—•••••—

The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants

WILLIAM J. COLLINS AND MARIANNE H. WANAMAKER

We construct datasets of linked census records to study internal migrants' selection and destination choices during the first decades of the "Great Migration" (1910–1930). We study both whites and blacks and intra- and inter-regional migration. While there is some evidence of positive selection, the degree of selection was small and participation in migration was widespread. Differences in background, including initial location, cannot account for racial differences in destination choices. Blacks and whites were similarly responsive to pre-existing migrant stocks from their home state, but black men were more deterred by distance, attracted to manufacturing, and responsive to labor demand.

At the turn of the twentieth century, real income per worker in the South was less than one-half of that in the rest of the United States (Easterlin 1960; Mitchener and McLean 1999). When WWI led to both a labor demand boom in northern industrial centers and the interruption

947

of immigration from Europe, southern workers moved away from their home region at high rates. This was the start of the "Great Migration," which waned during the Great Depression but surged again between 1940 and 1965. By 1970, 35 percent of southern-born black men and 19 percent of southern-born white men (age 25 and over) were living outside of the South.

The Great Migration was an important event in the history of U.S. labor market integration, and it has had long-lasting social ramifications. Social scientists have studied its causes and consequences for almost 100 years (inter alia, Scroggs 1917; U.S. Department of Labor 1919; Lewis 1931; and recently Hornbeck and Naidu 2014; Collins and Wanamaker 2014; Black et al. 2015). Scholars have focused primarily on the inter-regional movement of African Americans. Indeed, the term "Great Migration" has traditionally been applied solely to black migration to the North. In this article, we expand the frame of investigation in two ways to provide additional perspective. First, we examine the migration decisions of both white and black men. The mass migration of white southerners was important in its own right and provides a natural comparison for the migration patterns of blacks. Second, our analysis includes both intra- and inter-regional migrants, whereas much of the previous literature has focused entirely on the latter. We find that destination choices *within* regions, including the South, provide valuable information and a more complete picture of internal migration patterns during the early decades of the Great Migration.

In addition to expanding the frame of investigation, we develop new data that provide deeper insight than previously available into the careers of southern men between 1910 and 1930. The cross-sectional datasets that inform most quantitative studies of U.S. internal migration in the early twentieth century have a major limitation: researchers simply cannot observe the same person before and after migration.[1] The absence of ex ante information hinders the study of how individual and local character-istics influenced both selection into migration and the migrants' choices of destination. Furthermore, using ex post measures of migrants' outcomes or human capital that are available in cross-sectional data sources may give misleading impressions of their pre-migration status. To overcome this problem, we create linked census records for more than 26,000 men, providing a clearer view of the same men before and after the start of

---

[1] See Bodnar, Simon, and Weber (1982), Collins (2000), Maloney (2001), Logan (2009), Collins and Wanamaker (2014), and Black et al. (2015) for examples of research in which migrants are observed before and after moving.

the Great Migration. We build the dataset by starting with a sample of southern-resident males, ages 0 to 40, in 1910. We then locate the same men in the 1930 census manuscripts and transcribe data from the hand-written documents. In 1910, the younger males in the sample (generally less than age 18) still lived with their parents and siblings, and the older males were already in the southern labor force. In both cases (for younger and older men), the dataset contains valuable pre-Great Migration information on personal, household, and local background.

For African Americans, the linked census records used here are the same as in William J. Collins and Marianne H. Wanamaker (2014), which focused on measuring black men's income gains from inter-regional migration after accounting for selection. This article goes beyond Collins and Wanamaker (2014) in several ways. It studies the selection of both white and black migrants, which required the creation of a new set of linked census records for more than 20,000 southern white men. In addition, this article studies both intra-regional and inter-regional migration patterns, whereas Collins and Wanamaker (2014) ignored intra-regional movement. Intra-regional flows were large in this period and worthy of scholars' attention. Finally, as described later, much of this article is dedicated to studying the migrants' choices of destination and comparing black and white migration patterns across potential destinations, a topic that is not addressed in Collins and Wanamaker (2014).

The linked census records allow us to address several important topics in the history of Americans' internal migration. First, after documenting the outstanding features of southern black and white migration patterns and migrant characteristics, we investigate whether the migrants were strongly selected on the basis of their pre-Great Migration characteristics. Second, we examine how southern migrants sorted themselves across potential destinations and the extent to which personal characteristics, such as place of origin and family background, account for black-white differences in migration patterns. Third, we estimate the migrants' responsiveness to variation in labor market opportunities and migration costs across potential destinations, paying particular attention to racial differences in behavior.

We find that southern men's participation in inter-state and inter-regional migration was widespread in the sense that the migrants' background characteristics were not much different from the non-migrants' characteristics. There is evidence consistent with a degree of positive selection into inter-state migration among both whites and blacks, as measured by indicators of job status. It is also clear that farm residents in 1910 were less likely to move than non-farm residents. Yet overall the

differences between migrants and non-migrants were small within race categories.

In studying the patterns of inter-state migration, we see that there was some overlap in the most popular destinations for white and black migrants, but there were also notable differences. Approximately 28 percent of inter-state migrants would have to change their destination to equalize the white and black distributions over states. Differences in the men's background characteristics can account for surprisingly little of the overall black-white differences in destination choice, which leads us to study differences in responsiveness to economic variables across potential destinations. We find that black and white men were similarly responsive to pre-existing distributions of state-to-state migrant stocks, but that black men were more deterred by distance, more attracted to manufacturing centers, and more responsive to cross-state variation in aggregate labor demand growth, ceteris paribus. A theme in the Great Migration literature emphasizes racial oppression as a strong, independent motivation for leaving the South. We find only mixed evidence that black migrants were more inclined to leave the South than white migrants and no evidence that black migrants moved to non-southern locations more frequently than they did southern ones, conditional on the states' economic characteristics. There is stronger evidence, however, that black migrants moved more frequently than southern white migrants to the Northeast and Midwest, whereas southern whites moved more frequently to the West, conditional on the states' economic characteristics.

BACKGROUND ON SOUTHERN MIGRATION

Despite the sizable regional differences in average income cited earlier, relatively few southern-born men left the region before WWI. For perspective, the charts at the top of Figure 1 use the IPUMS cross-sectional data for each census from 1880 to 1960 (Ruggles et al. 2010) to depict the cumulative inter-regional migration rate for men born in the South between 1860 and 1919 (i.e., the share of southern-born men in each ten-year birth cohort who resided outside the South at each census date). This spans all the birth cohorts of men in our sample of linked census manuscripts (1870 to 1910), which is described in detail in the article's next section, as well as those born up to ten years before and after the men in the linked sample. The bottom panel of Figure 1 plots cumulative interstate migration *within* the South (the share of southern-born men residing outside their state of birth, but in the South). There is no IPUMS sample for 1890, and we simply connect 1880 to 1900
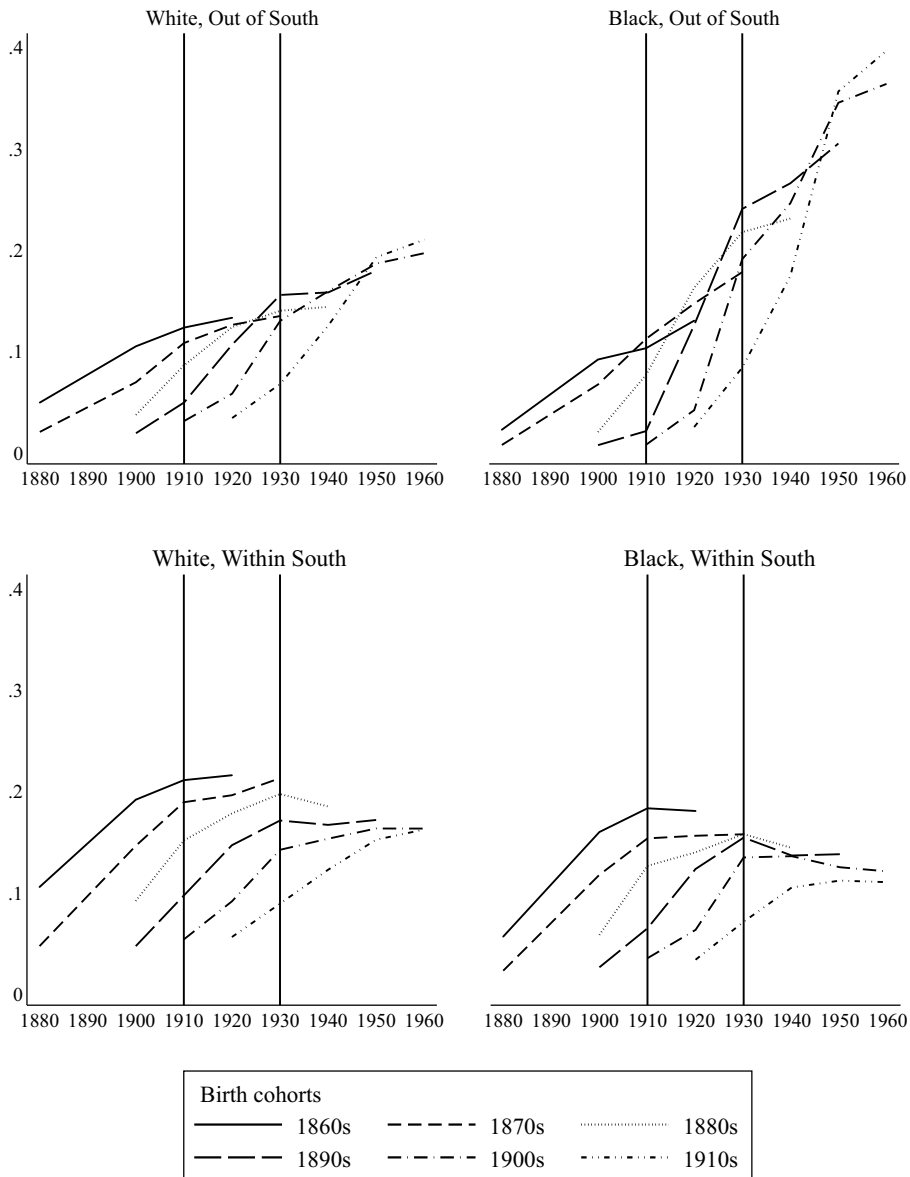
FIGURE 1

