The Great Migration in Black and White:
New Evidence on the Geographic Mobility of American Southerners

William J. Collins and Marianne H. Wanamaker

September 2013

Preliminary and incomplete draft: Please do not cite without permission

Abstract: We construct a new dataset of linked census records to follow southern men—both black and white—during the first decades of the "Great Migration" from the U.S. South. We find that detailed observable personal characteristics cannot account for black-white differences in migration choices in this period. Rather, black and white men responded differently to variation in the characteristics of potential destinations. Discrete choice models show that black men were relatively strongly drawn to states with high shares of manufacturing and rapid labor demand growth, were less likely to follow in the footsteps of previous migrants, were more likely to choose a northern destination, and were more deterred by distance than southern whites.

Acknowledgements: The authors are grateful for suggestions from Robert Margo, Peter McHenry, Suresh Naidu, Abigail Wozniak, and seminar participants at the University of Georgia, the ASSA Meetings (2013), and participants in the World Congress of Cliometrics (2013). Ye Gu and Justin Roush provided excellent research assistance. The Grey and Dornbush Funds at Vanderbilt University and Office of Research at the University of Tennessee have provided generous research support. NSF support (SES 1156085 and 1156057) is gratefully acknowledged. Collins is the Terence E. Adderley Jr. Professor of Economics at Vanderbilt University (william.collins@vanderbilt.edu). Wanamaker is an Assistant Professor of Economics at the University of Tennessee (wanamaker@utk.edu).

In the early twentieth century, real income per capita, labor productivity, and wages varied widely across the United States.[1] When World War I led to both a labor demand boom in industrial centers and an interruption of European immigration, southern workers moved away from the region at high rates. This was the start of the "Great Migration," which waned during the Great Depression but surged again between 1940 and the mid-1960s. By 1970, 35 percent of southern-born black men and 19 percent of white men (age 25 and over) were living outside of the region.

The Great Migration was a key event in the history of U.S. labor market integration, and it had long-lasting and wide-ranging social ramifications. Consequently, an extensive literature on its causes and consequences spans almost 100 years.[2] Unfortunately, the cross-sectional data that underpin nearly all of the quantitative studies of internal migration in this period are severely limited; researchers simply cannot observe the same person before and after migration in cross-sectional data. The absence of *ex ante* information on southern men not only obscures the place of origin and timing of migration (beyond birth state and birth year), but it also precludes efforts to study how individual and local characteristics may have influenced selection into inter-state migration and the choice of destination.

In this paper we construct a new individual-level dataset of linked census records for more than 26,000 southern males, which allows us to observe men before and after the start of the Great Migration. We start with a microdata sample of southern males, ages 0 to 40 in 1910, before the onset of mass migration. We then search for and locate the same men in the 1930 census manuscripts and transcribe data from the handwritten documents. In 1910, we observe the younger men in our sample (generally less than 18) when they still lived with their parents and siblings, and we observe the older men when they were already in the southern labor force. By 1930, many of these men had left the South or relocated within the region.

In addition to the dataset's valuable pre- and post-World War I structure, there are two novel aspects of this paper's scope. One is its inclusion of white southerners in the frame of investigation. The overwhelming focus of the previous literature has been on African American migrants. While this is understandable given the historical salience of black migration, the simultaneous mass migration of white southerners is important in its own right and provides a natural comparison for the

[1] See Easterlin (1960), Mitchener and McLean (1999), Caselli and Coleman (2001), and Margo (2004) for evidence and discussion of the regional differences.

[2] Recent work by economists includes Boustan (2009), Black et al. (2011), Hornbeck and Naidu (2012), Chay and Munshi (2012), Collins and Wanamaker (2014). Less recent but still important empirical work on black migration includes U.S. Department of Labor (1919), Lewis (1931), Higgs (1976), Vickery (1977), Gill (1979), Tolnay and Beck (1990), Carrington et al. (1996), Maloney (2001), and Vigdor (2002).

migration patterns of black southerners. It is readily apparent that black and white men followed different migration patterns in this period. The linked dataset can help us quantify and understand those differences.

The second novel aspect is that intra-regional migration and inter-regional migration are incorporated into the same econometric framework. This is a natural way to approach a multinomial choice setting, but it adds a dimension to the analysis as much of the literature has focused on the binary outcome of leaving the South. We do pay special attention to movement beyond the regional border, but treating the South and North as economically diverse places allows us to see whether the border itself (or more precisely, unobservables associated with crossing the border) influenced migration choices beyond what can be accounted for by differences in economic characteristics across states.

The new dataset allows us to answer fundamental and interconnected questions about migrants and non-migrants in the first decades of the Great Migration. First, what were the outstanding features of migration in this period, and how different were black and white migration patterns? Second, to the extent black and white migration patterns differ, can those differences be explained largely by black-white differences in *ex ante* personal and local characteristics? Third, what does a model of location choice that considers both person-specific and choice-specific variables (e.g., distance or labor demand growth) reveal about the responsiveness of southern blacks and whites to variation in economic conditions? Fourth, given estimates of the model's parameters, what might migration patterns have looked like if southerners had had higher levels of educational attainment, or different arrays of previous migrant "networks," or lower distance-related costs to migration?

We find that observable personal characteristics and place-of-origin fixed effects simply cannot account for black-white differences in migration choices in this period. Rather, observationally similar black and white men responded differently to variation in the characteristics of potential destinations during the first decades of the Great Migration. In particular, black men were more deterred by distance than southern whites, were less likely to follow in the footsteps of previous migrants than whites, and were more strongly attracted to places with rapid employment growth and high pre-war levels of manufacturing employment. Black men were more attracted to non-southern states than whites, but this seems to reflect avoidance of the North by southern whites more than a special attraction of the North for blacks (conditional on observable state economic characteristics). While there is evidence that higher levels of education diminished the deterrent effect of distance, racial differences in education cannot account for much of the racial difference in

migration patterns (e.g., whites' relative attraction to California, or blacks' relative attraction to Pennsylvania). On the other hand, it appears likely that advances in transportation technology prior to World War I had significant effects on southerners' interstate and inter-regional migration rates.

1. Background

Despite sizable regional differences in average income, relatively few southern-born men left the region prior to World War I, at least in comparison to later periods. For perspective, in 1910, about 90 percent of both black and white southern-born males, ages 30 to 40, still resided in the South. This does not imply that the southern labor force was stagnant or immobile, as there is considerable evidence of migration *within* the South. But it does suggest a regional separation that was jarred, if not fully dislodged, by the exogenous shock of World War I. In 1930, about 77 percent of southern-born black males, ages 30 to 40, still resided in the South, and 84 percent of whites.

Scholars have suggested several reasons for the relatively low rates of black migration prior to World War I: low levels of human and financial capital; discrimination by northern employers who readily hired immigrant laborers from Europe (Myrdal 1944, Collins 1997); and weak integration of regional labor markets, a legacy of slavery, combined with relatively strong economic ties between the industrial North and European labor markets (Wright 1986 and 1987, Rosenbloom 2002). Institutional impediments to economic mobility among southern agricultural workers also may have been significant (Ransom and Sutch 1977, Naidu 2010). But there were no formal policy barriers to internal migration in the United States.

The story for southern whites has been far less explored.[3] Some of the same factors that inhibited black movement may have also affected white movement, especially among the poorest whites. But it seems clear that the black and white stories differ in important ways. Most obviously, late nineteenth-century whites were not recently removed from slavery, were far more likely than blacks to have acquired some wealth and literacy, and were less concentrated geographically in the Cotton Belt. In 1870, for instance, nearly 40 percent of white southern men, age 20 to 60, owned some amount of real property, compared to less than 5 percent of blacks. Approximately 74 percent of southern whites (over age 9) could read and write, compared to 15 percent of blacks. Arguably, in the early twentieth century, southern whites were less subject to discrimination from northern

[3] Gregory (2005) provides a broad discussion of the social impacts of white and black migration from the South. Also see Fligstein (1981). There is some research specifically on southern white migrants and their assimilation in northern cities (Akers 1936, Killian 1953, Luebke and Hart 1958, Berry 2000), but not much on the period we are studying or written by economists.

employers and employees than were blacks, but they also might have had more scope for economic advance in the South than black workers.[4] It is also possible that lingering resentment from the Civil War influenced white southerners' perceptions of the North.

To the extent that ignorance and poverty precluded large-scale, long-distance internal migration, this constraint should have receded with each generation's educational and economic advances (Higgs 1982, Margo 1990, Collins and Margo 2006). Improving transportation and communication networks in the South also may have facilitated out-migration. For African-Americans specifically, political disenfranchisement, mob violence, de jure segregation, and, in general, the ascendance of the Jim Crow regime may have provided an increasingly strong incentive to migrate from the region.[5]

Taken together, these trends may have yielded a southern labor force circa 1910 that was more capable of long-distance movement than ever before, but it was the exogenous eruption of World War I that decisively altered the rates and patterns of inter-state migration. It created both a major demand boom in industrial centers (predominantly northern) and a temporary halt to European immigration, which was later reinforced by immigration restrictions. Many northern employers recruited southern migrants for the first time, gained experience in hiring and evaluating them, and established networks to draw on the southern labor supply (Whatley 1990; Berry 2000; Foote, Whatley and Wright 2003). As is commonly found in studies of migration (Hatton and Williamson 1998, McKenzie and Rapoport 2007), networks of previous migrants helped perpetuate migration patterns (Carrington et al. 1996). Consequently, the Great Migration continued long after the impetus of World War I.

2. New Data and an Overview of Migration Patterns

Linking Micro-level Census Data from 1910 to 1930

Our approach merges a new panel dataset of black and white males who resided in the South in 1910 with county and state-level census data (Haines 2010). To construct the panel data, of course, we must locate each individual at two points in time. We began with IPUMS sample of the 1910 United States *Census of Population* and selected all southern-resident males between the ages

[4] Killian (1949, pp. 201-217), for example, reports on less-than-favorable views of southern whites held by several employers in Chicago. But this seems mild in comparison to views held about African Americans.

[5] See Kousser (1974) and Woodward (1974) on disenfranchisement. See Tolnay and Beck (1990) on mob violence and lynching. See Margo (1990) on widening racial gaps in school quality. Myrdal emphasized that blacks were shut out of growing employment in southern manufacturing (1944, p. 188). The term "Jim Crow" is a commonly used embodiment of the segregated South.

of 0 and 40 (Ruggles et al. 2010).[6] We then attempted to link these men "forward" by locating them in the handwritten manuscripts of the 1930 census. We used each individual's name, place of birth, and age from the 1910 Census as search criteria for location in the 1930 census.[7] From an initial sample of 111,524 individuals, the linking process successfully located 26,829 individuals, a 24 percent match rate.[8] The final sample is 20 percent black.

