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

Bryan A. Stuart
University of Michigan
bastuart@umich.edu

Evan J. Taylor
University of Michigan
evanjt@umich.edu

July 4, 2017

Abstract

This paper examines the role of social interactions in location decisions. We study over one million long-run location decisions made during two landmark migration episodes by African Americans from the U.S. South and whites from the Great Plains. We develop a new method to estimate the strength of social interactions for each receiving and sending location. Social interactions strongly influenced the location decisions of black migrants, but were less important for white migrants. Social interactions were particularly important in providing African American migrants with information about attractive employment opportunities and played a larger role in less costly moves.

JEL Classification Codes: J61, N32, O15, R23, Z13

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

*Thanks to 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 seminar participants at the Economic Demography Workshop, Trans-Atlantic Doctoral Conference, University of Michigan, and Urban Economics Association for helpful comments and discussions. 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. Any errors are our own.

1 Introduction

A large and growing literature finds that social interactions influence many economic outcomes, including crime, education, and employment (for recent reviews, see Blume et al., 2011; Epple and Romano, 2011; Munshi, 2011; Topa, 2011). While research has long-recognized the importance of migration, there is little evidence that isolates the role of social interactions in location decisions, and even less evidence on the types of individuals or economic conditions for which social interactions are most important. Evidence on social interactions in location decisions can inform theoretical models of migration, the equilibrium of local labor markets, and the impacts of policies that affect migration incentives.

This paper provides new evidence on the magnitude and nature of social interactions in 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. The millions of moves in these episodes yield particularly valuable settings for studying the long-run effects of social interactions on location decisions. We use confidential 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. Detailed geographic information allows us to distinguish birth town-level social interactions from other determinants of location decisions, such as expected wages or moving costs. 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.

We develop a new, intuitive method of characterizing social interactions in location decisions. The social interactions (SI) index allows us to estimate the strength of social interactions for each receiving and sending location, which we then relate to locations' economic characteristics. We show that existing methods may mischaracterize the strength of social interactions in our setting. In particular, the influential approach of Bayer, Ross and Topa (2008) could estimate strong social interactions for popular destinations even if true social interactions are relatively weak, and as a result could misstate the overall strength of social interactions. Our method does not suffer from this problem. Under straightforward and partly testable assumptions, the SI index identifies the effect of social interactions and maps directly to structural social interactions models.

We find very strong social interactions among Southern black migrants and smaller interactions among whites. Our estimates imply that if we observed one randomly chosen African American move from a birth town to a destination county, then on average 1.9 additional black migrants from that birth town would make the same move. For white migrants from the Great Plains, the average is only 0.4, and results for Southern whites are similarly small. Interpreted through the social interactions model of Glaeser, Sacerdote and Scheinkman (1996), our estimates imply that 49 percent of African American migrants chose their long-run destination because of social interactions, while 16 percent of Great Plains whites were similarly influenced.

To understand the nature of social interactions in location decisions, we examine whether economic characteristics of receiving and sending locations are associated with stronger social interactions. Social interactions among African Americans are stronger in destinations with a higher share of 1910 employment in manufacturing, a particularly attractive sector for black workers in our sample. This evidence highlights an important role for network-based information about employment opportunities or job referrals in determining location decisions. We also find that social interactions are stronger in destinations that were closer and more connected by railroads, pointing to the importance of access to information and low moving costs. In addition, social interactions are stronger in sending counties with higher literacy rates in 1920, suggesting that education and related factors facilitated social interactions. Social interactions among whites are less sensitive to employment opportunities and moving costs.

Several pieces of evidence support the validity of our empirical strategy. Our research design 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. This design implies that destination-level SI index estimates should not change when controlling for observed 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 social interactions in certain locations, like Rock County, Wisconsin, for which rich qualitative work supports our findings (Bell, 1933; Rubin, 1960; Wilkerson, 2010).

This paper makes three contributions. First, we develop a new method of characterizing the magnitude and nature of social interactions. 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 social interactions in a variety of other settings. Second, we provide new evidence on the importance of social interactions in location decisions and the types of individuals and economic conditions for which social interactions are most important. Previous work shows that individuals tend to migrate to the same areas, often broadly defined, as other individuals from the same town or country, but does not isolate the role of social interactions (Bartel, 1989; Bauer, Epstein and Gang, 2005; Beine, Docquier and Ozden, 2011; Giuletti, Wahba and Zenou, 2014; Spitzer, 2016).[1] Third, our results inform landmark migration episodes that have drawn interest from economists for a century (Scroggs, 1917; Smith and Welch, 1989; Carrington, Detragiache and Vishwanath, 1996; Collins, 1997; Boustan, 2009, 2010; Hornbeck, 2012; Hornbeck and Naidu, 2014; Johnson and Taylor, 2016; Black et al., 2015; Collins and Wanamaker, 2015). Our results complement the small number of interesting but unrepresentative historical accounts suggesting that social interactions were important in these migration episodes (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 non-linear relationship accords with a network formation model with fixed costs of participation. We find some evidence that social interactions were stronger in denser sending communities, consistent with the results in Chay and Munshi (2015). We differ in our empirical methodology, study of white migrants from the Great Plains and South, and examination of how social interactions vary across destinations.

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

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 1910-1930, slowed during the Great Depression, and then resumed forcefully from 1940 to the 1970s.

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). Underlying causes included 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: 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 their role diminished sharply after the 1920s, and most individuals paid for the relatively expensive train fares themselves (Gottlieb, 1987; Grossman, 1989).[2] African American newspapers from the largest destinations circulated throughout the South, providing information on life in the North (Gottlieb,

[2] 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).

1987; Grossman, 1989).[3]

A small number of historical accounts suggest a role for social interactions 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.[4] These accounts motivate our focus on social interactions among migrants from the same town.

The experience of John McCord captures many important features of early black migrants' location decisions.[5] Born in Pontotoc, Mississippi, nineteen-year-old McCord traveled in search of higher wages in 1912 to Savannah, Illinois, where a fellow Pontotoc-native connected him with a job. McCord moved to Beloit, Wisconsin in 1914 after hearing of employment opportunities and quickly began working as a janitor at the manufacturer Fairbanks Morse and Company. After two years in Beloit, McCord spoke to his manager about returning home for a vacation. The manager asked McCord to recruit workers during the trip, and McCord returned with 18 unmarried men, all of whom were soon hired. Thus began a persistent flow of African Americans from Pontotoc to Beloit: among individuals born from 1916-1936, 14 percent of migrants from Pontotoc lived in Beloit's county at old age (Table 2, discussed below).

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 social interactions in location decisions (Jamieson, 1942; Gregory, 1989). For example,

[3]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).

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

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

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. These episodes, which took place around the same time, saw a similar share of the population undertake long-distance moves. Figure 2 shows that 97 percent of blacks born in the South and 90 percent of whites born in the Great Plains lived in their birth region in 1910, and out-migration reduced this share to 75 percent for both groups by 1970. In addition, both African American and white migrants experienced discrimination in many destinations, although African Americans faced more severe discrimination and had less wealth (Gregory, 2005).

3 Estimating Social Interactions in Location Decisions

3.1 Data on Location Decisions

We use confidential administrative data to measure location decisions made during these historical migration episodes. In particular, 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, 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.[6] This sample lies at the center of both mass migration episodes, and out-migration rates for the 1916-1936 cohorts are among the highest of all cohorts for both episodes (Appendix Figure A.1). As seen in Figure 1, 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

[6]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.

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 3 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.[7] 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.[8] 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 count that person as a migrant. 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 below.

3.2 Econometric Model: The Social Interactions Index

We first introduce some notation and discuss the basic idea underlying our approach to estimating social interactions.[9] Let $D_{i,j,k} = 1$ if migrant i moves from birth town j to destination county k and $D_{i,j,k} = 0$ if migrant i moves elsewhere. The probability of a migrant born in town j choosing destination k is $P_{j,k} \equiv \mathbb{E}[D_{i,j,k}]$. 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 number of people who move from birth town j to destination k is $N_{j,k} \equiv \sum_{i \in j} D_{i,j,k}$, and the number of migrants from birth town j is $N_j \equiv \sum_k N_{j,k}$.

Positive social interactions increase the variance of individuals' decisions (Glaeser, Sacerdote

[7] These include the seven states already listed, plus Arkansas, Tennessee, Texas, and Virginia.

[8] This border region includes Arkansas, Colorado, Iowa, Minnesota, Missouri, Montana, New Mexico, Texas, and Wyoming.

[9] See Brock and Durlauf (2001) and Blume et al. (2011) for comprehensive discussions of various approaches to estimating social interactions.

and Scheinkman, 1996; Bayer, Ross and Topa, 2008; Graham, 2008). To see this, imagine that we observed multiple realizations of $N_{j,k}$ from a fixed data generating process. The across-realization variance of location decisions for a single birth town-destination county pair would be

$$\begin{aligned}\mathbb{V}[N_{j,k}] &= \sum_{i\in j}\mathbb{V}[D_{i,j,k}] + \sum_{i\neq i'\in j}\mathbb{C}[D_{i,j,k}, D_{i',j,k}] \\ &= N_j P_{j,k}(1-P_{j,k}) + N_j(N_j-1)C_{j,k},\end{aligned} \tag{1}$$

where $C_{j,k} \equiv \sum_{i\neq i'\in j}\mathbb{C}[D_{i,j,k}, D_{i',j,k}]/(N_j(N_j-1))$ is the average covariance of location decisions for two migrants from the same town. Positive social interactions ($C_{j,k} > 0$) clearly increase the variance of location decisions. If, counterfactually, we observed multiple realizations of $N_{j,k}$, we could directly estimate $\mathbb{V}[N_{j,k}]$ and $P_{j,k}$, and then estimate of $C_{j,k}$ using equation (1). Because we observe a single realization of location decisions for each (j,k) pair, we use an econometric model to estimate social interactions.

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 social interactions. Extending their model to our setting yields

$$D_{i,j(i),k}D_{i',j(i'),k} = \alpha_{g,k} + \sum_{j\in g}\beta_{j,k}1[j(i) = j(i') = j] + \epsilon_{i,i',k}, \tag{2}$$

where $j(i)$ is the birth town of migrant i, and both i and i' live in birth town group g. As described below, 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 propensity of migrants from the same birth town j to co-locate in k.[10] Equation (2) allows location decision determinants to vary arbitrarily at the birth town group-destination level through $\alpha_{g,k}$ (e.g., because of differences in migration costs

[10]Bayer, Ross and Topa (2008) study the propensity of workers that live in the same census block to work in the same census block, beyond the propensity of workers living in the same block group (a larger geographic area) to work in the same block. 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').

due to railroads or highways).

To better understand the reduced-form model in equation (2), we map the parameters of the extended Bayer, Ross and Topa (2008) model, $(\alpha_{g,k}, \beta_{j,k})$, into classic parameters governing social interactions, $(P_{j,k}, C_{j,k})$. 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. For example, this assumes away the possibility that migrants from Pigeon Creek, Alabama had preferences or human capital particularly suited for Niagara Falls, New York relative to migrants from a nearby town, such as Oaky Streak, which is six miles away. This assumption attributes large differences in realized moving propensities across nearby towns to social interactions.

Assumption 1 is plausible in our setting. Preferences for destination features (e.g., wages or climate) likely did not vary sharply across nearby birth towns, and individuals had little information about most destinations outside of what was relayed through social networks. Furthermore, migrants tended to work in different industries (Appendix Table A.1), suggesting a negligible role for human capital specific to a destination county that differed across nearby towns. Conditional on migrating, the cost of moving to a given destination likely did not vary sharply across nearby towns.[11] We do not restrict the probability of moving from birth town group g to destination k, $P_{g,k}$, which allows destinations to vary in their attractiveness to migrants for myriad reasons.

The second assumption is that social interactions occur only among individuals from the same birth town:

Assumption 2. $\mathbb{C}[D_{i,j,k}, D_{i',j',k}] = 0$ for individuals 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 classic parameters governing social interactions, $(P_{j,k}, C_{j,k})$. Positive social interactions across nearby towns, which violate Assumption 2, would lead us to underesti-

[11] 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.

mate the strength of town-level social interactions.

Under Assumptions 1 and 2, the slope coefficient in equation (2) equals the covariance of location decisions from birth town j to destination k: $\beta_{j,k} = C_{j,k}$.[12] In addition, the fixed effect in equation (2) 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 social interactions.

In certain settings, the Bayer, Ross and Topa (2008) model could mischaracterize the strength of social interactions. 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 birth town j making the same move. Slight manipulation of the definition of the covariance of location decisions yields

$$C_{j,k} = P_{g,k}\left(\mu_{j,k} - P_{g,k}\right). \tag{3}$$

Equation (3) shows that variation in $C_{j,k}$ arises from two sources: the probability of moving to a destination, $P_{g,k}$, and the "marginal social interaction effect," $\mu_{j,k} - P_{g,k}$. For example, $C_{j,k}$ could be large for a popular destination like Chicago 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, $C_{j,k}$, or any aggregation of $C_{j,k}$ is not an attractive measure of social interactions.[13]

To characterize the strength of social interactions, we propose an intuitive social interactions (SI) index that equals the expected increase in the number of people from birth town j that move

[12] Proof:

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

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

[13] This issue applies in general when using the covariance of decisions to estimate social interactions. For example, there is considerable variation in the probability of working at specific locations or establishments.

to destination county k when an arbitrarily chosen person i is observed to make the same move,

$$\Delta_{j,k} \equiv \mathbb{E}[N_{-i,j,k}|D_{i,j,k}=1] - \mathbb{E}[N_{-i,j,k}|D_{i,j,k}=0], \tag{4}$$

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 positive social interactions in moving from j to k, while $\Delta_{j,k}=0$ indicates no social interactions.

The SI index, $\Delta_{j,k}$, possesses several attractive properties as a method of measuring social interactions. The SI index permits meaningful comparisons of social interactions across heterogeneous receiving and sending locations. The SI index also requires minimal assumptions about the specific behaviors that lead to social interactions. For example, correlated location decisions could arise because individuals value living near their friends and family or because social networks provide information about job opportunities. The SI 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 SI index for each destination k depends only on the structural parameter s ($\Delta_{j,k}=s-1$), while the covariance of location decisions depends on the moving probability ($C_{j,k}=(s-1)P_{g,k}(1-P_{g,k})/(N_j-1)$). As another example, we connect the SI index to the model of Glaeser, Sacerdote and Scheinkman (1996) in Section 4.5. In addition, the SI index can be estimated non-parametrically with increasingly available data.

In Appendix A, we show that the SI index can be written as

$$\Delta_{j,k} = \frac{(\mu_{j,k}-P_{g,k})(N_j-1)}{1-P_{g,k}} = \frac{C_{j,k}(N_j-1)}{P_{g,k}-P_{g,k}^2}. \tag{5}$$

Several features of equation (5) are noteworthy. First, the SI index depends on the classic parameters governing social interactions, $(P_{g,k}, C_{j,k})$. Second, the SI index increases in the marginal social interaction effect, $\mu_{j,k}-P_{g,k}$. If migrants move independently of each other, then $\mu_{j,k}-P_{g,k} = \Delta_{j,k}=0$. Third, the SI index scales down $C_{j,k}$ for more popular destinations, as $P_{g,k} << 0.5$ is the relevant range in our setting. Finally, the SI index does not necessarily increase in the number of migrants from birth town j, N_j, as the marginal social interaction effect might decrease in N_j.[14]

The SI index captures social interactions that generate location decisions of migrants from the same birth town that are correlated more strongly than is predicted by the location decisions of migrants from nearby towns. While other forms of social interactions exist, the SI index does not measure them. For example, if social interactions affected whether individuals migrated, but not where they moved, then the SI index would equal zero. Relatedly, the SI 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.[15]

3.3 Estimating the Social Interactions Index

As suggested by equation (5), estimating the SI index is straightforward. We first define birth town groups, and then non-parametrically estimate the underlying parameters $P_{g,k}$, $P^2_{g,k}$, 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 cross validation.[16] Given x^*, the location of the grid is determined by a single latitude-longitude reference point. SI index estimates are very similar across four different reference points, so we average estimates across them.[17]

[14] 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.

[15] We allow social interactions to influence out-migration, but we do not directly examine this channel.

[16] 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.3 reports the values of x^* chosen by cross-validation.

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

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 widely in average county size. We emphasize results based on cross validation in the main text and include results based on counties in the appendix.[18]

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,

$$\widehat{P_{g,k}} = \frac{\sum_{j \in g} N_{j,k}}{\sum_{j \in g} N_j}. \tag{6}$$

We estimate the squared moving probability using the closed-form solution implied by equation (2),[19]

$$\widehat{P^2_{g,k}} = \frac{\sum_{j \in g} \sum_{j' \neq j \in g} N_{j,k} N_{j',k}}{\sum_{j \in g} \sum_{j' \neq j \in g} N_j N_{j'}}, \tag{7}$$

and the covariance of location decisions using the closed-form solution implied by equation (2),

$$\widehat{C_{j,k}} = \frac{N_{j,k}(N_{j,k}-1)}{N_j(N_j-1)} - \widehat{P^2_{g,k}}. \tag{8}$$

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

Given $(\widehat{P_{g,k}}, \widehat{P^2_{g,k}}, \widehat{C_{j,k}}, N_j)$, we can estimate the SI index, $\Delta_{j,k}$, using equation (5). However,

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

[19] Equation (7) 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.

each estimate $\widehat{\Delta_{j,k}}$ depends primarily on a single birth town observation. To conduct inference, increase the reliability of our estimates, and decrease the number of parameters reported, we aggregate SI index estimates across all birth towns in each state,

$$\widehat{\Delta_k} = \sum_j \left(\frac{\widehat{P_{g(j),k}} - \widehat{P^2_{g(j),k}}}{\sum_{j'} \widehat{P_{g(j'),k}} - \widehat{P^2_{g(j'),k}}} \right) \widehat{\Delta_{j,k}}, \tag{9}$$

where $g(j)$ is the group of town j. The weights in equation (9) arise naturally from assuming that $\Delta_{j,k}$ does not vary across birth towns within a state.[20] The destination-level SI 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 SI index estimates by aggregating across destinations and towns within birth county c,

$$\widehat{\Delta_c} = \sum_k \sum_{j \in c} \left(\frac{\widehat{P_{g(j),k}} - \widehat{P^2_{g(j),k}}}{\sum_{k'} \sum_{j' \in c} \widehat{P_{g(j'),k'}} - \widehat{P^2_{g(j'),k'}}} \right) \widehat{\Delta_{j,k}}. \tag{10}$$

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

To facilitate exposition, we have described estimation of the SI index in terms of four distinct components, $(\widehat{P_{g,k}}, \widehat{P^2_{g,k}}, \widehat{C_{j,k}}, N_j)$. However, SI index estimates depend only on observed population flows, and equation (9) forms the basis of an exactly identified generalized method of moments (GMM) estimator. To estimate the variance of $\widehat{\Delta_k}$, we treat the birth town group as the unit of observation and use a GMM variance estimator. This is akin to calculating heteroskedastic robust standard errors clustered by birth town group.[21] Appendix B contains details.