INTER- AND INTRA-REGIONAL MIGRATION OF SOUTHERN-BORN MEN, BY RACE
AND BIRTH COHORT, 1880 TO 1960

Notes: The figure plots the share of southern-born men in each cohort who reside outside their
state of birth, either "Out of South" or "Within South." The South is defined as elsewhere in the
article (excluding Delaware, Washington DC, and Maryland). Cohorts are observed at most six
times, corresponding to ages 0–9, 10–19, and so on. Year of birth is defined as year-of-observation
minus (age+1). The 1860s birth cohort is first plotted for 1880 (at ages 10–19). Vertical lines at
1910 and 1930 indicate the timeframe examined using the linked dataset.
Sources: Plots are based on state-of-birth and state-of-residence from the 1880–1960 IPUMS
cross-sections (Ruggles et al. 2010). There is no sample for 1890.

for the relevant cohorts. The vertical lines at 1910 and 1930 correspond to this article's main window of analysis, reflecting the structure of the dataset of linked census records and the first decades of the Great Migration. Throughout the article, the "southern" states include: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. We categorize Delaware, Maryland, and the District of Columbia as "non-southern."

In 1910, prior to the onset of the Great Migration, only 11 percent of black and white southern-born males in their 30s (the 1870s birth cohort) resided outside the South. For whites, this is approximately the same level of cumulative inter-regional migration as observed ten years earlier (in 1900) for men at a similar point in the lifecycle (the 1860s birth cohort). Later cohorts of whites undertook substantially more inter-regional migration: 15 percent of the 1890s birth cohort had left the South by 1930, and closer to 20 percent had left among those exposed to the WWII boom (observed in 1950). For blacks, the 1870s birth cohort had a slightly higher rate of cumulative inter-regional migration by 1910 than the 1860s cohort observed in 1900 (11 compared to 9 percent). That difference is dwarfed by subsequent changes in black inter-regional migration rates, with nearly one-quarter of the 1890s birth cohort leaving the South by 1930 and even higher rates for cohorts exposed to WWII (e.g., 36 percent of the 1910s birth cohort observed in 1950).

The comparatively low rate of inter-regional migration before 1910 does not imply that the southern labor force was stationary, as there is considerable evidence of mobility *within* the South (Wright 1986). The lower panels of Figure 1 indicate that 15 percent of blacks and nearly 20 percent of whites in the 1870s birth cohort had moved away from their birth state by 1910 but stayed within the South. In this sense, intra-regional mobility was more common than inter-regional mobility prior to 1910. In particular, there were sizable flows of southern-born whites into Texas and Oklahoma and sizable flows of southern-born blacks into Florida, Arkansas, and Mississippi.

After 1910, flows within the South continued, especially to Texas, Oklahoma, and Florida, but destinations like California, Ohio, Illinois, Pennsylvania, and New York became much more prominent than previously. Between 1910 and 1930, Figure 1 shows slight declines in within-South migration but significant increases in inter-regional migration, reflecting a shift in the relative attraction of non-southern residence for both blacks and whites.

Scholars have suggested several reasons for the relatively low rates of African American migration from the South prior to WWI: low levels of human and financial capital, which have been found to impede long-distance migration in this and other settings (Logan 2009; Hatton and Williamson 2002); discrimination by northern employers who readily hired immigrant laborers from Europe (Myrdal 1944; Collins 1997); and weak integration of northern and southern labor markets compared to the strong ties between northern and European labor markets, a legacy of mass migration from Europe that began in the 1840s (Thomas 1954; Wright 1986; Hatton and Williamson 1998; Rosenbloom 2002). Impediments to the mobility of southern agricultural workers also may have been significant (Ransom and Sutch 1977; Grossman 1989; Naidu 2010).

Despite sizable inter-state flows of southern whites, the history of their internal migration has been far less explored than that of African Americans (Akers 1936; Killian 1953; Berry 2000; Gregory 2005). Of course, some of the same factors that inhibited blacks' movement prior to WWI also may have affected southern whites. But it seems clear that the black and white stories differ in important ways. Most obviously, whites were not recently removed from slavery, were less concentrated than blacks in the Cotton Belt, were more likely than blacks to have acquired some wealth and literacy, and likely faced less discrimination in distant labor markets. For perspective, in 1870 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 (Ruggles et al. 2010). However, while southern whites might have found it easier to afford long-distance moves than blacks (on average), they also might have found it less attractive. For instance, opportunities to advance in southern labor markets, whether by ascending the agricultural ladder or moving into skilled non-agricultural work, may have seemed more plentiful to whites than to blacks. It is also possible that lingering hostility from the Civil War influenced white southerners' perceptions of the North.

If ignorance and poverty inhibited long-distance migration by black and white southerners, this constraint may have receded with each generation's educational and economic advances in the late nineteenth and early twentieth centuries (Higgs 1982; Margo 1984, 1990; Collins and Margo 2006). In addition, improving transportation and communication networks in the South may have facilitated migration by lowering the associated costs and uncertainties. Surfaced roads more than doubled

between 1904 and 1914 in the South (U.S. Department of Agriculture 1917). Nearly all southerners lived in counties with railroad access by 1911 (Atack 2013 and personal communication). The circulation of northern newspapers, such as the Chicago *Defender*, increased in the South during the 1910s (Grossman 1989). Environmental shocks also may have driven some migration (Higgs 1976; Lange, Olmstead, and Rhode 2009; Boustan, Kahn, and Rhode 2012; Hornbeck and Naidu 2014). Most prominently, the boll weevil spread from Texas, where it gained a foothold in the 1890s, to Mississippi by 1907 and North Carolina by 1919. The boll weevil disrupted cotton production and imparted a long-lasting negative productivity shock, perhaps making southern agriculture a less attractive option than before, at least within the Cotton Belt (Higgs 1976; Lange, Olmstead, and Rhode 2009; Baker 2015). Finally, specifically for African Americans, political disenfranchisement, mob violence, *de jure* segregation, and, in general, the ascendance of the Jim Crow regime may have provided a strong incentive to leave the region (Myrdal 1944; Woodward 1955; Kousser 1974; Tolnay and Beck 1990; Margo 1990). Taken together, these trends may have yielded a southern labor force circa 1910 that was more able and more inclined to migrate long distances than ever before.