As mentioned above, the linked data offer several advantages relative to the state-level aggregates or micro-level cross-sections that have typically underpinned quantitative studies of the Great Migration. The key distinction is that we observe the same person both before and after the onset of the Great Migration's first wave. In 1910, we see the younger men (aged roughly 0 to 18) in our sample when they still resided with their family. Therefore, we observe several characteristics about the household in which they grew up—what their parents did for a living, where they were located, whether they attended school, whether they had siblings, and so on. The older men in our sample (18 and older) in 1910 are observed after they have left their parents, but also after they have entered the southern labor force. Therefore, we know what kind of job they held in 1910, in addition to whether they were literate, where they lived, whether they were married, whether they owned a home, and so on. In 1930, we see whether the men have moved since 1910 and, if so, where they have moved and what kind of job they have. It is straightforward to include county-of-origin information from Haines (2010), which characterizes each individual's environment, beyond personal and household variables.

A major concern for any dataset that is constructed by linking individuals across census dates is that the linked sample might exhibit selection bias that interferes with subsequent analyses and interpretations. In our case, the men in our linked sample are very similar to those in the full, randomly selected base sample from 1910. Table 1 reports the summary statistics for linked and full samples of blacks and whites separately. Although there are some small differences in observable

[6] Southern states in this paper include: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. We treat Delaware, Maryland, and the District of Columbia as non-southern.

[7] Our search criteria include a SOUNDEX version of the individual's last name, the first three letters of the first name, the state of birth, and birth year within two years. SOUNDEX is an algorithm used to generate alternative spellings of a surname. SOUNDEX matches include the exact last name and any reasonably close approximation to that name.

[8] Our linkage rates are similar to those in Ferrie and Long (2013) and Abramitzky, Boustan, and Eriksson (2012) who also create samples of linked census records. We deleted cases where more than one individual in 1910 matched to the same individual in 1930. Match rates for blacks (19%) and whites (26%) differ primarily because of advances in search technology over time. (The samples were matched sequentially.) These advances have no systematic relationship to the outcomes of interest. For all analysis in this paper, we weight observations to reflect the differential match rates.

characteristics (e.g., white men in the linked sample are slightly more likely to be attending school and living in owner-occupied housing in 1910 than men in the full IPUMS sample), the overall results in Table 1 reveal no strong evidence that the linked samples exhibit biased selection. Therefore, we take our sample and the results to be fairly representative of the southern male population.

A separate check with the 1930 IPUMS cross-sectional sample of southern-born black men, age 20 to 60 (to correspond to those 0 to 40 in 1910), reveals that 22 percent resided outside the South at the time of the 1930 census. This is close to the 20 percent of our matched sample who resided in the South in 1910 but not in 1930. The corresponding numbers for white migrants are 15 percent and 17 percent.[9]

Migration Patterns in the Linked Data

Table 2 provides an overview of black and white migration patterns, with focus on the propensities to migrate (Panel A), distance and direction of migration (Panel B), and characteristics of chosen states among migrants (Panel C). By 1930, a large fraction of men in our sample (35 to 40 percent) had left their 1910 state of residence, with approximately an even split between "within-South" and "outside South" migration.[10] Among those who left the South, the Midwest was the most common destination. Blacks also selected strongly into the Northeast, whereas whites selected relatively strongly into the West. We calculate measures of distance-travelled based on the latitude and longitude of the center of each individual's 1910 and 1930 counties of residence (Panel B). In the full sample, average migration distances (including zeros for those who did not move) are remarkably similar across the black and white samples. Among the inter-state migrants, whites moved further than blacks on average both within the South (322 versus 266 miles) and when leaving the South (696 versus 577 miles).

Panel B of Table 2 also shows differences in migration directions, as reflected in the change in latitude and longitude of county of residence. A positive value for change-in-latitude is northward migration, and a positive value for change-in-longitude is eastward migration. For all non-migrants, the values of both are zero. In the full sample, the average black male moved northward and eastward, whereas the average white male moved northward and westward (though not as far north

[9] We do not expect these numbers to be exactly the same because of interregional migration prior to 1910 and sample variability.

[10] Again, we classify DC, Maryland, and Delaware as non-southern. The within-South movement has received far less attention from scholars than inter-regional migration, and it is not the focus of this draft of the paper, but it is worthy of more investigation and appreciation (see Wright 1986, p. 65).

as blacks). These differences are amplified as we narrow the sample to interstate southern and inter-regional migrants. Among white inter-regional movers, the migrants to the far West appear to strongly influence the average change of longitude

Figure 1 maps the distribution of interstate migrants for blacks and whites separately. Black migrants were drawn heavily to the industrial and urban centers of the North: Pennsylvania, Illinois, Ohio, New York, and Michigan. Within the South, black migrants most frequently chose Florida and North Carolina. White migrants were strongly drawn to Ohio and, to a lesser extent, Illinois in the North, but Texas, Oklahoma, and California were also major destinations.[11]

These differences in location choice are echoed in Panel C of Table 2, which presents the average characteristics of destination choices as measured in 1910. On average, black interstate migrants chose states that were more urban and had higher manufacturing employment, higher average output per worker, and lower agricultural employment than white choices. In addition, blacks were more likely to migrate to states with higher black populations in 1910.

3. Approaches to Characterizing Migration Patterns and Black-White Differences

Can ex ante characteristics account for black-white migration differences?

To this point, we have characterized the differences in migration patterns between blacks and whites as "racial" differences. But of course blacks and whites differed in a number of ways that may have influenced migration decisions. In 1910, blacks and whites had different distributions across places within the South, had different levels of education and literacy, had different occupational distributions, and so on.[12] It is possible that adjusting for these ex ante characteristics may explain, in a proximate sense, the racial differences in migration patterns.

To fix ideas, if black and white workers with similar characteristics and from similar environments responded in the same way to the economic shock of World War I, then we would be able to explain a large share of black-white differences in migration patterns with controls for

[11] More detailed breakdowns of the data reveal that white inter-regional migrants were drawn heavily from states near the regional border—Kentucky, Virginia, Arkansas, and Oklahoma had the highest white out-migration rates. Blacks in these states had high inter-regional migration rates, too, but they also had high rates of migration from states that were far from the regional boundary, such as Florida, Georgia, North Carolina, and South Carolina. (These detailed breakdowns are reported in Appendix A1.) Moreover, the black population was relatively concentrated in those "far" states of the Deep South.

[12] To a large extent these differences in observables are rooted in the legacy of slavery and plantation agriculture, including ongoing discrimination in the provision of education and protection under the law. In this sense, there is a very real "racial" dimension to differences in observables like literacy and initial location. The point of the exercise described in the text, however, is to study the migration decisions of observationally similar men circa 1910 and to see whether adjustments for observables narrow (or widen) the black-white differences in migration patterns.

detailed background characteristics.[13] This motivates two simple and preliminary approaches to examining the data. First, we estimate linear probability and OLS models of migration outcomes. Then, we estimate multinomial logit discrete choice models that treat each state as a potential destination. These approaches provide a useful starting point and help answer the basic question posed in the subsection's heading. Moreover, they make full use of the extensive set of background information that we have about each migrant. A richer model of location choice, however, would incorporate variation in benefits and costs across alternatives. We develop such a model in the next subsection.

For the first approach (linear probability and OLS models), the fully specified models are described in Equation 1:

$$P_{iac} = \tau R_{iac} + \beta X_{iac} + \gamma_a + \theta_c + e_{iac}. \qquad (1)$$

P_{iac} is a migration outcome variable, such as "inter-regional migrant," "miles moved," or "change in latitude." R_{iac} is the indicator variable for race (1=black). γ_a and θ_c are age and county-of-residence (in 1910) fixed effects, and e_{iac} is the error term.

X_{iac} is a vector of detailed individual and household characteristics recorded in the 1910 census manuscripts. A different set of X_{iac} variables is available depending on whether the person had left his parents' home, and so we estimate separate regressions for those under 18 and residing with father and those 18 and over. For the younger men in our sample, X_{iac} includes indicators for the father's industry, whether the family resides in owner-occupied housing, the individual's literacy status (only available for those aged 10 and older), whether the person is a first-born male, a measure of skill (only available for those aged 11 and older), and whether the individual attended school in 1910.[14] For the older men in our sample, X_{iac} includes indicators for one's own industry, owner-occupied housing status, marital status, literacy status, and a measure of skill based on occupation of employment.[15] For both young and old men, we include indicators for residence in small cities (<25,000 residents) and large cities.

When the regression is estimated without covariates or fixed effects other than race, τ simply measures the difference in the outcome for blacks relative to whites. We will refer to this as the

[13] Implicit in this statement is the idea that observationally similar blacks and whites would have had access to the same variety of labor market opportunities at home and in other states, which of course might not be true. This is why our main analysis incorporates more information about potential destinations.

[14] There are 19 industry categories based on Lebergott (1964) and IPUMS industry coding (usa.ipums.org/usa/). For individuals under age 18, we estimate the years of schooling that individual will acquire by age 18 – a number that depends on age, current school attendance, and race. More details can be found in the Appendix A3.

[15] The variable is similar to the IPUMS variable edscor, but is based on the 1940 education level of individuals in the same occupation. We convert this skill measure into a standardized Z-score, by age, so that skill is measured relative to peers of the same age. See the Appendix A3 for further detail.

"unadjusted" racial difference. When run with the full specification, τ will measure the "adjusted" difference in choice probabilities, conditional on all these observable characteristics, including age and county-of-origin fixed effects. The hypothesis is that differences in observable personal characteristics and unobservable county-level fixed effects explain black-white differences in migration behavior. If so, then our estimates of τ in the fully specified regressions will approach zero.

Table 3 reports the estimates of τ for a variety of outcomes: inter-regional migration, migration within the South, distance travelled, and the change in latitude and longitude.

Black men who were young in 1910 were substantially more likely to leave the South by 1930 than were white men (by about 7 pp).[16] Adjusting for observables in 1910 tends to widen the racial difference for young men to about 9 pp (though the standard error also increases), and opens a black-white gap among older men (to about 4 pp). Relative to the baseline regional outmigration levels of about 20 percent (in Table 2), these are relatively large racial differences in inter-regional migration rates, and it is clear that accounting for background characteristics does not account for the racial difference in outmigration.

The propensity for interstate movement within the South, however, is fairly similar for blacks and whites, and controlling for background characteristics fully accounts for that difference among older men. Differences in distance travelled are negligible, but differential changes in latitude and longitude are significant (economically and statistically), and again background characteristics do not account for the racial differences. On average, blacks move substantially further north and east relative to whites, even after conditioning on starting location.

Of course, these differences in change-in-latitude, change-in-longitude, distance moved, and region of residence reflect differences in the underlying choice of a specific location. To examine the choice of location more directly, we estimate multinomial logit models that are similar in spirit to the specification described in Equation 1. But in this case, instead of binary outcomes (e.g., inter-regional migration) or continuous outcomes (e.g., miles moved), each state is treated as a discrete choice.[17] This allows us to characterize black-white differences in migration choices across states,

[16] The sample reported in the unadjusted column 1 of Table 3 is only those individuals for which all control variables employed in column 2 are available.

