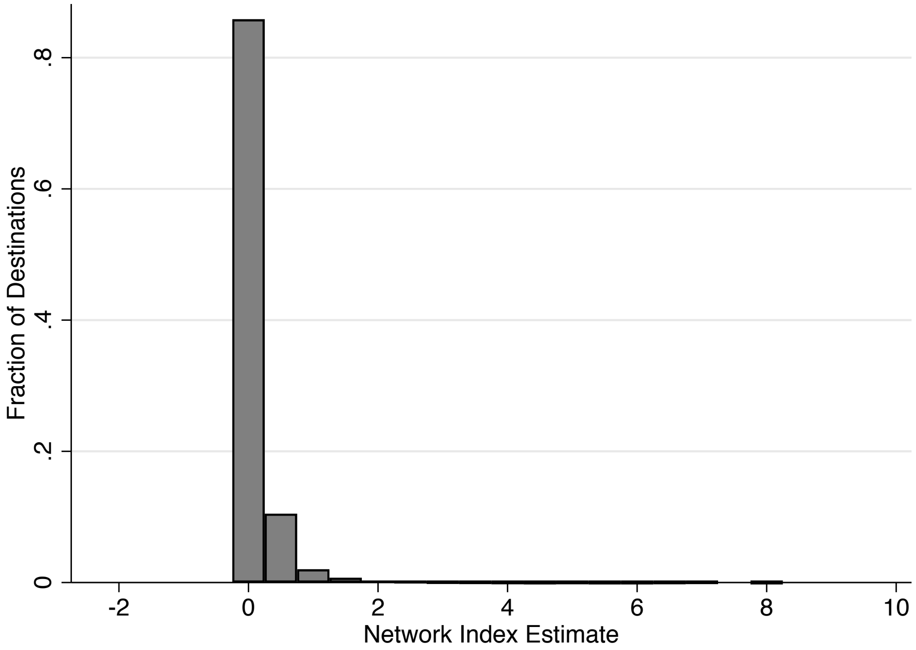
Notes: For the South, our sample includes migrants born in the seven states in dark grey (Alabama, Georgia, Florida, Louisiana, Mississippi, North Carolina, South Carolina). A migrant is someone who at old age lives outside of the former Confederate states, which are the dark and light grey states. For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

Figure 3: Distribution of Destination-Level Network Index Estimates

(a) Southern Black Migrants

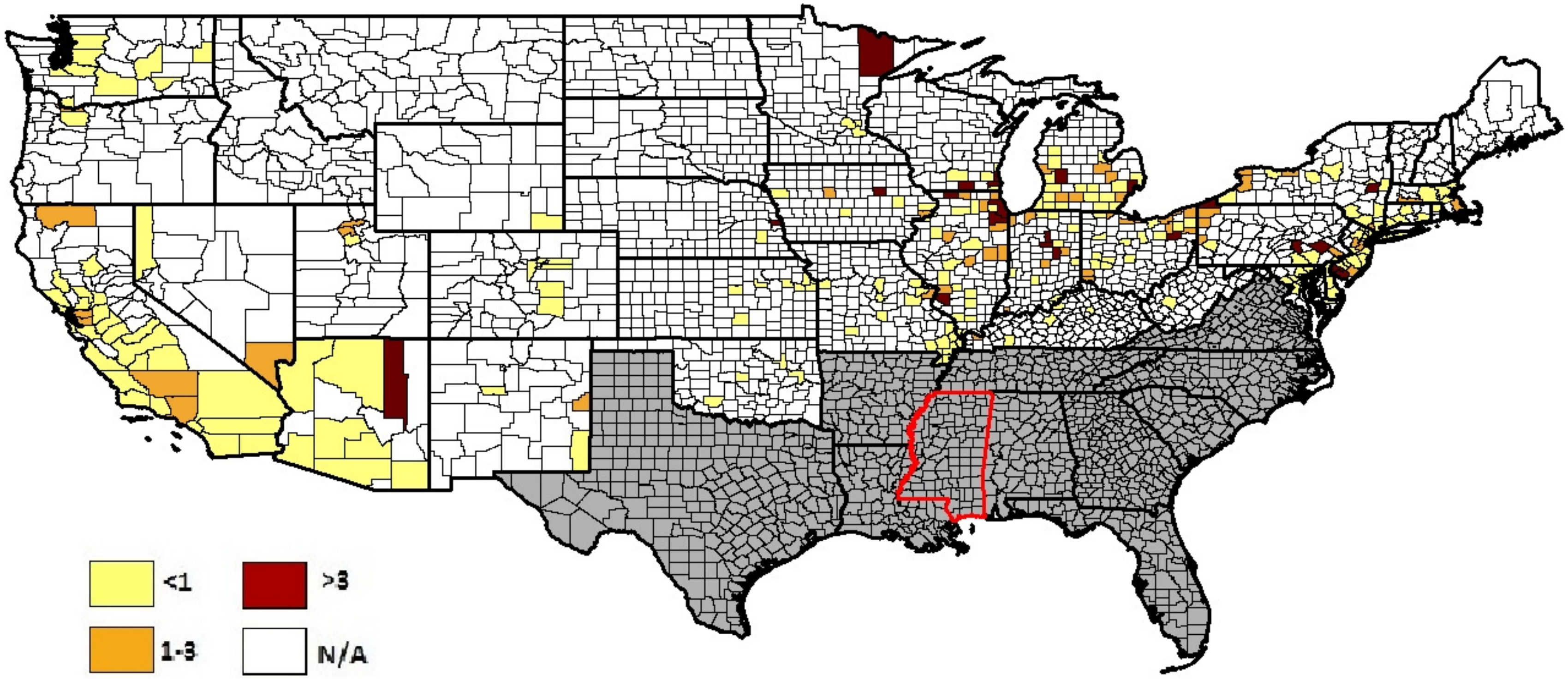


(b) Great Plains White Migrants



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, WI, $\hat{\Delta}_k = 15.2$ from South Carolina to Rensselaer County, NY, and $\hat{\Delta}_k = 18.1$ from Florida to St. Joseph County, IN.
Source: Duke SSA/Medicare data

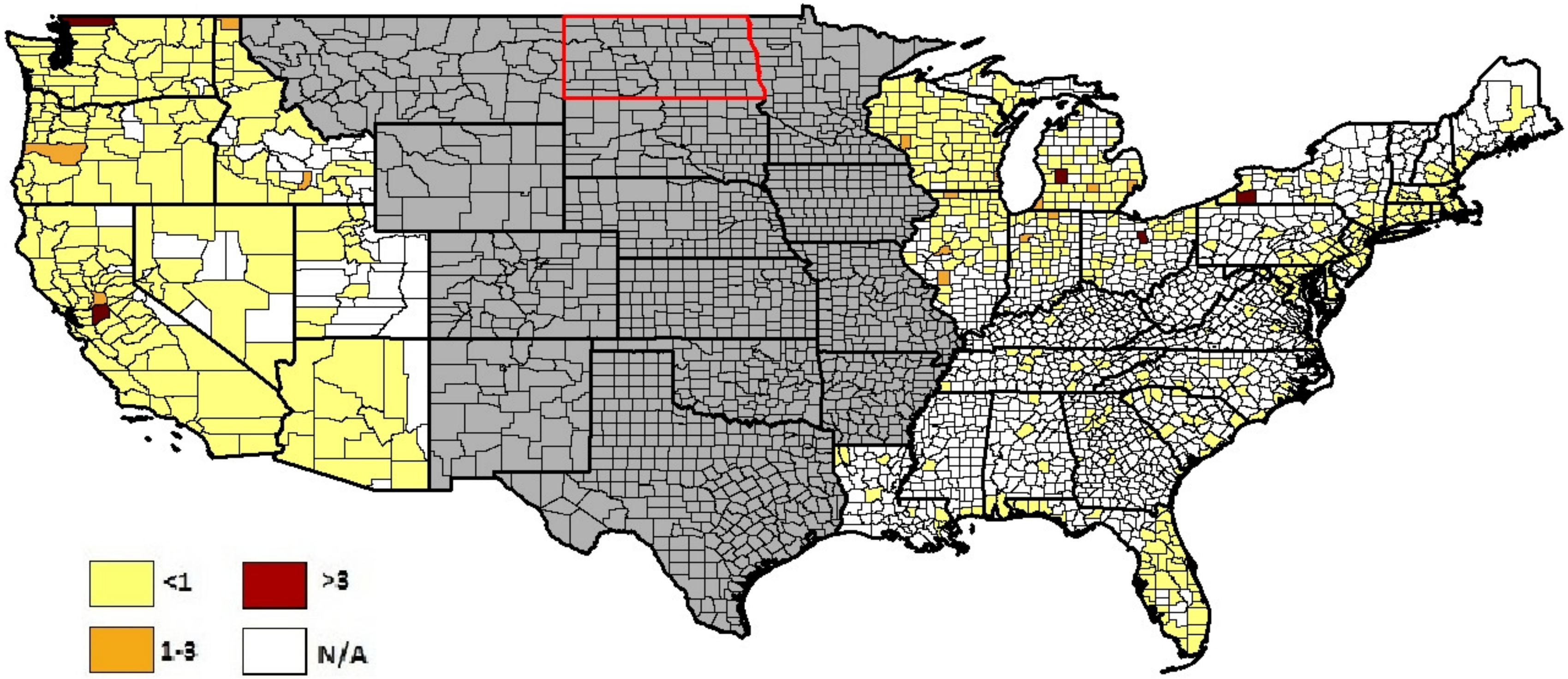
Figure 4: Spatial Distribution of Destination-Level Network Index Estimates, Mississippi-born Black Migrants



Notes: Figure displays destination-level network index estimates, $\hat{\Delta}_k$, across U.S. counties for Mississippi-born black migrants. The South is shaded in grey, with Mississippi outlined in red. Destinations to which less than 10 migrants moved are in white. Among all African American estimates, $\hat{\Delta}_k = 3$ corresponds to the 95th percentile, and $\hat{\Delta}_k = 1$ corresponds to the 81st percentile.

Source: Duke SSA/Medicare data

Figure 5: Spatial Distribution of Destination-Level Network Index Estimates, North Dakota-born White Migrants



Notes: See note to Figure 4. Among all Great Plains white estimates, $\hat{\Delta}_k = 3$ is greater than the 99th percentile, and $\hat{\Delta}_k = 1$ corresponds to the 98th percentile.
Source: Duke SSA/Medicare data

Appendices

A Derivation of Network Index

Appendix A derives the expression for the network index in equation (4).

First, recall the definition of the network index, $\Delta_{j,k} \equiv \mathbb{E}[N_{-i,j,k}|D_{i,j,k} = 1] - \mathbb{E}[N_{-i,j,k}|D_{i,j,k} = 0]$. Because $\mathbb{E}[N_{-i,j,k}|\cdot] = (N_j - 1)\,\mathbb{E}[D_{i',j,k}|\cdot]$ for $i' \neq i$, we can rewrite this as

$$\Delta_{j,k} = (N_j - 1)\left(\mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1] - \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 0]\right),\ i' \neq i. \tag{A.1}$$

The law of iterated expectations implies that the probability of moving from birth town j to destination k can be written

$$P_{j,k} = \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1]P_{j,k} + \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 0](1 - P_{j,k}). \tag{A.2}$$

Using the definition $\mu_{j,k} \equiv \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1]$ and rearranging equation (A.2) yields

$$\mathbb{E}[D_{i',j,k}|D_{i,j,k} = 0] = \frac{P_{j,k}(1 - \mu_{j,k})}{1 - P_{j,k}}. \tag{A.3}$$

Hence, we have

$$\begin{aligned}\mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1] - \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 0] &= \mu_{j,k} - \frac{P_{j,k}(1 - \mu_{j,k})}{1 - P_{j,k}} && \text{(A.4)}\\ &= \frac{\mu_{j,k} - P_{j,k}}{1 - P_{j,k}}. && \text{(A.5)}\end{aligned}$$

Substituting equation (A.5) into equation (A.1) yields

$$\Delta_{j,k} = (N_j - 1)\left(\frac{\mu_{j,k} - P_{j,k}}{1 - P_{j,k}}\right). \tag{A.6}$$

Applying the law of iterated expectations to the first term of the covariance of location decisions, $C_{j,k}$, yields

$$\begin{aligned}C_{j,k} &\equiv \mathbb{E}[D_{i',j,k}D_{i,j,k}] - \mathbb{E}[D_{i',j,k}]\,\mathbb{E}[D_{i,j,k}] && \text{(A.7)}\\ &= \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1]P_{j,k} - (P_{j,k})^2 && \text{(A.8)}\end{aligned}$$

Using the definition of $\mu_{j,k}$ and rearranging yields $\mu_{j,k} - P_{j,k} = C_{j,k}/P_{j,k}$. Substituting this expression into (A.6), and noting that Assumption 1 implies that $P_{j,k} = P_{g,k}$, yields equation (4).

B Generalized Method of Moments Formulation

B.1 Basic Model

As described in the text, we can derive the destination-level network index, Δ_k, in two ways: as a weighted average of $\Delta_{j,k}$ or by assuming that for each destination $\Delta_{j,k}$ is constant across birth towns within a birth state. Both approaches lead to the same point estimate of the destination-level network index, but the latter approach allows us to use GMM to estimate standard errors.

If we assume that the network index, $\Delta_{j,k}$, is constant across birth towns within a birth state, the destination-level network index, Δ_k, can be written

$$\Delta_k = \Delta_{j,k} = \frac{C_{j,k}(N_j - 1)}{P_{j,k} - P_{j,k}^2}. \tag{A.9}$$

It is useful to rewrite this as

$$\Delta_k \left(P_{j,k} - P_{j,k}^2\right) - C_{j,k}(N_j - 1) = 0. \tag{A.10}$$

To conduct inference, we treat the birth town group as the unit of observation. Aggregating across towns within a birth town group yields

$$\Delta_k Y_{g,k} - X_{g,k} = 0, \tag{A.11}$$

where

$$X_{g,k} \equiv \sum_{j \in g} C_{j,k}(N_j - 1) \tag{A.12}$$

$$Y_{g,k} \equiv \sum_{j \in g} P_{j,k} - P_{j,k}^2. \tag{A.13}$$

In the text, we describe how we construct our estimates $\widehat{P_{j,k}}$, $\widehat{P_{j,k}^2}$, and $\widehat{C_{j,k}}$. These estimates immediately lead to estimates $\widehat{X_{g,k}}$ and $\widehat{Y_{g,k}}$, which can be written as deviations from the underlying parameters,

$$\widehat{X_{g,k}} = X_{g,k} + u_{g,k}^X \tag{A.14}$$

$$\widehat{Y_{g,k}} = Y_{g,k} + u_{g,k}^Y. \tag{A.15}$$

This allows us to rewrite equation (A.11),

$$\Delta_k \widehat{Y_{g,k}} - \widehat{X_{g,k}} + (\Delta_k u_{g,k}^Y - u_{g,k}^X) = 0. \tag{A.16}$$

Because we have unbiased estimates of $P_{j,k}$, $P_{j,k}^2$, and $C_{j,k}$, we have unbiased estimates of $X_{g,k}$ and $Y_{g,k}$. This implies that

$$\mathbb{E}\left[\Delta_k \widehat{Y_{g,k}} - \widehat{X_{g,k}}\right] = 0. \tag{A.17}$$

Equation (A.17) is the basis of our GMM estimator. The sample analog is

$$\frac{1}{G}\sum_g \left(\widehat{\Delta_k}\widehat{Y_{g,k}} - \widehat{X_{g,k}}\right) = 0, \tag{A.18}$$

where G is the number of birth town groups in a state. This can be rewritten

$$\widehat{\Delta_k} = \frac{\sum_j \widehat{C_{j,k}}(N_j - 1)}{\sum_{j'} \widehat{P_{j',k}} - \widehat{P^2_{j',k}}}. \tag{A.19}$$

Equation (A.19) is identical to equation (8).

The above derivation is for a single destination-level network index, but can easily be expanded to consider all K destination-level network index parameters. The aggregated moment condition is

$$\mathbb{E}\begin{bmatrix} \Delta_1\widehat{Y_{g,1}} - \widehat{X_{g,1}} \\ \vdots \\ \Delta_K\widehat{Y_{g,K}} - \widehat{X_{g,K}} \end{bmatrix} \equiv \mathbb{E}\left[\boldsymbol{f}(\boldsymbol{w}_g, \boldsymbol{\Delta})\right] = \boldsymbol{0}, \tag{A.20}$$

where $\boldsymbol{w}_g$ is observed data used to construct $\widehat{\boldsymbol{X_g}}$ and $\widehat{\boldsymbol{Y_g}}$ and $\boldsymbol{\Delta} \equiv (\Delta_1, \ldots, \Delta_K)'$ is a $K \times 1$ vector of destination-level network index parameters.

Under standard conditions (e.g., Cameron and Trivedi, 2005), the asymptotic distribution of $\boldsymbol{\Delta}$ is

$$\sqrt{G}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}) \xrightarrow{d} \mathcal{N}\left[\boldsymbol{0}, \hat{\boldsymbol{F}}^{-1}\hat{\boldsymbol{S}}(\hat{\boldsymbol{F}}')^{-1}\right], \tag{A.21}$$

where

$$\hat{\boldsymbol{F}} = \frac{1}{G}\sum_g \left.\frac{\partial \boldsymbol{f}_g}{\partial \boldsymbol{\Delta}'}\right|_{\hat{\boldsymbol{\Delta}}} \tag{A.22}$$

$$= \frac{1}{G}\sum_g \begin{bmatrix} \widehat{Y_{g,1}} & 0 & 0 & \cdots & 0 \\ 0 & \widehat{Y_{g,2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \widehat{Y_{g,K}} \end{bmatrix} \tag{A.23}$$

and

$$\hat{\boldsymbol{S}} = \frac{1}{G}\sum_g \boldsymbol{f}(\boldsymbol{W}_g, \hat{\boldsymbol{\Delta}})\boldsymbol{f}(\boldsymbol{W}_g, \hat{\boldsymbol{\Delta}})'. \tag{A.24}$$

While it is convenient to describe the asymptotic properties when grouping all destinations together into $\boldsymbol{\Delta}$, we estimate each destination-level network index parameter Δ_k independently.

B.2 Comparing Estimates from Two Models

The GMM framework facilitates a comparison of estimates from different models. Under the null hypothesis we wish to test, we have two unbiased estimates for $X_{g,k}$ and $Y_{g,k}$:

$$\widehat{X^1_{g,k}} = X_{g,k} + u^X_{g,k} \tag{A.25}$$

$$\widehat{Y^1_{g,k}} = Y_{g,k} + u^Y_{g,k} \tag{A.26}$$

$$\widehat{X^2_{g,k}} = X_{g,k} + v^X_{g,k} \tag{A.27}$$

$$\widehat{Y^2_{g,k}} = Y_{g,k} + v^Y_{g,k}. \tag{A.28}$$

We estimate the unrestricted version of the model using GMM, for which the sample analog of the moment condition is

$$\frac{1}{G}\sum_g \begin{pmatrix} \widehat{\Delta^1_k}\widehat{Y^1_{g,k}} - \widehat{X^1_{g,k}} \\ \widehat{\Delta^2_k}\widehat{Y^2_{g,k}} - \widehat{X^2_{g,k}} \end{pmatrix} \tag{A.29}$$

This simply stacks the two estimates of the destination-level network index, Δ_k into a single, exactly-identified system.

Let $\Delta^1 \equiv N^{-1}\sum_k N_k \Delta_k$ be the migrant-weighted average of the destination-level network index parameters, where $N \equiv \sum_k N_k$ is the total number of migrants from a birth state. We are interested in testing whether $\Delta^1 = \Delta^2$. To test this hypothesis, we form the test statistic

$$\hat{t} = \frac{\widehat{\Delta^1} - \widehat{\Delta^2}}{\left(\widehat{\text{Var}}[\widehat{\Delta^1} - \widehat{\Delta^2}]\right)^{1/2}}. \tag{A.30}$$

Given destination-level network index estimates $\widehat{\Delta^1_k}$ and $\widehat{\Delta^2_k}$, it is straightforward to construct the averages $\widehat{\Delta^1}$ and $\widehat{\Delta^2}$. To estimate the variance in the denominator of the test statistic, we assume that destination-level network index estimates are independent of each other. Given the large number of sending birth towns, and the large number of destinations, we believe that the covariance between two destination-level network index estimates is likely small. Furthermore, we are not confident in our ability to reliably estimate the covariance of the covariances of location decisions, as would be necessary if we did not assume independence. Under the independence assumption, we can estimate $\widehat{\text{Var}}[\widehat{\Delta^1} - \widehat{\Delta^2}]$ as the appropriately weighted sum of

$$\widehat{\text{Var}}[\widehat{\Delta^1_k} - \widehat{\Delta^2_k}] = \widehat{\text{Var}}[\widehat{\Delta^1_k}] + \widehat{\text{Var}}[\widehat{\Delta^2_k}] - 2\widehat{\text{Cov}}[\widehat{\Delta^1_k}, \widehat{\Delta^2_k}] \tag{A.31}$$

which we obtain from the GMM variance estimate.