Against this backdrop, the exogenous shock of WWI created both a major labor demand boom in industrial centers, predominantly located in the North, and a temporary halt to European immigration, which was later reinforced by immigration restrictions. Many industrial employers recruited southern migrants for the first time, gained experience in hiring, training, 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, networks of previous migrants helped perpetuate migration patterns in and from the South (Carrington, Detragiache, and Vishwanath 1996; Hatton and Williamson 1998; McKenzie and Rapoport 2007). Consequently, high rates of migration continued long after the impetus of WWI, with major repercussions for American economic and social history.

Within this historical setting, our thinking about migration and location choice is guided by theoretical and empirical work by Larry A. Sjaastad (1962), Jennifer Roback (1982), George J. Borjas, Stephen G. Bronars, and Stephen J. Trejo (1992), Jeffrey Grogger and Gordon H. Hanson (2011), and Enrico Moretti (2011). We suppose that an individual's location decision depends on expected income, amenities, and relocation costs (broadly defined), while recognizing that these expectations may vary across types of workers, by race, skill, or other initial conditions.

These differences across workers may give rise to interesting patterns of selection into inter-state migration and of sorting across potential destinations, which we explore later. For both selection and sorting, the dataset of linked census records provides new opportunities for analysis.

NEW DATA: LINKING CENSUS RECORDS, 1910–1930

To construct the linked dataset, we began with the IPUMS sample of manuscript data from the 1910 *Census of Population* and selected all southern-resident males between the ages of 0 and 40 (Ruggles et al. 2010). We then attempted to link these men "forward" by locating them in the handwritten manuscripts of the 1930 *Census of Population*. We used each individual's name, place of birth, and age from the 1910 records as search criteria for location in the 1930 records.[2] From an initial sample of 111,524 individuals, the linking process successfully located 26,829 individuals, a 24 percent match rate.[3] The final sample is 20 percent black. Additional details on the linking process and the variables available in each census year are provided in the Appendix.

As mentioned above, the linked data offer several advantages relative to the state-level aggregates or micro-level cross-sections that have typically supported 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 big wave. In 1910, we see the younger men in our sample when they still resided with their family. Therefore, we observe many characteristics about the household in which they grew up—what their parents did for a living, where they were located, whether they attended school, their literacy (if older than 9), whether they lived on a farm, 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 owned a home, and so on. In 1930, we see whether the men have moved since 1910 and, if so, to where. Given the 1910 county

[2] Our search criteria include a SOUNDEX version of the last name, first three letters of the first name, 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 reasonably close approximations to that name.

[3] Our linkage rates are similar to those in Long and Ferrie (2013) and Abramitzky, Boustan, and Eriksson (2012). We deleted cases where more than one individual in 1910 matched to the same individual in 1930. Match rates for blacks (19 percent) and whites (26 percent) differ, which we believe is partially attributable to advances in search technology over time. The samples were matched sequentially.

and state identifiers, it is straightforward to merge the linked dataset with 1910 county and state-level information from Michael Haines (2010), which helps characterize each individual's local economic and social environment.

A major concern for any such dataset is that the linked sample might exhibit selection bias that interferes with subsequent analyses and interpretations. Fortunately, the men in the linked sample are similar to those in the randomly selected base sample from 1910. Table 1 separately reports the summary statistics for the linked and base samples of blacks and whites. Although there are some small differences, their statistical and economic significance is limited. Similar conclusions follow when we estimate the probability of being in the linked sample as a function of observable characteristics, conditional on being in the base IPUMS sample. From that exercise (details of which are discussed in the Appendix), we conclude that literacy, farming occupations, and West Virginia residence are statistically significant predictors of being found in the 1930 census manuscripts, but the practical significance of these increased probabilities is small, generally less than a 3 percent increase.[4]

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. We do not expect these numbers to be exactly the same because of interregional migration of men prior to 1910 in the IPUMS cross-section (where migrants are defined using place of birth information) and sampling variation. The corresponding numbers for southern-born whites observed outside the South in 1930 are 15 percent and 17 percent. Across all states, the distribution of inter-state migrants in the linked sample is highly correlated with inter-state migrants in the 1930 IPUMS sample (0.96 for both the white and black samples), and rarely deviates by more than 1 percentage point. The correlation is especially high for non-southern destinations, where pre-1910 migration is less likely to confound the comparisons (Appendix Table 3). Overall, the results do not suggest that the linked sample is biased in a way that will confound our analyses, and we take the linked sample to be fairly representative.

---

[4] We evaluated the results' sensitivity to using stricter match criteria by restricting the sample to exact matches on last name and close matches on first name (Jaro-Winkler distance <0.3). The main results are unchanged. See the Online Appendix.

TABLE 1
COMPARISON OF LINKED AND BASE SAMPLE CHARACTERISTICS,
SOUTHERN MALES 1910

	White Males			Black Males		
	Matched Sample	Full IPUMS Sample	*p-value of Diff.*	Matched Sample	Full IPUMS Sample	*p-value of Diff.*
Panel A: Distribution of state of residence (percentages)						
Alabama	6.5	6.5	*1.00*	9.8	10.3	*0.23*
Arkansas	6.4	6.1	*0.09*	5.0	4.9	*0.61*
Florida	2.1	2.3	*0.09*	4.0	3.8	*0.43*
Georgia	7.5	7.5	*0.99*	13.5	13.9	*0.48*
Kentucky	10.8	10.4	*0.14*	3.2	3.0	*0.44*
Louisiana	4.7	5.2	*0.01*	8.2	8.7	*0.26*
Mississippi	4.2	4.1	*0.74*	12.9	12.2	*0.14*
North Carolina	7.9	7.6	*0.14*	8.9	8.0	*0.03*
Oklahoma	8.0	8.1	*0.84*	1.5	1.7	*0.31*
South Carolina	3.5	3.6	*0.28*	11.0	10.2	*0.10*
Tennessee	8.6	8.6	*0.86*	5.3	5.5	*0.54*
Texas	16.2	16.8	*0.02*	9.3	9.0	*0.60*
Virginia	6.9	7.0	*0.57*	6.8	7.9	*0.01*
West Virginia	6.9	6.2	*0.00*	0.7	1.0	*0.04*
Panel B: Personal characteristics (percent, except age variables)						
Attending school (age 0–20)	82.9	81.8	*0.04*	62.3	62.4	*0.94*
In owner-occupied housing	52.4	48.8	*0.00*	22.4	23.9	*0.05*
Literate (age 10–20)	91.8	90.1	*0.00*	62.4	62.6	*0.86*
Literate (age 10–40)	92.6	91.2	*0.00*	65.7	65.6	*0.91*
Father is farmer (age 0–20)	60.9	59.0	*0.00*	63.4	63.3	*0.30*
1910 city population						
Not in city	70.0	69.4	*0.09*	74.5	73.2	0.05
City pop. <=25,000	20.4	20.8	*0.25*	16.3	17.3	*0.07*
City pop. >25,000	9.6	9.8	*0.31*	9.2	9.5	*0.53*
Min age	0	0	—	0	0	—
Max age	40	40	—	40	40	—
Median age	16	15	*0.00*	16	15	*0.08*
Mean age	17.1	16.8	*0.01*	17.0	16.7	*0.06*
St. dev.	11.1	11.5	*0.00*	11.2	11.3	*0.50*

Notes: A variance-ratio test is used to compare sample standard deviations and a Wilcoxon matched-pairs signed-rank test is used to compare sample medians. All others comparison of means are done with standard t-tests. The matched samples contain 21,367 white men and 5,462 black men.

Sources: The linked sample is created by taking the 1910 IPUMS sample of white and black males, age 0–40, who reside in the South and searching for these men in the 1930 census manuscripts. The text and Appendix describe sample construction in detail. The IPUMS data are from Ruggles et al. (2010).

MIGRANT SELECTION

Patterns of migrant selection are important for both sending regions (e.g., whether the region tends to lose highly skilled workers) and receiving regions (e.g., whether migrants are likely to assimilate quickly and whether they are substitutes or complements for the area's native workers). The linked dataset is especially well suited to characterize selection into migration because it has such detailed background information on the men in 1910. These characterizations, in turn, provide a better basis than previously available for understanding the origins and consequences of the Great Migration, even though assessing the full range of possible implications from selection, such as the wage impact on sending and receiving areas (Boustan 2009), is beyond this article's scope.