[17] With a large number of choices and a large number of control variables, computation is cumbersome—a separate coefficient is estimated for each variable for each potential destination. Therefore, we consolidate some states that receive very few migrants in our dataset and reduce the set of control variables. The control variables include state of origin fixed effects, age fixed effects (if over 18), whether the person lived in owner-occupied housing, whether the person (or their father, if under 18) was literate, an industry of employment fixed effect (if over 18), whether

net of differences in observable characteristics such as state of origin, literacy, industry of employment, and whether residing in owner-occupied housing.[18]

For each state, we calculate a conditional and unconditional measure, δ_s, which is the difference in black and white average migration probabilities to state *s*. In the unconditional case, δ_s is simply the black-white difference in the average probability of migration to a particular state. In the conditional case, δ_s reflects the black-white difference in the predicted probability of migration to each state after adjusting for differences in observable characteristics.[19] Because a different coefficient on race is estimated for every state in the multinomial logit, these results are not easily summarized in a table, but can be conveyed in a figure. Figure 2 plots the difference in black and white migration to each state as represented by $\delta_{unconditional}$ and $\delta_{conditional}$. If observable characteristics largely account for the differences in state choices, the $\delta_{conditional}$ measures will be closer to zero than the raw differences—the plot should look relatively flat. If not, the plotted points will be close to the 45-degree line.

Plots of the conditional and unconditional difference in migration probabilities in Figure 2 closely track the 45-degree line, indicating a quite limited ability of observable characteristics to explain black-white migration differences. Further, deviations from the 45-degree line are roughly symmetric, indicating that controlling for initial conditions is just as likely to increase racial differences in migration probabilities as decrease them. This conclusion is confirmed by Figure 3, which shows the estimated probabilities of migration to each state for blacks and whites, net of the effect of differences in other observable characteristics between racial groups. Again, we find substantial differences in migration probabilities by race, and the overall patterns are similar to the unadjusted patterns plotted in Figure 1.

Using both a simple regression and a discrete choice framework, we find that racial differences in observable background characteristics in 1910 simply cannot account for much of the black-white difference in migration choices. Instead, similarly situated blacks and whites were making substantially different choices about whether and where to migrate.

Modeling Choice of Location

---

attending school (if under 18), whether the individual was a first-born child (if under 18), and indicators for small (<25,000 residents) and large cities of residence in 1910.

[18] The "home" state is a separate choice category in this analysis.

[19] Conditional migration probabilities are calculated by using the coefficients from the multinomial logit analysis to predict black and white migration probabilities to each state *after equalizing all other observable characteristics across races.* In effect, we are netting out the differences in other observable characteristics and then re-calculating the racial difference in migration probabilities to each state. See Appendix A2 for further details.

To better understand the distribution of migrants across locations, it is natural, and apparently essential, to incorporate variation in the characteristics of potential choices into the econometric framework. Importantly, in the early 20th century, many of these place-specific characteristics may have varied by race. For instance, we will attempt to take account of variation in race-specific income levels and race-specific networks of previous migrants. We also include measures of industry composition and a simple variable "North" to capture some region-specific amenities (such as more secure civil rights or less rigid social segregation) that could have differentially attracted black and white migrants.

By moving toward a framework that emphasizes cross-state variation in economic characteristics, our work moves closer to that of Vickery (1977) and Collins (1997), who worked with state level aggregates in studying cross-state variation in black (and only black) migration rates. But because we continue to work with linked micro data we can continue to incorporate some key individual-level information into our analysis, such as the distance from county-of-origin to prospective destinations, estimates of individual-level schooling that draw on detailed personal and local information, and 1910 farm or urban status. Moreover, we can do so in a way that allows people to choose their home state or any other, rather than excluding non-migrants from consideration.[20] Finally, because the sample includes both white and black men, we can measure racial differences in the responsiveness to certain state characteristics.

In this setting, we are guided by Sjaastad (1962), McFadden (1974), Roback (1982), and Moretti (2011). In particular, we suppose that people relocate in response to spatial differences in expected income, amenities, and relocation costs. In a setting with many choices, we assume that individual *i* chooses to migrate to state *j* if

$$U(Y_{ij}) > U(Y_{ik}), \forall\, k \in \{1, \ldots, J\}\ and\ j \neq k$$

where U(.) represents utility over a vector Y_{ik} containing cost, income, and amenities for individual *i* in state *k*.

If the utility over each state choice includes a random component with an Extreme Value Type 1 distribution, the probability of choosing any particular state is represented by:

$$P(D_i = j) = \frac{\exp(\gamma' Y_{ij})}{\sum_{k=1}^{J} \exp(\gamma' Y_{ik})}, \quad \forall j \in [1,2,\ldots J] \qquad (2)$$

[20] In this way, our work is similar to Wozniak (2009). Also see Davies et al. (2001) for an application of conditional logit estimation to interstate migration with modern data. In theory, an alternative-specific conditional logit model (STATA's asclogit) would allow us to simultaneously incorporate a variety of individual and destination-specific characteristics, but in practice, this approach posed computational problems.

where D_i is the location choice of individual i. This is the familiar conditional logit framework described in McFadden (1974). An important characteristic of this approach is that any component of Y_{ij} that does not vary across locations, including individual-specific variables like age or race, will fall out of the estimating equation.[21] However, interactions of individual-specific variables with variables that differ across states are identified.

We include variables in Y_{ij} that reflect the expected costs and the benefits of potential destination choices. First, we calculate the distance from individual i's county of residence in 1910 to each potential choice state j (the value is zero for the home state).[22] This captures relocation costs that are proportional to distance, and our specification allows for nonlinearities in those costs. In addition, we calculate the share of all people born in state i who resided in state j in 1910, separately by race, using the 1910 IPUMS (Ruggles et al. 2010). Essentially, these are pre-1910 migration rates from state i to state j for blacks and whites. This helps capture the influence of unobserved relationships between states i and j, including the influence of networks that facilitate migration by providing a cultural home, assistance with finding employment and housing, and so on, all of which lower the expected cost of migration.[23]

Next, we include a number of state-level economic characteristics that may have influenced expected earnings. Some of these variables reflect conditions circa 1910 and are meant to capture differences across states in pre-Great Migration wage levels (by race) and industrial structures; others attempt to capture labor demand shifts after 1910. We estimate state and race-specific wage levels circa 1910 by combining measures of average real output per worker from Mitchener and McLean (1999) with race-specific adjustment factors calculated from the 1940 census microdata (the first census year with wage data).[24] We also include a measure of the percentage of the 1910 labor force employed in agriculture and manufacturing industries (separately) in the estimation. Southern blacks and whites may have had different sets of skills (or preferences or opportunities reflecting variation

[21] In addition, for specifications with state-alternative fixed effects, components of Y_{ij} that vary across state choices but not across individuals within state choices cannot be separately identified.

[22] We average the latitude and longitude of the actual 1930 locations of southern migrants in our sample in that state and measure the distance to that "average" (synthetic) location.

[23] For the home state, this is essentially the rate of non-migration, which does not correspond well to the "network" interpretation suggested in the text. In practice, we allow the variable to have a different coefficient for the home state by interacting it with an indicator for the home state, as discussed below.

[24] We average the Mitchener and McLean productivity levels from 1900 and 1920 for each state. Then, we use the 1940 microdata to calculate ratios of black (or white) wages to all wages in each state. We use those ratios to scale the Mitchener and McLean-based productivity measure up or down. For race-state pairs with small counts in the 1910 IPUMS, we use an adjustment factor based on region.

in discrimination) that drew them differentially to different types of jobs, and measures of overall employment growth or pre-war average wages may not fully capture these differences.

We estimate post-1910 labor demand shifts with a Bartik-style measure of predicted employment growth in each state:

$$B_j = \sum_{l=1}^{L} e_{jl} * g_l$$

where *l* indexes industries. The measure is generated by multiplying the size of the 1910 labor force in each industry in each state (e_{jl}) by a 1910-1930 nationwide employment growth rate in that industry (g_l) and summing across industries.[25] The advantage of the Bartik measure is that it gauges employment growth in any particular state without utilizing ex post (and endogenous) state-specific growth rates.

Finally, we include a dummy variable for the "home state" in the analysis to capture the strong draw of the initial state of residence for potential migrants.[26] For regressions that are run without state fixed effects, we include state-level controls for the percent of the state population in urban areas and a cubic in the state's total population simply because larger states are likely to attract a larger share of migrants from any particular source, all else the same.

We interact each state-specific variable with a binary variable, *Race* (=1 if black). The main specifications also include a dummy variable for non-Southern states, *North*. The coefficient on *North* will be negative (or less than 1 in odds ratio form) if southern whites avoided the North, conditional on cross-state observables. The interaction *North* × *Race*, reveals whether southern blacks were differentially attracted to the North (relative to whites), perhaps reflecting race-specific amenities associated with greater civil and social liberties.

Results

In columns 1-2 of Table 4, we include all the variables mentioned above. Column 1 is our baseline estimate. Column 2 incorporates state-alternative fixed effects. Variables that do not vary across individuals within a state-alternative (e.g., state population, urban percentage, manufacturing and agriculture shares, Bartik employment growth) are not identified because they are absorbed by

[25] See Bartik (1991).

[26] Each individual in the data has the same home category choice. For example, an individual living in Alabama in 1910 does not have "move to Alabama" as an option in our framework. Rather, they may move to "home", which is the same option as moving to Louisiana for an individual living in Louisiana. The latter does not have Louisiana as a separate option. The HOME indicator simply allows the home state to have a draw separate from that of all other states. Results are quantitatively similar if we keep the original set of state choices (allowing those in Alabama to choose to move to Alabama), and simply include the HOME indicator.

the state fixed effect. Interactions on these variables with race or skill variables are still identified. The key results are largely consistent across specifications.

We report estimates as odds ratios since logit coefficients are not directly interpretable. Odds ratios greater than 1 indicate higher probabilities of migration to states with higher values of the covariate in question, whereas values lower than 1 indicate the opposite. To facilitate interpretation, we convert continuous variables (*Distance*, *Wage*, and *Employment Growth*) to units of standard deviation. For instance, an odds ratio of 1.2 for such variables implies that one-standard deviation increase in the variable in a state raises the odds of someone choosing that state by a factor of 1.2. Percentage measures (*Agriculture*, *Manufacturing*, and *Network*) are multiplied by 100%. All other variables are binary.