One issue with calculating this test statistic is that, when estimating the variance of our network index estimates under the extension in Section 3.4, we ignore the variance that arises because $\widetilde{P_{j,k}}$ and $\widetilde{P^2_{j,k}}$ rely on OLS estimates. We could account for this variance using a bootstrap, but the computational cost is very high, as it takes about 48 hours to construct the estimates in Column 3 of Table 4. Not accounting for this variance means that the p-values are too low. When pooling states for Southern black migrants, the p-value for the comparison between columns 1 and 2 of Table 4 is 0.33; it is 0.44 for columns 1 and 3. For Great Plains white migrants, the p-values are both 0.00.

C Estimating Cross-Group Network Indices

When estimating cross-group network indices, we are interested in the expected increase in the number of type b people from birth town j that move to destination county k when an arbitrarily chosen person i of type w is observed to make the same move,

$$\Delta_{j,k}^{b|w} \equiv \mathbb{E}[N_{j,k}^{b}|D_{i,j,k}^{w}=1]-\mathbb{E}[N_{j,k}^{b}|D_{i,j,k}^{w}=0]. \tag{A.32}$$

The steps described in Appendix A yield

$$\Delta_{j,k}^{b|w}=\frac{C_{j,k}^{b,w}N_{j}^{b}}{P_{j,k}^{w}(1-P_{j,k}^{w})}, \tag{A.33}$$

where $C_{j,k}^{b,w}$ is the covariance of location decisions between migrants of type b and w, N_{j}^{b} is the number of type b migrants born in j, and $P_{j,k}^{w}$ is the probability that a migrant of type w moves from j to k.

We estimate $P_{j,k}^{w}$ as described in the text. To estimate $C_{j,k}^{b,w}$, consider the model

$$D_{i,j(i),k}^{b}D_{i',j(i'),k}^{w}=\alpha_{g,k}+\sum_{j\in g}\beta_{j,k}^{b,w}1[j(i)=j(i')=j]+\epsilon_{i,i',k}. \tag{A.34}$$

This model is analogous to equation (1) in the text and yields the following covariance estimator,

$$\widehat{C_{j,k}^{b,w}}=\frac{N_{j,k}^{b}N_{j,k}^{w}}{N_{j}^{b}N_{j}^{w}}-\frac{\sum_{j\in g}\sum_{j'\neq j\in g}N_{j,k}^{b}N_{j',k}^{w}}{\sum_{j\in g}\sum_{j'\neq j\in g}N_{j}^{b}N_{j'}^{w}}. \tag{A.35}$$

We estimate the destination-level network index as

$$\hat{\Delta}_{k}^{b|w}=\sum_{j}\left(\frac{\widehat{P_{j,k}^{w}}-\widehat{(P_{j,k}^{w})^{2}}}{\sum_{j'}\widehat{P_{j',k}^{w}}-\widehat{(P_{j',k}^{w})^{2}}}\right)\hat{\Delta}_{j,k}^{b|w}. \tag{A.36}$$

We only estimate network indices for destinations which received at least ten black and white migrants from a given state. When calculating weighted averages of $\hat{\Delta}_{k}^{b|w}$, we use the number of type w individuals who moved to each destination.

D Addressing Measurement Error due to Incomplete Migration Data

Network index estimates depend on population flows observed in the Duke SSA/Medicare data, which is incomplete because some individuals die before enrolling in Medicare and some individuals' birth town information is unavailable. We first address the consequences of measurement error due to incomplete migration data under a missing at random assumption. If we observe a random sample of migration flows for each birth town-destination pair, then measurement error does not bias estimates of the covariance of location decisions, $C_{j,k}$, or moving probabilities, $P_{g,k}$. As a result, equation (4) implies that network index estimates will be attenuated because we undercount the number of migrants from each town, N_j.

More specifically, let N_j^* be the true number of migrants from birth town j that live to age 65, let α be the coverage rate, and assume that

$$N_j = \alpha N_j^*. \tag{A.37}$$

We approximate the coverage rate by dividing the number of individuals in the Duke SSA/Medicare data by the number of individuals in decennial census data who are born from 1916-1936 and survive to age 65.[61] The overall coverage rate is 64.9 percent for African Americans from the South and 82.2 percent for whites from the Great Plains (Appendix Table A.12), which implies that $N_j^* \approx 1.54N_j$ for Southern African Americans and $N_j^* \approx 1.22N_j$ for Great Plains whites. As an approximate measurement error correction, network index estimates should be multiplied by factors of 1.54 and 1.22 for Southern black and Great Plains white migrants. Appendix Table A.13 presents results that reflect state-specific coverage rate adjustments. The weighted average of destination-level network index estimates is 3.06 for Southern African Americans and 0.46 for Great Plains whites. Adjusting for incomplete data under a missing at random assumption increases both the magnitude of network index estimates and the black-white gap.

An alternative approach is to define N_j^* as the true number of migrants that live to a younger age, such as 40. Under this benchmark, coverage rates would be lower, and the estimates that adjust for measurement error would be larger. We do not focus on this alternative because, as described in the text, our data are best-suited for measuring long-run location decisions for individuals who survive to age 65.

Without making a missing at random (MAR) assumption, we can derive a lower bound on the network index and show that estimates of this lower bound still reveal sizable migration networks. As described in the text, the network index, $\Delta_{j,k}$, depends on the covariance of location decisions for migrants from birth town j to destination k, $C_{j,k}$, the probability of moving from birth town group g to destination k, $P_{g,k}$, and the number of migrants from town j, N_j. To focus on the key issues, suppose that we have an unbiased estimate of $P_{g,k}$ and consider the consequences of measurement error in $C_{j,k}$ and N_j. Let $\Delta_{j,k}^*$ and $C_{j,k}^*$ be the true values of the network index and covariance of location decisions. The true parameters are connected through the equation

$$\Delta_{j,k}^* = \frac{C_{j,k}^*(N_j^* - 1)}{P_{g,k} - P_{g,k}^2}. \tag{A.38}$$

Using the definition of covariance, it is straightforward to show that

$$C_{j,k}^* = \alpha^2 C_{j,k} + 2\alpha(1-\alpha)C_{j,k}^{\text{in, out}} + (1-\alpha)^2 C_{j,k}^{\text{out, out}}, \tag{A.39}$$

where $C_{j,k}$ is the covariance of location decisions between migrants who are in our data, $C_{j,k}^{\text{in, out}}$ is

---

[61] We use the 1990 Census to construct coverage rates for individuals born from 1916-1925 and the 2000 Census for individuals born from 1926-1935.

the average covariance of location decisions between a migrant who is in our data and a migrant who is not, and $C_{j,k}^{\text{out, out}}$ is the average covariance of location decisions between migrants who are not in our data.

When not assuming that data are MAR, the covariance of location decisions among migrants not in our data ($C_{j,k}^{\text{in, out}}$ and $C_{j,k}^{\text{out, out}}$) could differ from the covariance of location decisions between migrants who are in our data ($C_{j,k}$). As a result, the network index based on our data, $\Delta_{j,k}$, might not simply be attenuated, as implied by the MAR assumption. In general, we cannot point identify the network index under this more general measurement error model. However, we can construct a lower bound for the strength of migration networks. In particular, we make the extreme assumptions that there are no interactions between migrants in and out of our data, so that $C_{j,k}^{\text{in, out}} = 0$, and that there are no interactions between migrants out of our data, so that $C_{j,k}^{\text{out, out}} = 0$. In this case, equations (A.37), (A.38), and (A.39) imply that

$$\Delta_{j,k}^{*} \geq \alpha \Delta_{j,k}, \tag{A.40}$$

so that we can estimate a lower bound on the true network index by multiplying the estimated network index by the coverage rate.[62] Combining the average coverage rates (64.9 and 82.2 percent) with the average destination-level network index estimates from Table 3, we estimate a lower bound for the network index of 1.26 for African Americans and 0.31 for whites. These lower bounds, which depend on extremely conservative assumptions about the migration of individuals not in our data, still reveal sizable networks, especially among African Americans.

E Differences in Family Size and the Black-White Gap

Appendix E provides a more detailed discussion of whether differences in family size explain the black-white network index gap.

To explore this issue, we decompose the network index into a component for migrants from the same family, Δ^{fam}, and a component for migrants not from the same family, Δ^{not}. To examine the importance of differences in family size, we assume that black and white network indices differ only because of differences in family size. Then we have

$$\Delta_b = \Delta^{fam} P_b^{fam} + \Delta^{not}(1 - P_b^{fam}) \tag{A.41}$$

$$\Delta_w = \Delta^{fam} P_w^{fam} + \Delta^{not}(1 - P_w^{fam}), \tag{A.42}$$

where Δ_b is the network index among black migrants, and P_b^{fam} is the probability that two ran-

---

[62] Proof: If $C_{j,k}^{\text{in, out}} = C_{j,k}^{\text{out, out}} = 0$, equations (A.37), (A.38), and (A.39) imply

$$\begin{aligned} \Delta_{j,k}^{*} &= \frac{\alpha^2 C_{j,k}\left(\frac{N_j}{\alpha} - 1\right)}{P_{g,k} - P_{g,k}^2} \\ &\geq \frac{\alpha^2 C_{j,k}\left(\frac{N_j}{\alpha} - \frac{1}{\alpha}\right)}{P_{g,k} - P_{g,k}^2} = \alpha \Delta_{j,k}, \end{aligned}$$

where the inequality comes from noting that $\alpha \in [0, 1]$ and assuming $C_{j,k} \geq 0$, and the final equality comes from equation (4) in the text. One could also construct upper bounds, but these are not particularly informative.

domly chosen black migrants are from the same family. Δ_w and P_w^{fam} are defined analogously. The black-white network index gap is

$$\Delta_b - \Delta_w = (\Delta^{fam} - \Delta^{not})(P_b^{fam} - P_w^{fam}). \tag{A.43}$$

Our data do not allow us to estimate $\Delta^{fam} - \Delta^{not}$, but we can use equation (A.43), along with estimates of $\Delta_b - \Delta_w$ and $P_b^{fam} - P_w^{fam}$ to explore whether it is reasonable to conclude that differences in family size explain the black-white gap. As described in the text, our average network indices for black and white migrants are 1.94 and 0.38. In the 1940 Census, the average within-household family size for individuals born from 1916-1936 is 6.16 for African Americans from the South and 5.25 for whites from the Great Plains. In the Duke SSA/Medicare dataset, there are 142 black migrants and 181 white migrants per town. However, as discussed in the text, the Duke data undercount the total number of migrants. If we inflate the migrant counts by 1.54 and 1.22, then we estimate 219 black migrants and 221 white migrants per town. Combining the Census family size estimates with the adjusted Duke migrant estimates, we have $P_b^{fam} = 6.16/219 = 0.028$ and $P_w^{fam} = 5.25/221 = 0.024$. With these estimates, $\Delta^{fam} - \Delta^{not}$ would have to equal 520 ($= 1.56/0.003$) people for differences in family size to fully explain the black-white gap. This is clearly implausible.

To construct an upper bound on the probability that two randomly chosen migrants are from the same family, we use the 100 percent sample of the 1940 Census to count the number of individuals in a county born from 1916-1936 with the same last name (Minnesota Population Center and Ancestry.com, 2013). On average, there are 54.5 African Americans with the same last name and 14.7 whites with the same last name. Using these numbers in the numerator leads to estimates of $P_b^{fam} = 54.5/219 = 0.249$ and $P_w^{fam} = 14.7/221 = 0.067$. In this case, $\Delta^{fam} - \Delta^{not}$ would have to equal 8.57 ($= 1.56/0.182$) people for differences in family size to fully explain the black-white gap. This approach considerably overestimates the extent of family connections, because many individuals with the same last name are not related, and we use counties, instead of towns, as the geographic unit in the numerator of P^{fam}. Even still, this gap seems too large to us. In sum, differences in family size might explain some, but not all, of the differences in migration networks between black and white migrants.

F Calculating County-Specific Relative Wages

Appendix F provides details on how we calculate county-specific relative wages, which we use as a correlate of destination network strength.

Consider the following model for the log hourly wage of individual i,

$$\ln(w_i) = X_i\theta_r + \phi_{r,c} + \epsilon_i, \tag{A.44}$$

where X_i is a vector of observed covariates—a constant term plus indicators for detailed educational attainment (of which there are 23), age, marital status (married or not), birth state, and birth state-by-age—and $\phi_{r,c}$ is a race-specific county fixed effect. We define the black-white relative wage in county c as

$$RW_{bw,c} = \phi_{b,c} - \phi_{w,c} + \bar{X}_b(\theta_b - \theta_w) \tag{A.45}$$

where $\bar{X}_b$ is the mode of X_i among black individuals.[63] A higher value of the black-white relative wage indicates less discrimination in county c. To construct the black-white relative wage, we estimate equation (A.44) on the sample of U.S. born men age 16–64 who are a wage/salary worker and have at least 26 weeks of work in the prior year in the 1940 complete count Census data. 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.

---

[63] We use the mode instead of the mean because the variables in X_i are categorical. This only affects the mean of the relative wage, and so our results would be identical from any other choice of $\bar{X}$.

Table A.1: Industry of Migrants and Non-Migrants, Southern Blacks and Great Plains Whites, 1950

	Percent of Group Working in Industry			
	Southern Black		Great Plains White	
	Migrants (1)	Non-Migrants (2)	Migrants (3)	Non-Migrants (4)
Agriculture, Forestry, and Fishing	1.23%	35.92%	9.38%	31.60%
Mining	1.33%	1.21%	2.02%	3.65%
Construction	10.19%	8.12%	11.98%	9.14%
Manufacturing	37.87%	22.09%	23.79%	10.98%
Transportation, Communication, and Other Utilities	11.80%	7.89%	9.58%	9.59%
Wholesale and Retail Trade	13.61%	10.46%	16.47%	16.87%
Finance, Insurance, and Real Estate	2.21%	0.78%	2.39%	2.20%
Business and Repair Services	2.98%	1.67%	4.11%	3.49%
Personal Services	6.30%	5.24%	2.16%	1.83%
Entertainment and Recreation Services	1.03%	0.63%	1.15%	0.76%
Professional and Related Services	3.95%	3.31%	5.67%	4.27%
Public Administration	6.57%	2.33%	11.08%	5.17%
Other	0.92%	0.35%	0.22%	0.43%

Note: Sample contains currently employed males, age 20-60 in the 1950 Census.
Source: Ruggles et al. (2019)

Table A.2: Size of Birth Town Groups Chosen by Cross Validation

Birth State	(1)
Panel A: Southern Black Migrants	
Alabama	52
Florida	138
Georgia	40
Louisiana	48
Mississippi	42
North Carolina	52
South Carolina	30
Panel B: Great Plains White Migrants	
Kansas	128
Nebraska	128
North Dakota	84
Oklahoma	68
South Dakota	112
Panel C: Southern White Migrants	
Alabama	156
Florida	270
Georgia	168
Louisiana	136
Mississippi	170
North Carolina	50
South Carolina	266

Notes: Column 1 displays the results of a cross validation procedure that chooses the length of the square grid used to define birth town groups. See text for details.
Source: Duke SSA/Medicare data

Table A.3: Number of Birth Towns and Migrants, by Birth State

Birth State	Birth Towns (1)	Migrants (2)	Migrants Per Town (3)
Panel A: Southern Black Migrants			
Alabama	693	96,269	138.9
Florida	203	19,158	94.4
Georgia	566	77,038	136.1
Louisiana	460	55,974	121.7
Mississippi	660	120,454	182.5
North Carolina	586	78,420	133.8
South Carolina	461	69,399	150.5
All States	3,629	516,712	142.4
Panel B: Great Plains White Migrants			
Kansas	883	139,374	157.8
Nebraska	643	134,011	208.4
North Dakota	592	92,205	155.8
Oklahoma	966	200,392	207.4
South Dakota	474	78,541	165.7
All States	3,558	644,523	181.1

Notes: Sample limited to towns with at least 10 migrants in the data.
Source: Duke SSA/Medicare data

Table A.4: Characteristics of Migrants and Non-Migrants from Linked Census Data, African Americans from South

	1920–1930 Censuses		1930–1940 Censuses		p-value, column differences		
	Non-Migrants (1)	Migrants (2)	Non-Migrants (3)	Migrants (4)	(1) – (2) (5)	(3) – (4) (6)	(2) – (4) (7)
Individual and family characteristics							
Attending school (age 6-17)	64.6	63.9	73.8	69.8	0.10	0.00	0.00
Literate (age 10+)	71.0	77.3	80.4	83.6	0.00	0.00	0.00
Age in 1920/1930	12.9	14.4	13.5	14.8	0.00	0.00	0.00
Father’s age in 1920/1930	45.7	46.7	45.5	46.9	0.00	0.00	0.06
Mother’s age in 1920/1930	38.7	39.9	39.3	40.6	0.00	0.00	0.00
Parent present	87.8	84.2	85.6	80.8	0.00	0.00	0.00
Parent literate	71.3	73.8	81.5	81.9	0.00	0.26	0.00
Owner-occupied housing	23.3	25.4	25.2	29.0	0.00	0.00	0.00
Number of siblings	3.6	3.3	3.5	3.0	0.00	0.00	0.00
Father occupation: professional	1.0	1.7	1.5	2.5	0.00	0.00	0.00
Father occupation: farmer	74.4	71.2	66.8	61.8	0.00	0.00	0.00
Father occupation: clerical or sales	0.2	0.5	0.5	0.8	0.00	0.00	0.00
Father occupation: craftsmen or operative	4.8	7.4	7.0	9.5	0.00	0.00	0.00
Father occupation: laborer or service worker	11.2	12.6	16.0	18.2	0.00	0.00	0.00
Father occupation: farm laborer	8.4	6.6	8.1	7.2	0.00	0.00	0.07
Town/city characteristics:							
Not in city	81.4	74.1	75.4	68.5	0.00	0.00	0.00
City population $\leq$ 25,000	13.1	17.4	14.6	19.1	0.00	0.00	0.00
City population $>$ 25,000	5.5	8.5	10.0	12.4	0.00	0.00	0.00

1910 County characteristics:							
Percent black	53.5	53.2	52.0	52.7	0.02	0.00	0.01
Percent of farmers who are black	50.3	50.3	48.4	49.5	0.78	0.00	0.00
Percent of black farmers who are owners	25.3	24.4	26.7	25.7	0.00	0.00	0.00
Percent of white farmers who are owners	58.3	57.2	58.9	58.2	0.00	0.00	0.00
Percent of farm acres in cotton	17.0	17.7	15.9	16.9	0.00	0.00	0.00
Percent of crop value in cotton	44.4	45.8	42.1	44.3	0.00	0.00	0.00
Black literacy rate (age 10+)	62.1	63.7	62.7	64.2	0.00	0.00	0.00
White literacy rate (age 10+)	92.0	92.5	91.8	92.5	0.00	0.00	0.37
Black school attendance rate (age 6-14)	55.3	57.3	55.6	57.7	0.00	0.00	0.00
White school attendance rate (age 6-14)	75.7	75.9	75.5	75.8	0.01	0.00	0.41
Black population density	34.7	37.7	39.4	39.1	0.00	0.62	0.02
White population density	44.0	46.0	57.3	50.0	0.07	0.00	0.01
Number of individuals	109,851	19,881	139,363	15,391			

Notes: Table reports averages of indicated variables. The 1920–1930 sample contains black men age 18-30 in 1930 who are born in Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, or South Carolina and can be matched to the 1920 Census. Migrants are individuals who live outside the former Confederate states in 1930. The 1930–1940 sample uses the same age range and definitions. Sources: Minnesota Population Center and Ancestry.com (2013), Haines and ICPSR (2010)

Table A.5: Characteristics of Migrants and Non-Migrants from Linked Census Data, Whites from Great Plains