[20]When assuming $\Delta_{j,k} = \Delta_k \forall j$, the derivation in Appendix A yields $\Delta_k = \left(\sum_j C_{j,k}(N_j - 1)\right) / \left(\sum_j P_{g(j),k}(1 - P_{g(j),k})\right)$, which leads directly to the estimator in equation (9).

[21]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.

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-level covariates to explain moving probabilities would not affect SI index estimates.

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

$$P_{j,k} = \rho_{g,k} + X_j \pi_k, \tag{11}$$

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. X_j contains an indicator for being along a railroad, an indicator for having above-median black population share, and four indicators corresponding to population quintiles.[22] These covariates, available from the Duke SSA/Medicare data and the railroad information used in Black et al. (2015), capture potentially relevant determinants of location decisions. For example, migrants born in larger towns might have had more human capital or information, and these resources might have made certain destinations more attractive, causing our SI index estimates to reflect variables correlated with birth town size instead of social interactions.

To implement this extension, we construct an alternative SI index estimate using an alternative moving probability estimate, $\widetilde{P_{j,k}}$, equal to the fitted value from the OLS regression

$$\frac{N_{j,k}}{N_j} = \rho_{g,k} + X_j \pi_k + e_{j,k}. \tag{12}$$

We use fitted values from a separate OLS regression, implied by equation (11), to form an alter-

[22] We construct percentiles for black population share and population separately for each birth state.

native squared moving probability estimate, $\widetilde{P^2_{j,k}}$.[23] We estimate all equations separately for each birth state.[24] Similarity between the baseline and alternative SI index estimates would provide support for our empirical strategy.[25]

4 Results: Social Interactions in Location Decisions

4.1 Social Interactions Index Estimates

Table 1 provides an overview of the long-run population flows that we use to estimate social interactions. 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 blacks born in the South and 35 percent of whites born in the Great Plains lived outside their birth region, while only nine percent of whites born in the South lived elsewhere.[26] We focus on Southern-born blacks and Great Plains-born whites, and leave results for Southern-born whites for the appendix. Appendix Table A.2 shows that, on average, there were 142 migrants per birth town for Southern African Americans and 181 migrants per birth town for Great Plains whites.

We begin with some examples that illustrate how we identify social interactions. Table 2 shows the birth town to destination county migration flows that would be most unlikely in the absence of social interactions. Panel A shows that 10-50 percent of African American migrants from each of 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

[23] 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, $j \neq j'$.

[24] When estimating the variance of our SI index estimates under this extension, we ignore the variance that arises because $\widetilde{P_{j,k}}$ and $\widetilde{P^2_{j,k}}$ rely on OLS estimates. Accounting for this variance would make our estimates with and without covariates appear even more similar when performing statistical tests.

[25] An alternative approach to assessing the validity of Assumption 1 is testing whether the parameter vector $\pi_k = 0$ in equation (12). We prefer to test the difference in SI index estimates because this approach allows us to consider the statistical and substantive significance of any differences.

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

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 SI index estimates. The results in Panel B for Great Plains whites are similar.

To summarize the importance of social interactions for all location decisions in our data, Table 3 reports averages of destination-level SI index estimates. For African Americans, unweighted averages of the destination-level SI index, $\widehat{\Delta_k}$ 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 social interactions are stronger 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 SI index estimates is 1.94; this means that when we observe one randomly chosen African American move from a birth town to a destination county, then on average 1.94 additional black migrants from that birth town would make the same move. Panel B contains results for white moves out of the Great Plains. The weighted average of destination-level SI index estimates for whites is 0.38, only one-fifth the size of the black average.[27] It appears that African American migrants relied more heavily on social networks in making their long-run location decisions. Historical context suggests that one explanation is that African Americans used social networks to overcome their lack of resources and the discrimination they faced in many destinations.

We provide a more complete picture of social interactions in Figure 4, which plots the distribution of destination-level SI index estimates.[28] Across the board, SI index estimates for African

[27] Appendix Table A.3 displays the lengths of the square grid chosen by cross validation. Appendix Table A.4 shows that results are similar when we define birth town groups using counties. For Southern blacks, the linear (rank) correlation between the destination-level SI 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.5 shows that average SI index estimates for whites from the South are somewhat smaller than for whites from the Great Plains.

[28] Appendix Figure A.7 displays the associated t-statistic distributions, and Appendix Figures A.8 and A.9 display

Americans are larger than those for whites. Social interactions are particularly strong for some destinations, especially for black migrants, and relatively weak for most destinations. Below, we examine whether destinations' economic characteristics can explain this heterogeneity.

To examine social interactions more closely, Figure 5 plots the spatial distribution of destination-level SI index estimates for Mississippi-born blacks. We estimate strong social interactions for several destinations: 23 counties have a SI index estimate greater than 3 and 58 counties have a SI 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.[29] We estimate strong social interactions ($\widehat{\Delta_k} > 3$) in Rock County, Wisconsin, consistent with historical accounts of blacks who moved from Mississippi to Beloit, which is located there (Bell, 1933; Rubin, 1960; Wilkerson, 2010). Figure 6 maps the destination-level SI index estimates for whites from North Dakota. We find little evidence of strong social interactions, although one exception is San Joaquin county ($\hat{\Delta}_k > 3$), an area described memorably in *The Grapes of Wrath* (Steinbeck, 1939).[30] Unlike black migrants, whites moved to a large number of destinations throughout the U.S. The difference between the number of destinations chosen by Mississippi blacks and North Dakota whites is striking, especially because our data contain almost 30,000 more migrants from Mississippi. Appendix Figures A.10 and A.11, for Southern Carolina-born blacks and Kansas-born whites, show similar patterns.

To assess the validity of our empirical strategy, we examine whether SI index estimates change when using birth town covariates to explain moving probabilities, as discussed in Section 3.4. Table 4 reports weighted averages of destination-level SI index estimates with and without covariates. When we examine birth states individually, there are no substantively or statistically significant differences between the two sets of estimates. When pooling all Southern states together, the esti-

analogous results for whites from the South. A destination county can appear multiple times in these figures because we estimate destination-level SI indices separately for each birth state.

[29]In Figure 5, the counties in white received less than 10 migrants.

[30]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.

mates are very similar in magnitude (1.94 and 1.92) and statistically indistinguishable ($p = 0.76$). When pooling all Great Plains states together, the estimates again are very similar in magnitude (0.38 and 0.36), but are statistically distinguishable ($p = 0.02$). The destination-level SI index estimates with and without covariates are highly correlated: the linear (rank) correlation is 0.914 (0.992) for blacks from the South and 0.939 (0.988) for whites from the Great Plains. On net, this evidence indicates that geographic proximity adequately controls for the relevant determinants of location decisions and supports the validity of our empirical strategy.

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 SI index estimates on these dimensions. Birth town size could be correlated with unobserved determinants of social interactions and location decisions, such as the level of social and human capital or information about destinations. However, it is not clear beforehand whether social interactions will vary with the size of receiving or sending locations. For reference, column 1 of Table 5 reports weighted averages of destination-level SI 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 SI index.[31] 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.[32] We exclude both large birth towns and large destinations in column 4. The average SI index estimates are similar across all four specifications for both Southern blacks and Great Plains whites.[33]

A widely noted feature of the Great Migration is the tendency of migrants to move along vertical pathways established by railroads, which reduced the cost of moving to destinations on the

[31]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

[32]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.

[33]Appendix Table A.6 reports similar results for Southern-born whites.

same line and increased the flow of information. Social interactions might have benefited from reduced moving costs and increased information, or social interactions might have facilitated moves to destinations requiring higher moving costs. Table 6 displays weighted averages of destination-level SI index estimates for different regions, demonstrating that social interactions among African Americans clearly follow vertical migration patterns. The largest SI index estimates in the Northeast come from the Carolinas, while the largest estimates in the Midwest are among migrants from Mississippi and Alabama, and the largest estimates in the West come from Louisiana.[34] Panel B displays results for Great Plains whites.[35] Social interactions among Great Plains whites were much stronger in the Midwest and West, where moving costs were lower, than the Northeast or South. These patterns suggest that lower moving costs and greater information facilitated social interactions.

To further understand the nature of social interactions, we examine whether the location decisions of African American migrants influenced white migrants from the same Southern birth town, and vice versa. While blacks and whites could have shared information about opportunities in the North, the high segregation in the Jim Crow South makes cross-race social interactions unlikely. Appendix C describes how we estimate cross-race social interactions. Appendix Table A.8 displays little evidence of cross-race interactions, indicating that social interactions operated within racial groups. In addition, there is little correlation between destination-level SI index estimates for blacks and whites from the South: the linear (rank) correlation is 0.076 (0.149). This implies that our SI index estimates do not simply reflect unobserved characteristics of certain Southern towns.

4.2 Addressing Measurement Error due to Incomplete Migration Data

SI 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

[34] We define regions slightly differently than the Census Bureau because we treat the former Confederate states as the South. The Census South region includes Delaware, D.C., Maryland, West Virginia, Kentucky, and Oklahoma. We include the first four states in the Northeast and the latter two in the Midwest.

[35] Appendix Table A.7 reports averages by region for Southern-born whites.

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 (5) implies that SI index estimates will be attenuated because we undercount the number of migrants from each town, N_j.

More specifically, suppose that we are interested in the effect of social interactions on location decisions at age 40. Denote the number of migrants that survive to age 40 by N_j^{40}, and assume for simplicity that this equals the observed number of migrants divided by a scaling factor, $N_j^{40} = N_j/\alpha$. To approximate the coverage rate α, we divide the number of individuals in the Duke/SSA Medicare data by the number of individuals in decennial census data.[36] The overall coverage rate is 52.3% for African Americans from the South and 69.3% for whites from the Great Plains (see Appendix Table A.9), which implies that $N_j^{40} \approx 1.91N_j$ for Southern blacks and $N_j^{40} \approx 1.44N_j$ for Great Plains whites. As an approximate measurement error correction, SI index estimates should be multiplied by a factor of 1.91 for Southern blacks and 1.44 for Great Plains whites. Appendix Table A.10 presents results that reflect state-specific coverage rate adjustments. The weighted average of destination-level SI index estimates is 3.71 for Southern blacks and 0.55 for Great Plains whites. Adjusting for incomplete data under a missing at random assumption increases both the magnitude of SI index estimates and the black-white social interactions gap.

Appendix D discusses the consequences of measurement error when we relax the missing at random assumption. We derive a lower bound on the SI index and show that estimates of this lower bound still reveal sizable social interactions.

4.3 The Role of Family Migration

The SI index might capture the influence 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 social interactions occur within the family. Unfortunately,

[36] We use the 1960 Census to construct coverage rates for individuals born from 1916-1925 and the 1970 Census for individuals born from 1926-1935.

we do not observe family relationships and have limited ability to study this question directly. However, we can examine whether our results stem entirely from the migration of male-female couples. If this were true, there would be no social interactions among men only or women only. Appendix Table A.10 shows that SI index estimates are similar in magnitude among men and women, implying that our results do not simply reflect the migration of couples.[37] 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 social interactions 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 vs. 5.25). This difference is too small to explain our finding that average SI index estimates are 410 percent larger among blacks than whites.[38] 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 social interaction strength. In sum, differences in family size might explain some, but not all, of the differences in social interactions between black and white migrants.[39]

4.4 Social Interactions and Economic Characteristics of Receiving and Sending Locations

To better understand why social interactions affected location decisions, we relate SI index estimates to economic characteristics of receiving and sending locations.

[37]The similarity between men and women is not surprising given the relative sex balance among migrants in this period (Gregory, 2005).

[38]The weighted average of SI index estimates in Table 3 is 1.938 for blacks and 0.380 for whites, and (1.938-0.380)/0.380 = 4.1. When adjusting for incomplete migration data under the missing at random assumption (Appendix Table A.10), social interactions among African Americans are 582 percent larger than among Great Plains whites.

[39]Conditional on family size, black and white migrants could have differed in the extent to which they tended to follow other family members. Our data do not allow us to examine this possibility.

We first consider the characteristics of receiving locations. Employment opportunities were among migrants' most important consideration, and relatively high wages made manufacturing jobs particularly attractive. In the presence of imperfect information among workers 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 almost certainly had more information about the largest destinations, the imperfect information channel suggests a stronger relationship between social interactions and manufacturing employment intensity in smaller destinations. In contrast, if information about employment opportunities was widespread, then social interactions might not be stronger in destinations with more manufacturing. Similar patterns could arise if workers relied on their social networks for job referrals instead of information.[40] 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 physical distance and railroads, represented another key consideration. Lower moving costs could have fostered social interactions 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 SI index estimates on county-level covariates. Column 1 of Table 7 shows that social interactions among African Americans are larger in destinations with a higher 1910 manufacturing employment share: a one standard deviation increase in the 1910 manufacturing employment share is associated with an increase in the SI index of 0.22 people.[41] Column 2 shows that the positive relationship between manufacturing employment and social interactions is almost seven times larger in smaller destinations.[42] There

[40]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).

[41]Appendix Table A.11 contains summary statistics. Appendix Figure A.12 plots the bivariate relationship between SI index estimates and 1910 manufacturing employment share, showing the considerable variation in the manufacturing employment share across destinations.

[42]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).

is little relationship between the strength of social interactions and the agriculture employment share. We also find that social interactions are stronger in destinations that were closer to and could be reached by rail directly or with one stop from migrants' birth state. Social interactions are stronger in destinations with a smaller black population share in 1900, suggesting that networks helped migrants find opportunities in new places. 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.[43] Columns 4-6 present results for white migrants from the Great Plains. For this group, there is little relationship between the strength of social interactions and the share of employment in manufacturing or agriculture.[44] Social interactions are again stronger in destinations that could be reached more easily by rail, but slightly weaker in destinations that were closer.

Overall, the results in Table 7 suggest that black social interactions responded more to attractive employment opportunities, especially in smaller destinations, and moving costs. This is consistent with black migrants relying more heavily on their social networks for information about employment opportunities or job referrals, possibly because they faced greater discrimination in labor markets or had fewer resources.

We next consider the relationship between social interactions and characteristics of sending counties. Social 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. Alternatively, resources that facilitated migration might have been a prerequisite for social interactions to influence location decisions. Another potentially important characteristic is population density, which could have strengthened social networks through frequent interactions (Chay and Munshi, 2015). We also consider literacy rates and, for African Americans, access to Rosenwald schools, which improved educational attainment among Southern blacks in this period (Aaronson and Mazumder,

[43]Results are qualitatively similar when using counties to define birth town groups (Appendix Table A.12). Results for Southern whites are in Appendix Table A.13.

[44]For destinations that intersect with the largest CMSAs, social interactions are actually weaker in destinations with more manufacturing.

2011). The relationship between education and social interactions is theoretically ambiguous, as education could promote social ties while also increasing the return to choosing a non-network destination. In addition, we examine whether social interactions were stronger in counties with greater access to railroads, which could have facilitated the transmission of information through both network and non-network channels.

Table 8 displays results from regressing birth county-level SI index estimates on birth county characteristics. Columns 1 and 2 contain results for black moves out of the South. Social interactions were stronger in counties with higher black farm ownership rates (which we use to proxy for assets), black population density, and black literacy rates. However, these estimates are relatively imprecise, and only the coefficient on the literacy rate is significant at the five percent level.[45] 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. The estimates in column 2 imply that a one standard deviation increase in log black density is associated with a 1.08 person increase in the SI index, and a one standard deviation increase in the black literacy rate is associated with a 0.48 person increase.[46] We find little evidence that social interactions varied with railroad access, although the standard errors are fairly large.

Columns 3 and 4 present results for white moves out of the Great Plains, where we use the white farm ownership rate, population density, and literacy rate as explanatory variables.[47] Overall, white social interactions appear to be less sensitive to birth county characteristics. The notable exception is that white social interactions were stronger in birth counties with lower literacy rates. This differs from results for blacks, for whom social interactions were stronger in birth counties with higher literacy rates. One possible explanation is that only whites with relatively little human capital relied on their social networks to obtain employment, while blacks with higher human capital relied on

[45]The positive correlation between social interactions and literacy rates is unlikely to be driven by black migrants reading the same newspaper, as only newspapers from the largest destinations (such as Chicago) circulated in the South.

[46]Appendix Table A.14 contains summary statistics for birth county characteristics.

[47]We do not include Rosenwald school exposure in columns 3 or 4 because Rosenwald schools existed primarily in the South.

their networks to overcome the more severe discrimination they faced.