In this section, we focus on characterizing who migrated and whether there are clear differences between migrants and non-migrants in terms of their background characteristics. We start by classifying all men in the linked sample into three mutually exclusive categories: "non-migrants" (who resided in the same state in 1910 and 1930), "within-South migrants" (who changed state of residence within the region), and "inter-regional migrants." Between 1910 and 1930, a large fraction of the men in the linked sample (35 percent of whites and 39 percent of blacks) had left their 1910 state of residence, with roughly even splits between "within-South" and "inter-regional" migration.

Table 2, column 1 reports the white men's average characteristics in 1910, before the start of the Great Migration, including their literacy rates, school attendance, occupational income and education scores, farm status and homeownership rates. The figures are tabulated separately by 1930 migration category. The second column reports *differences* in characteristics between the migrant categories and the non-migrant category, and the third column reports the differences that remain after controlling for age and county-of-origin fixed effects.[5] In practice, the county-level fixed effects in column 3 absorb local push factors, such as boll weevil destruction, and control for selection that derives from differences in place-of-origin characteristics (which is included in the simple comparison in

---

[5] "The "adjusted differences" are coefficients on migration category dummies (expressed relative to the non-migrant category) from regressions with age (γ_i) and county-of-origin (θ_j) fixed effects: $Y_i = \alpha + \sigma SM_i + \rho RM_i + \gamma_i + \theta_j + e_i$, where SM is a dummy for within-South migrants and RM is a dummy for inter-regional migrants. Y variables are the 1910 characteristics (e.g., literacy, occupation score, etc.).

column 2).[6] The next three columns repeat the tabulations for the sample of black men. Although this is very basic information about the migrants and non-migrants, none of it can be inferred from census cross-sections, and in this respect the linked manuscript data are crucial.

If ability or productivity were positively correlated with migration, then we would expect migrants to have better outcomes than non-migrants in terms of human capital, occupational status, or family background *before* leaving the South. While discrimination in the South slowed black men's economic and educational progress, there was considerable variation in literacy, education, occupation, property ownership, and other measures in 1910. For whites and blacks, there is some evidence of positive selection into inter-regional migration on the basis of 1910 literacy (panel A), but the differences are quantitatively small at 1 to 4 percentage points. They are statistically insignificant after controlling for age and place of origin. This contrasts with some of the previous literature's suggestion of strongly positive migrant selection in the early twentieth century based on census cross-sections that observe migrants ex post (e.g., Tolnay 1998). The discrepancy suggests that migrants might have improved their literacy after leaving the South, or that census enumeration of literacy might have been regionally biased. More generally, this raises concern regarding the practice of using ex post migrant characteristics from cross-sections to make inferences about selection into migration. As linked historical datasets become more common, scholars may be able to avoid this measurement problem.

For an alternative view of educational background, we examined school attendance in 1910 and found small differences in attendance rates between the migrant categories (panel B), with or without controls. From this perspective, migrant selection on the basis of formal education seems rather weak in the early decades of the Great Migration. We note, however, that in comparison to studies of migration in more recent decades (Borjas, Bronars, and Trejo 1992; Vigdor 2002; Grogger and Hanson 2011), which are often motivated by the Roy model (1951), our metric for educational attainment is fairly crude. We cannot follow such studies without better data for quality-adjusted educational attainment and variation in returns to education (by race) across locations in the early twentieth century, which to our knowledge does not exist.

---

[6] If individuals tend to migrate from particularly affluent locations, for example, we would infer positive selection in column 2 but not necessarily in column 3. Both perspectives are useful. Changes in coefficients between columns 2 and 3 are driven primarily by the inclusion of county fixed effects, not age effects.

TABLE 2
SELECTION INTO 1910–1930 MIGRATION ON BASIS OF 1910 CHARACTERISTICS

	White Males			Black Males		
	Sample Average	Difference vs. Non-migrants	Adjusted Difference	Sample Average	Difference vs. non-Migrants	Adjusted Difference
A. Literacy rate (*N*=14,761 white and 3,702 black)						
Non-migrants	0.924	—	—	0.650	—	—
Within-South migrants	0.925	0.001 (0.006)	–0.000 (0.006)	0.653 —	0.003 (0.021)	–0.008 (0.026)
Inter-regional migrants	0.936	0.013** (0.006)	–0.004 (0.006)	0.684 —	0.035* (0.020)	0.008 (0.024)
B. School attendance rate (age 6 to 15) (*N*=6,587/1,705)						
Non-migrants	0.828	—	—	0.613	—	—
Within-South migrants	0.823	–0.006 (0.013)	–0.005 (0.015)	0.612 —	–0.006 (0.031)	–0.013 (0.046)
Inter-regional migrants	0.838	0.010 (0.013)	–0.003 (0.015)	0.655 —	0.043 (0.026)	0.035 (0.036)
C. Occupation income score, in hundreds of 1950 dollars (*N*=11,960/3,259)						
Non-migrants	16.3	—	—	12.3	—	—
Within-South migrants	17.3	1.00*** (0.325)	0.80*** (0.292)	13.4 —	1.07*** (0.384)	0.53* (0.315)
Inter-regional migrants	18.2	1.86*** (0.379)	0.81*** (0.285)	13.2 —	0.85** (0.357)	0.65* (0.357)

D. Occupation education score (*N*=11,834/3,239)						
Non-migrants	10.77	—	—	7.90	—	—
Within-South migrants	10.87	0.104**	0.079	8.04	0.131**	0.082
		(0.047)	(0.048)	—	(0.054)	(0.067)
Inter-regional migrants	10.92	0.148***	0.062	8.02	0.120***	0.025
		(0.053)	(0.050)	—	(0.041)	(0.042)
E. Farm origin (*N*=21,367/5,462)						
Non-migrants	0.610	—	—	0.592	—	—
Within-South migrants	0.530	–0.080***	–0.083***	0.521	–0.071***	–0.029
		(0.012)	(0.009)	—	(0.021)	(0.019)
Inter-regional migrants	0.500	–0.11***	–0.071***	0.502	–0.090***	–0.033*
		(0.013)	(0.010)	—	(0.019)	(0.017)
F. Homeownership rate (*N*=21,367/5,462)						
Non-migrants	0.543	—	—	0.222	—	—
Within-South migrants	0.472	–0.071***	–0.056***	0.200	–0.022	–0.004
		(0.010)	(0.010)	—	(0.015)	(0.016)
Inter-regional migrants	0.505	–0.038***	–0.037***	0.251	0.030**	0.007
		(0.011)	(0.010)	—	(0.015)	(0.016)

* = Significant at the 90 percent level.
** = Significant at the 95 percent level.
*** = Significant at the 99 percent level.
Notes: "Adjusted differences" are regression coefficients that measure differences in 1910 characteristics among inter-regional, intra-region, and non-migrant categories (where non-migrants are the base category), controlling for age and county-of-origin fixed effects. Non-migrants are defined as residing in the same state in 1910 and 1930; within-South migrants left their state-of-origin but remained in the South. Literacy is recorded for those who are age 10 and over; school attendance is examined for 6 to 15 year olds. Occupational income and education scores are conditional on labor force participation in 1910. The standard deviation for occupation income score (occupation education score) is 12.0 (1.75) for whites and 7.3 (0.93) for blacks. Standard errors, clustered by county of origin, are in parentheses.
Sources: Data are from the sample of linked census records, as described in the text and Appendix.

Among those old enough to report occupations in 1910 (panels C and D), there is somewhat stronger evidence of positive selection on the basis of pre-Great Migration labor market outcomes. To the extent that better skills, ability, or motivation translated into better occupational standing before WWI, the occupational data provide a more sensitive indicator of selection than literacy alone. Individual income was not reported in the 1910 census, and so we first examine a modified version of the "occupational income score" (panel C).[7] This is a commonly used IPUMS variable that assigns an income to each detailed occupation category based on the median income observed in that occupation in the 1950 census (Ruggles et al. 2010). Pre-war occupational scores for "within-South" and "inter-regional" white migrants are 5 to 10 percent higher than for non-migrants. For blacks, most group differences are slightly smaller in magnitude than for whites, but the point estimates are consistent with positive selection.[8]

In panel D, we assign a race-specific "education score" to each detailed occupation in 1910 based on the average educational attainment of southern workers in the corresponding occupation categories in the 1940 census (Ruggles et al. 2010), which was the first census to inquire about educational attainment. The goal is simply to provide an alternative characterization of occupational status that is connected to formal educational attainment rather than income. Again, there is some evidence of positive selection into migration, approximately one-tenth of a grade, but the differences are small and statistically insignificant after controlling for age and initial location.