Distance strongly deterred migration, and more so for blacks. The specification includes higher order terms in distance to allow more flexibility, but this makes interpreting the odds ratios difficult. Therefore, Figure 4 plots the odds ratios for whites and blacks separately, over a range from 25 to 2,500 miles. The log odds of migrating shorter distances are greater for blacks than whites, but the comparison reverses at mileages of 980 and greater when blacks are deterred by distance more than white. The stronger deterrent of distance to blacks is consistent with differences in pre-migration wealth constraints on long-distance mobility. At a proximate level, this result also reflects whites' relatively high propensity for movement to California.[27]

As expected, pre-1910 migration rates are highly predictive of post-1910 migration patterns for both whites and blacks. More surprisingly, the evidence indicates that pre-existing connections were *less* important for black migrants than for whites, reflecting a stronger re-orientation of black migration patterns in this period relative to the one before. This is also interesting in light of work in development economics that suggests networks are particularly important for less-skilled migrants (McKenzie and Rapoport 2007), given that blacks had substantially lower education levels than whites on average. While we find some evidence consistent with differences across skill levels (discussed below), the black-white difference remains after including the skill variable. This might reflect the nature of the World War I demand shock, which likely had a bias toward unskilled workers, and the response of northern employers who turned to the South for new workers, often hiring blacks for the first time.

The significant coefficient on *North* indicates that southern whites were more likely to stay in the South than one would expect on the basis of cross-state differences in economic characteristics

[27] The result is not explained by Western migration itself as the pattern holds even in the absence of Western states in the choice set.

and whites' responsiveness to those characteristics. Black southerners were more likely than whites to move North, as reflected in the marginally statistically significant coefficient on *North* × *Race*. Ignoring for now that possibility that movement North entailed some unobserved disamenities (e.g. weather), the combined coefficients on *North* and *North*×*Race* (multiplying odds ratios to calculate the full effect) do not suggest that blacks were especially attracted to the North, beyond what would be explained by variation in economic characteristics. On the other hand, if one believes that the *North* coefficient (for whites) captures disamenities such as northern weather, then the *North*×*Race* coefficient might capture the race-specific amenities of northern residence (in a difference-in-difference sense). We plan to do more work to distinguish between these interpretations.

The results strongly indicate that black and white southerners pursued opportunities in states with relatively high employment growth, as measured with the Bartik-style variable defined above. Black workers appear to have been more strongly responsive to variation in demand growth than whites. Conditional on other state characteristics, whites avoided states with high pre-existing levels of manufacturing or agricultural employment, and they were surprisingly drawn to states with relatively low wage rates. This might reflect an affinity for the South (correlated with low wages) that is not fully captured by distance, the North dummy variable, and earlier migration flows. Relative to whites, we find that blacks differentially selected states with high pre-WWI employment shares in agriculture (the traditional center of black employment) and manufacturing (where new opportunities opened in the industrial North). Blacks were also differentially attracted to states with high levels of wages. But in each case (agriculture, manufacturing, and wages), and calculating roughly, the combination of main effect and race-specific effect yields an odds ratio for blacks that is below or near 1, implying no strong correlation between those characteristics and black migration.

Differences by Skill Level

Under certain conditions, the skill level of potential migrants may hold predictive power over their eventual inter-regional migration decision. [29] If returns to skill are higher in the North than in the South, high-skilled individuals would have migrated inter-regionally. Conversely, higher returns to skill in the South, where skill was relatively scarce, may have encouraged negative selection into the migrant stream. Although Collins and Wanamaker (2013) find little evidence of selection (of either sign) into the black migrant stream, we evaluate whether black-white differences in underlying skill can explain the differences in migration choices across states. For instance, perhaps blacks were

[29] See Roy (1951) and Borjas (1987).

more affected by distance or more attracted to states with high levels of manufacturing and agricultural because of differing skill levels.

If differences in the distribution of educational attainment across black and white males are driving the racial differences in responsiveness to destination characteristics, then the inclusion of skill interactions should drive the odds ratios on race interactions closer to one. In columns 3 and 4 of Table 4, we repeat the specifications from columns 1 and 2, but we limit the sample to individuals for whom a skill proxy is available (ages 11 and up). The sample size is reduced substantially, but the size and direction of coefficients are largely unchanged relative to the full sample. Then, for columns 5 and 6, we add an interaction of each state-level covariate with a skill proxy to the specifications (with and with state-alternative fixed effects).

Although the coefficients on race interactions attenuate somewhat after the addition of these skill measures, the reductions are usually small. We do note that blacks' stronger selection into northern states relative to whites (the coefficient on *North*×*Race*) is in part accounted for by differences in skill levels. The coefficient on *North*×*Skill* indicates that higher skilled southerners were less likely to choose inter-regional migration, consistent with a Roy model in which the return to skill was higher in the South than the North, perhaps due to the relative paucity of skill in the South. Blacks' differential selection into states with high shares of agriculture and manufacturing employment are largely unchanged after controlling for skill. There is a slight negative correlation between skill and the choice of states with high manufacturing employment, but the odds ratios on mfgpct and mfgpct×RACE are barely affected. Agriculture shows a similar pattern.

In addition, the coefficients on the skill interactions themselves indicate that the higher skilled were more likely to pursue labor demand growth, but less likely to migrate to states with higher average wages and more employment in manufacturing and agriculture. Networks were less important for higher-skilled individuals (column 5), although this result is not significant in a specification with state fixed effects (column 6). We explore the differences by skill further in the next section.

Differences in Choice Probabilities by Initial Condition

Although the conditional logit provides a simple framework for evaluating the importance of alternative characteristics in driving location choice, interacting these characteristics with individual-specific characteristics becomes computationally cumbersome very quickly. Yet there may be important differences in migration choices that depend on other aspects of individuals other than race and skill. In Table 5, we present coefficients from the baseline model (Table 4, Column 1) separately

for different sub-populations of the original sample. We can then examine the effects of each state-specific covariate conditional on individual-specific characteristics.

First, we evaluate the model separately for individuals living on and off a farm in 1910 (Columns 1 and 2 of Table 5). There are a few remarkable differences across the columns. Distance is a stronger deterrent to farm residents than non-farm residents, and networks are more important for farm residents as well. For black migrants, networks were a more important determinant of location choice for farm residents. We also find that black farm residents are more responsive to average wages, but surprisingly less likely to migrate to states with a heavy agriculture presence. Also, notably, the North held greater allure for white farm residents, but not black. And the draw of home was far greater for black farm residents relative to their non-farm peers.

Columns 3 and 4 repeat this exercise for urban and non-urban residents in 1910. Again, there are remarkable differences across columns. White urban residents were less likely to pursue employment growth, and black urban residents were less responsive to manufacturing and agricultural employment opportunities. The same is true for wages; black urban residents were less responsive to wave differentials than their non-urban peers. And although we find that urban white residents were (conditionally) less likely to move outside of the South, urban black residents were substantially more likely to do so.

Finally, Columns 5 and 6 bifurcate the 1910 sample by skill. Column 5 is based on only above-median skill individuals while Column 6 is based on below-median workers. We find that skilled white Southerners were more responsive to employment growth than their unskilled counterparts, while skilled black migrants were less likely to pursue states with agriculture and manufacturing opportunities. And although skilled whites pursued higher wages more so than their low-skilled peers, the opposite is true for black skilled and unskilled workers. Conditional on all the attributes of northern and southern states captured in the model, unskilled blacks were less likely to migrate North and more likely to remain in their home state.

5. Counterfactuals

The conditional logit estimates allow us to generate predicted distributions of black and white southerners across locations under counterfactual characteristics of individuals and states. These counterfactuals serve two purposes: they help us to appreciate the magnitudes and economic significance of the conditional logit coefficients, and they help us address some historical scenarios of particular interest.

First, we consider how the predicted migration patterns of blacks would have differed if their average education level more closely resembled that of whites, and vice versa. African American slaves had very low levels of literary—it was generally illegal for them to learn to read and for others to teach them. Despite significant gains in literacy and educational attainment among subsequent, post-emancipation generations, the deficit in human capital was one of the gravest legacies of slavery. Segregated schooling and widening gaps in quality between white and blacks schools perpetuated the deficit (Margo 1990). Other work has argued that geographic mobility was an important route to raising one's income in this period (Collins and Wanamaker 2014), and so to the extent that low levels of human capital attenuated migration, they also would have attenuated opportunities for income growth.

The net effect of an increase in "skill" on migration patterns is not obvious from the estimates in Tables 4 and 5 because skill is interacted with each covariate. Column 1 of Table 6, Panel A, reports black intra- and inter-regional migration rates in the raw data, which is our baseline for comparison. In Column 2, we increase black skill levels in the data by one standard deviation, decrease black skill levels by the same, and re-compute predicted migration patterns. In practice, increasing blacks' skill level by 1 unit effectively raises the black average to match the white average in our dataset, and decreasing white skill level by 1 unit lowers the white average to that of blacks. But despite this large increment to average black skill, the net impact on migration patterns is small. Higher skill levels would have resulted in slightly higher rates of intra-regional migration for blacks, but no change in the flow of black migrants out of the South. For whites, lower skill would have resulted in almost no change in migration propensities.[31]

Second, we note that for the purpose of inter-regional migration, blacks were at a locational disadvantage relative to whites. Blacks were disproportionately concentrated in the Deep South and, therefore, had to travel greater distances to reach northern locations. We re-weight the black sample by the ratio of white to black residents in each observation's 1910 state of origin. (For example, blacks who lived in Kentucky, where the white to black ratio was relatively high, receive greater weight than blacks in Alabama where the reverse was true.) In Column 3 of Table 6, Panel A, the result is a 4 percentage-point increase in the probability of inter-regional migration, largely at the expense of intra-regional migration. Blacks, then, would have been even more likely to migrate North had they been located in states closer to the regional border in 1910, which is consistent with the strong deterrent effects of distance seen in Table 4. This counterfactual inter-regional black

[31] We find little evidence of skill effects in the multinomial logit analysis, but in this case skill is allowed to interact with characteristics of the location choices.

migration rate is 7 points higher than the actual rate observed for whites (17 percent), indicating that geographically similar blacks and whites made far different choices regarding state choice in 1930. Reweighting the white observations in a similar way (to reflect black location) in Column 3 of Panel B, we find that whites would have substituted intra-South migration for inter-regional migration had their initial conditions matched blacks'.

Finally, we estimate migration flows under a counterfactual scenario of full economic convergence across the U.S. Among other effects, the Great Migration is frequently credited with playing a role in national labor market integration. The movement of workers away from the low-wage South to the high-wage North served to lower labor costs in the North, raise them in the South, and, eventually, even the distribution of industry across locations. But suppose the South was not economically differentiated from non-Southern regions from the beginning? Would migration patterns have been different? We assigned Southern states in our sample the average value of variables capturing economic conditions for non-Southern states: labor growth, average wages, percent in manufacturing, and percent in agriculture. The resulting predictions for inter-regional migration are substantially reduced for both blacks and whites. (Column 5 in both panels of Table 6.) Indeed, an integrated U.S. South would not have experienced nearly as heavy a population outflow before 1930.