4.5 Connecting the Social Interactions Index to a Structural Model

Finally, we connect the SI index to the structural social interactions model of Glaeser, Sacerdote and Scheinkman (1996). The additional assumptions in their model allow us to estimate the share of migrants that chose their long-run location because of social interactions, a parameter that complements our SI index in intuitively describing the size of social interactions. This connection also demonstrates how our SI index integrates the model of Glaeser, Sacerdote and Scheinkman (1996) and the general identification strategy of Bayer, Ross and Topa (2008).

Migrants, indexed on a circle by $i \in \{1, \ldots, N_j\}$, are either a "fixed agent" or a "complier." Fixed agents choose their location independently of other migrants, while a complier i chooses the same destination as his or her neighbor, $i-1$. The probability that a migrant is a complier equals χ, assumed for simplicity to be constant across birth towns and destinations for a given birth state. The covariance of location decisions for migrants i and $i+n$ is $\mathbb{C}[D_{i,j,k}, D_{i+n,j,k}] = P_{g,k}(1-P_{g,k})\chi^n$. Hence, the average covariance of location decisions implied by the model is

$$C_{j,k}(\chi; P_{g,k}, N_j) \equiv \frac{\sum_{i \in j} \sum_{i' \neq i \in j} \mathbb{C}[D_{i,j,k}, D_{i',j,k}]}{N_j(N_j-1)} \tag{13}$$

$$= \frac{2P_{g,k}(1-P_{g,k}) \sum_{a=1}^{N_j-1}(N_j-a)\chi^a}{N_j(N_j-1)}. \tag{14}$$

In the absence of social interactions, there are no compliers, and the covariance of location decisions equals zero.[48]

Substituting the expression for $C_{j,k}$ in equation (14) into the expression for the SI index in

[48] Glaeser, Sacerdote and Scheinkman (1996) measure social interactions using the normalized variance of outcomes, which in our model is

$$\mathbb{V}\left[\sum_{i=1}^{N_j} \frac{D_{i,j,k} - P_{g,k}}{N_j}\right] = \frac{P_{g,k}(1-P_{g,k})}{N_j} + \left(\frac{N_j-1}{N_j}\right) C_{j,k}(\chi; P_{g,k}, N_j).$$

equation (5) yields

$$\Delta_{j,k} = 2 \sum_{a=1}^{N_j - 1} (1 - a/N_j)\chi^a. \tag{15}$$

With a sufficiently large number of migrants, we obtain $\Delta_{j,k} = 2\chi/(1-\chi)$. Because the destination-level SI index, Δ_k, is just a weighted average of $\Delta_{j,k}$, and the average destination-level SI index, denoted Δ, is just a weighted average of Δ_k, we can estimate the probability that an individual is a complier as

$$\hat{\chi} = \frac{\hat{\Delta}}{2 + \hat{\Delta}}. \tag{16}$$

As seen in Table 9, we estimate that between 29 (Florida) and 57 percent (South Carolina) of black migrants chose their long-run location because of social interactions. There is considerable variation across destination regions.[49] For example, of Mississippi-born migrants, 32 percent of Northeast-bound, 57 percent of Midwest-bound, and 34 percent of West-bound migrants chose their location because of social interactions. Among whites from the Great Plains, between 11 (Kansas) and 19 percent (North Dakota) of migrants chose their destination because of social interactions. Although estimates of χ depend on stronger assumptions than are needed to estimate the SI index, they help illustrate the considerable impact of social interactions on location decisions for Southern blacks and the smaller impact among whites.[50]

Explicit connections to structural models also allow us to refine the interpretation of the SI index. One parameter of interest, which we denote $\theta_{j,k}$, is the number of additional people induced to move from birth town j to destination k by moving one migrant along this path. The relationship between $\Delta_{j,k}$ and $\theta_{j,k}$ depends on the underlying structural model. In the coalition model, where all migrants in birth town j form coalitions of size s, all members of a coalition move to the same

[49] Assuming that χ is constant across destinations implies that it should not vary across different regions. Nonetheless, we find the rescaled regional estimates to be informative. Appendix E contains a richer model that allows the probability of complying to vary with birth town and destination.

[50] Estimates of χ would be larger if we used estimates of the SI index that accounted for measurement error due to incomplete migration data under a missing at random assumption, as described in Section 4.2.

destination, and all coalitions move independently of each other, $\theta^{\text{COA}}_{j,k} = \Delta_{j,k} = s - 1$. In the model of Glaeser, Sacerdote and Scheinkman (1996), $\theta^{\text{GSS}}_{j,k} = 0.5\Delta_{j,k}$.[51] The difference between $\Delta_{j,k}$ and $\theta_{j,k}$ stems from the weak structural assumptions embodied in the SI index. The weakness of these assumptions, and the ability to map the SI index directly to several structural models, are valuable features of our approach.

5 Conclusion

This paper provides new evidence on the magnitude and nature of social interactions in location decisions. We use confidential 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 social interactions (SI) index that characterizes the strength of social interactions for each receiving and sending location. The SI index allows us to estimate the overall magnitude of social interactions and the degree to which social interactions were associated with economic characteristics of receiving and sending locations. The SI index can be used for other outcomes and settings to provide a deeper understanding of social interactions.

We find very strong social interactions among Southern black migrants and smaller interactions among whites. Estimates of our social interactions (SI) index imply that if we observed one randomly chosen African American move from a birth town to a destination county, then on average 1.9 additional black migrants from that birth town would make the same move. For white migrants from the Great Plains, the average is only 0.4, and results for Southern whites are similarly small.

[51] In the Glaeser, Sacerdote and Scheinkman (1996) model, migrant i has the following effect on migrant $i + n$,

$$\mathbb{E}[D_{i+n,j,k}|D_{i,j,k} = 1, D_{1,j,k}, \ldots, D_{i-1,j,k}] - \mathbb{E}[D_{i+n,j,k}|D_{i,j,k} = 0, D_{1,j,k}, \ldots, D_{i-1,j,k}] = \chi^n,$$

which implies that

$$\theta^{\text{GSS}}_{j,k}(i) = \mathbb{E}[N_{-i,j,k}|D_{i,j,k} = 1, D_{1,j,k}, \ldots, D_{i-1,j,k}] - \mathbb{E}[N_{-i,j,k}|D_{i,j,k} = 0, D_{1,j,k}, \ldots, D_{i-1,j,k}] = \sum_{a=1}^{N_j - i} \chi^a.$$

As $N_j \to \infty$, $\theta^{\text{GSS}}_{j,k}(i) \to \chi/(1-\chi) = 0.5\Delta_{j,k}$.

Interpreted through the social interactions model of Glaeser, Sacerdote and Scheinkman (1996), our estimates imply that 49 percent of African American migrants chose their long-run destination because of social interactions, while 16 percent of Great Plains whites were similarly influenced. In addition, our results suggest that social interactions among African Americans were particularly important in providing migrants with information about attractive employment opportunities, and that social interactions played a larger role in less costly moves. Our results also suggest that educational attainment and related factors in the South facilitated social interactions. One interpretation of our results is that African Americans relied on social networks more heavily to overcome a greater lack of resources and more intense discrimination in labor and housing markets.

These results shed new light on migration decisions. Social interactions play a major role in our setting, especially for migrants with fewer opportunities and resources. Our results suggest that social interactions help migrants mitigate the substantial information frictions in long-distance location decisions. Social interactions likely 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 more effectively shift labor 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 social interactions.

Our results also have implications for economic outcomes besides migration. Birth town social networks continued to operate after location decisions had been 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, 2017).

References

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

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., 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.

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: 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*.

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. 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.

Epple, Dennis, and Richard E. Romano. 2011. “Peer Effects in Education: A Survey of the Theory and Evidence.” In *Handbook of Social Economics*. Vol. 1, ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 1053–1163. Elsevier.

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.

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.

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: 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.

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: 685–718.

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.

Minnesota Population Center, and Ancestry.com. 2013. “IPUMS Restricted Complete Count Data: Version 1.0 [Machine-readable database].”

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

Munshi, Kaivan. 2011. “Labor and Credit Networks in Developing Economics.” In *Handbook of Social Economics*. Vol. 1, ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 1223–1254. Elsevier.

Rubin, Morton. 1960. “Migration Patterns of Negroes from a Rural Northeastern Mississippi Community.” *Social Forces*, 39(1): 59–66.

Ruggles, Steven, J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. “Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database].”

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.

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. 2017. “The Effect of Social Connectedness on Crime: Evidence from the Great Migration.”

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.

Topa, Giorgio. 2011. “Labor Markets and Referrals.” In *Handbook of Social Economics*. Vol. 1, ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 1193–1221. Elsevier.

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	
	People	Region	Birth State	Other State
Birth State	(1)	(2)	(3)	(4)
Panel A: Southern Blacks				
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: Great Plains Whites				
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: Southern Whites				
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 3 displays birth regions.
Source: Duke SSA/Medicare data

Table 2: Extreme Examples of Correlated Location Decisions, Southern Blacks and Great Plains Whites

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)	Social Interaction Index Estimate (9)
Panel A: Southern Blacks								
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 Whites								
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 SI 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 Social Interactions Index Estimates, by Birth State

Birth State	Number of Migrants (1)	Unweighted Average (2)	Weighted Average (3)
Panel A: Black Moves out of South			
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: White Moves out of Great Plains			
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 SI 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 Social Interactions Index Estimates, With and Without Birth Town Covariates

	Include Covariates		p-value of
	No	Yes	difference
Birth State	(1)	(2)	(3)
Panel A: Black Moves out of South			
Alabama	1.888	1.852	0.763
	(0.195)	(0.189)	
Florida	0.813	0.742	0.401
	(0.117)	(0.119)	
Georgia	1.657	1.689	0.658
	(0.177)	(0.175)	
Louisiana	1.723	1.651	0.862
	(0.478)	(0.474)	
Mississippi	2.303	2.295	0.967
	(0.313)	(0.306)	
North Carolina	1.539	1.482	0.149
	(0.130)	(0.127)	
South Carolina	2.618	2.636	0.827
	(0.301)	(0.304)	
All States	1.938	1.917	0.764
	(0.110)	(0.108)	
Panel B: White Moves out of Great Plains			
Kansas	0.255	0.233	0.112
	(0.024)	(0.024)	
Nebraska	0.361	0.349	0.504
	(0.082)	(0.082)	
North Dakota	0.464	0.445	0.456
	(0.036)	(0.035)	
Oklahoma	0.453	0.439	0.241
	(0.036)	(0.036)	
South Dakota	0.350	0.331	0.145
	(0.026)	(0.026)	
All States	0.380	0.363	0.021
	(0.022)	(0.022)	

Notes: All columns contain weighted averages of destination-level SI index estimates, $\hat{\Delta}_k$, where the weights are the number of people who move from each state to destination k. Column 2 controls for birth town-level covariates as described in the text. Column 3 reports the p-value from testing the null hypothesis that the two columns are equal. Birth town groups are defined by cross validation. Standard errors are in parentheses.
Source: Duke SSA/Medicare data

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

Exclude Largest Birth Towns:	No	Yes	No	Yes
Exclude Largest Destinations:	No	No	Yes	Yes
Birth State	(1)	(2)	(3)	(4)
Panel A: Black Moves out of South				
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: White Moves out of Great Plains				
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 SI 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: Average Social Interactions Index Estimates, by Destination Region

	Destination Region			
	Northeast (1)	Midwest (2)	West (3)	South (4)
Panel A: Black Moves out of South				
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: White Moves out of Great Plains				
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 SI index estimates, $\hat{\Delta}_k$, where the weights are the number of people who move from each state to destination k. See footnote 34 for region definitions. We do not estimate social interactions for blacks who move to the South. Birth town groups are defined by cross validation. Standard errors are in parentheses.
Source: Duke SSA/Medicare data

Table 7: Social Interactions Index Estimates and Destination County Characteristics

	Dependent variable: Destination-level SI index estimate					
	Black Moves out of South			White Moves out of Plains		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing employment share, 1910	1.583***	0.361	0.328	0.025	-0.238**	-0.238**
	(0.432)	(0.481)	(0.489)	(0.080)	(0.116)	(0.116)
Manufacturing employment share by small destination indicator		2.162***	2.190***		0.338**	0.342**
		(0.753)	(0.749)		(0.147)	(0.147)
Agriculture employment share, 1910	0.046	-0.335	-0.362	0.042	0.040	0.036
	(0.229)	(0.408)	(0.412)	(0.035)	(0.102)	(0.102)
Agriculture employment share by small destination indicator		0.562	0.486		0.010	0.018
		(0.459)	(0.474)		(0.108)	(0.108)
Small destination indicator		-0.515**	-0.521**		-0.015	-0.019
		(0.241)	(0.241)		(0.064)	(0.064)
Log distance from birth state	-0.362***	-0.334***	-0.315***	0.070*	0.083**	0.075*
	(0.061)	(0.063)	(0.061)	(0.037)	(0.038)	(0.039)
Direct railroad connection from birth state	0.315***	0.336***	0.349***	0.218***	0.218***	0.209***
	(0.111)	(0.112)	(0.128)	(0.044)	(0.044)	(0.046)
One-stop railroad connection from birth state	0.225***	0.218***	0.184**	0.091***	0.086***	0.086***
	(0.077)	(0.075)	(0.078)	(0.018)	(0.018)	(0.018)
Log population, 1900	0.102***	0.112***	0.116***	0.010	0.019**	0.018**
	(0.039)	(0.040)	(0.042)	(0.008)	(0.008)	(0.008)
Percent African American, 1900	-2.037***	-1.851***	-1.760***	-0.239***	-0.254***	-0.251***
	(0.332)	(0.330)	(0.327)	(0.034)	(0.036)	(0.035)
Birth state fixed effects			x			x
R2	0.093	0.103	0.115	0.031	0.035	0.035
N (birth state-destination county pairs)	1,469	1,469	1,469	3,822	3,822	3,822
Destination counties	371	371	371	1,148	1,148	1,148

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. Standard errors, clustered by destination county, are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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

Table 8: Social Interactions Index Estimates and Birth County Characteristics

	Dependent variable: Birth county-level SI index estimate			
	Black Moves out of South		White Moves out of Plains	
	(1)	(2)	(3)	(4)
Black/white farm ownership rate, 1920	1.854	2.123	0.948*	0.819
	(1.353)	(1.390)	(0.559)	(0.585)
Log black/white population density, 1920	1.099*	1.027*	0.219*	0.258**
	(0.562)	(0.565)	(0.112)	(0.121)
Rosenwald school exposure	-0.981	-1.202*		
	(0.656)	(0.687)		
Black/white literacy rate, 1920	3.680**	5.128**	-3.908	-8.238**
	(1.574)	(2.094)	(3.122)	(3.484)
Railroad exposure	-0.309	-0.268	-0.150**	-0.136*
	(0.423)	(0.442)	(0.073)	(0.078)
Percent African American, 1920	0.606	0.880	1.097	1.969
	(1.684)	(1.589)	(1.118)	(1.215)
Birth state fixed effects		x		x
R2	0.090	0.097	0.095	0.147
N (birth counties)	549	549	394	394

Notes: The dependent variable is the birth county level social interaction estimate. 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

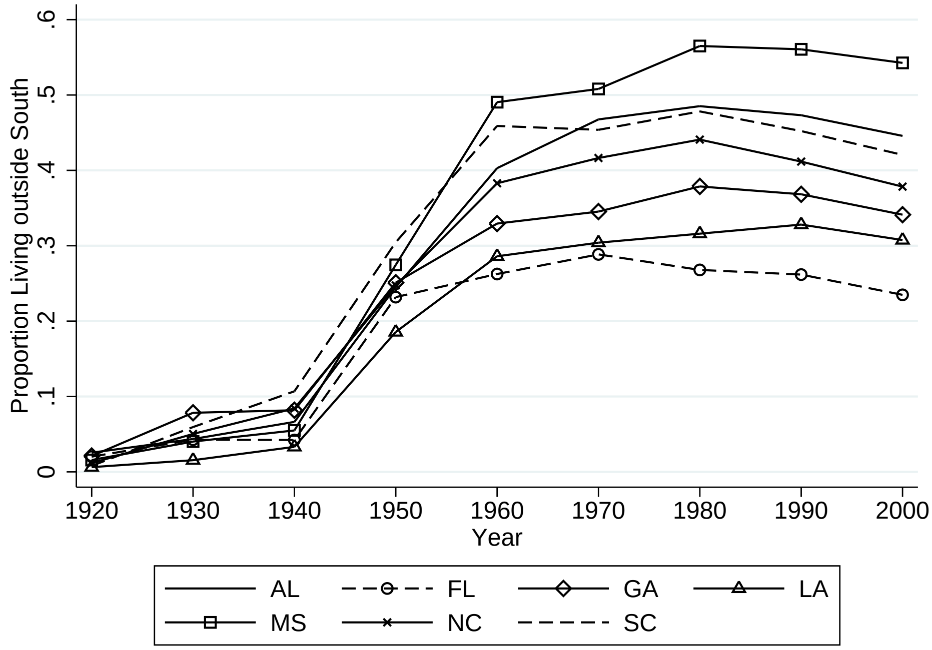
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010) data, Aaronson and Mazumder (2011) data, and Black et al. (2015) data

Table 9: Estimated Share of Migrants That Chose Their Destination Because of Social Interactions