In summary, our measures of skill—whether based on occupational or educational information—are generally consistent with a limited degree of positive selection for both whites and blacks in the early decades of the Great Migration. The differences across migrant groups are quantitatively small, however, and thus the degree of migrant selection from this perspective seems rather weak.

---

[7] Given the difference in status for tenant farmers relative to farm owners, we further adjusted the occupation score for farmers based on their homeownership. Homeownership is not reported in the 1950 microdata sample, the year on which occupation scores are originally based. Therefore, for farmers who reported owning (renting) their homes, we multiplied median farmer income in 1950 (the IPUMS-defined *occscore*) by the ratio of median income among farmer-owners (farmer-renters) in the 1960 IPUMS sample to median income of all farmers in the 1960 sample, restricting the 1960 sample to southern residents. This adjusted *occscore* is analyzed in the third panel of Table 2.

[8] For reference, a standard deviation in the occupation income score variable is 12.0 for whites and 7.3 for blacks (panel C of Table 2). A standard deviation in the occupation education score variable is 1.75 for whites and 0.93 for blacks (panel D).

Table 2, panel E examines selection on the basis of farm status in 1910. From this perspective, there are notable differences. For both whites and blacks, those who lived on farms in 1910 were less likely to migrate out of the state than others. In the "unadjusted" columns, the farm-status difference between the migrant groups and the non-migrant group is between 7 and 11 percentage points (whites and blacks). The addition of age and county-of-origin controls reduces the gap relative to non-migrants for white inter-regional migrants and for both categories of African Americans migrants, but non-trivial differences remain, especially for whites. Thus, it appears that an agricultural background tended to hinder long-distance migration, even when comparisons are based on within-county variation.

Owner-occupied housing status, which is examined in Table 2's last panel, is of particular interest because it is the only census variable in 1910 that reflects household assets (ownership of real property). In general, household wealth may facilitate long-distance migration, but in this historical context homeownership may also indicate a prior decision to settle down in a particular area or, for young men in our sample (residing with their parents in owner-occupied housing), it may reflect an expectation to inherit or receive a parental gift of local property (Abramitzky, Boustan, and Eriksson 2013). Among whites, residing in owner-occupied housing in 1910 is associated with substantially less long-distance migration (between 3 and 7 percentage points), even after adjustments for age and county of residence. African Americans were far less likely than whites to own their homes, and the pattern with respect to migration is different. Inter-regional black migrants were somewhat more likely to reside in owner-occupied housing in 1910. If the negative coefficients for whites are interpreted as indicators of relatively strong local attachments among property owners and their children, then it would appear that black property owners and their children did not share such strong attachments.

Overall, migration in the linked dataset does not conform to a simple characterization in terms of negative or positive selection. Migrants held jobs in 1910 with higher occupation scores than non-migrants on average, which is consistent with positive selection. But these differences were not large. Moreover, differences across migrant groups in terms of literacy, school attendance, or occupational-education scores were small or non-existent. Farm residence was the most robust (negative) predictor of subsequent migration for whites and blacks, a finding that is consistent with our expectations but novel in the sense that, to our knowledge,

no previously constructed dataset could observe the pre-migration farm status of individual men in this period. We interpret the overall results as a reflection of remarkably broad participation in internal migration. The differences between migrants and non-migrants in the early decades of the Great Migration appear less salient than the increased volume and new directions of migration. This leads us to investigate further *where* the migrants moved and why they decided to move there.

OVERVIEW OF MIGRATION PATTERNS AND SORTING

The linked dataset provides new opportunities for studying and comparing migrant sorting patterns, both as a function of individual and place-of-origin characteristics and as a function of labor market conditions and migration costs across space. In addition, the linked data allow us to be more specific about when people moved (between 1910 and 1930) than is possible with census cross-sections, where prior location is known only at the time of birth and migration could have occurred at any time afterwards.

Table 3 provides summary statistics of the migration patterns in the linked dataset, including the propensity to migrate (panel A) and distance and direction of migration (panel B). The table's third column reports simple black-white differences. The fourth column reports coefficients on a black indicator variable from regressions that control for state-of-origin fixed effects. As mentioned above, between 1910 and 1930, a large fraction of the men in the linked sample left their 1910 state of residence, with approximately an even split of migrants between "within-South" and "outside South" destinations. Among those who left the South, the Midwest was the most common destination for both whites and blacks, but black inter-regional migrants moved relatively strongly into the Northeast (compared to whites), whereas whites moved relatively strongly into the West.[9] Detailed state-to-state migration patterns are reported in the Online Appendix.

Panel B of Table 3 reports measures of distance travelled based on the latitude and longitude of the center of each individual's 1910 and 1930

---

[9] The most common destination counties for inter-regional black migrants were Cook County (Chicago), IL; Philadelphia, PA; Wayne (Detroit), MI and New York, NY. White inter-regional migrants most frequently relocated to Los Angeles, CA; Cook (Chicago), IL; Wayne (Detroit), MI and Summit (Akron), OH. Intra-regional black migrants moved most frequently to Shelby County (Memphis), TN; Duval (Jacksonville), FL; Hamilton (Chattanooga), TN; and Chatham (Savannah), GA. White inter-regional migrants relocated most frequently to Oklahoma City, OK; Shelby (Memphis), TN; Jefferson (Birmingham), AL; and Dallas, TX.

TABLE 3
MIGRATION PATTERN SUMMARY STATISTICS, BY RACE: 1910–1930

	White	Black	Unconditional Black-White Difference	Conditional Black-White Difference
Panel A: Migration propensities (shares)				
No state-to-state migration	0.65	0.61	–0.037***	–0.058***
			(0.008)	(0.008)
Within-South migrants	0.18	0.19	0.0088	–0.018***
			(0.006)	(0.006)
Inter-regional migrants	0.17	0.20	0.028***	0.075***
			(0.006)	(0.006)
South-to-Northeast migrants	*0.03*	*0.08*	*0.050****	*0.050****
			(0.003)	(0.003)
South-to-Midwest migrants	*0.09*	*0.11*	*0.016****	*0.049****
			(0.005)	(0.005)
South-to-West migrants	*0.05*	*0.01*	*–0.037****	*–0.024****
			(0.003)	(0.003)
Panel B: Geographic characteristics of migration				
Distance travelled (miles)				
Full sample	219.8	210.7	–9.06	3.67
			(5.56)	(5.76)
Within-South migrants	321.8	266.0	–55.81***	–41.70***
			(9.27)	(9.43)
Inter-regional migrants	695.7	576.7	–119.0***	–232.9***
			(19.15)	(19.83)
Change in latitude, home to destination (degrees)				
Full sample	0.69	1.38	0.69***	0.60***
			(0.046)	(0.047)
Within-South migrants	–0.09	0.47	0.56***	0.23***
			(0.116)	(0.109)
Inter-regional migrants	3.96	6.20	2.23***	0.56***
			(0.138)	(0.130)
Change in longitude, home to destination (degrees)				
Full sample	–0.90	0.30	1.21***	1.09***
			(0.102)	(0.106)
Within-South migrants	–0.51	0.20	0.71***	1.34***
			(0.220)	(0.213)
Inter-regional migrants	–4.88	1.42	6.30***	6.20***
			(0.472)	(0.540)

Notes: The last column reports regressions coefficients that compare black and white migration patterns controlling for state-of-origin fixed effects. The "full sample" includes non-migrants (defined as those who do not leave the state-of-origin). South-to-Northeast includes 1930 residents of the Northeast census regions and also Delaware, Maryland, and Washington, DC. Latitude differences are positive for south-to-north migration. Longitude differences are positive for west-to-east migration. Standard errors, clustered by county of origin, are in parentheses.
Sources: Data are from the linked sample of census records, as described in the text and Appendix.

county of residence. In the full sample, average migration distances, including zeros for those who did not change counties, are similar across the black and white samples, with or without controls for state-of-origin. It is clear, however, that conditional on migrating, whites moved farther than blacks on average, both within the South (322 versus 266 miles) and when leaving the South (696 versus 577 miles) (also see Tolnay et al. 2005). Controlling for state-of-origin amplifies the large difference in distance travelled by inter-regional migrants (column 4). The average black male moved northward and eastward (positive changes in latitude and longitude), whereas the average white male moved northward and westward, though not as far north as blacks. Among white inter-regional movers, the migrants to the West strongly influence the average change of longitude, and black-white differences in east-west mobility patterns are striking, even when controlling for state of origin.[10]

For the sake of concise description and to facilitate discrete-choice modeling, we focus on the migrants' choice of state. In our sample 51 percent of inter-state migrants chose non-urban locations, and therefore focusing solely on migrants to cities would omit a large share of the sample, distort the ex ante set of destination choices, and generally mischaracterize the period's internal migration patterns. Moreover, because the multinomial logit model, described below, estimates a large number of coefficients for each potential destination, working at the county- or city-level would be computationally prohibitive.