	1920–1930 Censuses		1930–1940 Censuses		p-value, column differences		
	Non-Migrants (1)	Migrants (2)	Non-Migrants (3)	Migrants (4)	(1) – (2) (5)	(3) – (4) (6)	(2) – (4) (7)
Individual and family characteristics							
Attending school (age 6-17)	86.0	85.4	89.0	87.9	0.04	0.00	0.00
Literate (age 10+)	99.3	99.6	99.5	99.6	0.00	0.05	0.66
Age in 1920/1930	12.9	13.8	13.5	14.3	0.00	0.00	0.00
Father's age in 1920/1930	46.2	47.0	46.1	46.5	0.00	0.00	0.00
Mother's age in 1920/1930	41.0	41.9	41.5	41.9	0.00	0.00	0.44
Parent present	96.4	93.9	95.2	92.7	0.00	0.00	0.00
Parent literate	98.0	97.8	98.2	98.2	0.03	0.42	0.00
Owner-occupied housing	67.1	62.6	57.5	51.5	0.00	0.00	0.00
Number of siblings	3.4	3.0	3.0	2.9	0.00	0.00	0.00
Father occupation: professional	8.7	14.5	10.1	11.5	0.00	0.00	0.00
Father occupation: farmer	68.5	52.2	59.9	53.4	0.00	0.00	0.03
Father occupation: clerical or sales	3.9	6.2	5.3	6.5	0.00	0.00	0.33
Father occupation: craftsmen or operative	10.9	17.8	13.9	16.4	0.00	0.00	0.00
Father occupation: laborer or service worker	5.3	6.4	7.9	8.7	0.00	0.00	0.00
Father occupation: farm laborer	2.6	2.7	2.9	3.4	0.26	0.00	0.00
Town/city characteristics:							
Not in city	64.8	48.9	61.1	55.4	0.00	0.00	0.00
City population $\leq$ 25,000	28.4	41.7	28.8	34.7	0.00	0.00	0.00
City population $>$ 25,000	6.7	9.4	10.1	9.8	0.00	0.09	0.12

1910 County characteristics:							
Percent black	2.7	2.9	3.0	3.2	0.00	0.00	0.00
Percent of white farmers who are owners	58.7	59.1	58.9	58.5	0.00	0.00	0.00
Percent of farm acres in cotton	2.0	1.9	2.1	2.5	0.34	0.00	0.00
Percent of crop value in cotton	6.3	6.3	6.7	8.1	0.81	0.00	0.00
White literacy rate (age 10+)	97.7	97.6	97.7	97.6	0.00	0.00	0.92
White school attendance rate (age 6-14)	86.2	86.0	85.8	85.5	0.00	0.00	0.00
Black population density	2.6	2.7	2.8	2.4	0.05	0.00	0.00
White population density	45.0	50.1	47.2	41.6	0.00	0.00	0.00
Number of individuals	188,700	19,620	215,457	34,893			

Notes: Table reports averages of indicated variables. The 1920–1930 sample contains white men age 18–30 in 1930 who are born in Kansas, Nebraska, Oklahoma, North Dakota, or South Dakota and can be matched to the 1920 Census. Migrants are individuals who live outside these states plus the light grey border region in Figure 2. The 1930–1940 sample uses the same age range and definitions. Sources: Minnesota Population Center and Ancestry.com (2013), Haines and ICPSR (2010)

Table A.6: Characteristics of Migrants and Non-Migrants from Linked Census Data, Whites from South

	1920–1930 Censuses		1930–1940 Censuses		p-value, column differences		
	Non-Migrants (1)	Migrants (2)	Non-Migrants (3)	Migrants (4)	(1) – (2) (5)	(3) – (4) (6)	(2) – (4) (7)
Individual and family characteristics							
Attending school (age 6-17)	82.0	82.2	84.5	85.1	0.68	0.12	0.00
Literate (age 10+)	95.0	97.3	97.0	98.3	0.00	0.00	0.00
Age in 1920/1930	13.1	14.2	13.6	14.4	0.00	0.00	0.00
Father's age in 1920/1930	45.5	46.8	45.8	46.8	0.00	0.00	0.95
Mother's age in 1920/1930	40.4	41.3	40.9	41.6	0.00	0.00	0.01
Parent present	94.9	92.5	93.8	91.4	0.00	0.00	0.00
Parent literate	93.5	95.3	95.4	96.2	0.00	0.00	0.00
Owner-occupied housing	56.6	58.4	49.7	51.8	0.00	0.00	0.00
Number of siblings	3.6	3.4	3.3	3.0	0.00	0.00	0.00
Father occupation: professional	9.0	12.9	10.7	15.5	0.00	0.00	0.00
Father occupation: farmer	66.0	57.1	54.8	46.6	0.00	0.00	0.00
Father occupation: clerical or sales	4.1	5.8	6.1	8.5	0.00	0.00	0.00
Father occupation: craftsmen or operative	12.8	16.8	18.1	20.3	0.00	0.00	0.00
Father occupation: laborer or service worker	5.3	5.1	7.3	7.0	0.43	0.34	0.00
Father occupation: farm laborer	2.8	2.3	3.0	2.1	0.01	0.00	0.25
Town/city characteristics:							
Not in city	70.1	60.5	66.0	56.5	0.00	0.00	0.00
City population $\leq$ 25,000	20.7	26.6	21.2	26.1	0.00	0.00	0.33
City population $>$ 25,000	9.2	12.9	12.8	17.5	0.00	0.00	0.00

1910 County characteristics:							
Percent black	35.7	35.1	35.7	35.8	0.00	0.88	0.00
Percent of farmers who are black	30.1	29.6	29.8	30.0	0.01	0.49	0.23
Percent of black farmers who are owners	33.7	36.1	34.6	36.7	0.00	0.00	0.04
Percent of white farmers who are owners	60.7	61.8	61.2	62.6	0.00	0.00	0.00
Percent of farm acres in cotton	11.6	11.0	11.1	10.9	0.00	0.00	0.41
Percent of crop value in cotton	34.8	32.6	33.5	32.4	0.00	0.00	0.45
Black literacy rate (age 10+)	65.3	66.8	65.7	67.0	0.00	0.00	0.06
White literacy rate (age 10+)	89.9	90.6	90.0	90.9	0.00	0.00	0.00
Black school attendance rate (age 6-14)	56.0	57.8	56.2	57.9	0.00	0.00	0.34
White school attendance rate (age 6-14)	74.3	74.9	74.1	74.8	0.00	0.00	0.53
Black population density	37.7	39.7	39.4	38.7	0.01	0.36	0.29
White population density	80.4	81.4	83.8	77.4	0.62	0.00	0.13
Number of individuals	293,678	14,167	355,197	12,004			

Notes: Table reports averages of indicated variables. The 1920–1930 sample contains white men age 18–30 in 1930 who are born in Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, or South Carolina and can be matched to the 1920 Census. Migrants are individuals who live outside the former Confederate states in 1930. The 1930–1940 sample uses the same age range and definitions. Sources: Minnesota Population Center and Ancestry.com (2013), Haines and ICPSR (2010)

Table A.7: Average Network Index Estimates, Southern White Migrants

Birth State	Number of Migrants (1)	Unweighted Average (2)	Weighted Average (3)
Alabama	43,157	0.204	0.516
		(0.014)	(0.052)
Florida	27,426	0.046	0.072
		(0.006)	(0.100)
Georgia	31,299	0.082	0.117
		(0.007)	(0.021)
Louisiana	31,303	0.122	0.269
		(0.011)	(0.071)
Mississippi	28,001	0.118	0.186
		(0.010)	(0.021)
North Carolina	47,146	0.179	0.412
		(0.012)	(0.040)
South Carolina	14,605	0.068	0.094
		(0.005)	(0.029)
All States	222,937	0.131	0.280
		(0.004)	(0.021)

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 in parentheses.
Source: Duke SSA/Medicare data

Table A.8: Average Network Index Estimates, By Size of Birth Town and Destination, Southern White Migrants

Exclude Largest Birth Towns: Exclude Largest Destinations: Birth State	No No (1)	Yes No (2)	No Yes (3)	Yes Yes (4)
Alabama	0.516	0.458	0.531	0.481
	(0.052)	(0.045)	(0.071)	(0.062)
Florida	0.072	0.074	0.134	0.030
	(0.100)	(0.012)	(0.082)	(0.009)
Georgia	0.117	0.101	0.119	0.088
	(0.021)	(0.012)	(0.019)	(0.013)
Louisiana	0.269	0.207	0.198	0.143
	(0.071)	(0.022)	(0.035)	(0.017)
Mississippi	0.186	0.185	0.135	0.134
	(0.021)	(0.022)	(0.013)	(0.013)
North Carolina	0.412	0.395	0.337	0.319
	(0.040)	(0.037)	(0.040)	(0.034)
South Carolina	0.094	0.090	0.058	0.055
	(0.029)	(0.023)	(0.013)	(0.012)
All States	0.280	0.254	0.262	0.223
	(0.021)	(0.013)	(0.021)	(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 D.C., San Francisco, Pittsburgh, and St. Louis, in addition to counties which received fewer than 10 migrants. Column 4 excludes large birth towns and large destinations. Birth town groups are defined by cross validation. Standard errors are in parentheses.
Source: Duke SSA/Medicare data

Table A.9: Average Cross-Race Network Index Estimates, Southern Black and White Migrants

Birth State	All Counties (1)	Excluding Largest CMSAs (2)
Panel A: Black Migrants Induced to Location by White Migrant		
Alabama	0.188	0.130
	(0.106)	(0.150)
Florida	0.026	0.005
	(0.059)	(0.036)
Georgia	-0.028	0.040
	(0.039)	(0.044)
Louisiana	-0.066	0.068
	(0.196)	(0.038)
Mississippi	0.246	0.049
	(0.185)	(0.033)
North Carolina	-0.010	-0.005
	(0.062)	(0.011)
South Carolina	0.197	-0.025
	(0.161)	(0.027)
All States	0.071	0.050
	(0.048)	(0.033)
Panel B: White Migrants Induced to Location by Black Migrant		
Alabama	0.052	0.038
	(0.048)	(0.042)
Florida	0.047	-0.018
	(0.064)	(0.036)
Georgia	-0.020	0.004
	(0.014)	(0.014)
Louisiana	-0.137	0.016
	(0.066)	(0.017)
Mississippi	-0.056	0.020
	(0.030)	(0.011)
North Carolina	0.021	-0.002
	(0.029)	(0.022)
South Carolina	-0.019	0.020
	(0.013)	(0.018)
All States	-0.019	0.019
	(0.015)	(0.013)

Notes: Table A.9 contains weighted averages of cross-race destination-level network index estimates. Birth town groups are defined by cross validation. Standard errors in parentheses.

Source: Duke SSA/Medicare data

Table A.10: Average Network Index Estimates, Birth Town Groups Defined by Cross Validation and Counties

Birth State	Cross Validation Unweighted (1)	Cross Validation Weighted (2)	Counties Unweighted (3)	Counties Weighted (4)
Panel A: Southern Black Migrants				
Alabama	0.770	1.888	0.616	1.393
	(0.049)	(0.195)	(0.034)	(0.170)
Florida	0.536	0.813	0.597	0.811
	(0.052)	(0.117)	(0.087)	(0.317)
Georgia	0.735	1.657	0.544	0.887
	(0.048)	(0.177)	(0.039)	(0.279)
Louisiana	0.462	1.723	0.399	2.209
	(0.039)	(0.478)	(0.039)	(0.920)
Mississippi	0.901	2.303	0.742	2.166
	(0.050)	(0.313)	(0.051)	(0.401)
North Carolina	0.566	1.539	0.402	1.022
	(0.039)	(0.130)	(0.028)	(0.123)
South Carolina	0.874	2.618	0.774	2.132
	(0.054)	(0.301)	(0.049)	(0.224)
All States	0.736	1.938	0.599	1.608
	(0.020)	(0.110)	(0.017)	(0.151)
Panel B: Great Plains White Migrants				
Kansas	0.128	0.255	0.106	0.194
	(0.007)	(0.024)	(0.008)	(0.028)
North Dakota	0.174	0.464	0.156	0.385
	(0.012)	(0.036)	(0.010)	(0.029)
Nebraska	0.141	0.361	0.121	0.399
	(0.008)	(0.082)	(0.009)	(0.117)
Oklahoma	0.112	0.453	0.102	0.372
	(0.008)	(0.036)	(0.007)	(0.036)
South Dakota	0.163	0.350	0.135	0.273
	(0.009)	(0.026)	(0.008)	(0.027)
All States	0.137	0.380	0.119	0.329
	(0.004)	(0.022)	(0.004)	(0.028)

Notes: Columns 1 and 3 are unweighted averages of destination-level network index estimates, $\hat{\Delta}_k$. Columns 2 and 4 are weighted averages, where the weights are the number of people who move from each state to destination k. In columns 1 and 2, we define birth town groups using cross validation, as described in the text. In columns 3 and 4, we use counties. Standard errors in parentheses.

Source: Duke SSA/Medicare data

Table A.11: Average Network Index Estimates, Birth Town Groups Based on Different Grid Sizes

	Weighted Average			Unweighted Average		
Grid Size:	50	100	200	50	100	200
Birth State	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Southern Black Migrants						
Alabama	1.869	2.256	2.398	0.759	0.846	0.913
	(0.203)	(0.198)	(0.196)	(0.046)	(0.046)	(0.045)
Florida	0.919	0.856	0.944	0.595	0.553	0.560
	(0.196)	(0.117)	(0.117)	(0.158)	(0.087)	(0.055)
Georgia	1.760	2.190	2.421	0.780	0.859	0.916
	(0.163)	(0.185)	(0.168)	(0.055)	(0.053)	(0.049)
Louisiana	1.887	2.097	2.660	0.469	0.508	0.549
	(0.542)	(0.507)	(0.717)	(0.038)	(0.034)	(0.035)
Mississippi	2.432	2.778	3.216	0.910	1.001	1.056
	(0.327)	(0.270)	(0.217)	(0.049)	(0.048)	(0.042)
North Carolina	1.557	1.719	1.877	0.566	0.629	0.678
	(0.133)	(0.149)	(0.139)	(0.041)	(0.043)	(0.037)
South Carolina	3.255	3.620	4.080	0.982	1.074	1.156
	(0.380)	(0.348)	(0.280)	(0.054)	(0.052)	(0.045)
All states	2.090	2.401	2.713	0.761	0.834	0.891
	(0.120)	(0.109)	(0.112)	(0.020)	(0.019)	(0.017)
Panel B: Great Plains White Migrants						
Kansas	0.256	0.256	0.253	0.122	0.127	0.130
	(0.028)	(0.026)	(0.019)	(0.010)	(0.008)	(0.006)
Nebraska	0.366	0.373	0.379	0.130	0.142	0.146
	(0.090)	(0.079)	(0.062)	(0.008)	(0.008)	(0.008)
North Dakota	0.424	0.490	0.529	0.164	0.177	0.186
	(0.032)	(0.037)	(0.038)	(0.011)	(0.011)	(0.011)
Oklahoma	0.425	0.488	0.514	0.107	0.115	0.119
	(0.038)	(0.038)	(0.034)	(0.008)	(0.008)	(0.007)
South Dakota	0.291	0.343	0.365	0.149	0.162	0.169
	(0.024)	(0.024)	(0.026)	(0.011)	(0.010)	(0.009)
All states	0.360	0.396	0.413	0.128	0.138	0.143
	(0.024)	(0.022)	(0.018)	(0.004)	(0.004)	(0.004)

Notes: Columns 1–3 are weighted averages of destination-level network index estimates, $\widehat{\Delta}_k$, where the weights are the number of people who move from each state to destination k. Columns 4–6 are unweighted averages. We define birth town groups as square grids, with the length of each square varying from 50 to 200 miles. Standard errors in parentheses.
Source: Duke SSA/Medicare data

Table A.12: Coverage Rates, Duke SSA/Medicare Dataset

Sample:	All	All	All	Men	Women	Cohort 1916–25	Cohort 1926–36
Birth State	Duke/SSA coverage rate, all (1)	Duke/SSA percent with town identified (2)	Duke/SSA coverage rate, town identified (3)	Duke/SSA coverage rate, town identified (4)	Duke/SSA coverage rate, town identified (5)	Duke/SSA coverage rate, town identified (6)	Duke/SSA coverage rate, town identified (7)
Panel A: Southern Black Individuals							
Alabama	86.4%	78.7%	68.0%	73.1%	64.6%	65.1%	70.4%
Florida	82.7%	83.6%	69.2%	72.3%	66.9%	65.5%	72.1%
Georgia	85.0%	73.1%	62.2%	65.2%	60.2%	57.0%	67.5%
Louisiana	85.2%	84.5%	72.0%	74.3%	70.3%	67.5%	76.0%
Mississippi	88.9%	74.7%	66.4%	69.7%	64.1%	63.9%	68.6%
North Carolina	88.5%	72.5%	64.2%	64.6%	63.9%	61.5%	66.5%
South Carolina	90.8%	61.9%	56.2%	57.3%	55.5%	53.6%	58.7%
All States	87.2%	74.4%	64.9%	67.6%	63.1%	61.3%	68.1%
Panel B: Great Plains White Individuals							
Kansas	88.1%	92.5%	81.5%	84.8%	78.6%	78.6%	84.4%
Nebraska	89.2%	93.3%	83.2%	87.5%	79.6%	80.8%	85.7%
North Dakota	88.1%	89.7%	79.0%	81.9%	76.7%	74.3%	84.0%
Oklahoma	93.1%	89.9%	83.7%	86.0%	81.8%	79.4%	87.7%
South Dakota	88.9%	91.2%	81.1%	82.6%	79.8%	78.7%	83.5%
All States	90.1%	91.3%	82.2%	85.2%	79.8%	78.8%	85.6%

Notes: Column 1 reports the number of individuals in the Duke SSA/Medicare dataset divided by the number of individuals in the 1990/2000 Census. Column 2 reports the share of individuals in the Duke SSA/Medicare dataset for whom birth town and destination county are identified. Columns 3–7 reports the number of individuals in the Duke SSA/Medicare dataset for whom birth town and destination county are identified divided by the number of individuals in the 1990/2000 Census. We use the 1990 Census for individuals born from 1916–1925 and the 2000 Census for individuals born from 1926–1936. The sample includes individuals living inside and outside their birth region.

Source: Duke SSA/Medicare data and Ruggles et al. (2019)

Table A.13: Average Network Index Estimates, Adjusted for Incomplete Migration Data

Sample: Birth State	All (1)	Men (2)	Women (3)	1916–25 Cohort (4)	1926–36 Cohort (5)
Panel A: Southern Black Migrants					
Alabama	2.776	1.160	1.621	1.274	1.672
	(0.287)	(0.120)	(0.175)	(0.145)	(0.165)
Florida	1.175	0.533	0.633	0.454	0.780
	(0.170)	(0.085)	(0.128)	(0.102)	(0.123)
Georgia	2.664	0.959	1.722	1.570	1.287
	(0.284)	(0.111)	(0.205)	(0.213)	(0.118)
Louisiana	2.393	1.184	0.991	0.973	1.676
	(0.664)	(0.316)	(0.389)	(0.225)	(0.573)
Mississippi	3.468	1.456	2.043	1.396	2.236
	(0.471)	(0.202)	(0.297)	(0.211)	(0.307)
North Carolina	2.398	1.029	1.404	1.215	1.326
	(0.203)	(0.100)	(0.121)	(0.115)	(0.109)
South Carolina	4.659	1.935	2.761	2.363	2.478
	(0.535)	(0.199)	(0.381)	(0.310)	(0.260)
All States	3.057	1.271	1.790	1.464	1.792
	(0.167)	(0.071)	(0.109)	(0.086)	(0.108)
Panel B: Great Plains White Migrants					
Kansas	0.313	0.151	0.178	0.204	0.167
	(0.029)	(0.016)	(0.017)	(0.020)	(0.014)
Nebraska	0.433	0.176	0.256	0.270	0.234
	(0.098)	(0.050)	(0.050)	(0.057)	(0.046)
North Dakota	0.587	0.250	0.338	0.387	0.277
	(0.046)	(0.021)	(0.031)	(0.032)	(0.021)
Oklahoma	0.541	0.250	0.291	0.285	0.319
	(0.043)	(0.023)	(0.023)	(0.025)	(0.026)
South Dakota	0.431	0.187	0.248	0.266	0.225
	(0.032)	(0.017)	(0.021)	(0.022)	(0.018)
All States	0.463	0.205	0.262	0.278	0.252
	(0.026)	(0.014)	(0.014)	(0.016)	(0.013)

Notes: Table A.13 reports weighted averages of destination-level network index estimates, adjusted for incomplete migration data using the coverage rates in Appendix Table A.12. Birth town groups are defined by cross validation. Standard errors in parentheses.