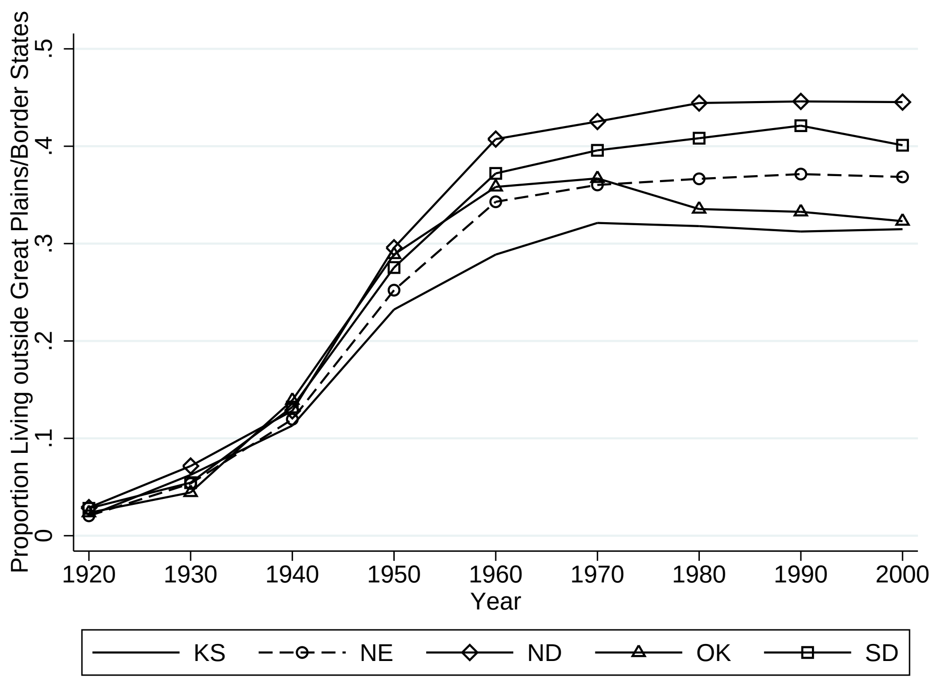
	Destination Region				
	All	Northeast	Midwest	West	South
Birth State	(1)	(2)	(3)	(4)	(5)
Panel A: Black Moves out of South					
Alabama	0.486	0.382	0.541	0.289	-
	(0.026)	(0.031)	(0.031)	(0.069)	-
Florida	0.289	0.328	0.284	0.117	-
	(0.030)	(0.039)	(0.043)	(0.042)	-
Georgia	0.453	0.436	0.508	0.170	-
	(0.026)	(0.039)	(0.038)	(0.070)	-
Louisiana	0.463	0.123	0.363	0.520	-
	(0.069)	(0.039)	(0.042)	(0.084)	-
Mississippi	0.535	0.316	0.571	0.341	-
	(0.034)	(0.025)	(0.036)	(0.028)	-
North Carolina	0.435	0.456	0.312	0.085	-
	(0.021)	(0.022)	(0.042)	(0.017)	-
South Carolina	0.567	0.592	0.379	0.095	-
	(0.028)	(0.029)	(0.032)	(0.023)	-
All States	0.492	0.482	0.530	0.412	-
	(0.014)	(0.016)	(0.022)	(0.060)	-
Panel B: White Moves out of Great Plains					
Kansas	0.113	0.038	0.184	0.123	0.025
	(0.009)	(0.009)	(0.032)	(0.012)	(0.003)
Nebraska	0.153	0.039	0.180	0.174	0.031
	(0.029)	(0.007)	(0.032)	(0.037)	(0.004)
North Dakota	0.188	0.051	0.168	0.208	0.023
	(0.012)	(0.012)	(0.020)	(0.015)	(0.004)
Oklahoma	0.185	0.025	0.163	0.213	0.036
	(0.012)	(0.003)	(0.032)	(0.015)	(0.003)
South Dakota	0.149	0.030	0.195	0.160	0.028
	(0.010)	(0.006)	(0.022)	(0.012)	(0.005)
All States	0.160	0.035	0.178	0.181	0.030
	(0.008)	(0.003)	(0.013)	(0.010)	(0.002)

Notes: Table contains estimates and standard errors of $\chi = \Delta/(2+\Delta)$, the share of migrants that chose their destination because of social interactions, based on weighted average estimates from column 3 of Table 3 and columns 1-4 of Table 6. Standard errors, estimated using the Delta method, are in parentheses.
Source: Duke SSA/Medicare data

Figure 1: Proportion Living Outside Birth Region, 1916-1936 Cohorts, by Birth State and Year



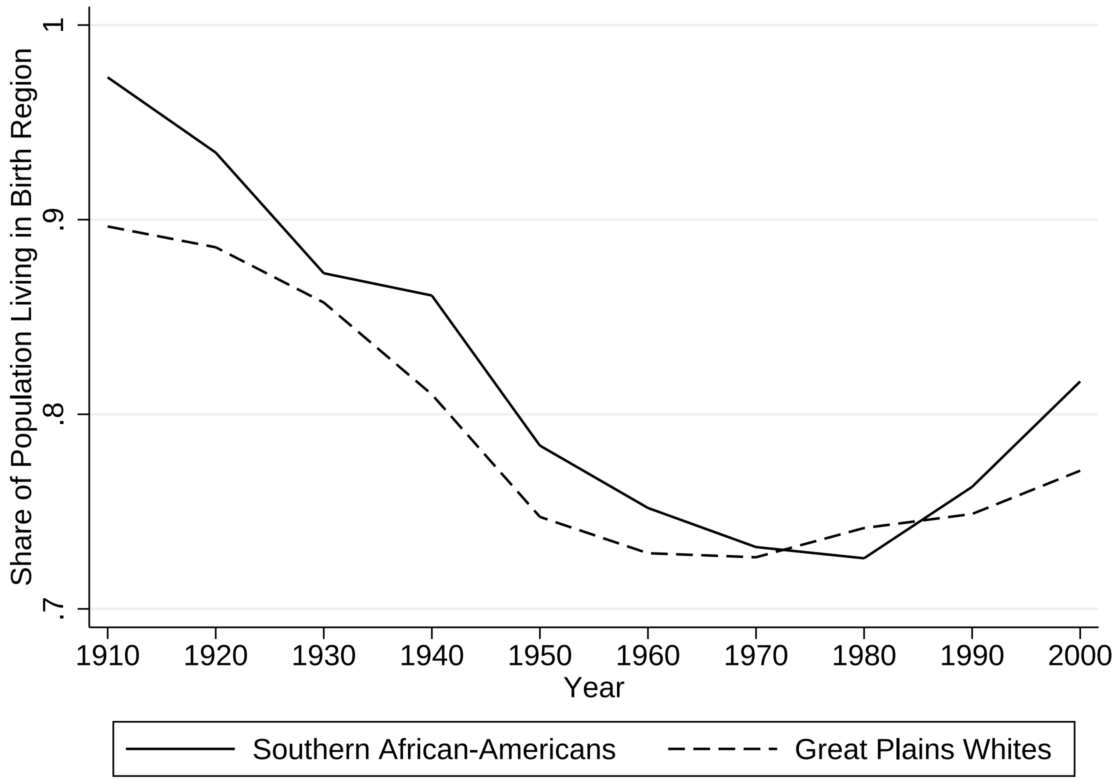
(a) Southern Blacks



(b) Great Plains Whites

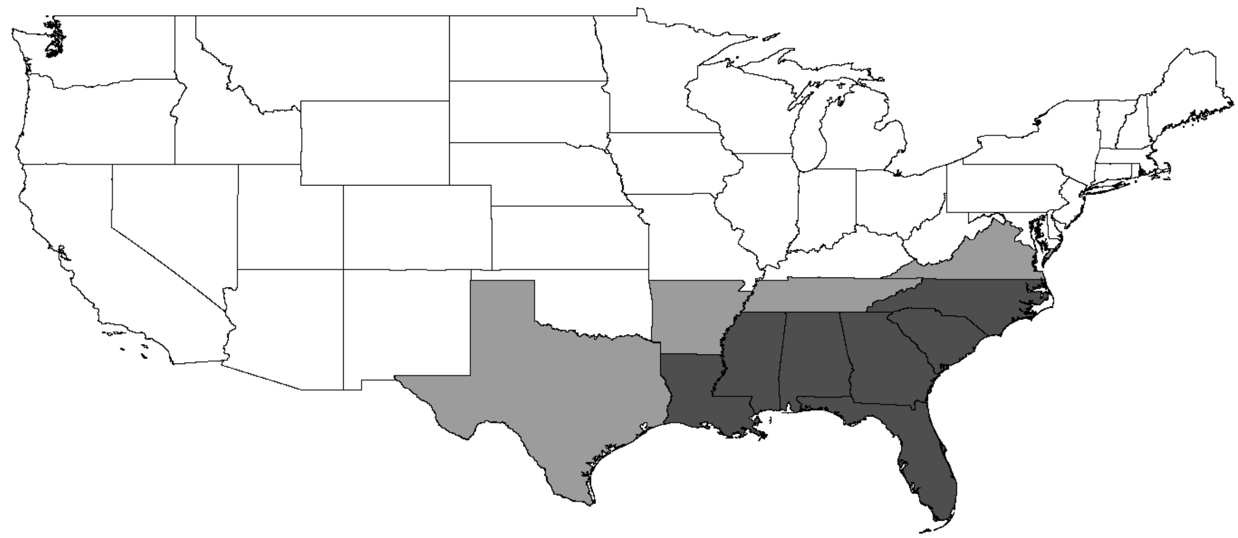
Notes: Figure 3 displays birth regions.
Source: Ruggles et al. (2010) data

Figure 2: Trajectory of Migrations out of South and Great Plains

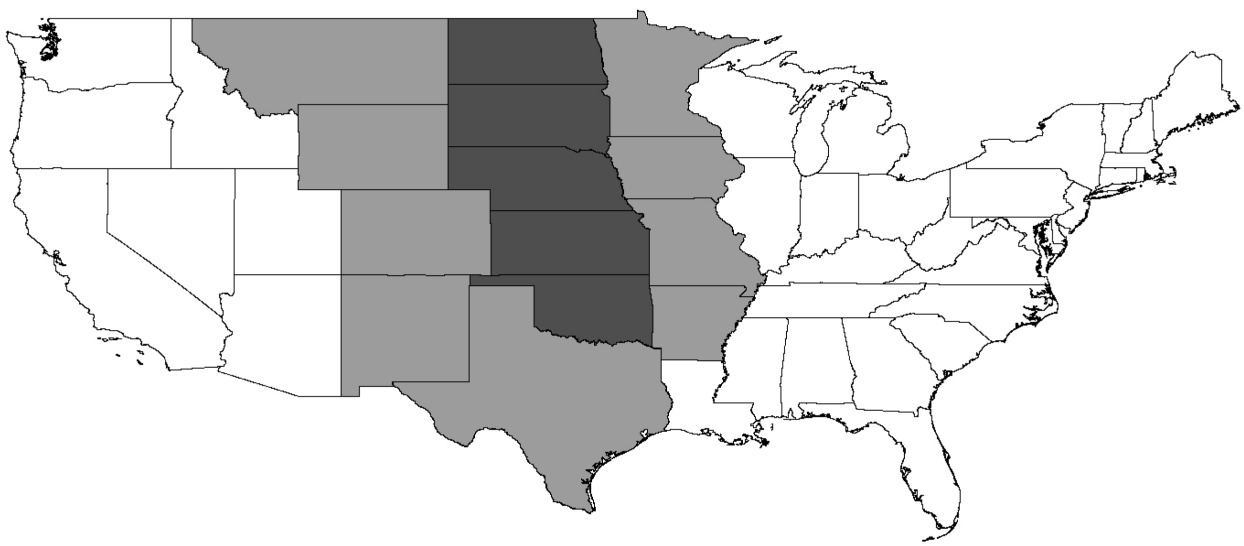


Notes: The solid line shows the proportion of blacks from the seven Southern birth states we analyze (dark grey states in Figure 3a) living in the South (light and dark grey states) at the time of Census enumeration. The dashed line shows the proportion of whites from the Great Plains states living in the Great Plains or Border States. We do not impose age or cohort restrictions for this graph.
Source: Ruggles et al. (2010) data

Figure 3: Geographic Coverage



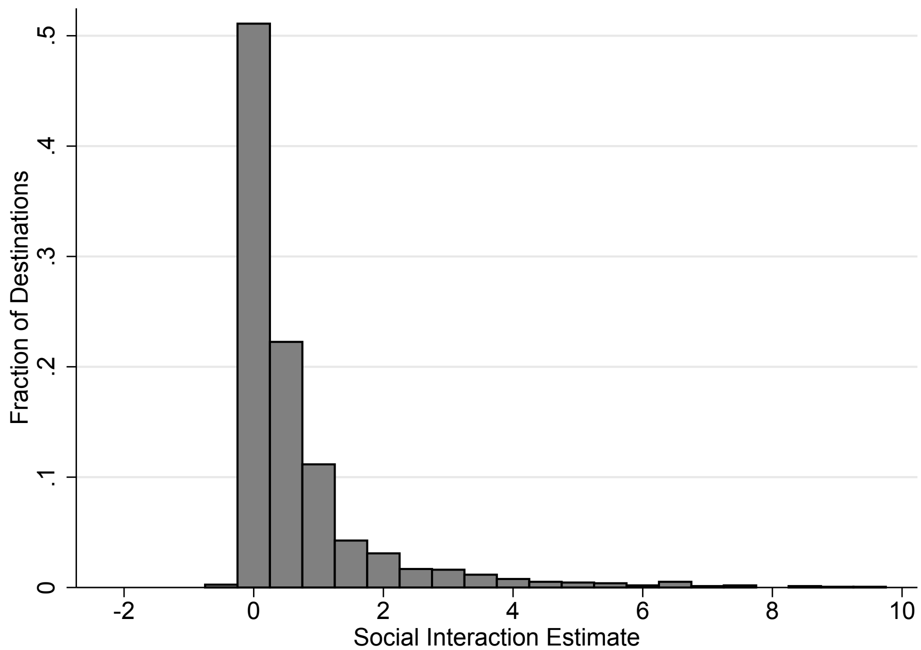
(a) South



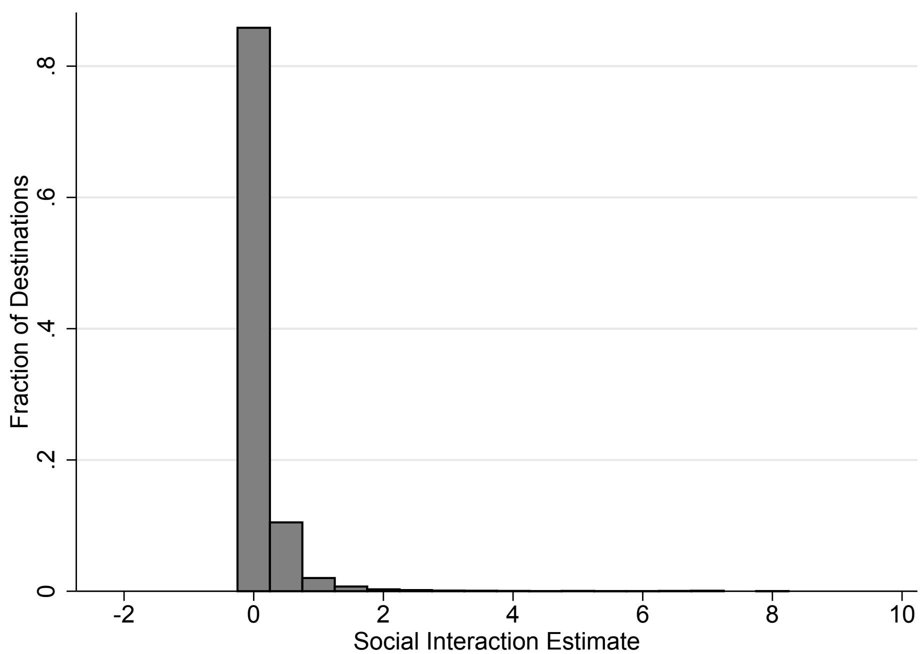
(b) Great Plains

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 4: Distribution of Destination-Level Social Interactions Index Estimates