6. Conclusions

We build a new dataset of more than 27,000 men who are linked from the 1910 to the 1930 census to study migration patterns during the first decades of the Great Migration. The dataset is novel in that it observes the same men before and after the start of the Great Migration, and the scope of our analysis is novel is that we include both white and black southerners and we treat states as alternative choices rather than focusing solely on South-to-North inter-regional migration. Summary statistics and simple analyses with linear probability, OLS, and mulitinomial logit models all indicate that there were differences in black and white migration patterns that cannot be attributed to differences in ex ante characteristics. A conditional logit model that features alternative-specific variables and that also allows for consideration of some individual-specific variables is a natural framework for studying migration choices and comparing the black and white patterns.

First, the human capital deficit in the South relative to the rest of the U.S., especially for African Americans, had many negative economic implications. But it appears that significantly slowing long-distance migration was not one of them. Our estimates suggest that sizable changes in the level of human capital would not have sizable effects on the predicted distribution of migrants.

Second, we confirm that previous state-to-state migration patterns (by race) were strong predictors of migration in the 1910 to 1930 period, consistent with a large literature on migration that emphasizes the importance of network ties in facilitating and perpetuating migration. Somewhat surprisingly, we did not find blacks to be more responsive than whites to the pre-existing patterns, which one might have expected for a poorer and less well-educated population. This may reflect the degree to which the Great Migration really was a break from the past for African Americans, as well as a response to the call from industrial employers who were willing to hire blacks for the first time. Third, when controlling for other state-specific economic characteristics, we find that blacks were more likely than whites to choose northern destinations, which may be evidence that they valued amenities associated with more secure civil rights. Finally, we find that distance was a significant deterrent to migration, especially for black workers. In one sense this is unsurprising, but its implications for the literature on the Great Migration are important and, to our knowledge, underappreciated. It implies that improvements in transportation technology promoted inter-regional migration from the South in this period. World War I and the end of mass European migration tend to get top billing when describing the impetus for the Great Migration, but better transport networks may have been an important prerequisite. We are working to fully quantify the implications of transportation costs for the migration patterns we observe.

Much more work needs to be done to confirm and refine the preliminary findings reported here.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, 5: 1832-1856.

Akers, Elmer. 1936. *Southern Whites in Detroit*. Manuscript. Ann Arbor, MI.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo: W.E. Upjohn Institute.

Berry, Chad. 2000. *Southern Migrants, Northern Exiles*. Urbana: University of Illinois Press.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2011. "The Impact of the Great Migration on Mortality of African-Americans: Evidence from the Deep South." Working paper.

Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants," *American Economic Review* 77: 531-553.

Boustan, Leah Platt. 2009. "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940-1970." *Journal of Economic History* 69,3: 755-782.

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

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

Chay, Kenneth, and Kaivan Munshi. 2012. "Black Networks after Emancipation: Evidence from Recronstruction and the Great Migration." Unpublished working paper.

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 Robert A. Margo. 2006. "Historical Perspectives on Racial Differences in Schooling in the United States." In *Handbook of the Economics of Education: Volume 1*, edited by E. Hanushek and F. Welch. New York: North-Holland: 107-154.

Collins, William J. and Marianne H. Wanamaker. 2014. "Selection and Income Gains in the Great Migration of African Americans." Forthcoming *AEJ: Applied*.

Davies, Paul S., Michael J. Greenwood, and Haizheng Li. 2001. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science* 41, 2: 337-360.

Easterlin, Richard A. 1960. "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950." In *Trends in the American Economy in the Nineteenth Century*. National Bureau of Economic Research, Conference on Research in Income and Wealth. Princeton, NJ: Princeton University Press: 73-140.

Eldridge, Hope T., and Dorothy Swaine Thomas. 1964. *Population Redistribution and Economic Growth, United States, 1870-1950, Vol. 3: Demographic Analyses and Interrelations*. Philadelphia: American Philosophical Society.

Foote, Christopher L., Warren C. Whatley, and Gavin Wright. 2003. "Arbitraging a Discriminatory Labor Market: Black Workers at the Ford Motor Company, 1918-1947." *Journal of Labor Economics* 21, 3: 493-532.

Gill, Flora. 1979. *Economics and the Black Exodus: An Analysis of Negro Emigration from the Southern United States, 1910-70*. New York: Garland Publishing.

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.

Haines, Michael R. and Inter-university Consortium for Political and Social Research. 2010. *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002* [computer file]. ICPSR02896-v3. Ann Arbor, MI: ICPSR [distributor].

Hatton, Timothy J. and Jeffrey G. Williamson. 1998. *The Age of Mass Migration: Causes and Economic Impact*. New York and Oxford: Oxford University Press.

Higgs, Robert. 1976. "The Boll Weevil, the Cotton Economy, and Black Migration 1910-1930." *Agricultural History* 50, 2: 335-50.

Higgs, Robert. 1982. "Accumulation of Property by Southern Blacks Before World War I." *American Economic Review* 72, 4: 725-737.

Hornbeck, Richard and Suresh Naidu. 2012. "When the Levee Breaks: Black Migration and Economic Development in the South." NBER Working Paper 18296.

Killian, Lewis M. 1953. "The Adjustment of Southern White Migrants to Northern Urban Norms." *Social Forces* 32: 66-69.

Killian, Lewis M. 1949. *Southern White Laborers in Chicago's West Side*. University of Chicago Ph.D. Dissertation.

Kousser, J. Morgan. 1974. *The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-party South, 1880-1910*. New Haven: Yale University Press.

Lebergott, Stanley. 1964. *Manpower in Economic Growth: the American Record Since 1800*. New York, NY: McGraw-Hill.

Lewis, Edward E. 1931. *The Mobility of the Negro: A Study in the American Labor Supply*. New York: Columbia University Press.

Long, Jason and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." *American Economic Review*. xxxx-xxxx.

Maloney, Thomas N. 2001. "Migration and Economic Opportunity in the 1910s: New Evidence on African-American Occupational Mobility in the North." *Explorations in Economic History* 38, 1: 147-165.

Margo, Robert A. 1990. *Race and Schooling in the South 1880-1950: An Economic History*. Chicago: University of Chicago Press.

Margo, Robert A. 2004. "The North-South Wage Gap Before and After the Civil War." In D. Eltis, F. Lewis, and K. Sokoloff (eds.), *Slavery in the Development of the Americas*, pp. 324-351. New York: Cambridge University Press.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press: 105-142.

McKenzie, David and Hillel Rapoport. 2007. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico." *Journal of Development Economics* 84, 1: 1-24.

Minnesota Population Center. 2011. National Historical Geographic Information System: Version 2.0 Minneapolis, MN: University of Minnesota.

Mitchener, Kris James and Ian W. McLean. 1999. "U.S. Regional Growth and Convergence, 1880-1980." *Journal of Economic History* 59: 1016-1042.

Moretti, Enrico. 2011. "Local Labor Markets." In *Handbook of Labor Economics*, Vol. 4B, edited by Orley Ashenfelter and David Card, 1237-1313. Amsterdam: North Holland.

Myrdal, Gunnar. 1944. *An American Dilemma: The Negro Problem and Modern Democracy*. New York: Harper & Brothers Publishers.

Naidu, Suresh. 2010. "Recruitment Restrictions and Labor Markets: Evidence from the Postbellum U.S. South." *Journal of Labor Economics* 28, 2: 413-445.

Ransom, Roger L. and Richard Sutch. 2001 [1977]. *One Kind of Freedom: The Economic Consequences of Emancipation*, Second Edition. New York: Cambridge University Press.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257-1278.

Rosenbloom, Joshua L. 2002. *Looking for Work, Searching for Workers: American Labor Markets during Industrialization*. New York: Cambridge University Press.

Roy, A.D. 1951. "Thoughts on the Distribution of Earnings." *Oxford Economics Papers* 3, 2: 135-146.

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]. Minneapolis: University of Minnesota.

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

Tolnay, Stewart E. and E.M. Beck. 1990. "Black Flight: Lethal Violence and the Great Migration, 1900 to 1930." *Social Science History* 14: 347-370.

U.S. Department of Labor. 1919. *Negro Migration in 1916-17*. Washington D.C. GPO.

Vickery, William E. 1977. *The Economics of Negro Migration, 1900-1960*. New York: Arno Press.

Vigdor, Jacob L. 2002. "The Pursuit of Opportunity: Explaining Selective Black Migration." *Journal of Urban Economics* 51, 3: 391-417.

Whatley, Warren. 1990. "Getting a Foot in the Door: 'Learning,' State Dependence, and the Racial Integration of Firms." *Journal of Economic History* 50, 1: 43-66.

Woodward, C. Vann. 1974. *The Strange Career of Jim Crow*, Third Revised Edition. New York: Oxford University Press.

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

Wright, Gavin. 1987. "Postbellum Southern Labor Markets." In *Quantity and Quiddity: Essays in US. Economic History*, edited by Peter Kilby, 98-134. Middletown, CT: Wesleyan University Press.

Table 1: Comparison of Linked and Full Sample Characteristics, Southern Males 1910

	White Males		Black Males	
	Matched Sample	Full IPUMS Sample	Matched Sample	Full IPUMS Sample
Panel A: Distribution of state of residence				
Alabama	6.5	6.5	9.8	10.3
Arkansas	6.4	6.1	5.0	4.9
Florida	2.1	2.3	4.0	3.8
Georgia	7.5	7.5	13.5	13.9
Kentucky	10.8	10.4	3.2	3.0
Louisiana	4.7	5.2	8.2	8.7
Mississippi	4.2	4.1	12.9	12.2
North Carolina	7.9	7.6	8.9	8.0
Oklahoma	8.0	8.1	1.5	1.7
South Carolina	3.5	3.6	11.0	10.2
Tennessee	8.6	8.6	5.3	5.5
Texas	16.2	16.8	9.3	9.0
Virginia	6.9	7.0	6.8	7.9
West Virginia	6.9	6.2	0.7	1.0
Panel B: Personal characteristics				
Attending school (age 0-20)	51.8	48.0	36.8	36.3
In owner-occupied housing	52.4	48.4	22.4	23.9
Literate (age 10-20)	91.8	90.1	62.4	62.6
Literate (age 10-40)	92.6	91.2	65.7	65.6
Father is farmer (age 0-20)	60.9	59.0	63.4	63.3
1910 city population				
Not in city	70.0	69.4	74.5	73.2
City pop. <=25,000	20.4	20.8	16.3	17.3
City pop. >25,000	9.6	9.8	9.2	9.5
Panel C: Age Distribution				
Min Age	0	0	0	0
Max Age	40	40	40	40
Median Age	16	15	16	15
Mean Age	17.1	16.8	17.0	16.7
Std. Dev.	11.1	11.5	11.2	11.3

Notes and sources: The linked sample is created by taking the 1910 IPUMS sample of white men, age 0-40, who reside in the South and searching for these men in the 1930 census manuscripts. The text contains more details on sample construction. The IPUMS data are from Ruggles et al. (2010).