Source: Duke SSA/Medicare data and Ruggles et al. (2019)

Table A.14: Summary Statistics, Destination County Characteristics

Variable	Mean	S.D.	N
Panel A: Southern Black Migrants			
Network index estimate, $\widehat{\Delta}_k$	0.722	1.358	1515
Manufacturing employment share, 1910	0.178	0.112	1515
Agriculture employment share, 1910	0.242	0.173	1515
Log distance from birth state	6.688	0.517	1515
Direct railroad connection from birth state	0.092	0.290	1515
One-stop railroad connection from birth state	0.547	0.498	1515
Log population, 1910	11.230	1.155	1515
Percent African-American, 1910	0.043	0.077	1515
Percent rural, 1910	0.464	0.296	1515
Black church members per capita, 1916	0.190	0.195	1515
Small destination indicator	0.615	0.487	1515
Black-white relative wage, 1940	-0.132	0.194	1408
Logan-Parman segregation measure, 1940	0.493	0.213	1408
Panel B: Great Plains White Migrants			
Network index estimate, $\widehat{\Delta}_k$	0.140	0.438	4104
Manufacturing employment share, 1910	0.123	0.101	4104
Agriculture employment share, 1910	0.413	0.215	4104
Log distance from birth state	6.799	0.353	4104
Direct railroad connection from birth state	0.107	0.309	4104
One-stop railroad connection from birth state	0.486	0.500	4104
Log population, 1910	10.262	1.059	4104
Percent African-American, 1910	0.117	0.188	4104
Percent rural, 1910	0.701	0.284	4104
White church members per capita, 1916	0.422	0.185	4104
Small destination indicator	0.858	0.349	4104
Plains-not Plains relative wage, 1940	0.044	0.178	2311
Panel C: Southern White Migrants			
Network index estimate, $\widehat{\Delta}_k$	0.127	0.552	3357
Manufacturing employment share, 1910	0.139	0.111	3357
Agriculture employment share, 1910	0.340	0.198	3357
Log distance from birth state	6.765	0.594	3357
Direct railroad connection from birth state	0.083	0.276	3357
One-stop railroad connection from birth state	0.481	0.500	3357
Log population, 1910	10.596	1.137	3357
Percent African-American, 1910	0.037	0.073	3357
Percent rural, 1910	0.598	0.293	3357
White church members per capita, 1916	0.413	0.165	3357
Small destination indicator	0.766	0.424	3357

Notes: The unit of observation is a birth state-destination county pair. Sample includes destination counties for which we estimate a network index.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Logan and Parman (2017) data, Minnesota Population Center and Ancestry.com (2013), Ruggles et al. (2019)

Table A.15: Network Index Estimates and Destination County Characteristics, Birth Town Groups Defined by Counties

	Dependent variable: Destination-level network index estimate					
	Southern Black Migrants			Great Plains White Migrants		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing employment share, 1910	1.629	-0.091	-0.139	-0.068	-0.381	-0.383
	(0.731)	(0.520)	(0.519)	(0.091)	(0.166)	(0.165)
Manufacturing employment share by small destination indicator		3.036	3.024		0.415	0.421
		(1.224)	(1.168)		(0.190)	(0.189)
Agriculture employment share, 1910	0.087	-0.563	-0.610	0.055	0.077	0.073
	(0.255)	(0.401)	(0.404)	(0.036)	(0.125)	(0.125)
Agriculture employment share by small destination indicator		0.964	0.902		-0.010	-0.005
		(0.497)	(0.509)		(0.128)	(0.128)
Small destination indicator		-0.668	-0.661		-0.031	-0.035
		(0.294)	(0.287)		(0.069)	(0.069)
Log distance from birth state	-0.325	-0.288	-0.317	0.041	0.047	0.039
	(0.067)	(0.077)	(0.061)	(0.026)	(0.027)	(0.028)
Direct railroad connection from birth state	0.348	0.352	0.355	0.159	0.155	0.146
	(0.114)	(0.115)	(0.140)	(0.034)	(0.034)	(0.035)
One-stop railroad connection from birth state	0.227	0.216	0.189	0.058	0.052	0.051
	(0.092)	(0.088)	(0.097)	(0.014)	(0.013)	(0.013)
Log population, 1910	0.055	0.051	0.049	0.022	0.028	0.027
	(0.059)	(0.060)	(0.066)	(0.009)	(0.009)	(0.009)
Percent African-American, 1910	-1.566	-1.362	-1.384	-0.185	-0.201	-0.198
	(0.331)	(0.362)	(0.320)	(0.031)	(0.032)	(0.032)
Percent rural, 1910	-0.265	-0.264	-0.249	-0.033	-0.034	-0.034
	(0.195)	(0.198)	(0.215)	(0.036)	(0.037)	(0.037)
Black/white church members per capita, 1916	-0.400	-0.300	-0.315	-0.073	-0.064	-0.064
	(0.153)	(0.154)	(0.163)	(0.033)	(0.032)	(0.032)
Birth state fixed effects			x			x
R-squared	0.057	0.066	0.074	0.031	0.034	0.034
N (birth state-destination county pairs)	1,515	1,515	1,515	4,104	4,104	4,104
Destination counties	382	382	382	1230	1230	1230

Notes: See notes to Table 6. Birth town groups are defined by counties.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Ruggles et al. (2019)

Table A.16: Network Index Estimates and Destination County Characteristics, Southern White Migrants

Dependent variable: Destination-level network index estimate	(1)	(2)	(3)
Manufacturing employment share, 1910	0.465	-0.015	0.020
	(0.186)	(0.155)	(0.158)
Manufacturing employment share by small destination indicator		0.700	0.706
		(0.317)	(0.319)
Agriculture employment share, 1910	0.068	0.191	0.226
	(0.054)	(0.158)	(0.159)
Agriculture employment share by small destination indicator		-0.081	-0.099
		(0.169)	(0.167)
Small destination indicator		-0.115	-0.113
		(0.076)	(0.075)
Log distance from birth state	-0.033	-0.036	-0.004
	(0.018)	(0.018)	(0.018)
Direct railroad connection from birth state	0.058	0.060	0.074
	(0.038)	(0.038)	(0.037)
One-stop railroad connection from birth state	0.054	0.047	0.055
	(0.021)	(0.020)	(0.020)
Log population, 1910	0.016	0.015	0.022
	(0.016)	(0.016)	(0.015)
Percent African-American, 1910	-0.247	-0.327	-0.249
	(0.102)	(0.099)	(0.097)
Percent rural, 1910	0.043	0.012	0.010
	(0.063)	(0.064)	(0.063)
White church members per capita, 1916	-0.083	-0.089	-0.075
	(0.046)	(0.049)	(0.049)
Birth state fixed effects			x
R-squared	0.012	0.017	0.027
N (birth state-destination county pairs)	3,357	3,357	3,357
Destination counties	784	784	784

Notes: See notes to Table 6.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Ruggles et al. (2019)

Table A.17: Network Index Estimates and Destination County Characteristics, Additional Explanatory Variables

	Dependent variable: Destination-level network index estimate					
	Southern Black Migrants			Great Plains White Migrants		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing employment share, 1910	1.670	0.289	0.228	-0.085	-0.281	-0.283
	(0.585)	(0.703)	(0.709)	(0.129)	(0.162)	(0.161)
Manufacturing employment share by small destination indicator		2.539	2.601		0.283	0.288
		(1.010)	(1.001)		(0.230)	(0.228)
Agriculture employment share, 1910	0.322	-0.181	-0.248	0.121	0.224	0.217
	(0.331)	(0.488)	(0.491)	(0.080)	(0.148)	(0.146)
Agriculture employment share by small destination indicator		0.773	0.716		-0.109	-0.103
		(0.506)	(0.516)		(0.141)	(0.140)
Small destination indicator		-0.543	-0.552		0.039	0.035
		(0.273)	(0.272)		(0.076)	(0.075)
Log distance from birth state	-0.408	-0.370	-0.363	0.069	0.084	0.076
	(0.075)	(0.079)	(0.078)	(0.048)	(0.051)	(0.056)
Direct railroad connection from birth state	0.316	0.321	0.302	0.246	0.249	0.245
	(0.123)	(0.124)	(0.142)	(0.044)	(0.046)	(0.050)
One-stop railroad connection from birth state	0.215	0.204	0.160	0.103	0.097	0.098
	(0.082)	(0.080)	(0.084)	(0.024)	(0.024)	(0.023)
Log population, 1910	0.058	0.058	0.067	0.015	0.028	0.027
	(0.058)	(0.061)	(0.062)	(0.014)	(0.014)	(0.014)
Percent African-American, 1910	-2.193	-2.006	-1.923	-0.224	-0.242	-0.240
	(0.446)	(0.429)	(0.432)	(0.040)	(0.043)	(0.043)
Percent rural, 1910	-0.361	-0.345	-0.298	-0.090	-0.077	-0.076
	(0.209)	(0.227)	(0.231)	(0.057)	(0.054)	(0.054)
Black/white church members per capita, 1916	-0.438	-0.357	-0.338	-0.144	-0.126	-0.127
	(0.191)	(0.201)	(0.202)	(0.051)	(0.050)	(0.050)
Black-white / Plains-not Plains Relative wage, 1940	0.401	0.391	0.402	-0.120	-0.115	-0.115
	(0.183)	(0.187)	(0.191)	(0.043)	(0.043)	(0.042)
Logan-Parman segregation measure, 1940	0.274	0.281	0.276			
	(0.222)	(0.214)	(0.217)			
Birth state fixed effects			x			x
R-squared	0.102	0.109	0.123	0.048	0.052	0.053
N (birth state-destination county pairs)	1,408	1,408	1,408	2,311	2,311	2,311
Destination counties	335	335	335	642	642	642

Notes: See notes to Table 6. Columns 1–3 contain the black-white relative wage, and columns 4–6 contain the Plains-not Plains relative wage. See Appendix F for details on how these are constructed. Standard errors, clustered by destination county, are in parentheses.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Logan and Parman (2017) data, Minnesota Population Center and Ancestry.com (2013), Ruggles et al. (2019)

Table A.18: Summary Statistics, Birth County Characteristics

Variable	Mean	S.D.	N
Panel A: Southern Black Migrants			
Network index estimate, $\widehat{\Delta_c}$	1.729	3.555	546
Percent of black farmers who are owners, 1910	0.324	0.254	546
Percent of black individuals in owner-occupied housing, 1910	0.264	0.145	546
Percent of black workers in agriculture, 1910	0.665	0.224	546
Percent of black workers in manufacturing, 1910	0.071	0.083	546
Percent of farm acreage in cotton, 1910	0.135	0.106	546
Log black population density, 1910	2.581	1.005	546
Black church members per capita, 1916	0.391	0.188	546
Rosenwald school exposure	0.202	0.214	546
Black literacy rate (10+), 1910	0.632	0.100	546
Black school attendance rate (6–14), 1910	0.544	0.131	546
Railroad exposure	0.546	0.404	546
Percent African-American, 1910	0.443	0.213	546
Percent rural, 1910	0.899	0.168	546
Percent voting for Strom Thurmond, 1948	0.457	0.335	546
Panel B: Great Plains White Migrants			
Network index estimate, $\widehat{\Delta_c}$	0.354	0.642	383
Percent of white farmers who are owners, 1910	0.646	0.212	383
Percent of white individuals in owner-occupied housing, 1910	0.642	0.164	383
Percent of white workers in agriculture, 1910	0.643	0.140	383
Percent of white workers in manufacturing, 1910	0.021	0.028	383
Percent of farm acreage in cotton, 1910	0.016	0.045	383
Log white population density, 1910	2.539	0.926	383
White church members per capita, 1916	0.307	0.116	383
White literacy rate (10+), 1910	0.978	0.021	383
White school attendance rate (6–14), 1910	0.851	0.072	383
Railroad exposure	0.539	0.391	383
Percent African-American, 1910	0.020	0.05	383
Percent rural, 1910	0.892	0.177	383
Panel C: Southern White Migrants			
Network index estimate, $\widehat{\Delta_c}$	0.212	0.772	576
Percent of white farmers who are owners, 1910	0.618	0.147	576
Percent of white individuals in owner-occupied housing, 1910	0.554	0.107	576
Percent of white workers in agriculture, 1910	0.645	0.185	576
Percent of white workers in manufacturing, 1910	0.084	0.074	576
Percent of farm acreage in cotton, 1910	0.127	0.106	576
Log white population density, 1910	2.862	0.802	576
White church members per capita, 1916	0.469	0.197	576
Rosenwald school exposure	0.198	0.218	576
White literacy rate (10+), 1910	0.907	0.064	576
White school attendance rate (6–14), 1910	0.745	0.085	576
Railroad exposure	0.533	0.415	576
Percent African-American, 1910	0.424	0.219	576
Percent rural, 1910	0.904	0.165	576
Percent voting for Strom Thurmond, 1948	0.437	0.335	576

Notes: Sample includes birth counties containing at least one town with at least 10 migrants in the Duke data. Railroad exposure is the share of migrants in a county that lived along a railroad. Rosenwald school exposure is the average Rosenwald coverage experienced over ages 7–13.
Sources: Duke SSA/Medicare data, Aaronson and Mazumder (2011) data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), ICPSR (1999), Ruggles et al. (2019)

Table A.19: Network Index Estimates and Birth County Characteristics, Southern White Migrants

Dependent variable: Birth county-level network index estimate	(1)	(2)
Percent of white farmers who are owners, 1910	-1.587	-1.514
	(0.764)	(0.725)
Percent of white individuals in owner-occupied housing, 1910	1.907	1.966
	(0.988)	(0.958)
Percent of white workers in agriculture, 1910	-0.588	-0.477
	(0.366)	(0.427)
Percent of white workers in manufacturing, 1910	-0.014	0.143
	(0.950)	(1.002)
Percent of farm acreage in cotton, 1910	-0.263	-0.427
	(0.501)	(0.561)
Log white population density, 1910	-0.052	-0.052
	(0.105)	(0.116)
White church members per capita, 1916	-0.212	-0.216
	(0.186)	(0.195)
Rosenwald school exposure	0.379	0.431
	(0.166)	(0.210)
White literacy rate (10+), 1910	-0.479	-0.608
	(0.774)	(0.839)
White school attendance rate (6–14), 1910	-0.334	-0.126
	(0.442)	(0.447)
Railroad exposure	0.075	0.082
	(0.067)	(0.067)
Percent African-American, 1910	-0.851	-0.633
	(0.253)	(0.287)
Percent rural, 1910	0.910	0.880
	(0.370)	(0.366)
Percent voting for Strom Thurmond, 1948	0.108	-0.352
	(0.103)	(0.195)
Birth state fixed effects		x
R-squared	0.131	0.138
N (birth counties)	576	576

Notes: See notes to Table 7.

Sources: Duke SSA/Medicare data, Aaronson and Mazumder (2011) data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), ICPSR (1999), Ruggles et al. (2019)

Table A.20: Average Network Index Estimates, by Destination Region

	Destination Region			
	Northeast (1)	Midwest (2)	West (3)	South (4)
Panel A: Southern Black Migrants				
Alabama	1.237	2.356	0.813	-
	(0.161)	(0.295)	(0.272)	-
Florida	0.978	0.793	0.264	-
	(0.172)	(0.169)	(0.107)	-
Georgia	1.546	2.067	0.410	-
	(0.243)	(0.310)	(0.205)	-
Louisiana	0.282	1.138	2.169	-
	(0.101)	(0.206)	(0.734)	-
Mississippi	0.924	2.662	1.036	-
	(0.105)	(0.396)	(0.130)	-
North Carolina	1.678	0.908	0.185	-
	(0.149)	(0.176)	(0.040)	-
South Carolina	2.907	1.223	0.211	-
	(0.351)	(0.167)	(0.055)	-
All States	1.860	2.259	1.402	-
	(0.120)	(0.195)	(0.345)	-
Panel B: Great Plains White Migrants				
Kansas	0.079	0.452	0.281	0.051
	(0.019)	(0.095)	(0.031)	(0.006)
Nebraska	0.080	0.439	0.420	0.063
	(0.014)	(0.096)	(0.109)	(0.009)
North Dakota	0.107	0.405	0.524	0.047
	(0.027)	(0.057)	(0.046)	(0.009)
Oklahoma	0.051	0.390	0.542	0.074
	(0.007)	(0.091)	(0.047)	(0.007)
South Dakota	0.061	0.485	0.381	0.058
	(0.013)	(0.069)	(0.034)	(0.011)
All States	0.073	0.434	0.442	0.062
	(0.007)	(0.039)	(0.029)	(0.004)

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. We define destination regions slightly differently than the Census Bureau because we treat the former Confederate states as the South. The Census South region includes Delaware, the District of Columbia, Maryland, West Virginia, Kentucky, and Oklahoma. We include the first four states in the Northeast and the latter two in the Midwest. We do not estimate network indices for African Americans who move to the South. Birth town groups are defined by cross validation. Standard errors are in parentheses.
Source: Duke SSA/Medicare data

Table A.21: Average Network Index Estimates, by Destination Region, Southern White Migrants

	Destination Region			
	Northeast (1)	Midwest (2)	West (3)	South (4)
Alabama	0.140	1.048	0.208	-
	(0.021)	(0.123)	(0.034)	-
Florida	0.090	0.070	0.277	-
	(0.017)	(0.020)	(0.104)	-
Georgia	0.104	0.307	0.082	-
	(0.013)	(0.049)	(0.023)	-
Louisiana	0.159	0.450	0.331	-
	(0.027)	(0.100)	(0.100)	-
Mississippi	0.067	0.301	0.127	-
	(0.014)	(0.052)	(0.014)	-
North Carolina	0.549	0.489	0.302	-
	(0.063)	(0.122)	(0.048)	-
South Carolina	0.111	0.081	0.073	-
	(0.011)	(0.012)	(0.022)	-
All States	0.275	0.534	0.220	-
	(0.024)	(0.044)	(0.026)	-

Notes: See note to Appendix Table A.20.
Source: Duke SSA/Medicare data

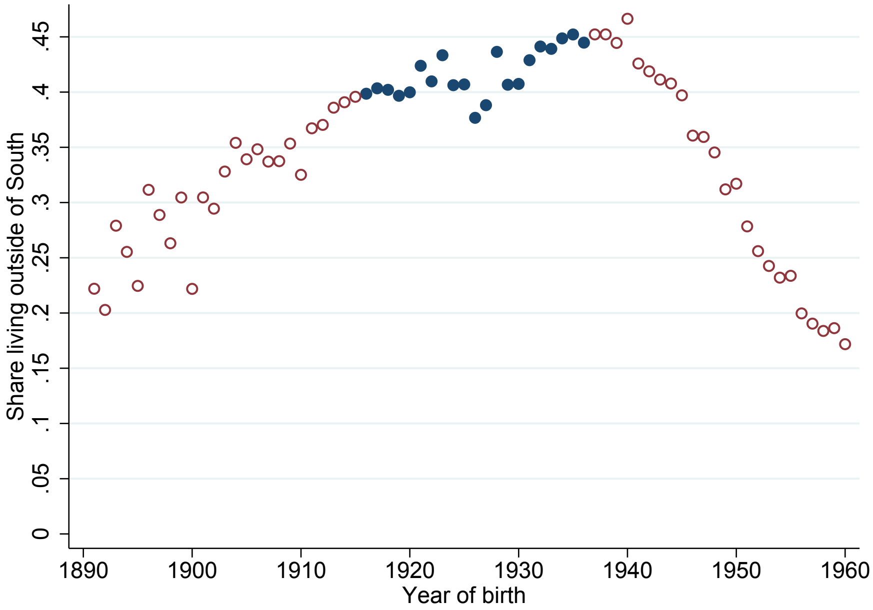
Table A.22: Changes in Regional Migration Patterns in a Counterfactual without Migration Networks