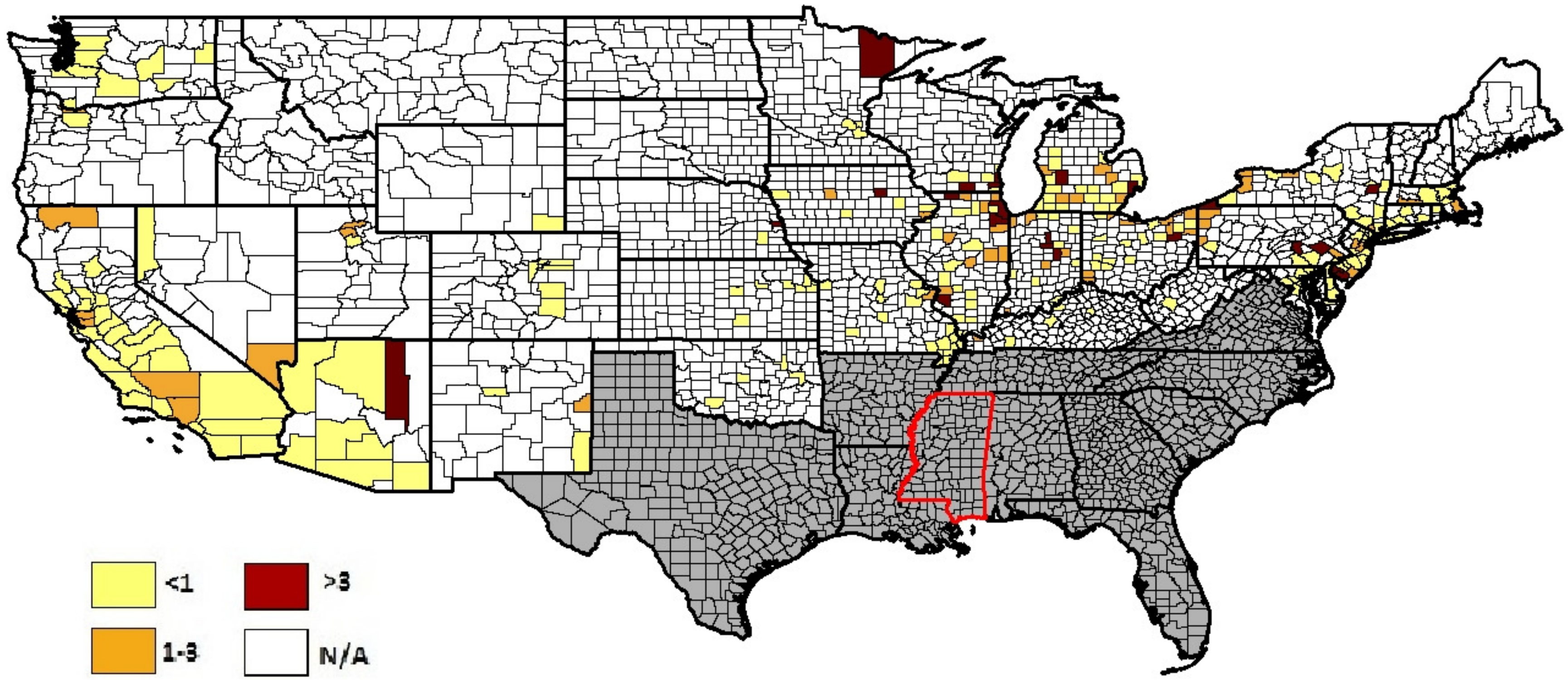
(a) Black Moves out of South



(b) White Moves out of Great Plains

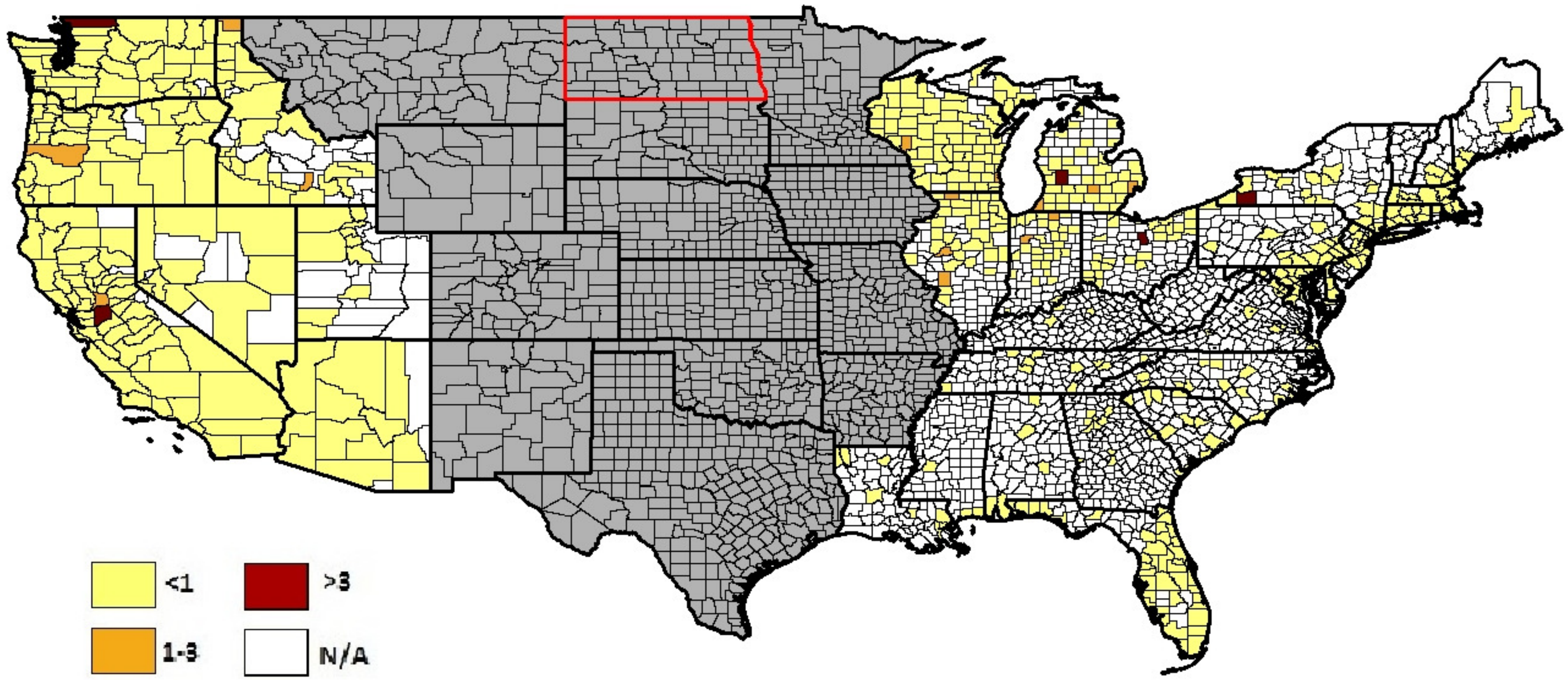
Notes: Bin width is 1/2. Birth town groups are defined by cross validation. Panel (a) omits the estimate $\hat{\Delta}_k = 11.4$ from Mississippi to Racine County, WI, $\hat{\Delta}_k = 15.2$ from South Carolina to Rensselaer County, NY, and $\hat{\Delta}_k = 18.1$ from Florida to St. Joseph County, IN.
Source: Duke SSA/Medicare data

Figure 5: Spatial Distribution of Destination-Level Social Interactions Index Estimates, Mississippi-born Blacks



Notes: Figure displays destination-level SI 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 6: Spatial Distribution of Destination-Level Social Interactions Index Estimates, North Dakota-born Whites



Notes: See note to Figure 5. 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 Social Interactions Index

Appendix A derives the expression for the social interactions (SI) index in equation (5).

First, recall the definition of the SI 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

$$\mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1] - \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 0] = \mu_{j,k} - \frac{P_{j,k}(1 - \mu_{j,k})}{1 - P_{j,k}} \tag{A.4}$$

$$= \frac{\mu_{j,k} - P_{j,k}}{1 - P_{j,k}}. \tag{A.5}$$

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

$$C_{j,k} \equiv \mathbb{E}[D_{i',j,k}D_{i,j,k}] - \mathbb{E}[D_{i',j,k}]\,\mathbb{E}[D_{i,j,k}] \tag{A.7}$$

$$= \mathbb{E}[D_{i',j,k}|D_{i,j,k} = 1]P_{j,k} - (P_{j,k})^2 \tag{A.8}$$

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 (5).

B Generalized Method of Moments Formulation

B.1 Basic Model

As described in the text, we can derive the destination-level SI 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 SI index, but the latter approach allows us to use GMM to estimate standard errors.

If we assume that the SI index, $\Delta_{j,k}$, is constant across birth towns within a birth state, the destination-level SI 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 (9).

The above derivation is for a single destination-level SI index, but can easily be expanded to consider all K destination-level SI 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 SI 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 SI 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 SI 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 SI 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

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

Given destination-level SI 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 SI 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 social interaction 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 $\widehat{\mathbb{V}}[\widehat{\Delta^1} - \widehat{\Delta^2}]$ as the appropriately weighted sum of

$$\widehat{\mathbb{V}}[\widehat{\Delta^1_k} - \widehat{\Delta^2_k}] = \widehat{\mathbb{V}}[\widehat{\Delta^1_k}] + \widehat{\mathbb{V}}[\widehat{\Delta^2_k}] - 2\widehat{\mathbb{C}}[\widehat{\Delta^1_k}, \widehat{\Delta^2_k}] \tag{A.31}$$

which we obtain from the GMM variance estimate.

C Estimating Cross-Group Social Interactions

When estimating cross-group social interactions, 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 (2) 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 SI 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 social interactions 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 Additional Detail on Measurement Error due to Incomplete Migration Data

Appendix D discusses the implications of measurement error due to incomplete migration data without making a missing at random (MAR) assumption. We derive a lower bound on the social interactions (SI) index and show that estimates of this lower bound still reveal sizable social interactions.

As described in the text, the SI 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, we assume 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 $\Delta_{j,k}^*$, $C_{j,k}^*$, and N_j^* be the true values of the SI index, covariance of location decisions, and number of migrants. The true parameters are connected through the equation

$$\Delta^*_{j,k} = \frac{C^*_{j,k}(N^*_j - 1)}{P_{g,k} - P^2_{g,k}}. \tag{A.37}$$

As in the text, we let α denote the coverage rate, defined by the relationship between the observed number of migrants, N_j, and the true number of migrants,

$$N_j = \alpha N^*_j. \tag{A.38}$$

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

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

where $C_{j,k}$ is the covariance of location decisions between migrants who are in our data, $C^{\text{in, out}}_{j,k}$ is the average covariance of location decisions between a migrant who is in our data and a migrant who is not, and $C^{\text{out, out}}_{j,k}$ 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 ($C^{\text{in, out}}_{j,k}$ and $C^{\text{out, out}}_{j,k}$) could differ from the covariance of location decisions between migrants who are in our data ($C_{j,k}$). As a result, the SI 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 SI index under this more general measurement error model. However, we can construct a lower bound for the strength of social interactions. In particular, we make the extreme assumptions that there are no social interactions between migrants in and out of our data, so that $C^{\text{in, out}}_{j,k} = 0$, and that there are no social interactions between migrants out of our data, so that $C^{\text{out, out}}_{j,k} = 0$. 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 SI index by multiplying the estimated SI index by the coverage rate.[52] The average coverage rate is 52.3% for African American migrants from the South and 69.3% for white migrants from the Great Plains. Combined with the average destination-level SI index estimates from Table 3, we estimate a lower bound for the SI index of 1.014 for African Americans and 0.263 for whites. These lower bounds, which depend on extremely conservative assumptions about the migration behavior of individuals not in our data, still

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

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

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 (5) in the text. One could also construct upper bounds, but these are not particularly informative.

reveal sizable social interactions, especially among African Americans.

E A Richer Model of Local Social Interactions

This section extends the local social interactions model in Section 4.5. In particular, we allow the probability that a migrant follows his neighbor to vary with birth town and destination.

Migrants from birth town j are indexed on a line by $i \in \{1, \ldots, N_j\}$, where N_j is the total number of migrants from town 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 social interactions, 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 his neighbor, $i-1$, chooses the same destination. A migrant chooses a destination in H_i if his neighbor chooses the same destination or his neighbor selects a destination in L_i.[54]

Migrants from the same birth town can 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 $m_{j,k} \equiv \mathbb{P}[k \in M_i | i \in j]$.

The probability that migrant i moves to destination k given that his neighbor moves there is

$$\rho_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1 | D_{i-1,j,k} = 1] = \mathbb{P}[k \in H_i] + \mathbb{P}[k \in M_i] \tag{A.41}$$

$$= h_{j,k} + m_{j,k}, \tag{A.42}$$

where $D_{i,j,k}$ equals one if migrant i moves from j to k and zero otherwise.

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 $m_{j,k} = \nu_{j,k}(1 - h_{j,k})$. We use $\nu_{j,k}$ to derive a simple sequential estimation approach.

[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 social interactions. 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.

[54] This model shares a similar structure as Glaeser, Sacerdote and Scheinkman (1996) in that some agents imitate their neighbors. However, we differ from Glaeser, Sacerdote and Scheinkman (1996) in that we model the interdependence between various destinations (i.e., this is a multinomial choice problem) and allow for more than two types of agents.

In equilibrium, the probability that a randomly chosen migrant i moves from j to k is

$$P_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1] = \mathbb{P}[D_{i-1,j,k} = 1, k \in H_i] + \mathbb{P}[D_{i-1,j,k} = 1, k \in M_i] + \sum_{k' \neq k} \mathbb{P}[D_{i-1,j,k'} = 1, k \in H_i, k' \in L_i] \tag{A.43}$$

$$= P_{j,k}h_{j,k} + P_{j,k}\nu_{j,k}(1 - h_{j,k}) + \sum_{k' \neq k} P_{j,k'}h_{j,k}(1 - \nu_{j,k'}) \tag{A.44}$$

$$= P_{j,k}\nu_{j,k} + \left(\sum_{k'=1}^{K} P_{j,k'}(1 - \nu_{j,k'}) \right) h_{j,k}. \tag{A.45}$$

The first term on the right hand side of equation (A.43) is the probability that a migrant's neighbor moves to k, and k is in the migrant's high preference group; in this case, social interaction reinforces the migrant's desire to move to k. The second term is the probability that a migrant follows his neighbor to k because of social interactions. The third term is the probability that a migrant resists the pull of social interactions because town k is in the migrant's high preference group and the neighbor's chosen destination is in the migrant's low preference group.

The average covariance of location decisions implied by the richer model is[55]

$$C_{j,k} = \frac{2P_{j,k}(1 - P_{j,k}) \sum_{s=1}^{N_j - 1} (N_j - s) \left(\frac{\rho_{j,k} - P_{j,k}}{1 - P_{j,k}} \right)^s}{N_j(N_j - 1)}. \tag{A.46}$$

Substituting equation (A.46) into equation (5) and simplifying yields[56]

$$\Delta_{j,k} = \frac{2(\rho_{j,k} - P_{j,k})}{1 - \rho_{j,k}}, \tag{A.47}$$

which can be rearranged to show that

$$\rho_{j,k} = \frac{2P_{j,k} + \Delta_{j,k}}{2 + \Delta_{j,k}}. \tag{A.48}$$

We could use equation (A.48) to estimate $\rho_{j,k}$ with our estimates of $P_{j,k}$ and $\Delta_{j,k}$.

Equations (A.42) and (A.45), plus the fact that $m_{j,k} = \nu_{j,k}(1 - h_{j,k})$, imply that

$$\rho_{j,k} = \nu_{j,k} + \frac{P_{j,k}(1 - \nu_{j,k})^2}{\sum_{k'=1}^{K} P_{j,k'}(1 - \nu_{j,k'})}. \tag{A.49}$$

We could use equation (A.49) to estimate $\nu_j \equiv (\nu_{j,1}, \ldots, \nu_{j,K})$ using our estimates of $(P_{j,1}, \ldots, P_{j,K}, \rho_{j,1}, \ldots, \rho_{j,K})$. In addition, we could use equation (A.45) to estimate $h_{j,k}$ with our estimates of $\rho_{j,k}$ and $\nu_{j,k}$. Finally, we could estimate $m_{j,k}$ using the fact that $m_{j,k} = \rho_{j,k} - h_{j,k}$.

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

[56]Equation (A.47) results from taking the limit as $N_j \to \infty$, and so relies on N_j being sufficiently large.

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

	Percent of Group Working in Industry			
	Southern Blacks		Great Plains Whites	
	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. (2010)

Table A.2: Number of Birth Towns and Migrants, by Birth State

Birth State	Birth Towns (1)	Migrants (2)	Migrants Per Town (3)
Panel A: Black Moves out of South			
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: White Moves out of Great Plains			
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.3: Size of Birth Town Groups Chosen by Cross Validation

Birth State	(1)
Panel A: Southern Blacks	
Alabama	52
Florida	138
Georgia	40
Louisiana	48
Mississippi	42
North Carolina	52
South Carolina	30
Panel B: Great Plains Whites	
Kansas	128
Nebraska	128
North Dakota	84
Oklahoma	68
South Dakota	112
Panel C: Southern Whites	
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.4: Average Destination-Level Social Interactions Index Estimates, Birth Town Groups Defined by Cross Validation and Counties

	Cross Validation		Counties	
Type of Average: Birth State	Unweighted (1)	Weighted (2)	Unweighted (3)	Weighted (4)
Panel A: Black Moves out of South				
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: White Moves out of Great Plains				
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 SI 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.5: Average Social Interactions Index Estimates, White Moves out of South

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 SI 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.6: Average Social Interactions Index Estimates, By Size of Birth Town and Destination, White Moves out of South

Exclude Largest Birth Towns:	No	Yes	No	Yes
Exclude Largest Destinations:	No	No	Yes	Yes
Birth State	(1)	(2)	(3)	(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 SI 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.7: Average Social Interactions Index Estimates, by Destination Region, White Moves out of South

	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: All columns contain weighted averages of destination-level SI index estimates, $\hat{\Delta}_k$, where the weights are the number of people who move from each state to destination k. See footnote 34 for region definitions. We do not estimate social interactions for Southern-born whites 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.8: Average Cross-Race Social Interactions Index Estimates, Southern White and Black Migrants

Birth State	All Counties (1)	Excluding Largest CMSAs (2)
Panel A: Blacks 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: Whites 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.8 contains weighted averages of cross-race destination-level SI index estimates. Birth town groups are defined by cross validation. Standard errors in parentheses.
Source: Duke SSA/Medicare data

Table A.9: 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 Blacks							
Alabama	70.2%	78.6%	55.2%	55.0%	55.4%	47.7%	62.8%
Florida	62.3%	83.3%	51.9%	53.2%	50.9%	45.8%	57.4%
Georgia	67.2%	72.8%	48.9%	47.5%	50.1%	43.2%	55.5%
Louisiana	67.9%	84.4%	57.3%	57.4%	57.2%	51.3%	63.2%
Mississippi	77.3%	74.6%	57.7%	57.7%	57.6%	50.4%	65.2%
North Carolina	68.5%	72.4%	49.6%	46.7%	51.9%	42.9%	56.5%
South Carolina	75.3%	61.6%	46.4%	43.6%	48.8%	39.3%	55.3%
All States	70.4%	74.2%	52.3%	51.2%	53.2%	45.5%	59.5%
Panel B: Great Plains Whites							
Kansas	75.9%	92.3%	70.1%	68.9%	71.3%	64.8%	76.0%
Nebraska	75.2%	93.2%	70.0%	69.8%	70.3%	65.6%	74.8%
North Dakota	76.1%	89.6%	68.1%	64.6%	71.7%	64.6%	71.8%
Oklahoma	75.8%	89.8%	68.1%	67.2%	69.0%	62.8%	73.2%
South Dakota	78.3%	91.0%	71.3%	70.5%	72.1%	64.3%	79.6%
All States	76.0%	91.2%	69.3%	68.1%	70.4%	64.2%	74.7%