Table 2: Migration Summary Statistics, by Race: 1910-1930

	White Males	Black Males
Panel A: Migration Categories		
No state-to-state migration	0.65	0.61
Within-South migrants	0.18	0.19
Inter-regional migrants	0.17	0.20
South-to-Northeast migrants	*0.03*	*0.08*
South-to-Midwest migrants	*0.09*	*0.11*
South-to-West migrants	*0.05*	*0.01*
Panel B: Geographic Characteristics of Migration		
Distance Travelled		
Full Sample	219.8	210.7
Within-South migrants	321.8	266.0
Inter-regional migrants	695.7	576.7
Change in Latitude – Home to Destination		
Full Sample	0.69	1.38
Within-South migrants	-0.09	0.47
Inter-regional migrants	3.96	6.20
Change in Longitude – Home to Destination		
Full Sample	-0.90	0.30
Within-South migrants	-0.51	0.20
Inter-regional migrants	-4.88	1.42
Panel C: Characteristics of 1930 Destinations in 1910 (All Migrants)		
Percent Urban Residents	35.6	39.3
Percent Labor Force in Agriculture	41.7	38.8
Percent Labor Force in Mfg	22.0	24.9
Percent Black	15.0	17.9
Average Wages Paid*	974.9	976.5

Notes: Within-South migrants are those moving outside of the home state but within the South. Within-South migration excludes 1930 residents of DC, MD and DE. South-to-Northeast includes 1930 residents of the Northeast census regions and also DC, MD and DE. South-to-Midwest (West) includes 1930 residents of the Midwest (West) region. "All migrants" are all state-to-state migrants. Latitude differences are positive for south-to-north migration. Longitude differences are positive for west-to-east migration.

*Taken as the average of 1900 and 1920 state estimates from Mitchener and Mclean (1999).

Table 3: Black-White Differences in Migration

		Unadjusted Black-White Difference	Adjusted Black-White Difference
Panel A: Full Sample			
Inter-Regional Migration	Under 18	0.07	0.09
		(0.02)	(0.06)
	18 and Over	0.01	0.04
		(0.01)	(0.01)
Within-South Migration	Under 18	0.02	0.03
		(0.02)	(0.06)
	18 and Over	0.03	-0.00
		(0.01)	(0.01)
Distance	Under 18	26.4	-0.7
		(13.0)	(52.5)
	18 and Over	-12.7	-11.3
		(7.1)	(9.7)
Change in Latitude	Under 18	1.16	1.22
		(0.14)	(0.41)
	18 and Over	0.52	0.31
		(0.08)	(0.09)
Change in Longitude	Under 18	1.33	1.52
		(0.21)	(0.95)
	18 and Over	0.97	0.90
		(0.12)	(0.17)
Panel B: Inter-Regional Migrants			
Distance	Under 18	-66.5	-280.5
		(34.7)	(368.0)
	18 and Over	-153.9	-272.8
		(24.3)	(53.8)
South-to-NE Migration	Under 18	0.21	-0.39
		(0.04)	(0.24)
	18 and Over	0.17	0.08
		(0.03)	(0.04)
South-to-MW Migration	Under 18	0.02	0.51
		(0.04)	(0.31)
	18 and Over	0.07	0.13
		(0.03)	(0.05)
South-to-West Migration	Under 18	-0.24	-0.12
		(0.03)	(0.25)
	18 and Over	-0.24	-0.21
		(0.02)	(0.04)
Change in Latitude	Under 18	2.80	0.02
		(0.28)	(1.79)

	18 and Over	2.07	0.38
		(0.20)	(0.35)
Change in Longitude	Under 18	6.54	4.65
		(0.86)	(8.54)
	18 and Over	6.04	6.60
		(0.60)	(1.47)

Notes: For columns 1 and 2, each entry corresponds to the coefficient on tau (coefficient on race where black=1) from a separate OLS regression. In column 1, there are no controls other than race, but the sample is limited to match that available for Column 2. Column 2 contains controls as specified in the text. Standard errors are clustered by 1910 household. Sample sizes for under 18/18 and over are 4440/10614. Regional Migrants are 802/1715.
***Represents significance at the 1% level, ** at the 5% level and * at the 10% level.

TABLE 4: Estimates from the Conditional Logit State Choice Model

	(1)	(2)	(3)	(4)	(5)	(6)
bartik_1930	5.313***		5.879***		5.875***	
	(15.76)		(13.99)		(14.00)	
bartik_1930xRACE	1.122***	1.157***	1.105***	1.146***	1.124***	1.176***
	(3.73)	(4.51)	(2.58)	(3.29)	(2.91)	(3.70)
bartik_1930xskill					1.020	1.029
					(1.13)	(1.53)
mfg_pct	0.956***		0.952***		0.952***	
	(-9.52)		(-8.84)		(-8.86)	
mfg_pctxRACE	1.052***	1.048***	1.051***	1.048***	1.048***	1.046***
	(6.33)	(5.75)	(5.16)	(4.43)	(4.53)	(3.91)
mfg_pctxskill					0.996	0.998
					(-1.28)	(-0.61)
ag_pct	0.971***		0.968***		0.969***	
	(-9.85)		(-8.95)		(-8.91)	
ag_pctxRACE	1.020***	1.020***	1.025***	1.024***	1.023***	1.022***
	(3.60)	(3.81)	(3.58)	(3.54)	(3.14)	(3.04)
ag_pctxskill					0.998	0.998
					(-0.93)	(-1.02)
ln(wage)	0.812***	0.707***	0.743***	0.657***	0.746***	0.663***
	(-4.78)	(-7.68)	(-5.43)	(-7.59)	(-5.32)	(-7.42)
ln(wage)xRACE	1.259***	1.388***	1.344***	1.477***	1.314***	1.400***
	(3.33)	(4.77)	(3.39)	(4.43)	(3.07)	(3.69)
ln(wage)xskill					0.973	0.939*
					(-0.92)	(-1.94)
distance	0.00192***	0.000890***	0.00246***	0.00107***	0.00230***	0.000975***
	(-24.80)	(-27.23)	(-19.93)	(-22.52)	(-19.91)	(-22.39)
distancexRACE	9.986***	3.275**	7.723***	2.983*	9.976***	4.207**
	(3.82)	(2.13)	(2.96)	(1.68)	(3.15)	(2.10)
distancexskill					1.313	1.450
					(1.05)	(1.51)
distance^2	6.7e+06***	2.5e+08***	5.0e+06***	2.2e+08***	5.9e+06***	2.8e+08***
	(15.61)	(19.59)	(12.90)	(16.63)	(12.83)	(16.52)
distance^{2}xRACE	9.5e-06***	0.00051***	1.6e-05***	0.00048***	7.9e-06***	0.00018***
	(-4.68)	(-3.40)	(-3.88)	(-2.91)	(-3.89)	(-3.10)
distance^{2}xskill					0.474	0.347
					(-0.73)	(-1.10)
network	1.298***	1.237***	1.304***	1.245***	1.306***	1.244***
	(31.78)	(18.49)	(28.16)	(15.56)	(28.34)	(15.45)
networkxRACE	0.948**	0.907***	0.923***	0.887***	0.913***	0.888***
	(-2.02)	(-3.89)	(-2.76)	(-4.23)	(-3.05)	(-4.19)
networkxskill					0.986**	1.001
					(-2.13)	(0.25)
NORTH	0.792***		0.769***		0.782***	
	(-4.28)		(-3.73)		(-3.47)	
NORTHxRACE	1.179	1.219*	1.171	1.168	1.089	1.112
	(1.48)	(1.78)	(1.13)	(1.12)	(0.58)	(0.72)
NORTHxskill					0.922	0.944
					(-1.64)	(-1.16)
HOME	1.034	0.477*	1.034	0.511	1.079	0.481
	(0.09)	(-1.87)	(0.07)	(-1.39)	(0.17)	(-1.53)
homexRACE	289.0***	1002.0***	163.9***	576.2***	115.8***	620.0***
	(6.76)	(8.89)	(5.56)	(7.20)	(4.74)	(6.84)
homexskill					0.629	1.009
					(-1.48)	(0.03)
State FE?	N	Y	N	Y	N	Y
N	1287792	1287792	811968	811968	811968	811968

Notes: Odds ratios are reported; standard errors in parentheses. All specifications include controls for urban population percentage, a cubic in total population, and 3rd and 4th-order distance terms (with RACE and skill interactions). Standard errors are clustered at the county of origin in 1910. Columns 1 and 2 contain the full sample. Columns 3 and 4 are limited to the sample where skill measures are available (11 and older), but do not include skill interactions for covariates. Columns 5 and 6 utilize the same sample as Columns 3 and 4 with skill interactions for covariates. Samples are weighted to reflect differences in match rates across black and white subsamples.

** indicates statistical significance at 99% confidence (with respect to zero), ** at 95%, and * at 90%.

TABLE 5: Estimates from the Conditional Logit State Choice Model - By Initial Conditions