	Destination Region			
Birth State	Northeast (1)	Midwest (2)	West (3)	South (4)
Panel A: Southern Black Migrants				
Alabama	4,354	-6,706	2,353	
	[19.2%]	[-10.8%]	[20.5%]	
Florida	-224	-68	291	
	[-1.8%]	[-1.7%]	[12.2%]	
Georgia	621	-2,147	1,526	
	[1.6%]	[-6.6%]	[25.4%]	
Louisiana	1,267	1,809	-3,076	
	[31.2%]	[11.7%]	[-8.4%]	
Mississippi	2,951	-7,303	4,352	
	[38.9%]	[-7.6%]	[25.5%]	
North Carolina	-2,252	1,033	1,220	
	[-3.3%]	[14.9%]	[36.9%]	
South Carolina	-2,175	1,056	1,119	
	[-3.6%]	[14.8%]	[43.5%]	
All States	4,541	-12,325	7,785	
	[2.1%]	[-5.5%]	[9.8%]	
Panel B: Great Plains White Migrants				
Kansas	485	-179	-1,645	1,339
	[6.9%]	[-1.1%]	[-1.7%]	[6.6%]
Nebraska	673	284	-2,462	1,505
	[10.7%]	[2.0%]	[-2.5%]	[10.9%]
North Dakota	305	527	-1,508	676
	[10.0%]	[4.6%]	[-2.1%]	[10.4%]
Oklahoma	819	995	-5,000	3,186
	[12.2%]	[7.4%]	[-3.3%]	[11.3%]
South Dakota	292	171	-1,126	662
	[9.7%]	[1.5%]	[-2.0%]	[9.4%]
All States	2,574	1,799	-11,740	7,368
	[9.9%]	[2.7%]	[-2.5%]	[9.7%]

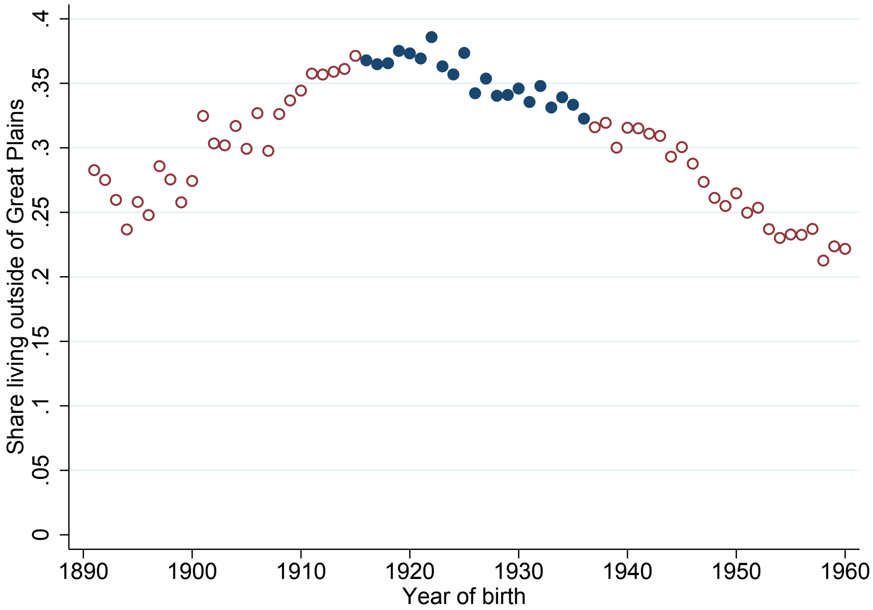
Notes: Table contains estimates of N_k^{cf}, the number of migrants that would have chosen destination county k in the absence of migration network, aggregated over all counties in each region. Percent changes of the number of migrants in the counterfactual are in brackets. See the text for details. Source: Duke SSA/Medicare data

Figure A.1: Migration Rates Around Ages 40–49

(a) Southern Black Migrants



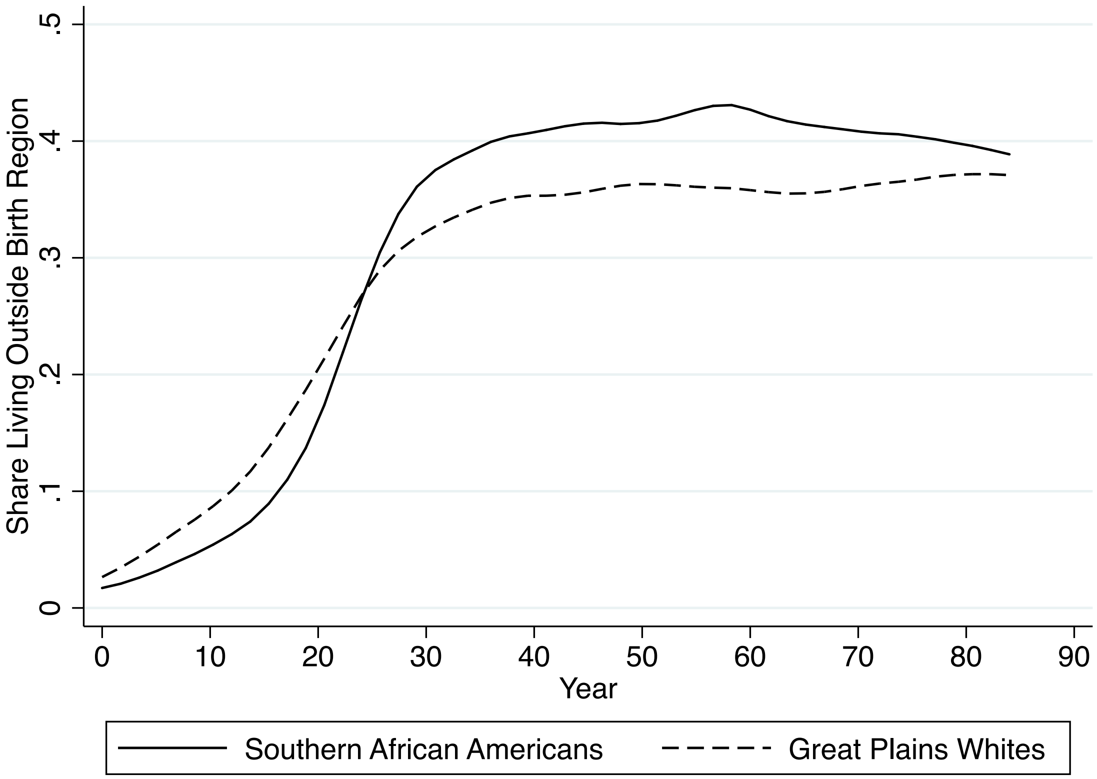
(b) Great Plains White Migrants



Notes: Panel A reports the share of African Americans born in AL, FL, GA, LA, MS, NC, and SC living outside of the former Confederate States. Panel B reports the share of whites born in KS, NE, ND, OK, and SD living outside of the Great Plains and border area shaded in light grey in Figure 2. For individuals born from 1891–1900, we measure their location using the 1900 Census. For individuals born from 1901–1910, we use the 1910 Census, and so forth. The shaded circles correspond to individuals born from 1916–1936, who comprise our sample from the Duke SSA/Medicare data.

Sources: 1940–2000 Census data from Ruggles et al. (2019)

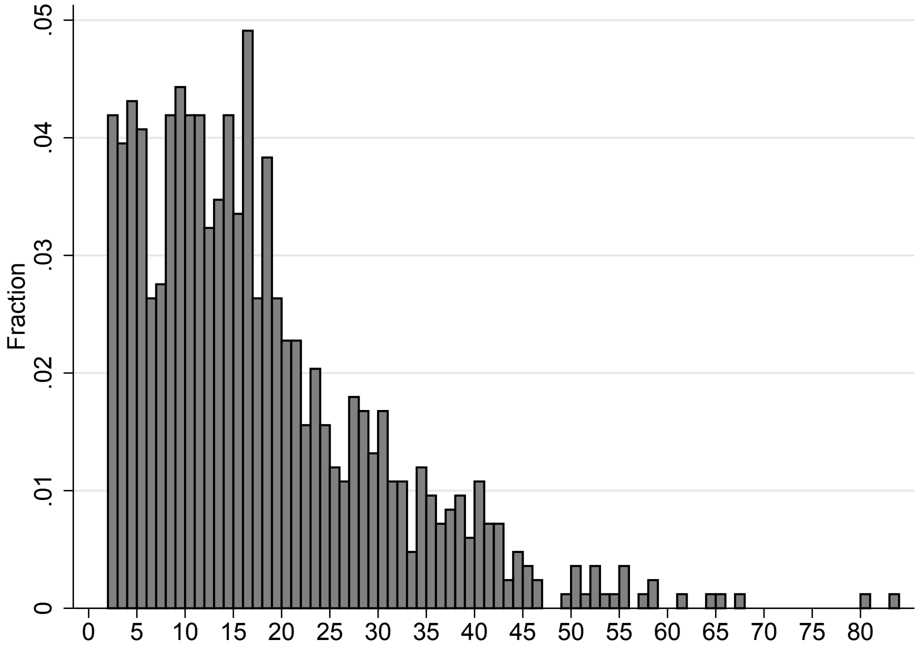
Figure A.2: Share Living Outside Birth Region, 1916–1936 Cohorts, by Age



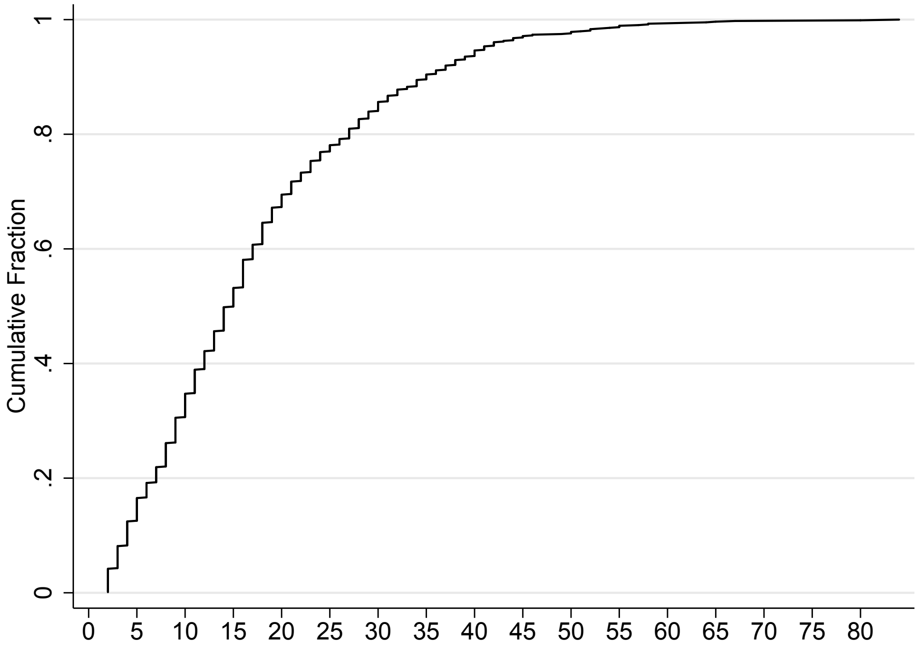
Notes: The solid line shows the percent of African Americans born from 1916–1936 in the seven Southern birth states we analyze (dark grey states in Figure 1a) living outside the South (light and dark grey states) at the time of Census enumeration. The dashed line shows the percent of whites born from 1916–1936 from the Great Plains states living outside the Great Plains or Border States. Both lines are locally mean-smoothed relationships of the underlying observations.
Source: Ruggles et al. (2019)

Figure A.3: Number of Towns per Birth Town Group, Cross Validation, Southern Black Migrants

(a) Histogram



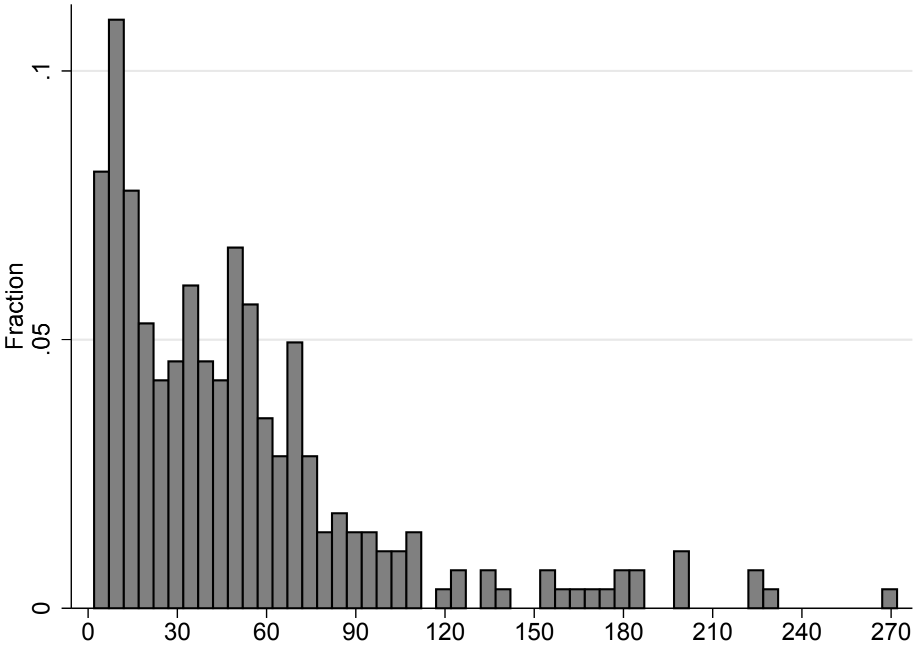
(b) Cumulative Distribution



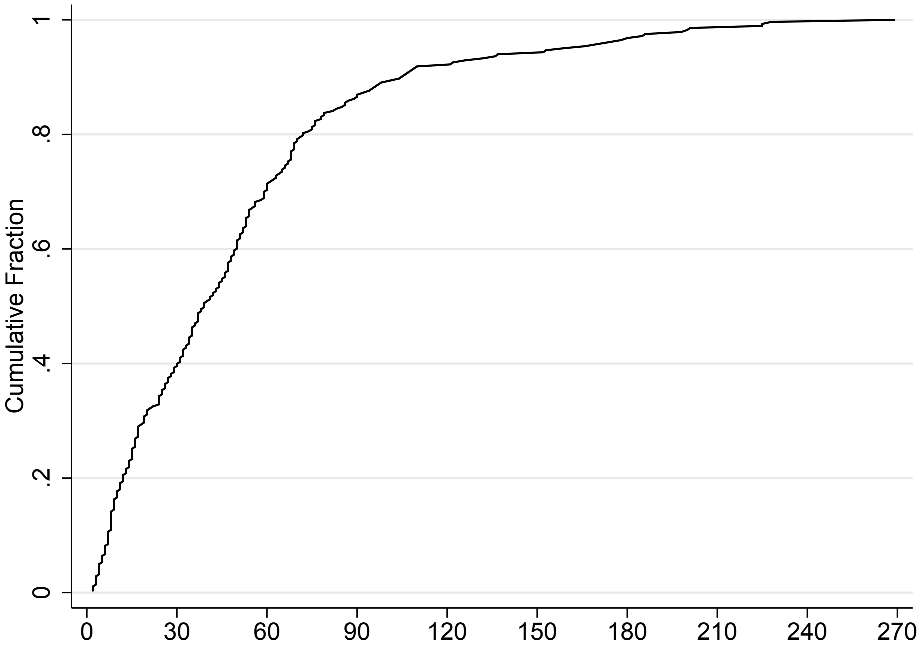
Notes: Figure excludes groups with a single town, as these are not used in the analysis. Bin width in panel (a) is 1.
Source: Duke SSA/Medicare data

Figure A.4: Number of Towns per Birth Town Group, Cross Validation, Great Plains White Migrants

(a) Histogram



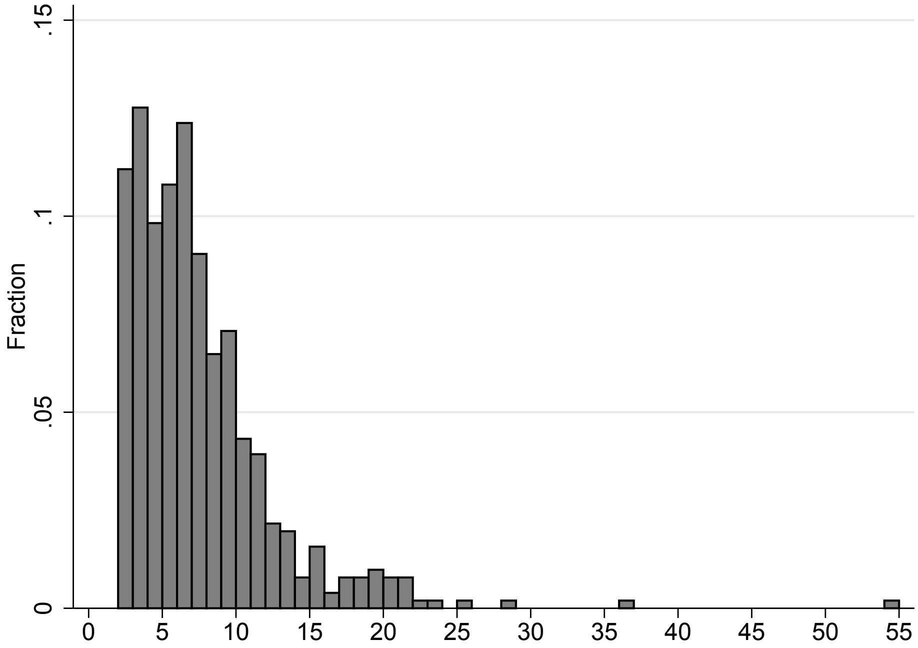
(b) Cumulative Distribution



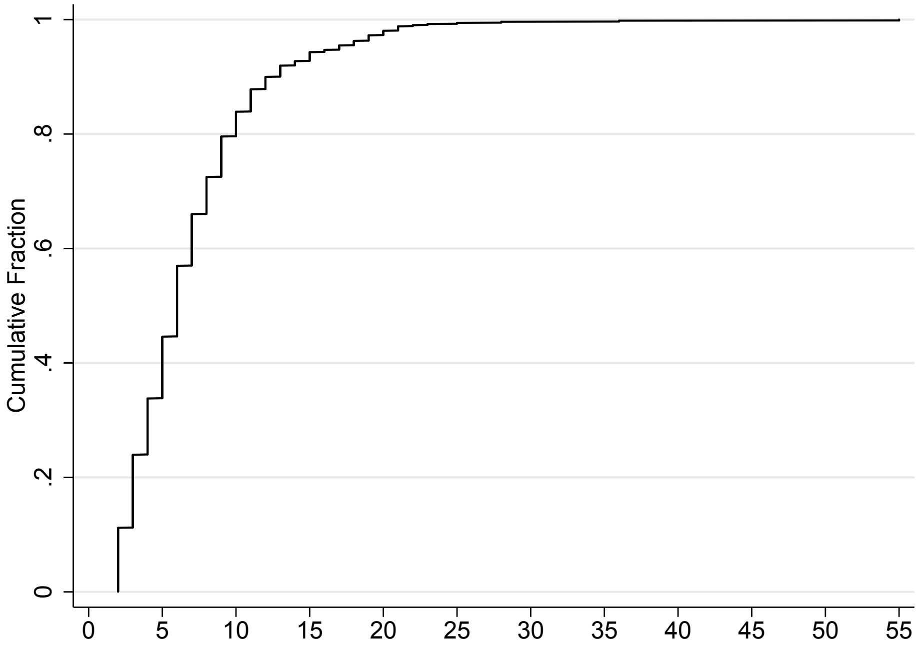
Notes: Figure excludes groups with a single town, as these are not used in the analysis. Bin width in panel (a) is 5.
Source: Duke SSA/Medicare data