Notes: Column 1 reports the number of individuals in the Duke SSA/Medicare dataset divided by the number of individuals in the 1960/1970 Census. Column 2 reports the share of individuals in the Duke SSA/Medicare dataset for whom birth town and destination county is identified. Columns 3-7 reports the number of individuals in the Duke SSA/Medicare dataset for whom birth town and destination county is identified divided by the number of individuals in the 1960/1970 Census. In all columns, we use the 1960 Census for individuals born from 1916-1925 and the 1970 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. (2010) data

Table A.10: Average Social Interactions 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: Black Moves out of South					
Alabama	3.420	1.542	1.891	1.739	1.874
	(0.353)	(0.160)	(0.204)	(0.197)	(0.185)
Florida	1.567	0.725	0.832	0.650	0.980
	(0.226)	(0.116)	(0.168)	(0.145)	(0.154)
Georgia	3.389	1.317	2.069	2.072	1.566
	(0.362)	(0.153)	(0.246)	(0.281)	(0.144)
Louisiana	3.007	1.533	1.218	1.280	2.015
	(0.834)	(0.408)	(0.478)	(0.296)	(0.689)
Mississippi	3.990	1.759	2.273	1.769	2.353
	(0.542)	(0.244)	(0.331)	(0.267)	(0.323)
North Carolina	3.104	1.414	1.729	1.742	1.561
	(0.263)	(0.137)	(0.150)	(0.164)	(0.128)
South Carolina	5.643	2.543	3.141	3.223	2.630
	(0.648)	(0.262)	(0.433)	(0.423)	(0.276)
All States	3.713	1.648	2.064	1.965	1.972
	(0.197)	(0.088)	(0.123)	(0.113)	(0.118)
Panel B: White Moves out of Great Plains					
Kansas	0.364	0.185	0.197	0.248	0.185
	(0.034)	(0.020)	(0.018)	(0.025)	(0.015)
Nebraska	0.515	0.221	0.290	0.333	0.268
	(0.117)	(0.063)	(0.056)	(0.071)	(0.053)
North Dakota	0.681	0.317	0.361	0.445	0.324
	(0.054)	(0.027)	(0.034)	(0.037)	(0.024)
Oklahoma	0.665	0.320	0.345	0.361	0.382
	(0.053)	(0.029)	(0.028)	(0.031)	(0.031)
South Dakota	0.491	0.220	0.274	0.325	0.236
	(0.037)	(0.020)	(0.023)	(0.027)	(0.018)
All States	0.552	0.258	0.297	0.338	0.294
	(0.031)	(0.017)	(0.016)	(0.019)	(0.016)

Notes: Table A.10 reports weighted averages of destination-level SI index estimates, adjusted for incomplete migration data using the coverage rates in Appendix Table A.9. Birth town groups are defined by cross validation. Standard errors in parentheses.
Source: Duke SSA/Medicare data

Table A.11: Summary Statistics, Destination County Characteristics

Variable	Mean	S.D.
Panel A: Black Moves out of South (N=1469)		
SI index estimate, $\widehat{\Delta_k}$	0.732	1.373
Manufacturing employment share, 1910	0.240	0.140
Agriculture employment share, 1910	0.223	0.168
Direct railroad connection from birth state	0.093	0.291
One-stop railroad connection from birth state	0.557	0.497
Log distance from birth state	6.684	0.517
Log population, 1900	11.004	1.105
Percent African American, 1900	0.045	0.082
Small destination indicator	0.608	0.488
Panel B: White Moves out of Great Plains (N=3822)		
SI index estimate, $\widehat{\Delta_k}$	0.140	0.441
Manufacturing employment share, 1910	0.169	0.134
Agriculture employment share, 1910	0.400	0.232
Direct railroad connection from birth state	0.112	0.315
One-stop railroad connection from birth state	0.504	0.500
Log distance from birth state	6.788	0.355
Log population, 1900	10.122	1.080
Percent African American, 1900	0.121	0.197
Small destination indicator	0.849	0.358
Panel C: White Moves Out of South (N=3153)		
SI index estimate, $\widehat{\Delta_k}$	0.131	0.566
Manufacturing employment share, 1910	0.195	0.141
Agriculture employment share, 1910	0.312	0.199
Direct railroad connection from birth state	0.084	0.278
One-stop railroad connection from birth state	0.492	0.500
Log distance from birth state	6.766	0.593
Log population, 1900	10.418	1.143
Percent African American, 1900	0.038	0.077
Small destination indicator	0.752	0.432

Notes: The unit of observation is a birth state-destination county pair. Sample includes destination counties that existed from 1900-2000 and for which we estimate a SI index.

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

Table A.12: Social Interactions Index Estimates and Destination County Characteristics, Birth Town Groups Defined by Counties

	Dependent variable: Destination-level SI index estimate					
	Black Moves out of South			White Moves out of Plains		
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing employment share, 1910	1.492**	0.053	0.034	0.021	-0.315**	-0.316**
	(0.636)	(0.412)	(0.416)	(0.069)	(0.137)	(0.137)
Manufacturing employment share by small destination indicator		2.535**	2.519***		0.424***	0.427***
		(0.983)	(0.946)		(0.154)	(0.154)
Agriculture employment share, 1910	0.044	-0.507	-0.531	0.027	-0.032	-0.037
	(0.202)	(0.392)	(0.398)	(0.030)	(0.109)	(0.109)
Agriculture employment share by small destination indicator		0.802*	0.739		0.074	0.080
		(0.485)	(0.501)		(0.114)	(0.114)
Small destination indicator		-0.650**	-0.643**		-0.085	-0.089
		(0.287)	(0.279)		(0.069)	(0.069)
Direct railroad connection from birth state	0.349***	0.372***	0.396***	0.170***	0.167***	0.160***
	(0.115)	(0.118)	(0.146)	(0.035)	(0.036)	(0.037)
One-stop railroad connection from birth state	0.222**	0.215**	0.195**	0.068***	0.062***	0.063***
	(0.092)	(0.089)	(0.096)	(0.015)	(0.014)	(0.014)
Log distance from birth state	-0.245***	-0.214***	-0.223***	0.049*	0.052*	0.045
	(0.072)	(0.080)	(0.065)	(0.028)	(0.029)	(0.029)
Log population, 1900	0.084*	0.092**	0.089*	0.012*	0.017**	0.017**
	(0.046)	(0.045)	(0.051)	(0.007)	(0.007)	(0.007)
Percent African American, 1900	-1.541***	-1.315***	-1.317***	-0.196***	-0.206***	-0.203***
	(0.289)	(0.341)	(0.305)	(0.029)	(0.030)	(0.030)
Birth state fixed effects			x			x
R2	0.055	0.065	0.074	0.029	0.033	0.033
N (birth state-destination county pairs)	1,469	1,469	1,469	3,822	3,822	3,822
Destination counties	371	371	371	1,148	1,148	1,148

Notes: The sample contains only counties that received at least 10 migrants. Birth town groups are defined by counties. We measure distance from the centroid of destination counties to the centroid of birth states. Standard errors, clustered by destination county, are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010) data, and Black et al. (2015) data

Table A.13: Social Interactions Index Estimates and Destination County Characteristics, White Moves out of South

Dependent variable: Destination-level SI index estimate			
	(1)	(2)	(3)
Manufacturing employment share, 1910	0.433***	-0.004	0.027
	(0.164)	(0.130)	(0.131)
Manufacturing employment share by small destination indicator		0.602**	0.595**
		(0.250)	(0.250)
Agriculture employment share, 1910	0.080*	0.123	0.157
	(0.045)	(0.144)	(0.144)
Agriculture employment share by small destination indicator		-0.039	-0.061
		(0.156)	(0.154)
Small destination indicator		-0.150**	-0.143**
		(0.074)	(0.073)
Direct railroad connection from birth state	0.069*	0.074*	0.087**
	(0.041)	(0.041)	(0.040)
One-stop railroad connection from birth state	0.059***	0.052**	0.060***
	(0.022)	(0.021)	(0.021)
Log distance from birth state	-0.042**	-0.047**	-0.011
	(0.019)	(0.019)	(0.020)
Log population, 1900	-0.010	-0.009	-0.001
	(0.013)	(0.012)	(0.012)
Percent African American, 1900	-0.264***	-0.354***	-0.272***
	(0.095)	(0.099)	(0.096)
Birth state fixed effects			x
R2	0.013	0.018	0.028
N (birth state-destination county pairs)	3,153	3,153	3,153
Destination counties	728	728	728

Notes: The sample contains only counties that received at least 10 migrants. Birth town groups are defined by cross validation. Standard errors, clustered by destination county, are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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

Table A.14: Summary Statistics, Birth County Characteristics

Variable	Mean	S.D.
Panel A: Black Moves out of South (N=549)		
SI index estimate, $\widehat{\Delta_c}$	1.721	3.544
Black farm ownership rate, 1920	0.318	0.246
Log black density, 1920	2.534	1.055
Rosenwald school exposure	0.204	0.217
Black literacy rate, 1920	0.705	0.093
Railroad exposure	0.542	0.405
Percent black, 1920	0.408	0.209
Panel B: White Moves out of Great Plains (N=394)		
SI index estimate, $\widehat{\Delta_c}$	0.352	0.636
White farm ownership rate, 1920	0.576	0.131
Log white density, 1920	2.476	1.006
White literacy rate, 1920	0.992	0.012
Railroad exposure	0.524	0.395
Percent black, 1920	0.017	0.041
Panel C: White Moves Out of South (N=560)		
SI index estimate, $\widehat{\Delta_c}$	0.207	0.774
White farm ownership rate, 1920	0.605	0.155
Log white density, 1920	3.028	0.776
White literacy rate, 1920	0.935	0.054
Railroad exposure	0.535	0.413
Percent black, 1920	0.397	0.212

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, Haines and ICPSR (2010) data, Aaronson and Mazumder (2011) data, and Black et al. (2015) data

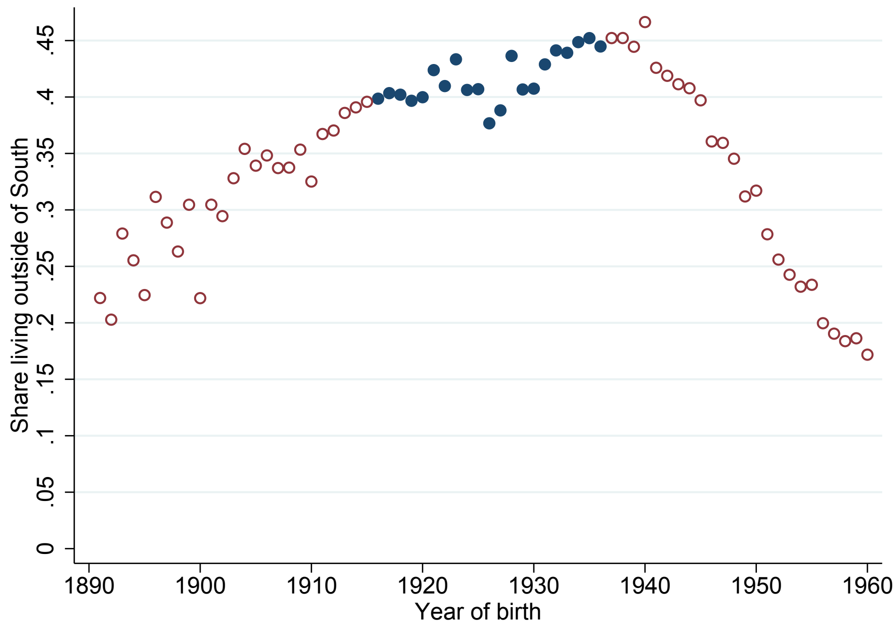
Table A.15: Social Interactions Index Estimates and Birth County Characteristics, White Moves out of South

Dependent variable: Birth county-level SI index estimate	(1)	(2)
White farm ownership rate, 1920	-0.488	-0.545*
	(0.332)	(0.330)
Log white population density, 1920	-0.216**	-0.290**
	(0.098)	(0.123)
White literacy rate, 1920	-0.108	0.555
	(0.523)	(0.578)
Railroad exposure	0.039	0.094
	(0.067)	(0.069)
Percent black, 1920	-1.279***	-1.492***
	(0.274)	(0.309)
Birth state fixed effects		x
R2	0.081	0.104
N (birth counties)	560	560

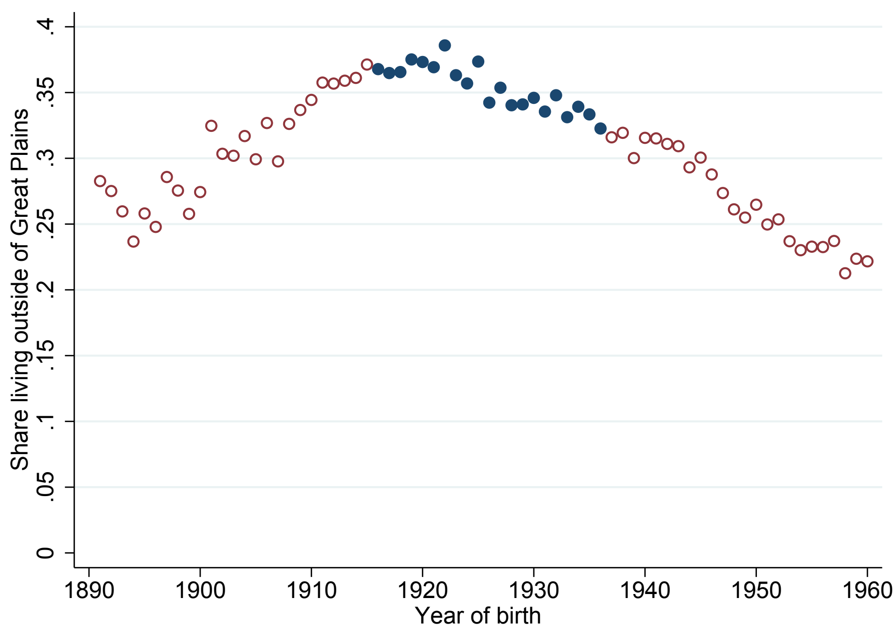
Notes: The dependent variable is the birth county level social interaction estimate. Railroad exposure is the share of migrants in a county that lived along a railroad. Heteroskedastic robust standard errors are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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

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



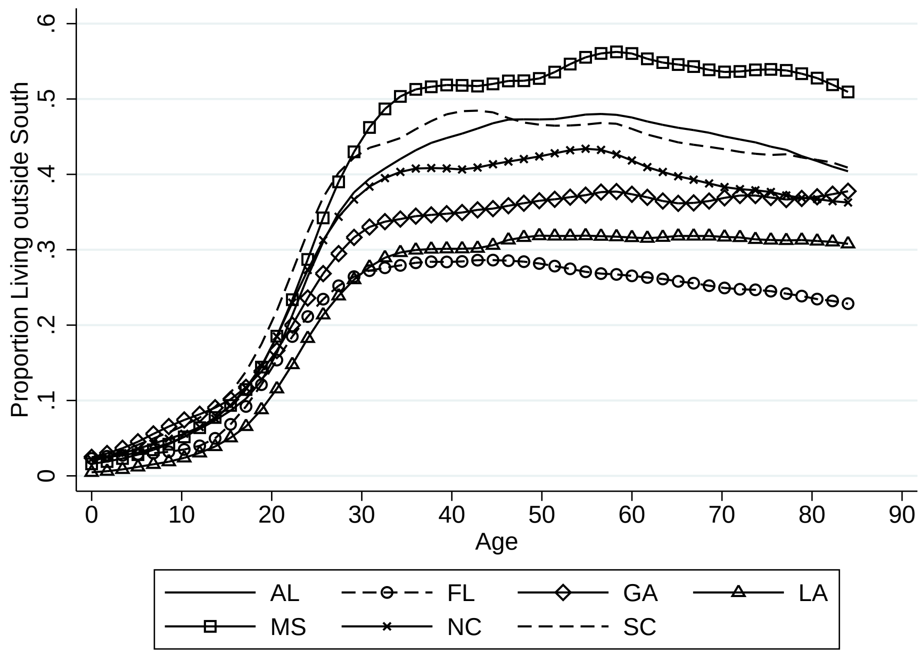
(a) South



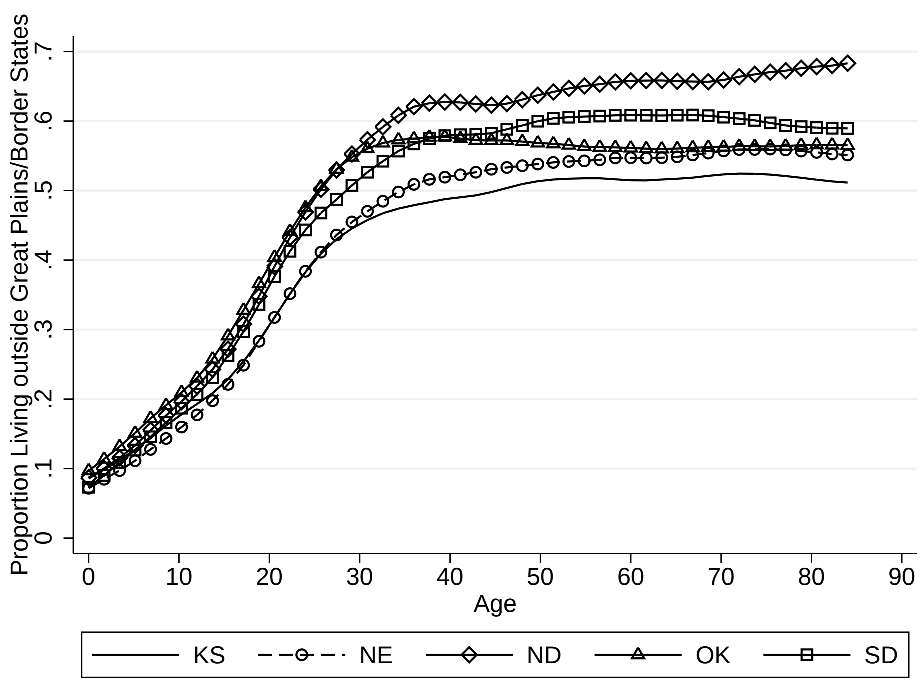
(b) Great Plains

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 3. 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. Source: IPUMS Census data, 1940-2000.