Figure 2 maps the distribution of inter-state southern migrants across destinations between 1910 and 1930. Continuing pre-1910 migration trends, migrants within the South tended to favor Texas, Oklahoma, and Florida, with Texas and Oklahoma being particularly important destinations for whites. In addition, and breaking with pre-1910 patterns, black migrants frequently chose locations in the industrial North; Pennsylvania, Illinois, Ohio, and New York all received more southern black migrants after 1910 than did any southern state. Pennsylvania was the most common destination for blacks, and it drew heavily from Virginia, Georgia, and South Carolina. Southern white migrants were also drawn strongly to Ohio and, to a lesser extent, Illinois. But California also emerged as a major destination for whites, drawing heavily from the 1910 residents of Texas, Oklahoma, Arkansas, and Kentucky.

---

[10] Appendix Figure 1 in the Online Appendix graphs scatterplots of changes in latitude and longitude, separately for southern whites and blacks. The main visual impression is one of strong northward movement for blacks but a much more diffuse pattern for whites.
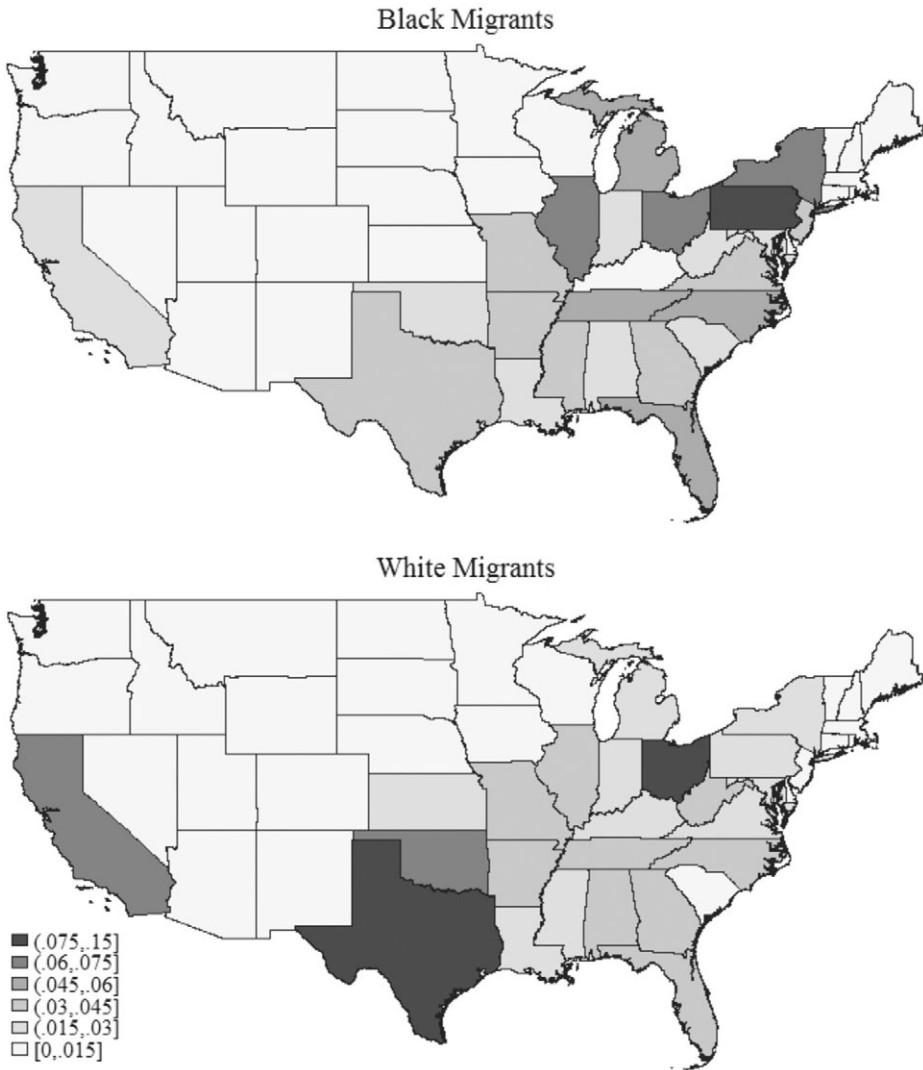
FIGURE 2
DISTRIBUTION OF INTER-STATE MIGRANTS IN LINKED SAMPLE, 1910–1930

Notes: The maps show the share of inter-state migrants choosing each destination state, separately by race. The shares add to 100 percent for each race category.
Sources: Data are from the linked sample of census records, as described in the text and Appendix.

Although there was a substantial degree of overlap in black and white migration flows, there were also notable differences among the most frequently chosen locales. Only Ohio and Illinois rank among the top five destination states for both groups, and in some cases the black-white differences are quite large (e.g., California and Pennsylvania). A

dissimilarity index provides a simple way to summarize the magnitude of black-white differences in the migrants' location choices. It indicates the share of migrants that would have to choose a different location for the distribution of choices to be the same across race categories. The index is calculated as $\frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_i}{B}-\frac{w_i}{W}\right|$ where i denotes a state, $b_i(w_i)$ is the number of black (white) men moving to state i and $B(W)$ is the total number of black (white) migrants in the sample. From this perspective, approximately 28 percent (index value of 0.28) of black or white inter-state migrants would have to choose a different destination for the black and white post-migration distributions to be equivalent.

BACKGROUND CHARACTERISTICS AND DIFFERENCES IN BLACK AND WHITE MIGRATION PATTERNS

Because black and white men differed in their observable characteristics and starting locations circa 1910, it is natural to ask whether such differences can account for black-white differences in migration patterns. We take two different approaches to this question. First, looking deeper into the summary statistics reported in Table 3, we estimate a series of ordinary least squares (OLS) regressions that include a race indicator variable and a rich set of background variables, such as age, literacy, industry of employment (or father's industry of employment), place of origin, and so on. Comparing "unadjusted" black-white migration differences to those "adjusted" for background characteristics, we find that the background characteristics generally cannot account for the differences in migration outcomes. In fact, adding control variables tends to widen, not narrow, some black-white differences in migration choices. For the sake of brevity, these results and additional details are provided in the Online Appendix.

Second, similar in spirit to the above but with a sharper focus on the actual choice of destination, we estimate multinomial logit models to characterize the location choices of inter-state migrants as a function of race and background characteristics. The multinomial framework treats each state as a potential destination, with the caveat that we combined some less populous states to facilitate estimation. For the subsample of individuals age 17 and under in 1910, the model includes indicator variables for race, father's literacy and industry of employment, own school attendance, place in the family's birth order, owner-occupied housing status, state of origin, and city size. For those 18 and older in 1910, the list

of independent variables is similar, but includes own literacy and industry of employment (rather than father's) and marital status, but omits school attendance and birth order. The Online Appendix provides more details.

Using the model's parameter estimates, the importance of black-white disparities in observable characteristics is revealed by comparing two counterfactual migration distributions in which men have the same characteristics except for race. Specifically, we take the full sample of white and black men and predict destination choices when all are assigned "black" status and then again when all are assigned "white" status, such that the differences across the two sets of predictions are attributable only to differences in race as all other personal attributes are equivalent across counterfactuals (StataCorp 2009). For each man, this procedure estimates probabilities of choosing each state under white and black model parameters; the probabilities are averaged to get the predicted "all white" or "all black" counterfactual distributions of migrants across states. If differences in background characteristics largely explained black-white differences in destination choice, then the value of the dissimilarity index between the all black and all white counterfactual distributions would be substantially less than the index value based on the actual distributions. In practice, however, recalculating the dissimilarity index between the "all black" and "all white" counterfactual distributions yields an index value of 0.27. From this perspective, only a small portion of black-white differences in destination choice, on net, can be accounted for by the background characteristics available in the census. Underlying the dissimilarity index results, we see that background characteristics are helpful in narrowing the black-white difference in migration to some states (most notably California), but they widen the black-white difference in other states (such as Illinois and Missouri), such that the overall dissimilarity index changes little.[11]

In sum, we find that there was substantial overlap in black and white migration patterns, but black-white *differences* in migration patterns were economically significant and generally were not a reflection of the migrants' background characteristics. Observationally similar southern men circa 1910 tended to make different migration decisions depending on their race. This finding leads us to focus on how strongly (and how differently) white and black migrants responded to variation in the costs and benefits of relocating to each potential destination. Conditional logit models are particularly useful for studying such issues.