Figure A.5: Number of Towns per Birth County, Southern Black Migrants

(a) Histogram



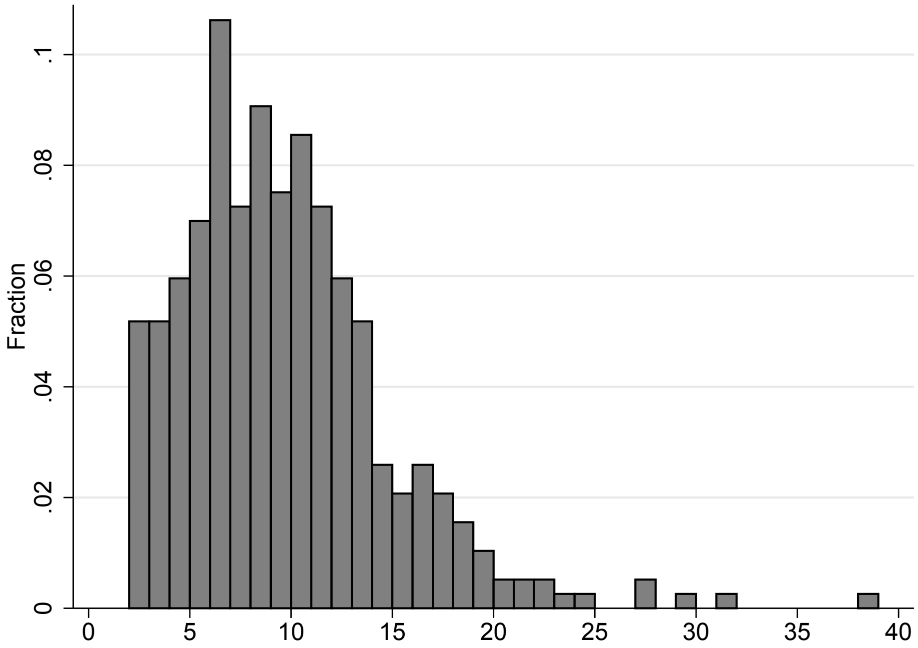
(b) Cumulative Distribution



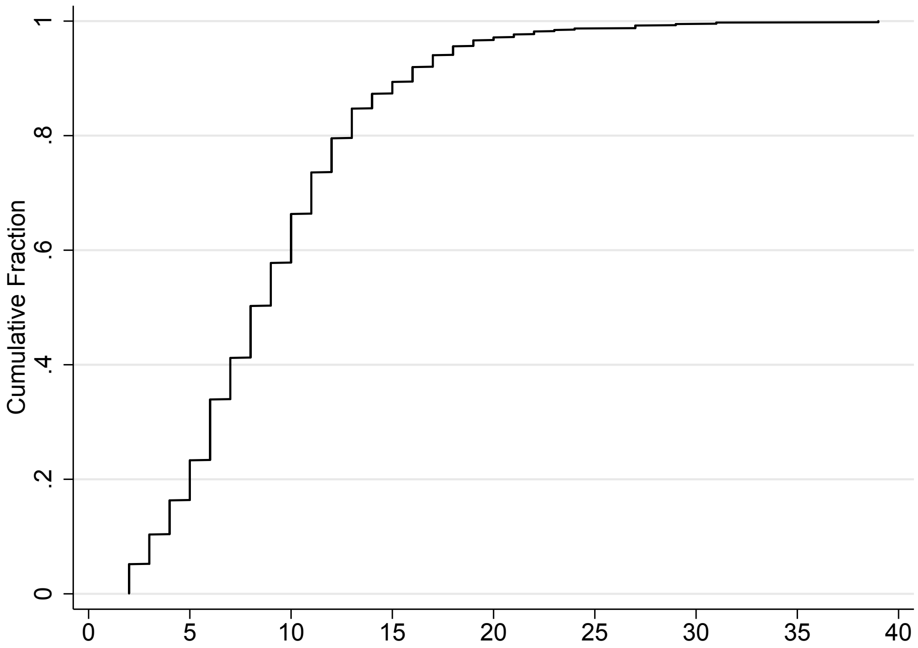
Notes: Figure excludes groups with a single town, as these are not used in the analysis. Bin width in panel (a) is 1.
Source: Duke SSA/Medicare data

Figure A.6: Number of Towns per Birth County, Great Plains White Migrants

(a) Histogram



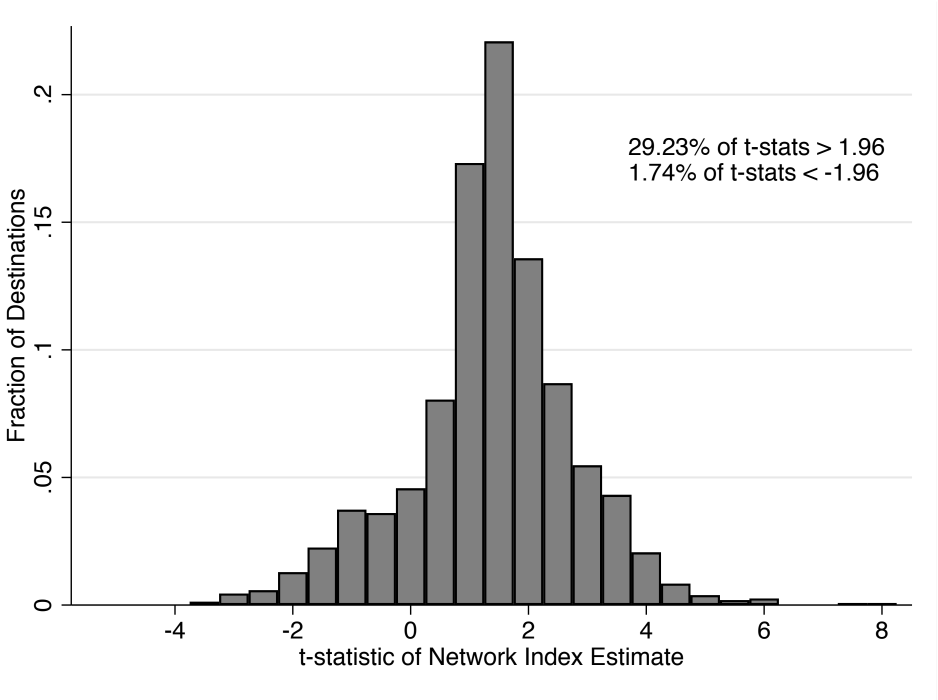
(b) Cumulative Distribution



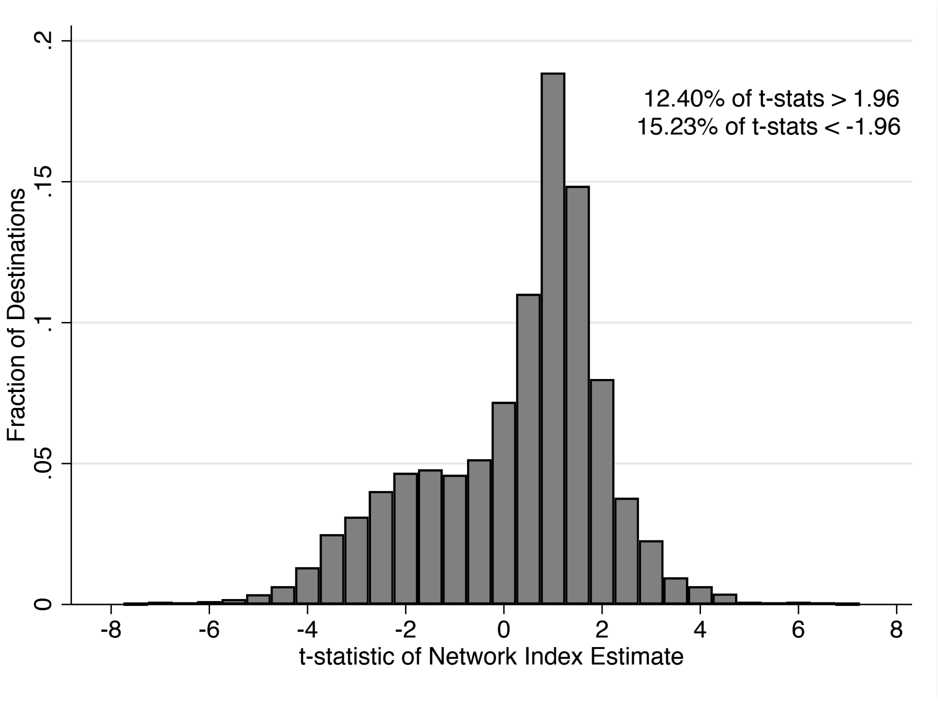
Notes: Figure excludes groups with a single town, as these are not used in the analysis. Bin width in panel (a) is 1.
Source: Duke SSA/Medicare data

Figure A.7: Distribution of Destination-Level Network Index t-statistics

(a) Southern Black Migrants

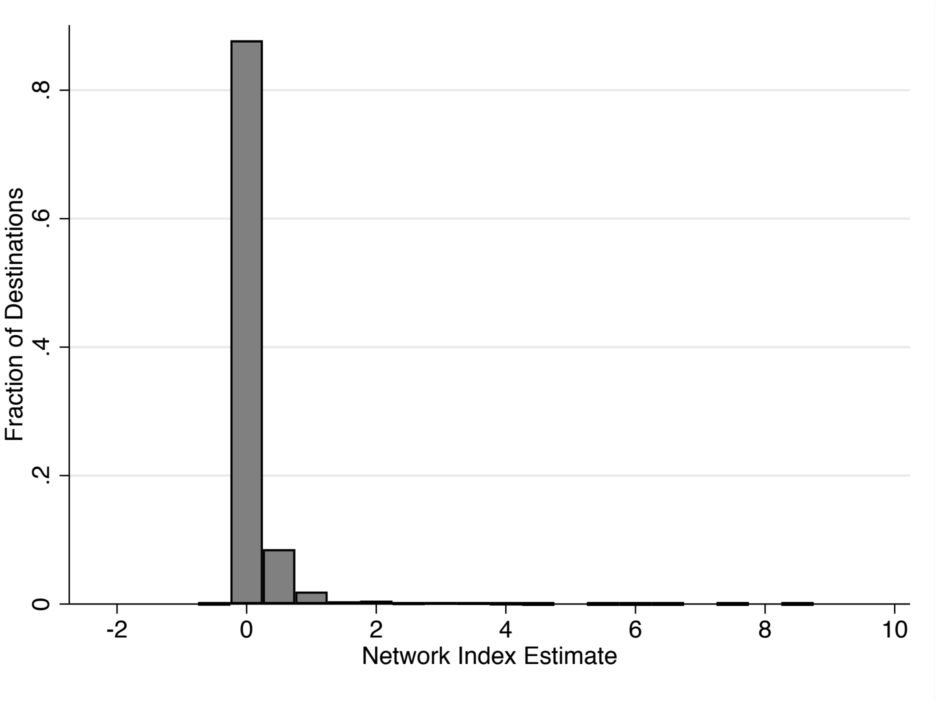


(b) Great Plains White Migrants



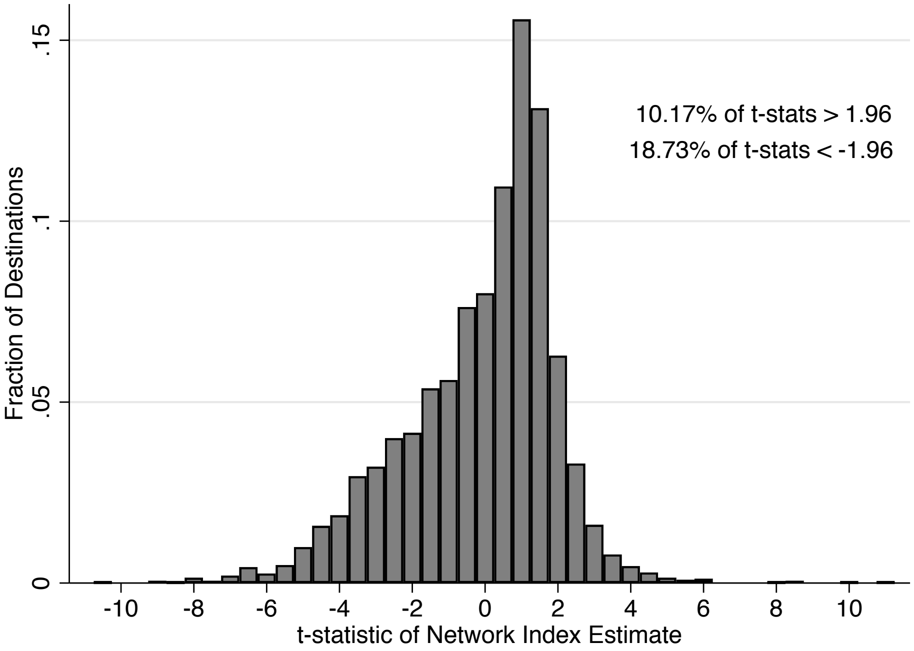
Notes: Bin width is 1/2. Birth town groups are defined by cross validation. Panel (a) omits the t-statistic of 13.7 from South Carolina to Hancock, WV.
Source: Duke SSA/Medicare data

Figure A.8: Distribution of Destination-Level Network Index Estimates, Southern White Migrants



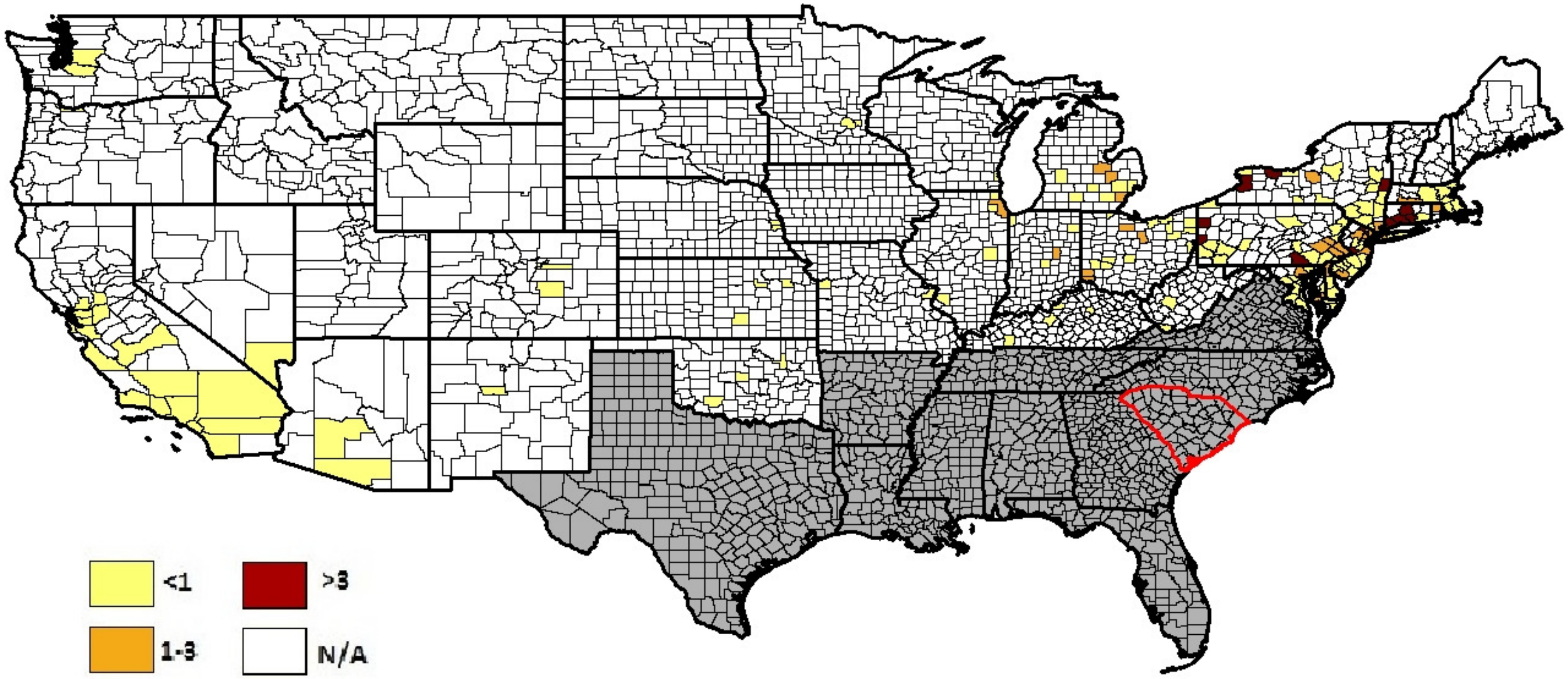
Notes: Bin width is 1/2. Figure omits estimate of $\hat{\Delta}_k = 19.3$ from Alabama to St. Joseph County, IN.
Source: Duke SSA/Medicare data

Figure A.9: Distribution of Destination-Level Network Index t-statistics, Southern White Migrants



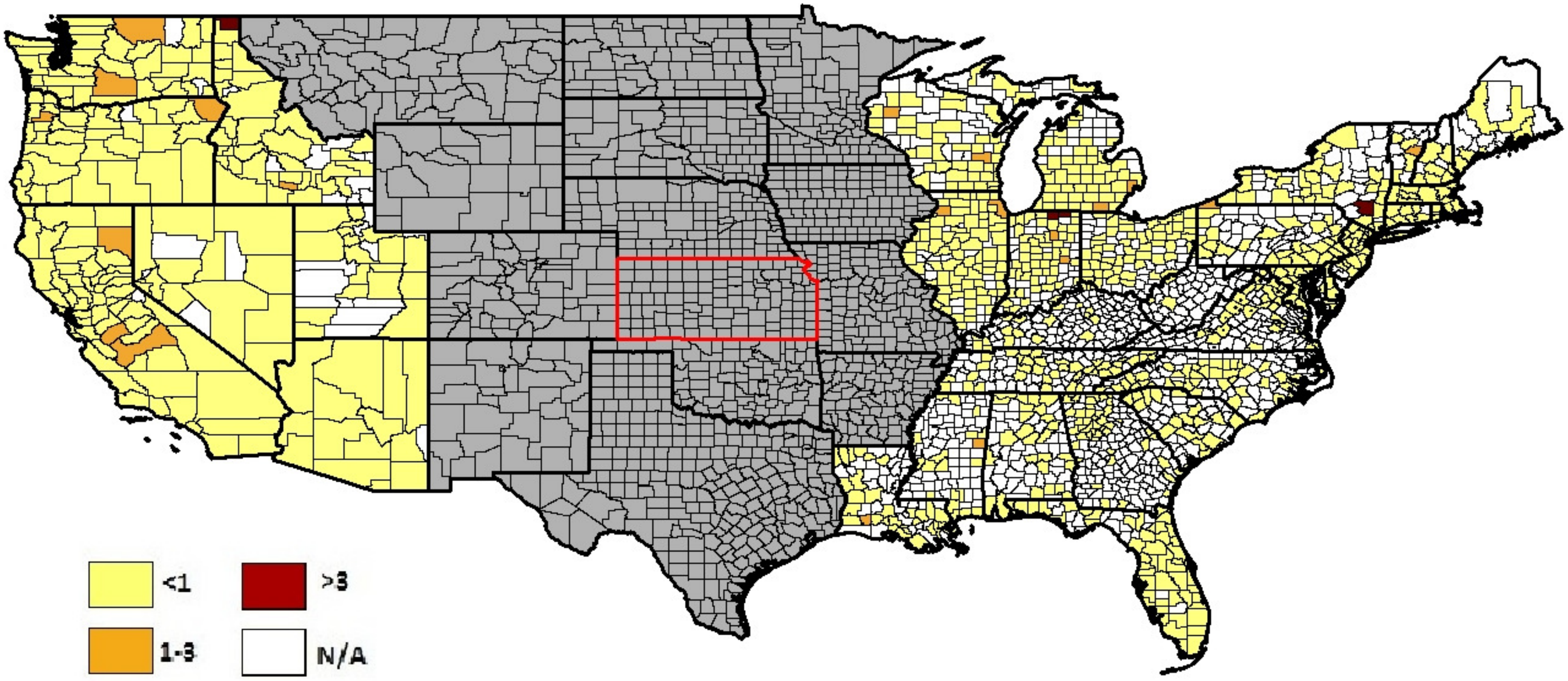
Note: Bin width is 1/2.
Source: Duke SSA/Medicare data

Figure A.10: Spatial Distribution of Destination-Level Network Index Estimates, South Carolina-born Black Migrants



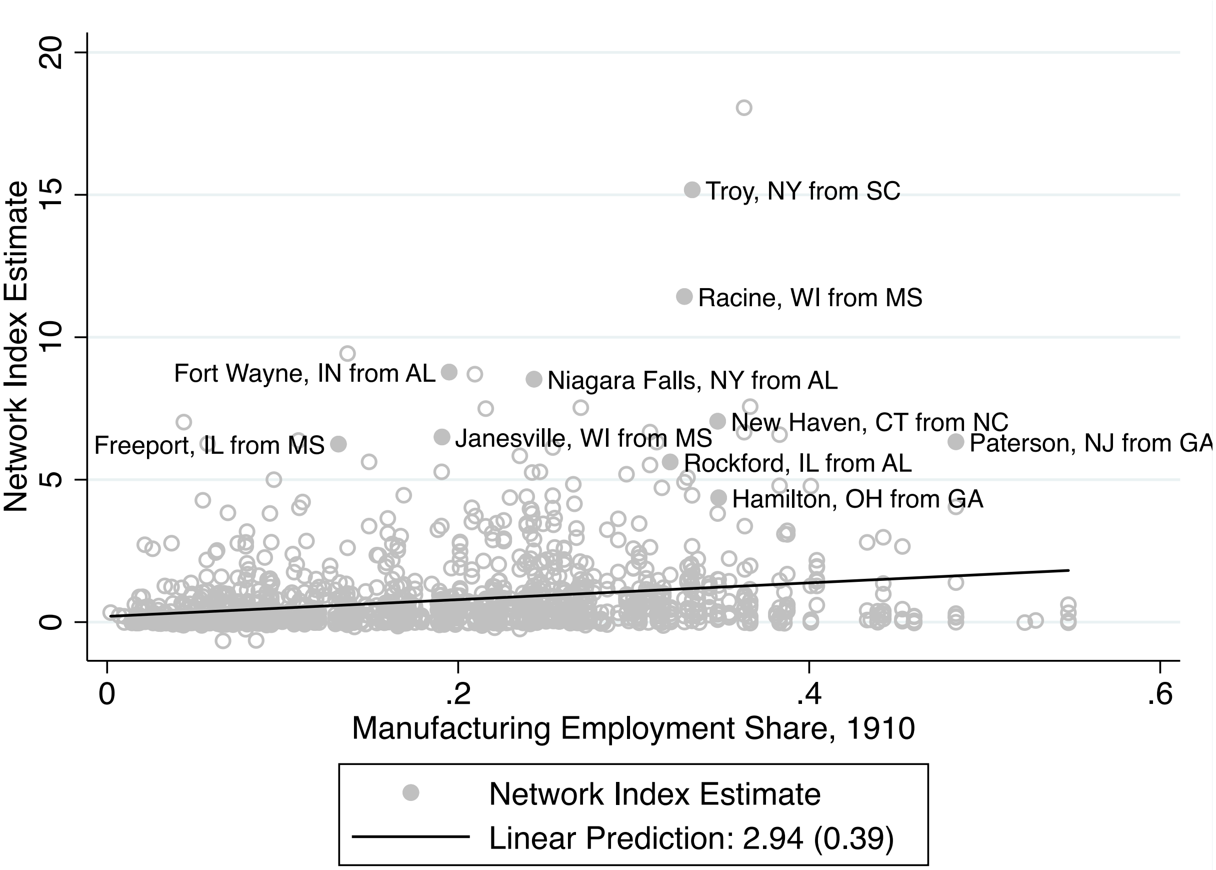
Notes: See note to Figure 4.