Figure A.2: Proportion Living Outside Birth Region, 1916-1936 Cohorts, by Birth State and Age



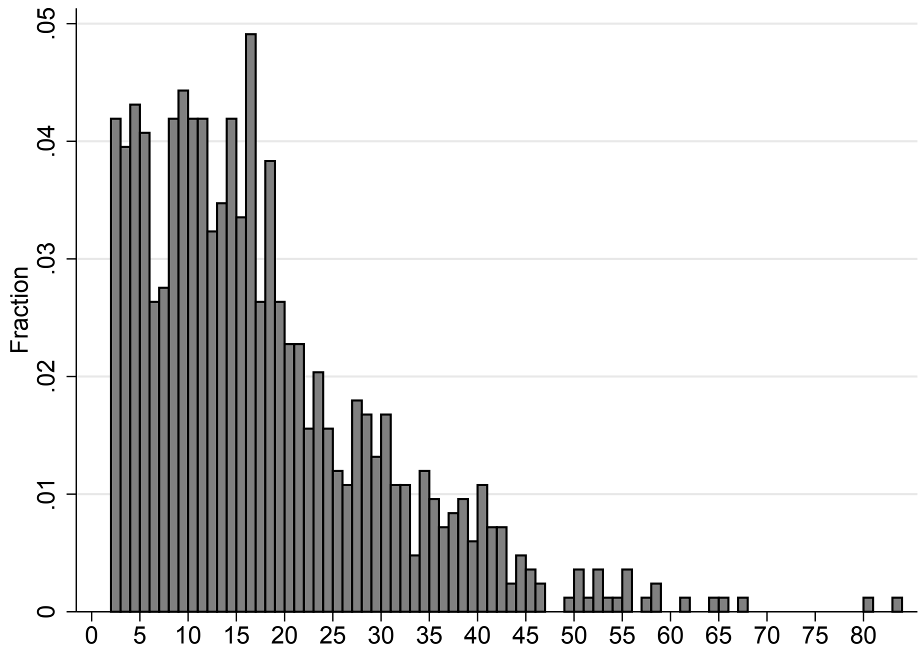
(a) Southern Blacks



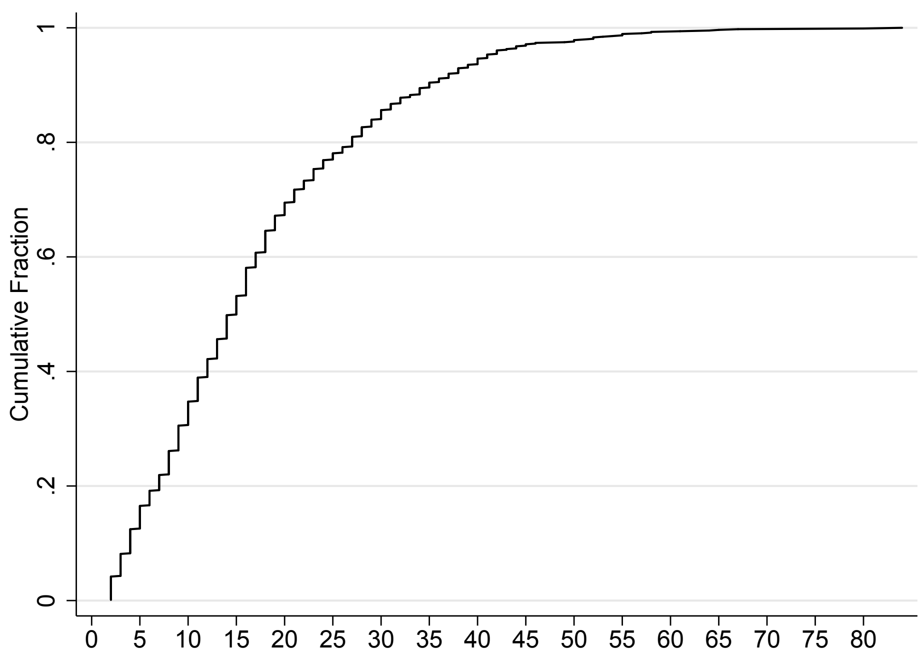
(b) Great Plains Whites

Notes: Figure A.2 displays the locally mean-smoothed relationships. Figure 3 displays birth regions.
Source: Ruggles et al. (2010) data

Figure A.3: Number of Towns per Birth Town Group, Cross Validation, Black Moves out of South



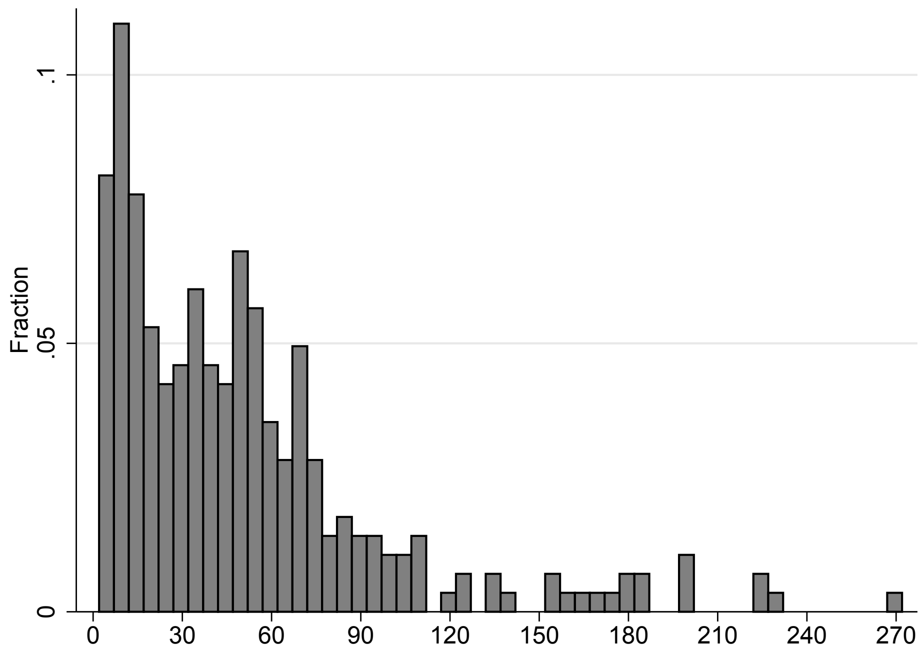
(a) Histogram



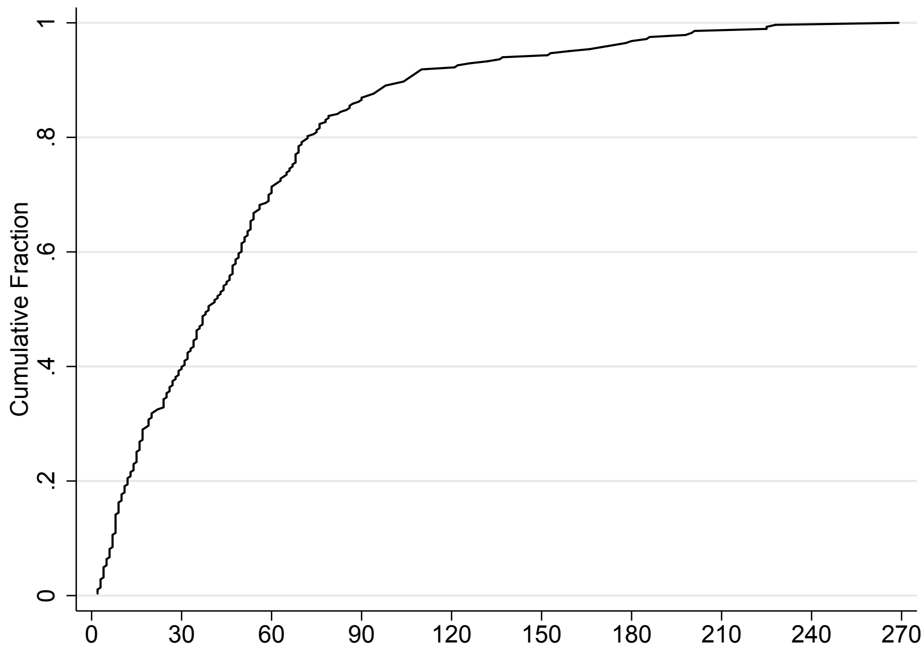
(b) Cumulative Distribution

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, White Moves out of Great Plains



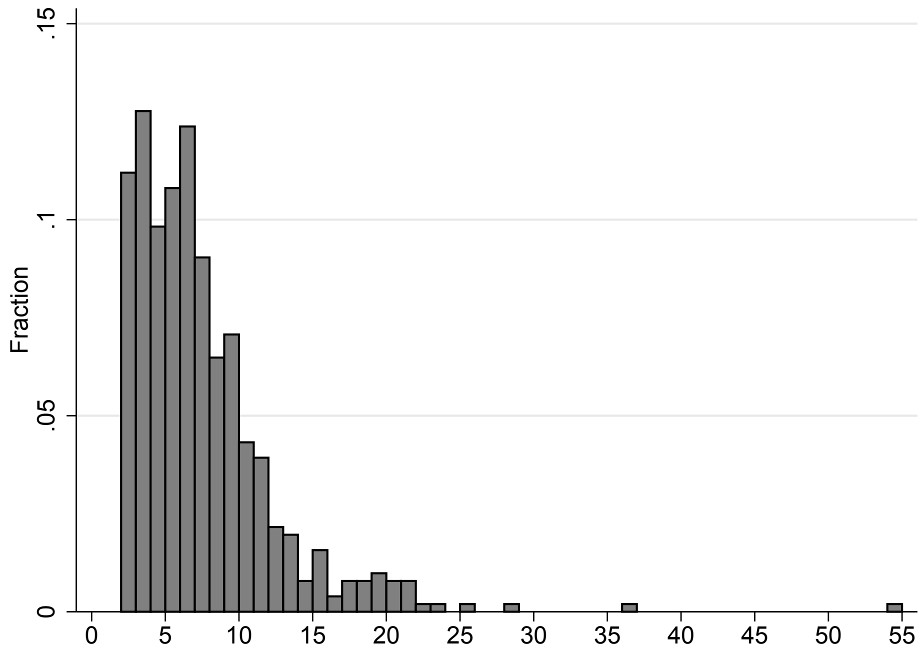
(a) Histogram



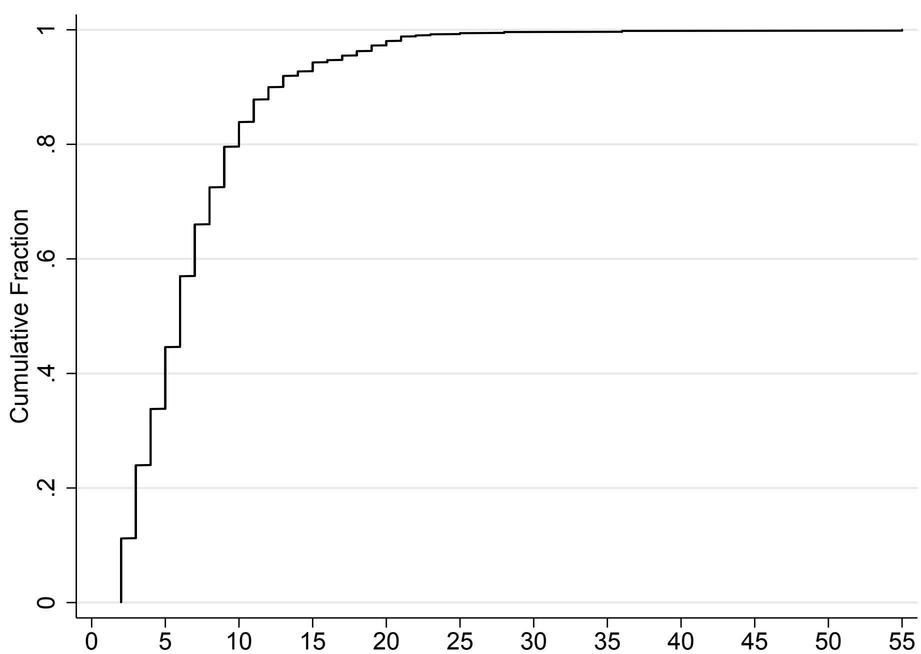
(b) Cumulative Distribution

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 County, Black Moves out of South



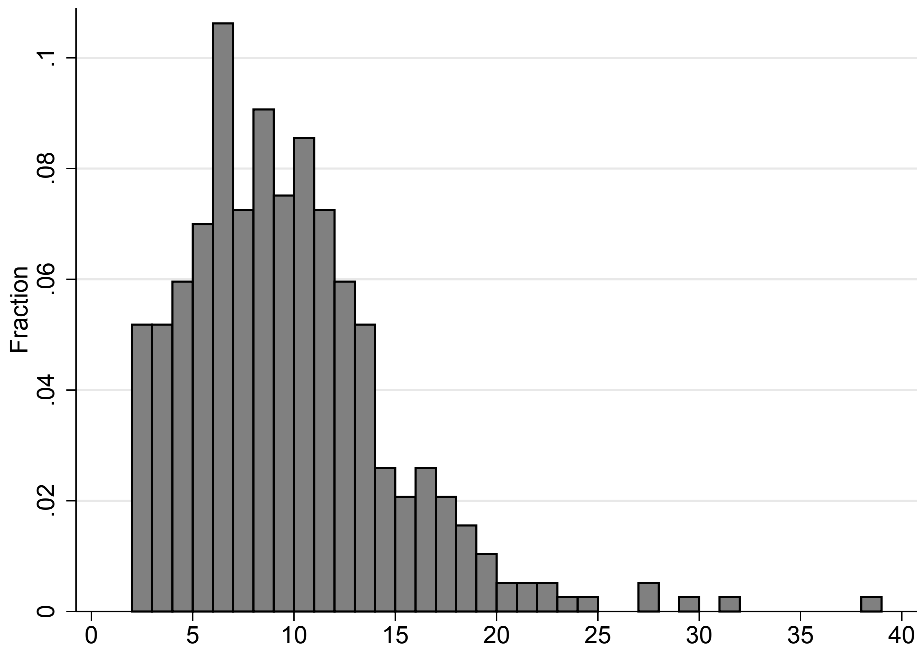
(a) Histogram



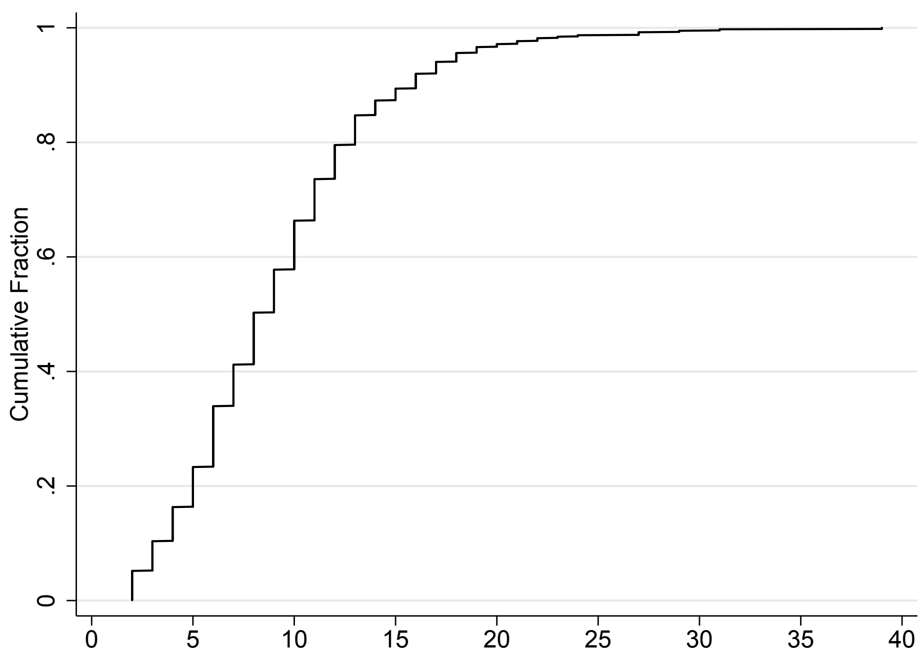
(b) Cumulative Distribution

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 County, White Moves out of Great Plains