	(1) FARM	(2) NON-FARM	(3) URBAN	(4) NON-URBAN	(5) SKILL	(6) NON-SKILL
bartik_1930	5.521***	5.234***	3.840***	6.766***	7.157***	4.982***
	(12.78)	(10.84)	(7.73)	(14.94)	(11.51)	(9.62)
bartik_1930xRACE	1.122***	1.119***	1.082*	1.049	1.088	1.115**
	(2.63)	(2.81)	(1.89)	(0.65)	(1.13)	(2.11)
mfg_pct	0.955***	0.955***	0.954***	0.954***	0.950***	0.950***
	(-7.95)	(-7.24)	(-5.93)	(-8.82)	(-7.36)	(-7.31)
mfg_pctxRACE	1.048***	1.056***	1.025***	1.118***	1.034*	1.060***
	(4.07)	(4.98)	(2.67)	(8.41)	(1.67)	(4.94)
ag_pct	0.970***	0.968***	0.956***	0.972***	0.972***	0.962***
	(-8.03)	(-8.76)	(-8.52)	(-8.77)	(-6.90)	(-7.77)
ag_pctxRACE	1.011	1.028***	0.989	1.065***	1.015	1.031***
	(1.46)	(3.44)	(-1.61)	(7.95)	(1.02)	(3.70)
ln(wage)	0.767***	0.851***	0.862*	0.799***	0.851**	0.597***
	(-4.96)	(-2.74)	(-1.72)	(-4.96)	(-2.52)	(-6.92)
ln(wage)xRACE	1.325***	1.226**	1.153	1.462***	1.077	1.675***
	(3.11)	(2.07)	(1.34)	(4.18)	(0.52)	(4.83)
distance	0.00144***	0.00252***	0.00245***	0.00173***	0.00225***	0.00288***
	(-18.26)	(-18.07)	(-13.07)	(-21.89)	(-15.88)	(-12.33)
distancexRACE	9.861***	7.560**	12.27***	6.922**	2.599	8.266**
	(2.60)	(2.45)	(2.81)	(2.18)	(0.70)	(2.54)
distance^2	1.8e+07***	2.7e+06***	3.6e+06***	8.6e+06***	7.2e+06***	2.5e+06***
	(11.65)	(11.11)	(8.35)	(13.75)	(10.24)	(7.71)
distance^{2}xRACE	3.5e-05***	8.5e-06***	5.8e-07***	.00011**	.00023	1.9e-05***
	(-2.75)	(-3.51)	(-4.17)	(-2.16)	(-1.49)	(-3.10)
network	1.320***	1.274***	1.257***	1.306***	1.281***	1.339***
	(27.54)	(21.73)	(12.47)	(30.86)	(22.76)	(23.62)
networkxRACE	1.038	0.900***	0.968	0.990	0.847***	0.923***
	(1.11)	(-3.27)	(-0.86)	(-0.33)	(-3.07)	(-2.62)
NORTH	0.869**	0.690***	0.660***	0.816***	0.719***	0.851
	(-2.12)	(-4.62)	(-3.47)	(-3.53)	(-4.03)	(-1.60)
NORTHxRACE	1.074	1.303	2.270***	0.254***	1.436	1.038
	(0.49)	(1.60)	(4.82)	(-6.27)	(1.21)	(0.23)
HOME	0.586	0.966	0.606	0.852	1.011	0.843
	(-1.15)	(-0.07)	(-0.52)	(-0.40)	(0.02)	(-0.28)
homexRACE	2033.4***	181.2***	4864.7***	30.75***	16.27	401.1***
	(6.98)	(4.94)	(7.35)	(3.27)	(1.59)	(5.71)
State FEs?	N	N	N	N	N	N
N	738720	549072	289056	998736	399264	412704

Notes: Odds ratios are reported; standard errors in parentheses. All specifications include controls for urban population percentage, a cubic in total population, and 3rd and 4th-order distance terms (with RACE interactions). Standard errors are clustered at the county of origin in 1910. Columns 1 and 2 represent those individuals residing and not residing on a farm in 1910, respectively. Columns 3 and 4 represent those living in urban locations and non-urban locations, respectively. Columns 5 and 6 represent those with above- and below-median skill levels relative to their age group in 1910, respectively.
** indicates statistical significance at 99% confidence (with respect to zero), ** at 95%, and * at 90%.

Table 6: Counterfactual Estimates of Migration Probabilities, By Region and Race

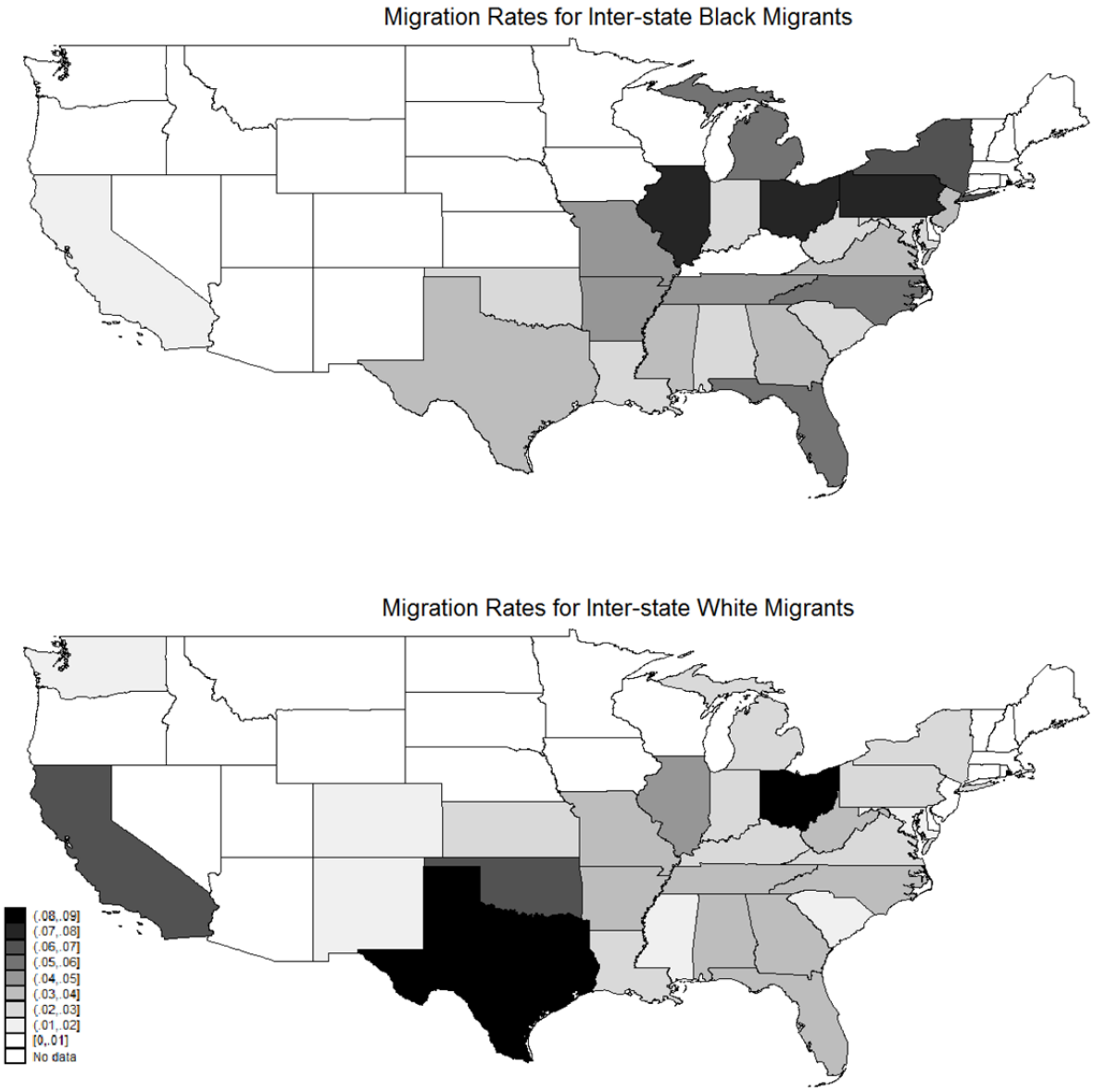
Panel A: BLACK MALES

		Counterfactuals		
	(1) Baseline	(2) White Avg Skill	(3) White Avg Location	(4) Regional Integration
No state-to-state migration	0.61	0.60	0.62	0.69
Within-South migrants	0.19	0.21	0.14	0.22
Inter-regional migrants	0.20	0.19	0.24	0.09
South-to-Northeast migrants	*0.08*	*0.08*	*0.08*	*0.04*
South-to-Midwest migrants	*0.11*	*0.11*	*0.15*	*0.05*
South-to-West migrants	*0.01*	*0.01*	*0.01*	*0.00*

Panel B: WHITE MALES

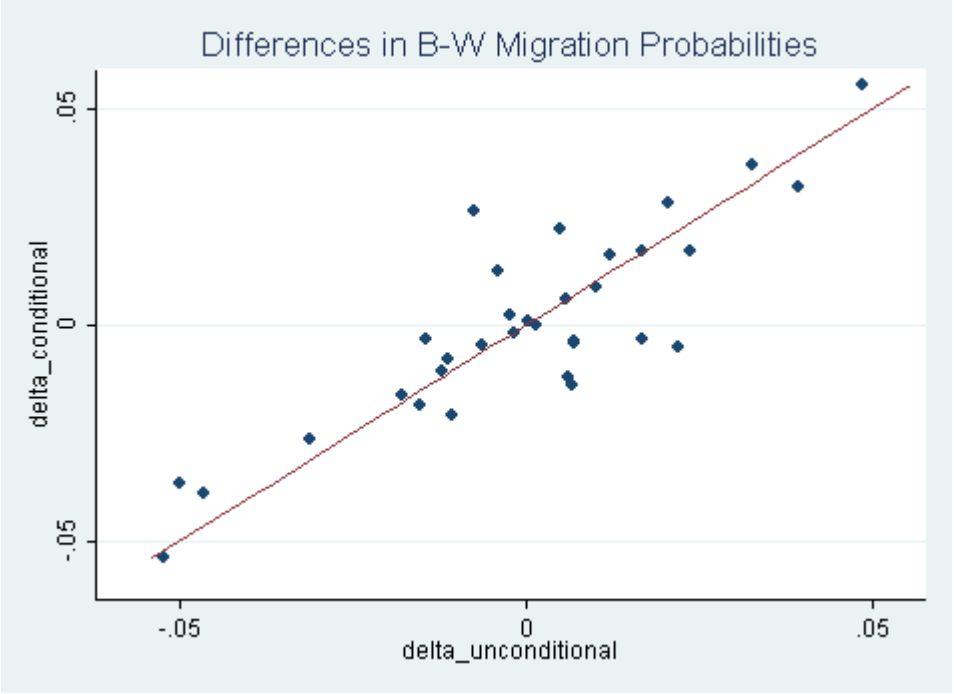
		Counterfactuals		
	(1) Baseline	(2) Black Avg Skill	(3) Black Avg Location	(4) Regional Integration
No state-to-state migration	0.65	0.67	0.65	0.68
Within-South migrants	0.18	0.17	0.21	0.22
Inter-regional migrants	0.17	0.16	0.14	0.11
South-to-Northeast migrants	*0.03*	*0.04*	*0.04*	*0.03*
South-to-Midwest migrants	*0.09*	*0.09*	*0.07*	*0.06*
South-to-West migrants	*0.05*	*0.03*	*0.03*	*0.02*

Figure 1: Migration Rates by State and Race, Conditional on Inter-State Migration