Figure A.11: Spatial Distribution of Destination-Level Network Index Estimates, Kansas-born White Migrants



Notes: See note to Figure 5.

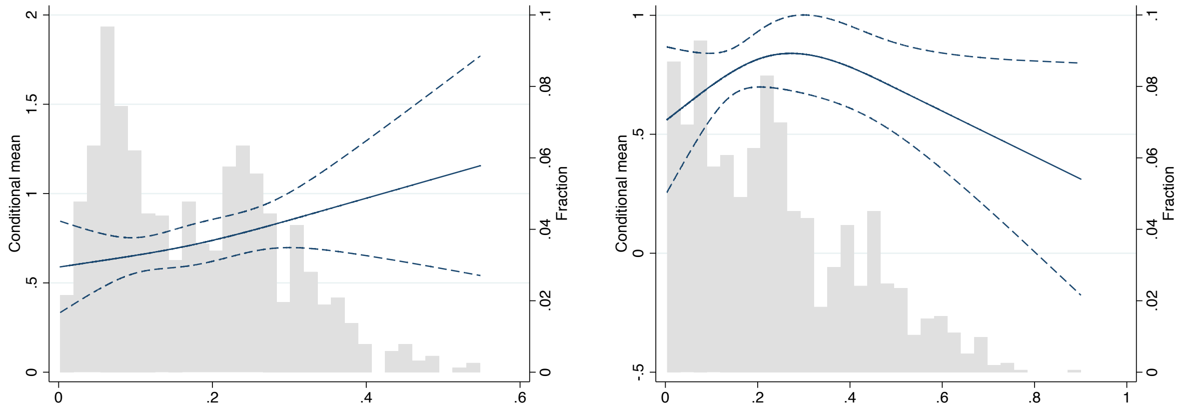
Figure A.12: Relationship between Southern Black Destination-Level Network Index Estimates and 1910 Manufacturing Employment Share



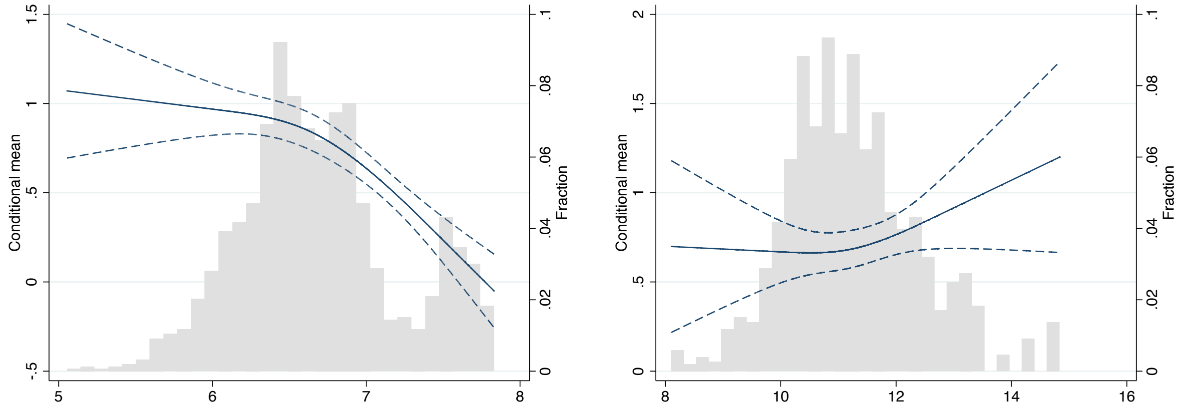
Notes: Linear prediction comes from an OLS regression that includes a constant and 1910 manufacturing employment share. Listed are the cities in Table 2.

Sources: Duke SSA/Medicare data and Haines and ICPSR (2010)

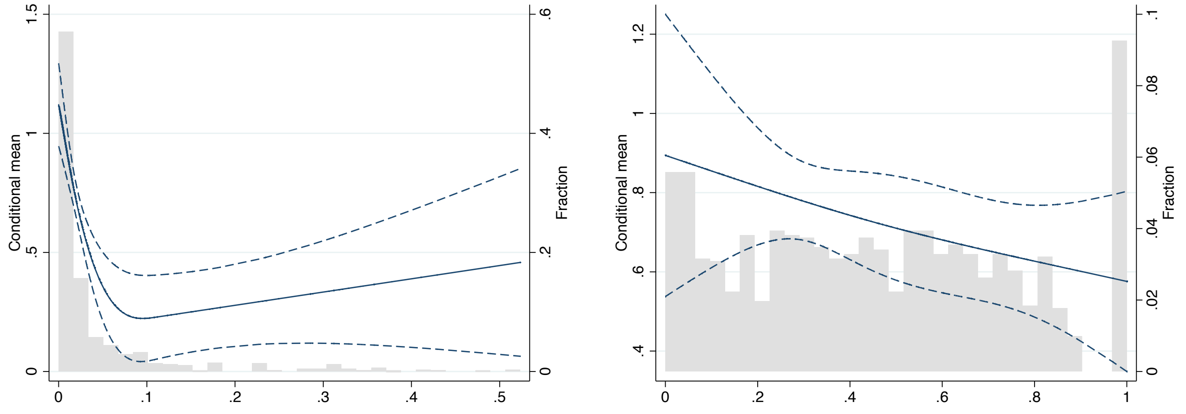
Figure A.13: Nonlinear Relationship between Covariates and Destination County Network Index Estimates, Southern Black Migrants


(a) Manufacturing employment share, 1910

(b) Agriculture employment share, 1910

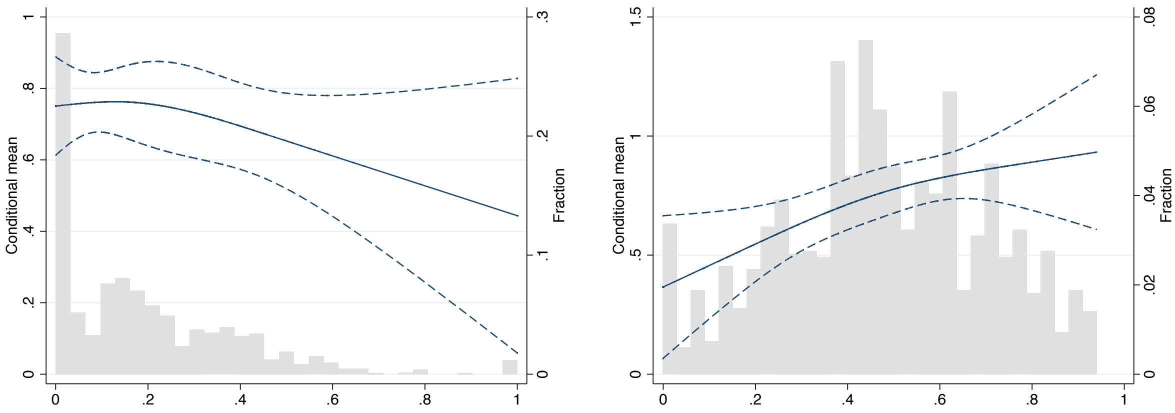

(c) Log distance from birth state

(d) Log population, 1910


(e) Percent African-American, 1910

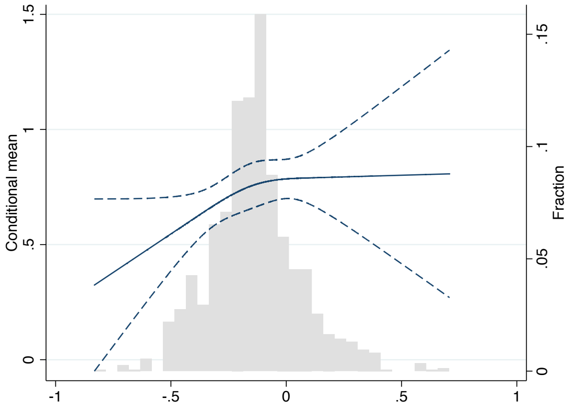
(f) Percent rural, 1910

Figure A.13: Nonlinear Relationship between Covariates and Destination County Network Index Estimates, Southern Black Migrants


(g) Black church members per capita, 1916

(h) Logan-Parman segregation measure, 1940

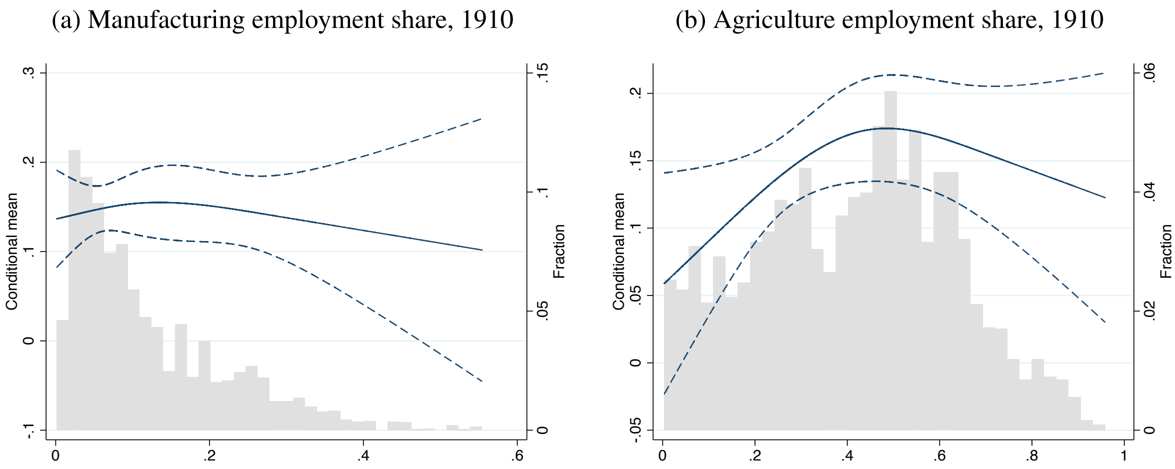
(i) Black-white relative wage, 1940



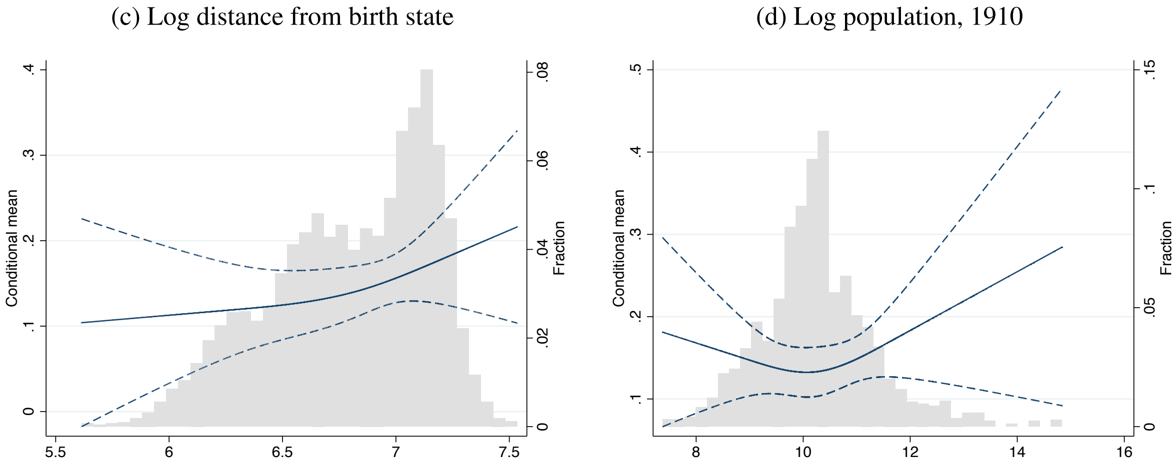
Notes: The solid blue line is the conditional mean of the birth county network index as a function of the indicated independent variable. Dashed lines are 95 percent confidence intervals. Results come from regressing destination county network index estimates against restricted cubic splines in the nine indicated variables, plus indicators for whether the destination has a direct or one-stop connection from the birth state. Grey bars are histograms of the underlying independent variable (right scale). See notes to Table 6.

Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Logan and Parman (2017), Minnesota Population Center and Ancestry.com (2013)

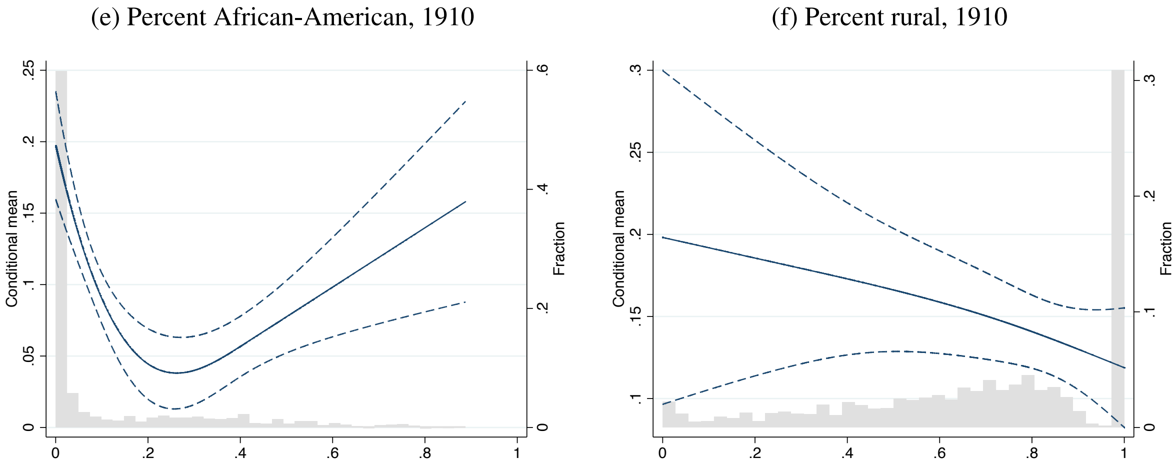
Figure A.14: Nonlinear Relationship between Covariates and Destination County Network Index Estimates, Great Plains White Migrants


(a) Manufacturing employment share, 1910

(b) Agriculture employment share, 1910

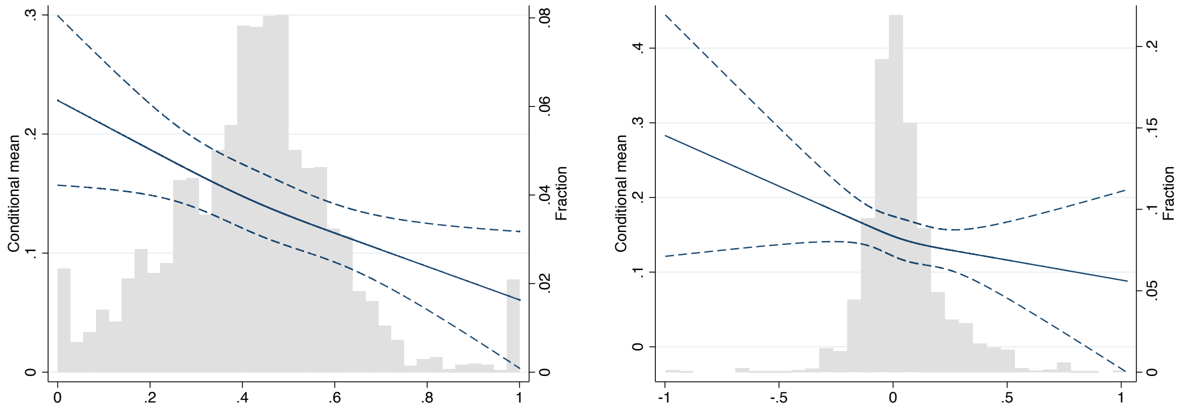

(c) Log distance from birth state

(d) Log population, 1910


(e) Percent African-American, 1910

(f) Percent rural, 1910

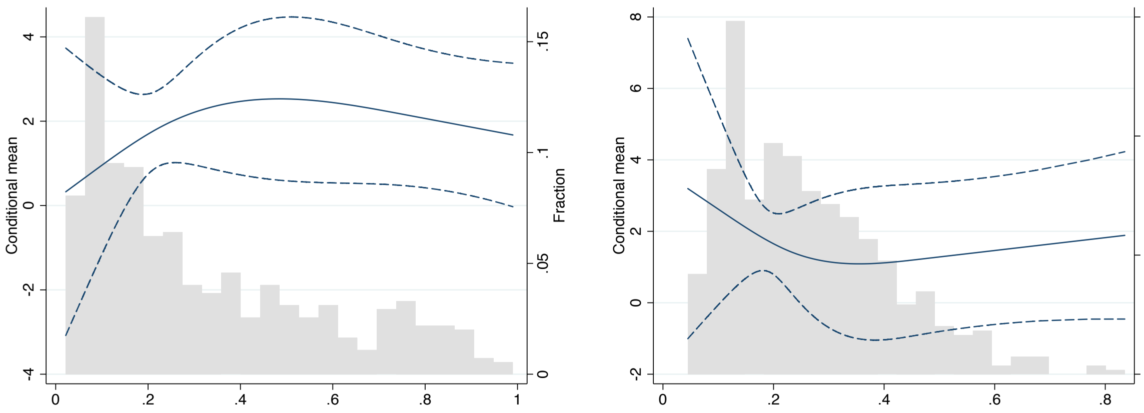
Figure A.14: Nonlinear Relationship between Covariates and Destination County Network Index Estimates, Great Plains White Migrants


(g) White church members per capita, 1916

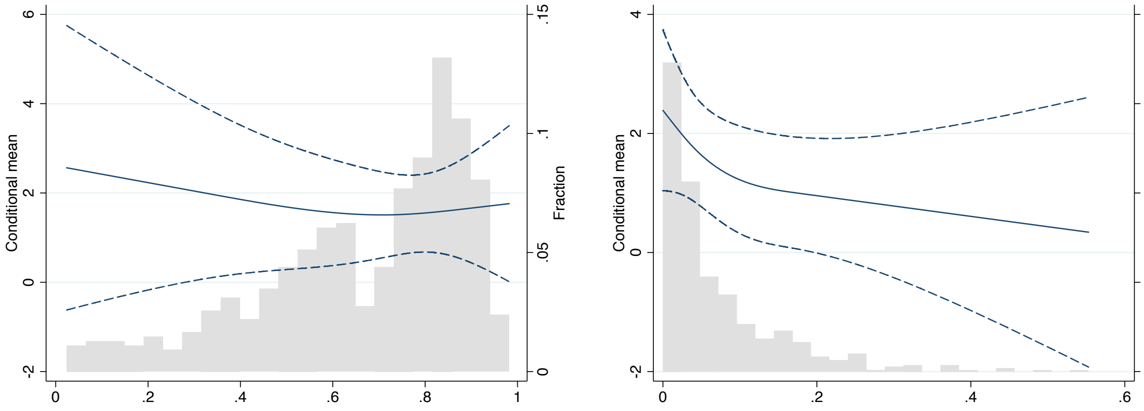
(h) Plains-not Plains relative wage, 1940