---

[11] The state-specific results are illustrated in the Online Appendix Figure 3.

CONDITIONAL LOGIT MODEL OF DESTINATION CHOICE

Suppose that individual i chooses to migrate to state j if

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

where U(.) represents utility over a vector X_{ik}, which contains variables that reflect the expected income, amenities, and migration costs for individual i in potential destination k. If the utility over each state choice includes a random component with an Extreme Value Type I distribution, the probability of choosing any particular state is represented by:

$$P(D_i = j) = \frac{\exp(\gamma X_{ij})}{\Sigma_{k=1}^{J} \exp(\gamma X_{ik})}, \ \ \forall j \in [1,...,J] \tag{1}$$

where D_i is the location choice of individual i. This is the familiar conditional logit framework for discrete choice described in Daniel McFadden (1974), applied to a setting in which migrants choose destinations (Davies, Greenwood, and Li 2001; Vigdor 2002; Wozniak 2010). As cited earlier in the article, our interpretation of the model in this setting is strongly influenced by previous research on the economics of migration and location choice (Sjaastad 1962; Roback 1982; Borjas, Bronars, and Trejo 1992; Grogger and Hanson 2011; Moretti 2011).

Maximum-likelihood estimates of γ describe how differences in economic characteristics across potential destination states are correlated with the choices of inter-state migrants. Note that in conditional logit models, any variable that does not vary across potential destinations for individual i, such as an individual-specific variable or place-of-origin-specific variable, falls out of the estimating equation and is not identified. Coefficients on destination-specific characteristics and on interactions between destination characteristics and individual-specific attributes can be identified (e.g., interactions of labor demand and race).

We estimate the conditional logit model with the sample of inter-state migrants. We do not include non-migrants in the analysis because doing so poses conceptual problems with key variables such as "distance" and "migrant stock" and with the treatment of "home" as a potential destination choice. Since this section of the article is primarily concerned with describing the migrants' choices of destinations rather than their selection into migration, we believe that concentrating on the migrants

simplifies the analysis and the interpretation of variables in a way that is helpful.[12]

Several variables comprise X_{ij}, starting with those that may have influenced expected employment opportunities and earnings. We construct a measure of post-1910 labor demand growth (B_j), following Timothy Bartik (1991). In Bartik's specification, employment growth in each state is predicted 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 then summing across industries within states.

$$B_j = \Sigma_{l=1}^{L} e_{jl} \times g_l.$$

The Bartik-style measure gauges total employment growth in each state without utilizing ex post state-specific labor force growth rates, which are endogenous to migration. We also estimate specifications that include the percentages of the 1910 labor force employed in agriculture and manufacturing separately rather than the Bartik measure, which combines information over all industries. In the specifications with agriculture and manufacturing employment variables, we also include a control variable for state population size because larger states are likely to attract more migrants all else equal.

The vector X_{ij} also includes race-specific income estimates for each state circa 1910. These measures combine state-level estimates of real income per worker from Kris James Mitchener and Ian W. McLean (1999) with race-specific adjustment factors for each state that are calculated from the 1940 census microdata (the first census year with wage data).[13] In essence, the output-per-worker benchmark in each state circa 1910 is scaled up or down depending on the ratio of black or white men's wages to all wages in that state in the 1940 census. The Online Appendix provides more detail.

---

[12] We use the multinomial logit where the emphasis is on individual-specific variables and the conditional logit where the emphasis is on destination-specific variables. A model that combines individual and destination-specific variables in the same framework, as implemented in Stata's *asclogit* command, can be estimated successfully only after stripping out nearly all the personal background variables. We have also explored nested logit models in which the census regions and the home state serve as nests.

[13] Implicitly, we assume that black and white men's wages relative to state averages were not strongly affected by the migration itself. The best relevant evidence is from Boustan (2009), which finds a modest negative effect (approximately 7 percent) of southern migrants on black wages in the North between 1940 and 1970, when the volume of migration was larger than before 1930. On this basis, we expect that migration between 1910 and 1930 had relatively small effects on overall wage levels.

Variables related to the cost of migration are also in X_{ij}. We calculate the log distance from each individual's county of residence in 1910 to each potential destination state to capture relocation costs that are proportional to (log) distance. In addition, we calculate the share of all people born in person i's home state who resided in state j in 1910, separately by race, using the 1910 IPUMS (Ruggles et al. 2010). These pre-1910 migration rates help account for the influence of pre-existing and race-specific relationships between states including, but not limited to, networks that facilitate migration by providing a cultural home and assistance with finding employment and housing.

A regional dummy variable for "Non-South" captures the influence of region-specific amenities, some of which may have differentially influenced decisions of black and white migrants, such as more secure civil rights or less rigid social segregation. The coefficient reflects migration above or below what would be expected on the basis of the economic variables included in X_{ij}. This is a simple way to test whether areas outside the South were especially attractive to southern migrants, conditional on other X variables. Of course, identifying this border effect relies upon the sample's inclusion of both intra-regional and inter-regional migrants, and identifying a race-specific border effect relies upon having both white and black migrants. The theme of escaping the South is prominent in many narrative descriptions of African Americans' motives for inter-regional migration (e.g., Wilkerson 2010), but others have noted that racial oppression long preceded the onset of the Great Migration (Higgs 1976; Vickery 1977). Whether areas outside the South were especially attractive to black migrants and whether that influence can be detected after controlling for economic conditions are open empirical questions. In subsequent analysis, we estimate separate coefficients for the Northeast, Midwest, and West (rather than treating the Non-South as a single region) or with a full set of destination fixed effects.

The baseline analysis also includes a variable for each state's level of urbanization to see whether migrants were drawn to highly urbanized states for reasons that are not captured by other independent variables. Robustness checks with different specifications and functional form assumptions are discussed later and in further detail in the Online Appendix.

CONDITIONAL LOGIT RESULTS

The baseline results are presented in Table 4. Columns 1 to 3 pertain to the sample of white men, columns 4 to 6 pertain to black men, and columns 7 to 9 pool the samples and report coefficients on X variables

interacted with a race dummy variable (black=1) to highlight differences. Models in columns 3, 6, and 9 include destination-state fixed effects. The presence of state fixed effects limits the set of coefficients we can identify to those that vary across individuals within potential destinations, which means we cannot identify most of the X variables' coefficients in this specification. Nonetheless, because distance and migrant-stock variables do vary within destination categories across individuals, their coefficients are identified, and the specification provides a useful robustness check.

A positive logit coefficient indicates that an increase in that variable for a particular state is associated with an increase the probability of migration to that state. Interpreting the magnitude of logit coefficients is not straightforward, and therefore we present some counterfactuals to illustrate the results. For reference, marginal effects for each variable for each destination state and race are reported in the Online Appendix.

Across all specifications and both races, migrants responded negatively to distance and positively to pre-existing stocks of migrants from the same state. Black migrants appear to have been more strongly deterred by distance than whites, and this difference is statistically significant even when controlling for pre-existing migrant stocks and destination fixed effects. The relatively strong response to distance may reflect southern blacks' lower average levels of wealth and educational attainment, which could affect their access to information about distant opportunities and their ability to finance long-distance travel. It is noteworthy, however, that splitting the black sample into literate and illiterate men (in 1910) does not reveal a statistically significant difference in their migration behavior with respect to distance (Online Appendix Table 8), nor does splitting the sample by 1910 homeownership status. For perspective on the magnitude of the distance coefficients in columns 1 and 4 of Table 4, the results suggest that if southern migrants had been located one standard deviation farther away from Ohio (all else held constant), then the share of migrants going to Ohio would have declined by 3 percentage points for whites and 4 percentage points for blacks, relative to a base of 8 and 7 percentage points respectively.[14] The average "marginal effect" over all destination states (equally weighted) is –1.15 percentage points for whites and –1.43 points for blacks for a one-standard deviation change in log distance. The empirical importance of distance suggests

[14] These figures are calculated as the difference between the predicted share of migrants choosing Ohio in our base regression and, given the parameter estimates, the predicted share choosing Ohio when distance to Ohio is increased by one standard deviation (0.64 in log points). Since the model is non-linear, the effect is different for every state.