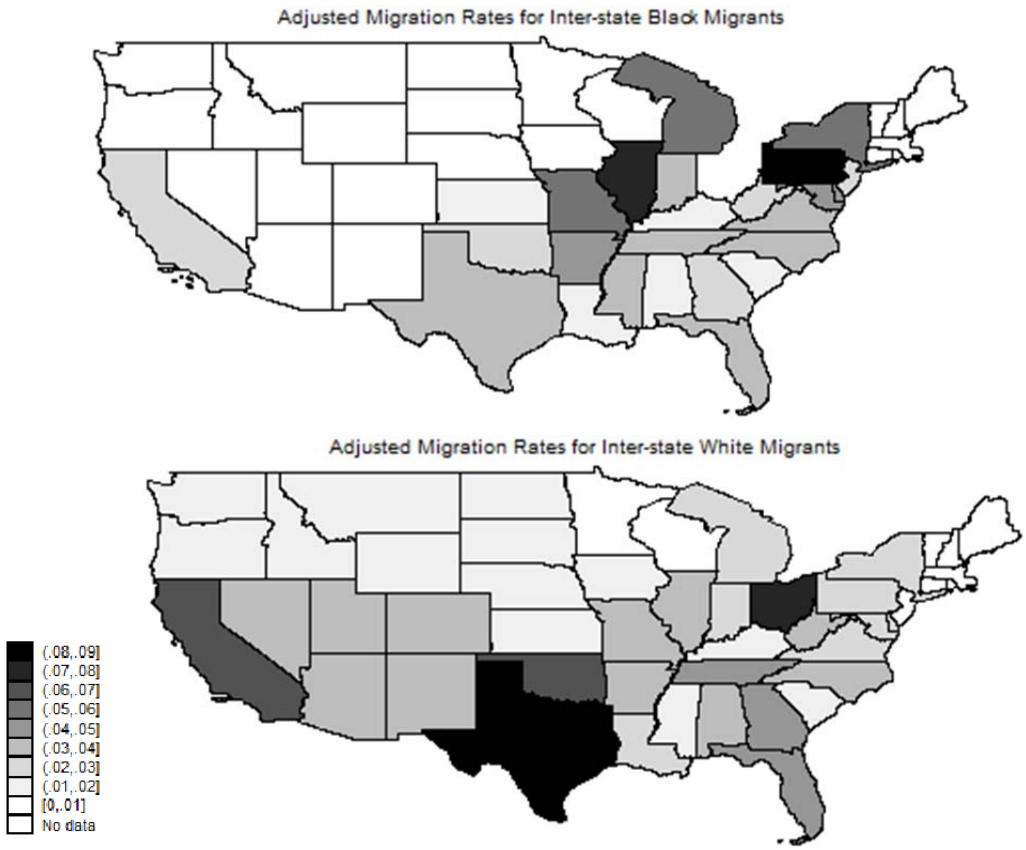
Notes: Migration rates are unadjusted, conditional on inter-state migration.
Source: See text.

Figure 2



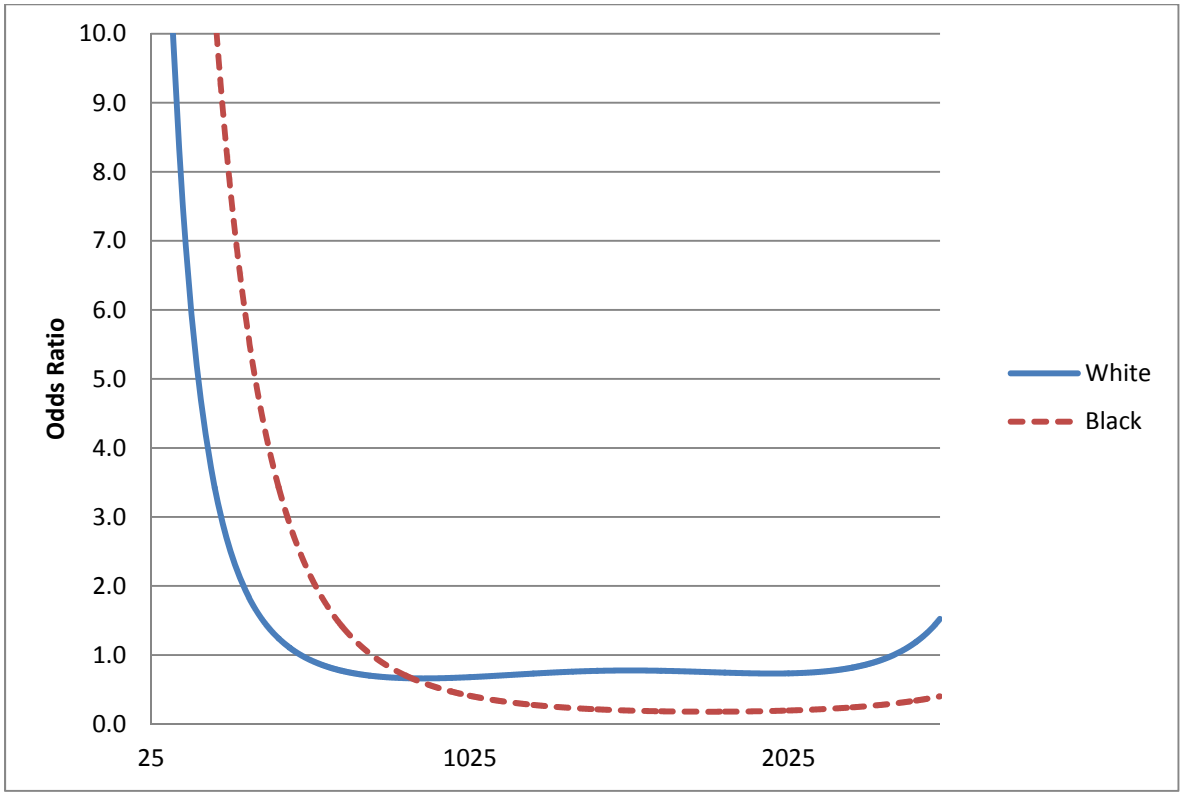
Notes: Each point represents a state. If background characteristics accounted for black-white differences in choice of destination, then "tau unconditional" would be near zero for each state.

Figure 3: Conditional Migration Rates by State



Notes: Migration rates are adjusted based on multinomial logit model, conditional on inter-state migration. See Appendix A2 for more information.
Source: See text.

Figure 4: Effect of Distance (in miles) on Migration Odds, Black versus White



10. Appendix

A1. Inter-regional Migration Rates by Home State

Table A1: Rate of inter-regional migration by region, state of origin and race: 1910-1930

	Black Sons <18), Fathers Present	Black Males, 18-40	White Sons (<18), Fathers Present	White Males, 18-40
South Atlantic	0.26	0.21	0.13	0.09
Florida	0.22	0.20	0.11	0.15
Georgia	0.25	0.23	0.07	0.08
North Carolina	0.21	0.12	0.10	0.06
South Carolina	0.26	0.17	0.06	0.04
Virginia	0.40	0.33	0.26	0.16
East South Central	0.23	0.18	0.20	0.16
Alabama	0.22	0.15	0.10	0.08
Kentucky	0.43	0.32	0.33	0.25
Mississippi	0.19	0.14	0.08	0.06
Tennessee	0.27	0.24	0.19	0.13
West South Central	0.13	0.12	0.18	0.20
Arkansas	0.17	0.21	0.24	0.21
Louisiana	0.10	0.06	0.08	0.12
Oklahoma	0.24	0.36	0.27	0.36
Texas	0.12	0.14	0.13	0.14
N	2,291	2,456	10,677	9,654

Notes and sources: Positive values for *dlatitude* represent northern migration. Positive (negative) values for *dlongitude* represent eastern (western) migration. Authors' calculations from data described in text.

A2. Multinomial Logit Analysis

As discussed in Section 3, we calculate a state-specific measure δ_s, which is the difference in black and white average migration probabilities to state s. $\delta_s^{unconditional}$ is just the racial difference in the average probability of migration to each state as observed in the data. For example, in our data, 2.4 percent of whites migrate to California while 0.7 percent of blacks do the same, making $\delta_{CA}^{unconditional} = 1.7$.

We then use a multinomial logit tool to calculate a conditional difference in black-migration, conditioning on observable characteristics of individuals and households. The control variables include race, state of origin fixed effects, age fixed effects (if over 18), whether the person lived in owner-occupied housing, whether the person (or their father, if under 18) was literate, an industry of employment fixed effect (if over 18), whether attending school (if under 18), whether the individual

was a first-born child (if under 18), and indicators for small (<25,000 residents) and large cities of residence in 1910. The analysis is performed separately for over and under-18 samples.

We see a counterfactual probability of migration to each state for blacks and whites if all other observable characteristics are equalized across races. To do so, we simply allow the entire sample to be counterfactually black or counterfactually white such that each race carries the average observable characteristics of the sample as a whole. (Sequentially, we replace the race variable to be "black" and then "white" for each individual in the sample, predicting the distribution of migration across states for each counterfactual.) Because the covariates are the same in the two counterfactuals, the differences in state migration probabilities calculated when the sample is assumed to be entirely black and when it is assumed to be entirely white reflects the conditional racial difference in migration probabilities ($\delta_s^{conditional}$). Figure 2 in the main text contains a comparison of $\delta_s^{unconditional}$ and $\delta_s^{conditional}$. Figure 3 shows the probability of migration across states for the counterfactual scenario where the entire sample is assumed to be black (top panel) and white (bottom panel).

Comparing Figure 3 to the maps in Figure 1, it is clear that equalizing observable characteristics across races does little to diminish differences in black-white migration patterns.

A3. Variable Definitions

Distance

Distance from each individual to each state is calculated in miles using the center-of-county latitude and longitude coordinates for the county of origin. For destination states other than the state of origin, the distance is taken to the mean latitude and longitude of actual migrants to that state. For the home state, the distance is 0. All distances are expressed in miles using STATA's "vincenty" function, whose documentation assures us that the distance is calculated to an "insane precision".

Skill

For individuals 18 and older and observed in occupations in 1910, we calculate an education score equal to the average highest grade completed of individuals in the same occupation in 1940, when this information was first reported in the U.S. Census. Because occupation and age are highly

correlated, we then calculate a Z-Score for the occupation score within each age from 18 to 40, thereby generating a measure of skill relative to others in the sample of the same age.

For individuals 11 through 17, we use school attendance to infer the probability of being enrolled at age 18. In addition to differences in the rate of school attendance (a characteristic observed in the Census), black and white youths differed in the probability of remaining in school going forward. For example, two 11-year-old individuals in the data who are both attending school, one black and one white, have different expected values for total years of schooling at age 18 because of differences in attrition.

For each individual in the sample, we assume a minimum of four years of school attendance. For an individual of age k who is observed attending school in the census, we use the age-specific rate of school attendance through age 18 to calculate a dropout rate at each age$> k$. The dropout rate can then be used to calculate the expected number of years that individual will attend school, conditional on attendance at age k. For an individual of age k who is observed not attending school in the census, we use a calculated dropout rate at each age$< k$ to infer the expected number of years of schooling the individual attained prior to dropping out.[34]

As a final adjustment, we multiply the estimated number of years for blacks by a factor of 0.752 to represent the ratio of average term lengths for four Southern states in 1910 where this statistic was reported (Alabama, Georgia, South Carolina, and North Carolina). School attendance rates peak at age 11 for both groups, and so we calculate this skill measure only for individuals aged 11 and older.

[34] This completed years of schooling inference is only possible if we make the assumption that dropout decisions are final – once individuals leave school they do not return. The validity of this assumption is debatable, and we use other measures of skill to verify our conclusions.