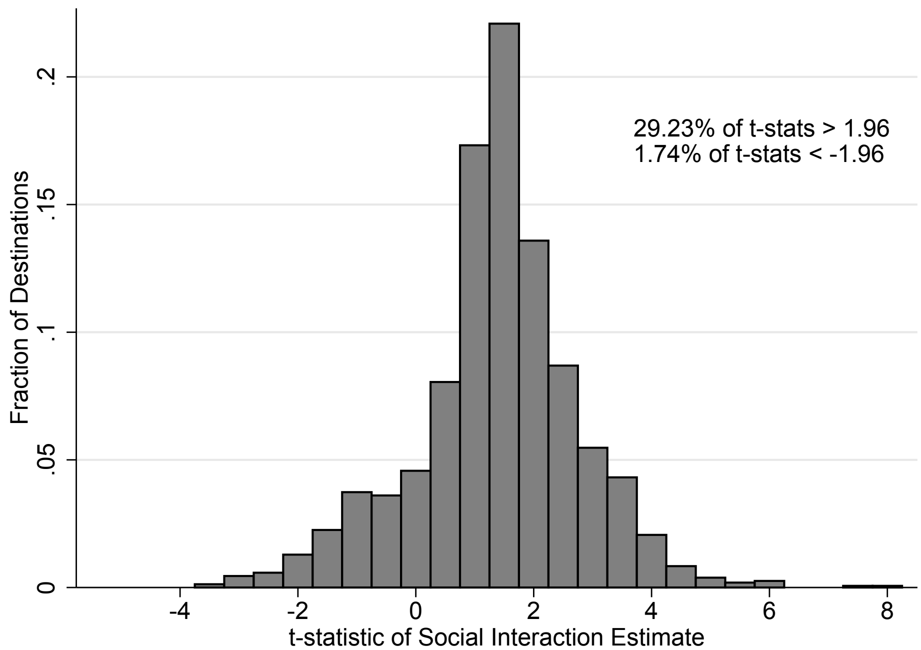
(a) Histogram



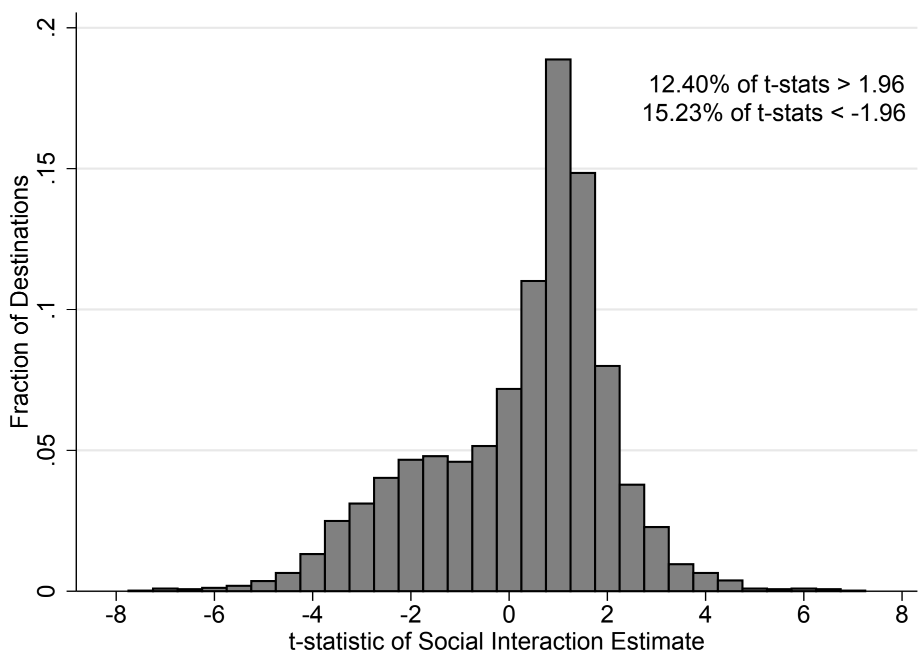
(b) Cumulative Distribution

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 Social Interactions Index t-statistics



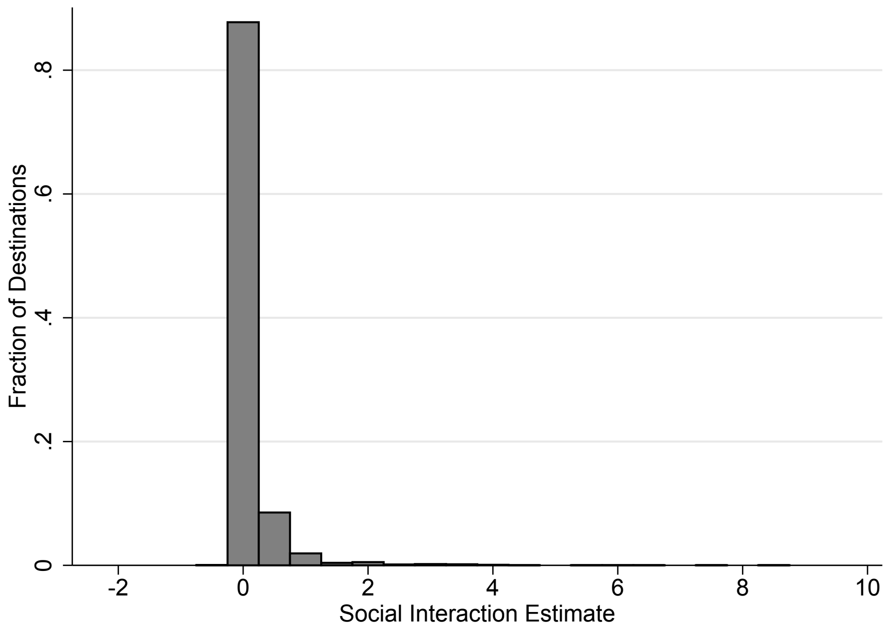
(a) Black Moves out of South



(b) White Moves out of Great Plains

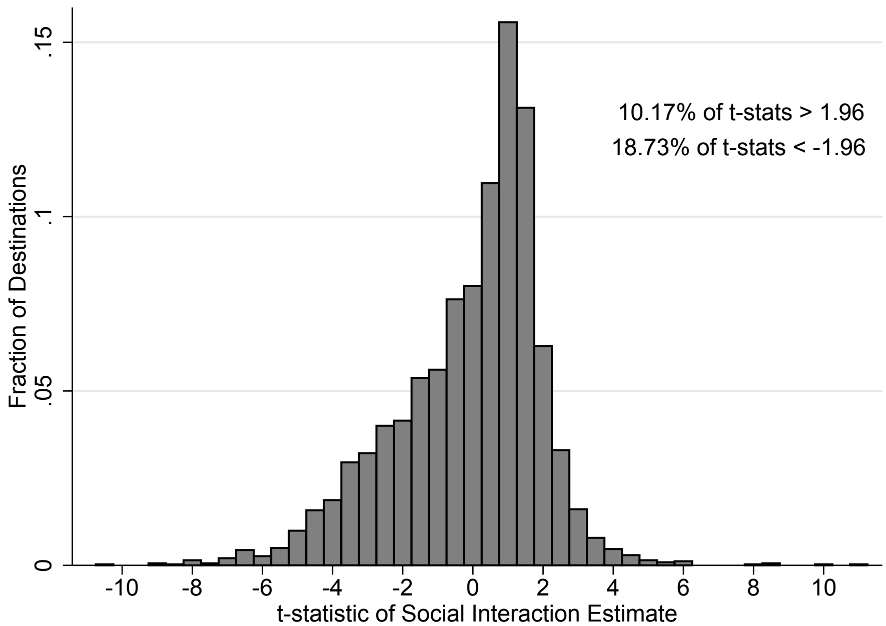
Notes: Bin width is 1/2. Birth town groups are defined by cross validation. Panel (a) omits the t-statistic of 13.7 from South Carolina to Hancock, WV.
Source: Duke SSA/Medicare data

Figure A.8: Distribution of Destination-Level Social Interactions Index Estimates, White Moves out of South



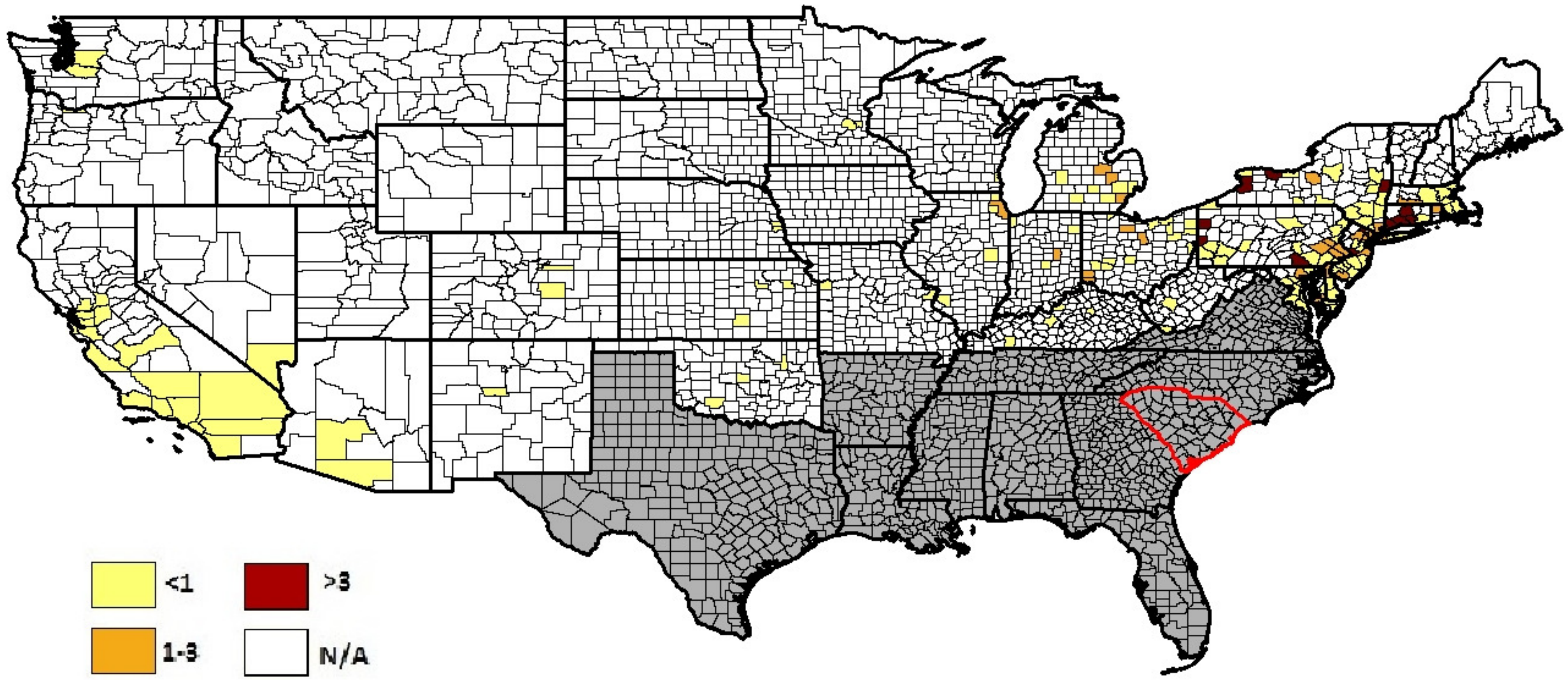
Notes: Bin width is 1/2. Figure omits estimate of $\hat{\Delta}_k = 19.3$ from Alabama to St. Joseph County, IN.
Source: Duke SSA/Medicare data

Figure A.9: Distribution of Destination-Level Social Interactions Index t-statistics, White Moves out of South



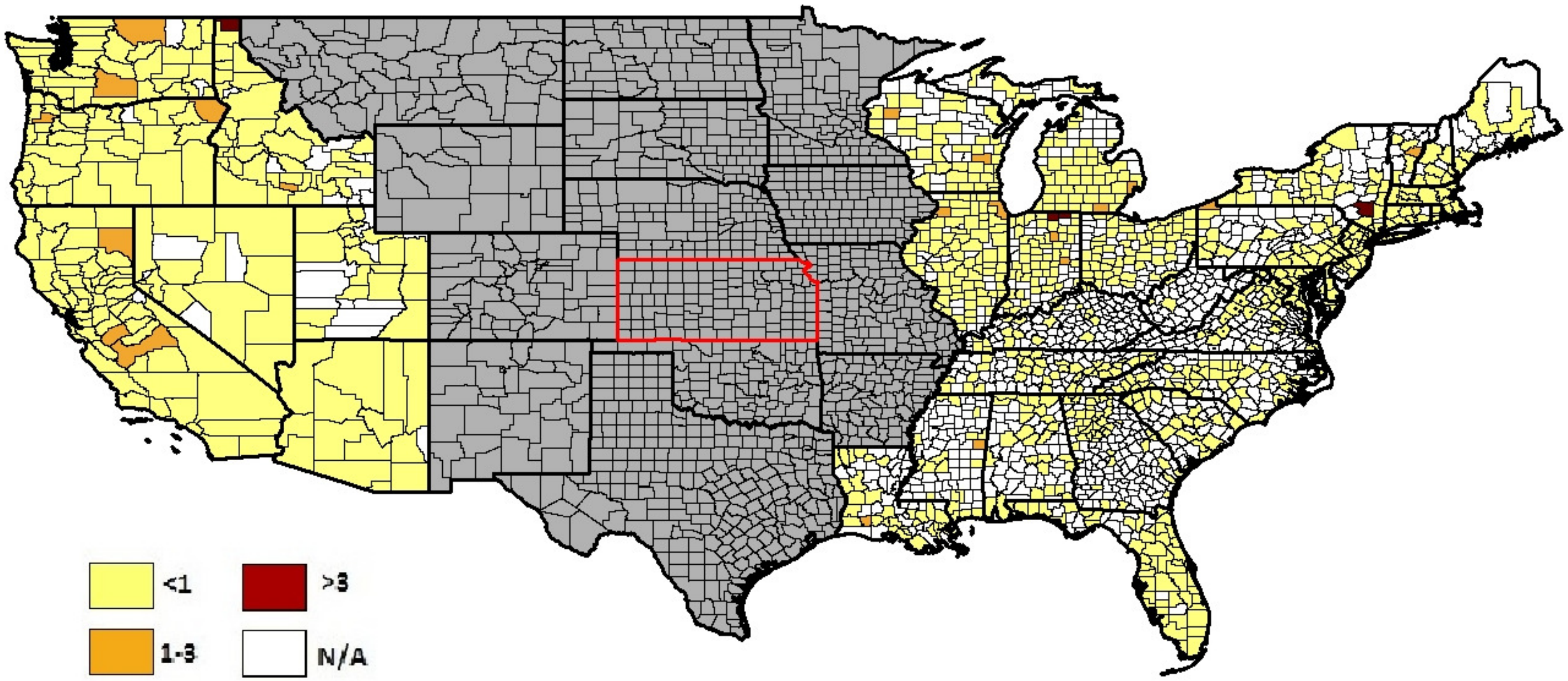
Note: Bin width is 1/2.
Source: Duke SSA/Medicare data

Figure A.10: Spatial Distribution of Destination-Level Social Interactions Index Estimates, South Carolina-born Blacks



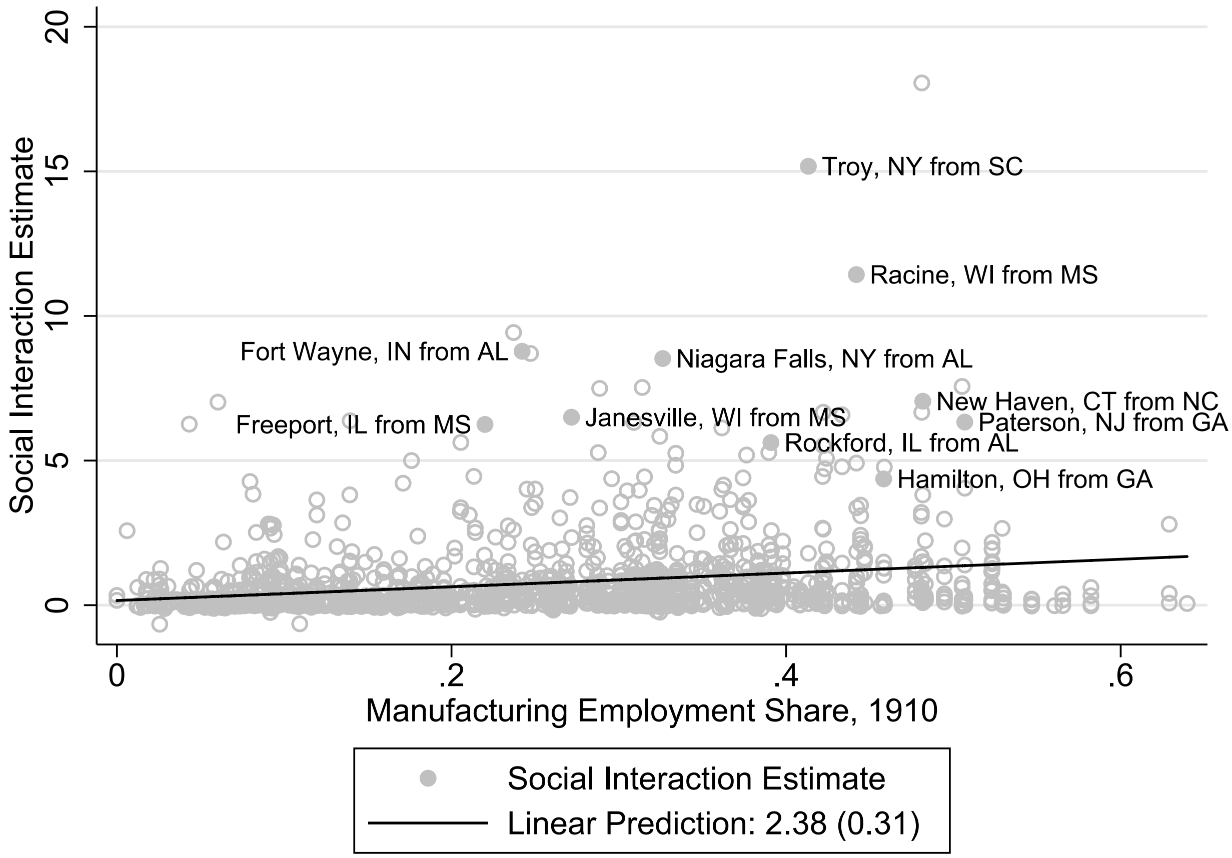
Notes: See note to Figure 5.

Figure A.11: Spatial Distribution of Destination-Level Social Interactions Index Estimates, Kansas-born Whites



Notes: See note to Figure 6.

Figure A.12: Relationship between Southern Black Destination-Level Social Interactions 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) data