Notes: See notes to Figure A.13.
Sources: Duke SSA/Medicare data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), Logan and Parman (2017), Minnesota Population Center and Ancestry.com (2013)

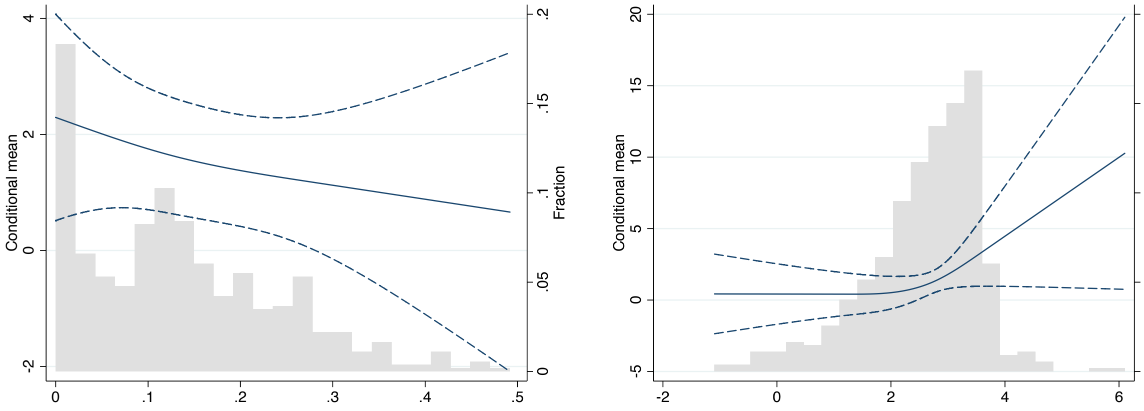
Figure A.15: Nonlinear Relationship between Covariates and Birth County Network Index Estimates, Southern Black Migrants


(a) Percent of farmers who are owners, 1910

(b) Percent in owner-occupied housing, 1910

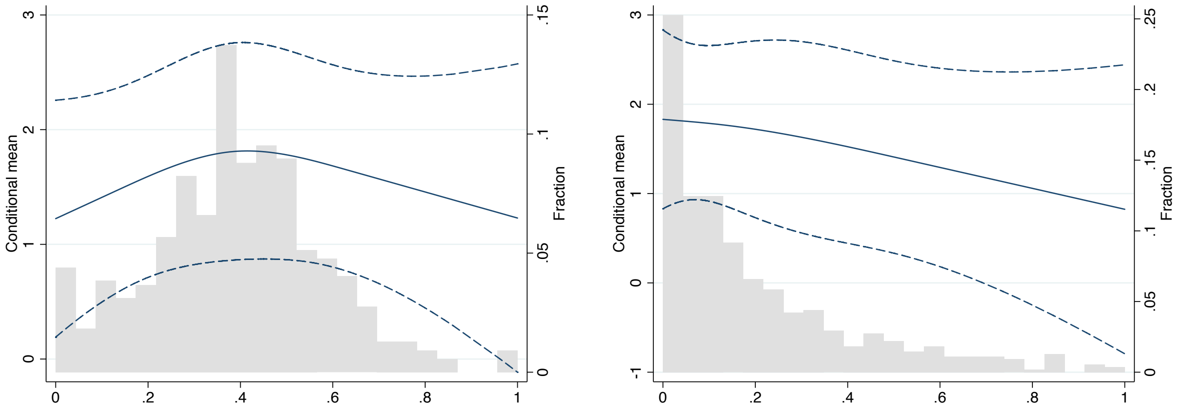

(c) Agriculture employment share, 1910

(d) Manufacturing employment share, 1910

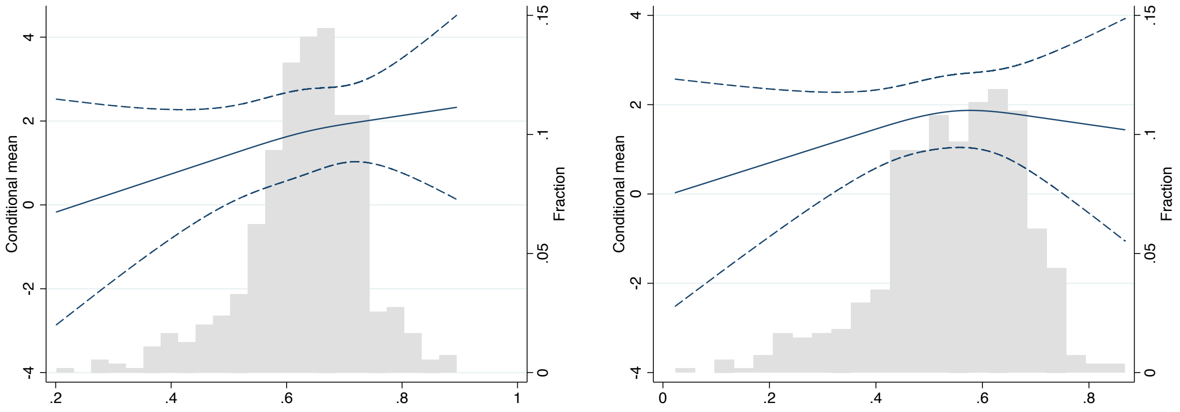

(e) Percent of farm acreage in cotton, 1910

(f) Log population density, 1910

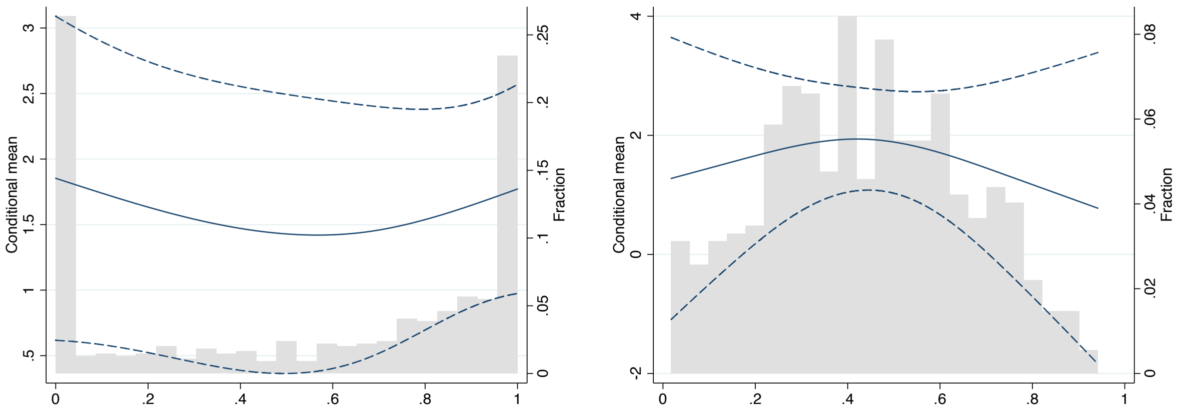
Figure A.15: Nonlinear Relationship between Covariates and Birth County Network Index Estimates, Southern Black Migrants


(g) Church members per capita, 1916

(h) Rosenwald school exposure

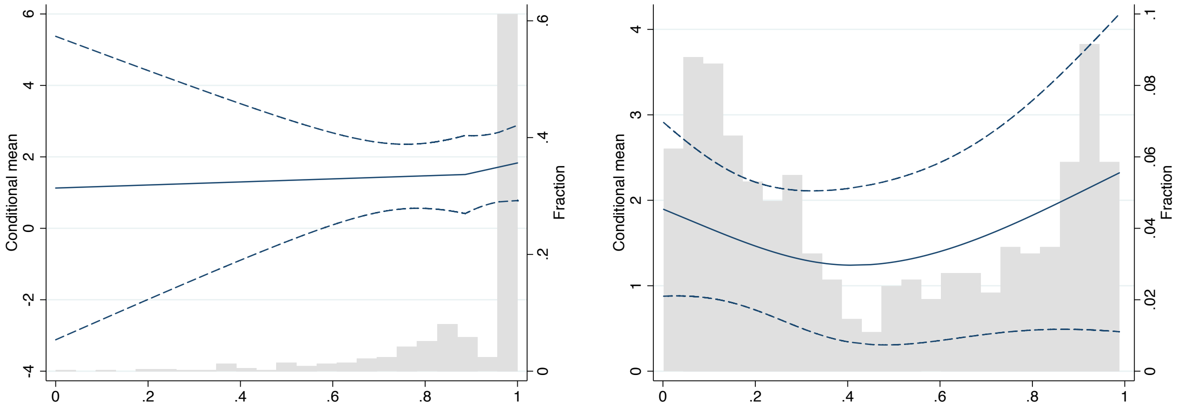

(i) Literacy rate (10+)

(j) School attendance rate (6–14)


(k) Railroad exposure

(l) Percent African-American, 1910

Figure A.15: Nonlinear Relationship between Covariates and Birth County Network Index Estimates, Southern Black Migrants



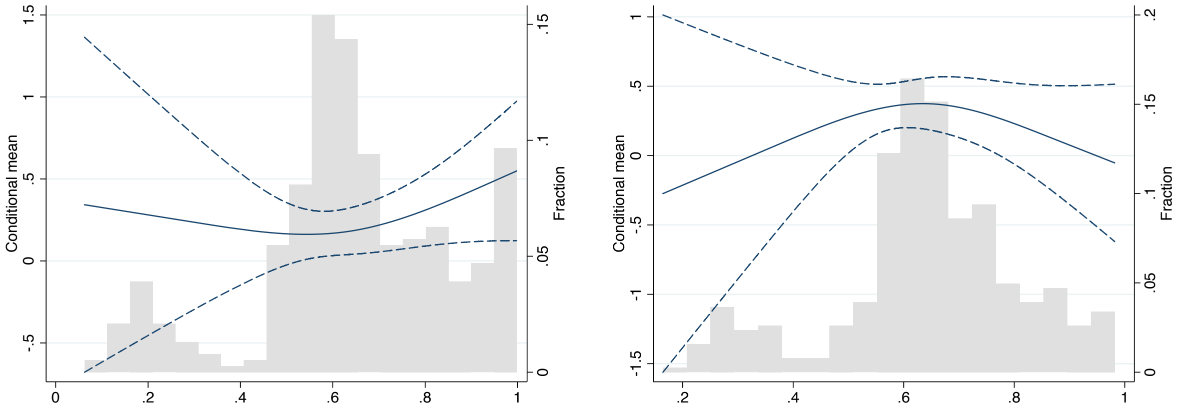
(m) Percent rural, 1910

(n) Percent voting for Strom Thurmond, 1948

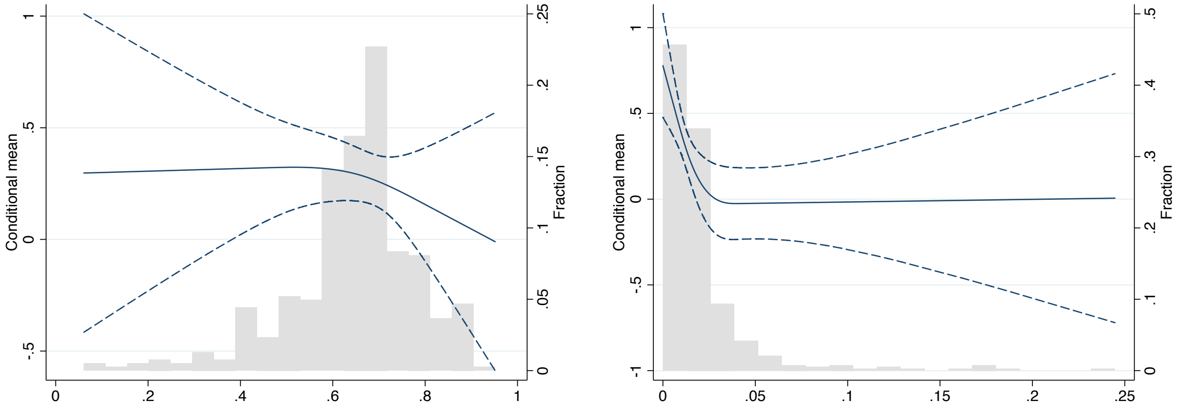
Notes: The solid blue line is the conditional mean of the birth county network index as a function of the indicated independent variable. Dashed lines are 95 percent confidence intervals. These results come from regressing birth county network index estimates against restricted cubic splines in the 14 indicated variables. Grey bars are histograms of the underlying independent variable (right scale). For panels A, B, C, D, F, G, I, and J, the explanatory variables are measured for African Americans.

Sources: Duke SSA/Medicare data, Aaronson and Mazumder (2011) data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), ICPSR (1999)

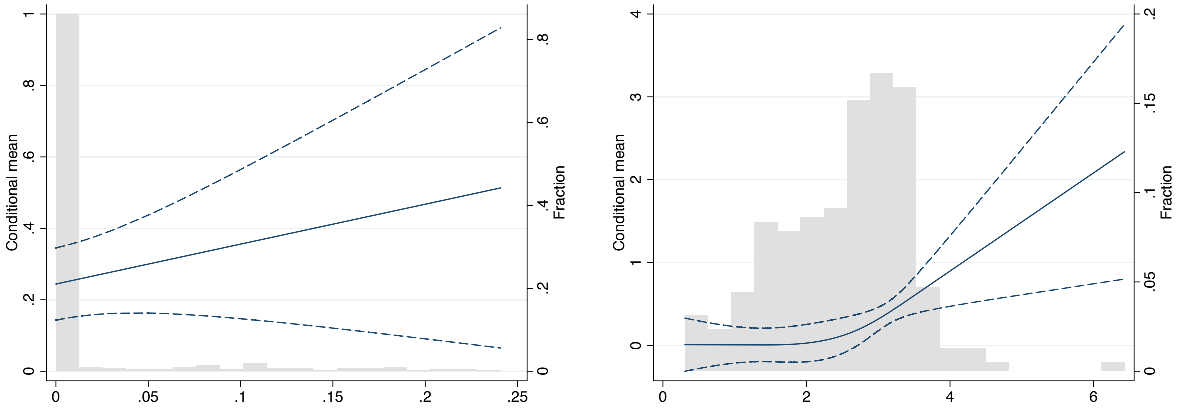
Figure A.16: Nonlinear Relationship between Covariates and Birth County Network Index Estimates, Great Plains White Migrants


(a) Percent of farmers who are owners, 1910

(b) Percent in owner-occupied housing, 1910

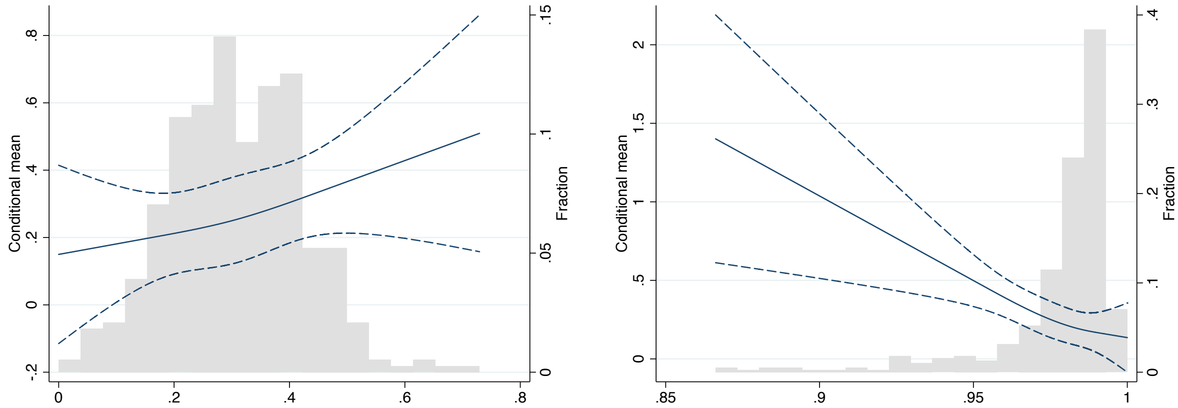

(c) Agriculture employment share, 1910

(d) Manufacturing employment share, 1910


(e) Percent of farm acreage in cotton, 1910

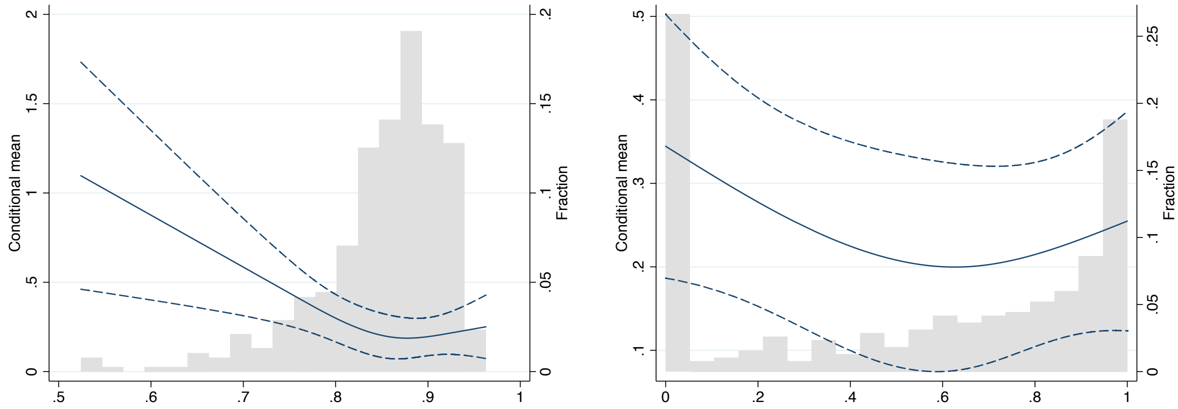
(f) Log population density, 1910

Figure A.16: Nonlinear Relationship between Covariates and Birth County Network Index Estimates, Great Plains White Migrants



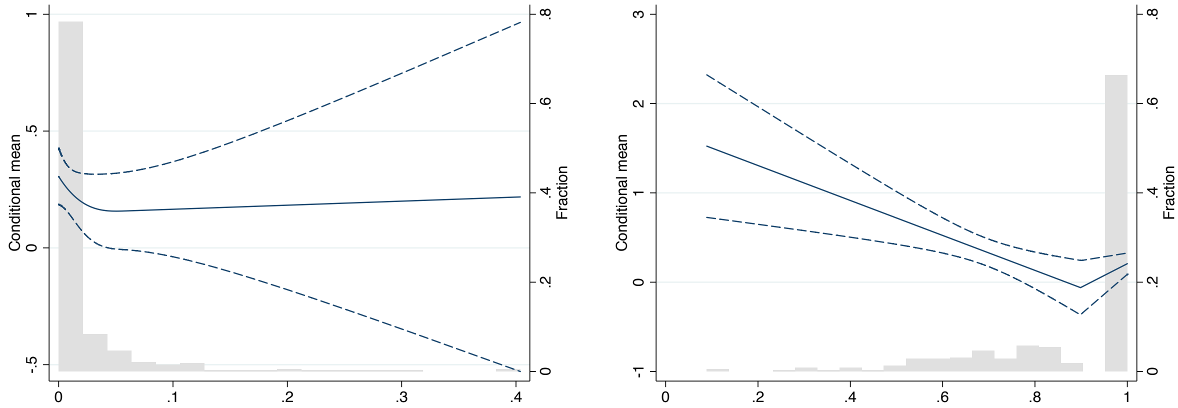
(g) Church members per capita, 1916

(h) Literacy rate (10+)



(i) School attendance rate (6–14)

(j) Railroad exposure



(k) Percent African-American, 1910

(l) Percent rural, 1910

Notes: See notes to Figure A.15. For panels A, B, C, D, F, G, I, and J, the explanatory variables are measured for whites.

Sources: Duke SSA/Medicare data, Aaronson and Mazumder (2011) data, Black et al. (2015) data, Census (1992), Haines and ICPSR (2010), ICPSR (1999)