TABLE 4
MIGRANT SORTING, CONDITIONAL LOGIT COEFFICIENTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Pooled, Coefficients on Race Interaction	(8) Pooled, Coefficients on Race Interaction	(9) Pooled, Coefficients on Race Interaction
	White (A)	White (B)	White (C)	Black (A)	Black (B)	Black (C)	(A)	(B)	(C)
Cost variables									
Log distance	−1.27*** (0.0278)	−1.23*** (0.0285)	−1.56*** (0.0339)	−1.78*** (0.0728)	−1.73*** (0.0738)	−1.84*** (0.0900)	−0.512*** (0.0759)	−0.500*** (0.0772)	−0.275*** (0.093)
Migrant stock	0.193*** (0.00799)	0.201*** (0.00877)	0.182*** (0.0137)	0.193*** (0.0251)	0.198*** (0.0252)	0.157*** (0.0312)	−0.000375 (0.0257)	−0.00330 (0.0265)	−0.0256 (0.0336)
Labor market variables									
Log average income	1.40*** (0.0859)	1.10*** (0.0955)	— —	0.556*** (0.139)	0.0478 (0.180)	— —	−0.848*** (0.159)	−1.05*** (0.201)	— —
Log labor demand	0.436*** (0.0180)	— —	— —	0.789*** (0.0377)	— —	— —	0.352*** (0.0413)	— —	— —
Percent manufacturing	— —	−0.0259*** (0.00284)	— —	— —	0.0101** (0.00494)	— —	— —	0.0360*** (0.00560)	— —
Percent agriculture	— —	−0.0271*** (0.00280)	— —	— —	−0.0281*** (0.00614)	— —	— —	−0.000976 (0.00667)	— —

Region and other control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Non-South	−0.452***	−0.412***	—	−0.415***	−0.291**	—	0.0369	0.121	—
	(0.0465)	(0.0474)	—	(0.126)	(0.127)	—	(0.134)	(0.136)	—
Urban	−0.00842***	−0.0151***	—	0.00587***	−0.0111***	—	0.0143***	0.00404	—
	(0.00108)	(0.00172)	—	(0.00182)	(0.00360)	—	(0.00214)	(0.00400)	—
Log population	—	0.503***	—	—	0.879***	—	—	0.376***	—
	—	(0.0207)	—	—	(0.0396)	—	—	(0.0451)	—
Pseudo R^2	0.21	0.21	0.26	0.26	0.26	0.29	0.22	0.22	0.27
N	7,498	7,498	7,498	2,114	2,114	2,114	9,612	9,612	9,612

* = Significant at the 90 percent level.
** = Significant at the 95 percent level.
*** = Significant at the 99 percent level.
Notes: The sample includes inter-state migrants from the linked dataset of census records. Columns 1–3 include white men. Columns 4–6 include black men. Columns 7–9 report interaction terms from a pooled regression with each regressor interacted with a race dummy (1=black). Specifications A and B differ in the inclusion/exclusion of labor demand, log population, percent manufacturing and percent agriculture variables only. Specification C includes potential destination state fixed effects. Distance refers to county of resident in 1910 to population-weighted center of each potential destination state. Migrant stock refers to the share of persons born in state *i* who are residing in state *j* in 1910 (calculated with IPUMS, Ruggles et al. 2010). The labor demand variable follows Bartik (1991) and uses 1910 industry composition at the state level to form a weighted average of national-level industry-specific growth rates. Log average income combines information from Mitchener and McLean (1999) and the 1940 census microdata (Ruggles et al. 2010), as described in the text and Appendix. Non-South equals one for destination states outside the South, including Maryland, Delaware, and Washington, DC. Percent of labor force in manufacturing and agriculture, percent urban, and total population in 1910 are from Haines (2010). Standard errors, clustered at the county of origin, are in parentheses.
Sources: Data are from the linked sample of census records, as described in the text and Appendix.

that improvements in transportation and information networks played an important role in facilitating internal migration and integrating U.S. labor markets.

The black and white coefficients on the pre-existing migrant stock variable are positive, similar across specifications, and not statistically different from one another in columns 7 to 9. Thus, responsiveness to migrant networks measured in this manner did not distinguish blacks' migration behavior from that of whites. However, blacks and whites' pre-existing migrant stocks were distributed differently. Therefore, networks might still have been an important determinant of differences in flows.[15] The results are fairly similar in specifications with destination fixed effects (columns 3 and 6), suggesting that the migrant stock coefficients in other specifications are not simply picking up destination-specific unobservables that had drawn previous migrants. For some states, especially those that had relatively small stocks of previous southern migrants, plausibly sized changes in the migrant-stock variable do not imply large changes in the predicted flow of migrants between 1910 and 1930. In California, for example, increasing the black migrant-stock variable's value to equal that of whites (within each state of origin) would raise California's share of black migrants from 1.8 to 1.9 percent. This accounts for little of the sizable racial difference in migration to California (5 percentage points). However, in some other states, the implications of scaling up the stock of previous migrants are larger. Increasing the black variable's value to equal that of whites for Texas would raise Texas's share of black migrants from 3.8 to 7.8 percent (compared to a share of 10.3 percent for whites). The average marginal effect of a one standard-deviation change in pre-existing migrant stock over all states is 0.45 percentage points for whites and 0.33 points for blacks.

The results in columns 1 and 4 of Table 4 indicate that white and black migration was strongly correlated with exogenous variation in aggregate labor demand growth. The positive response to this variable is no great surprise because booming labor markets are commonly cited as motivation

[15] Stuart and Taylor (2014) use data on place of birth (1916–1936 birth cohorts) and place of residence (after 1976) from Medicare and Social Security records to examine the role of local networks in black and white inter-regional migration, primarily in the post-1940 period. They find that networks were more important for black inter-regional migrants from the Deep South than for white inter-regional migrants from the Great Plains or the Deep South. Although all results point towards networks being important for blacks, their results contrast with ours regarding the comparison of blacks and whites. Whether the contrast is due to our relatively rough proxy for networks or differences in econometric methods, sample coverage, definitions of migrants, or changes over time will require more research. Also see Chay and Munshi (2012) on migration networks.

for long-distance migration and, in general, large states tend to have large changes in employment and, therefore, attract large shares of migrants. It is interesting that black men appear to have been highly responsive to this signal relative to white men (column 7), perhaps reflecting the relative importance of finding employment quickly upon arrival or extra weight given to this signal as a correlate of changing employment opportunities for blacks specifically (i.e., relative to previous discriminatory practices in northern and western labor markets). For perspective on the coefficient, if Tennessee's labor demand had grown one standard deviation faster, the estimates suggest that its share of black migrants would have increased by 5 percentage points relative to a base of 5 percent; for whites, the predicted increase is 2 percentage points relative to a base of 4 percent. Given a one-standard-deviation change in labor demand, the average marginal effect over all potential destination states is 1.10 percentage points for whites and 2.27 points for blacks.

In columns 2, 5, and 8, we enter separate variables for manufacturing and agricultural employment shares in 1910, rather than aggregate employment growth, to provide a different perspective on the economic characteristics of states that may have influenced migration patterns. Black men were inclined to select states with high levels of manufacturing employment, but disinclined to select agricultural states, all else the same. It is notable that the coefficient on manufacturing is positive even when a "Non-South" dummy variable, urban share, and average income are included as independent variables in the specification. In other words, the positive manufacturing coefficient for blacks suggests that the sector itself was an important determinant of black migration patterns. For white men, both the manufacturing and agricultural coefficients are negative in column 2, suggesting that they tended to seek residence in states with higher levels of employment in "other" industries, such as construction, mining, trade, transportation, and services.

We also estimate models that distinguish among parts of the South that were differentially affected by the spread of the boll weevil. We defined three groups of states depending on their cotton intensity and the boll weevil's time of arrival. We considered states "cotton intensive" if at least 20 percent of total crop value came from cotton production in 1910 (Haines 2010). Based on the map from Walter Hunter and B.R. Coad (1923), we coded Texas, Arkansas, Louisiana, and Mississippi as "early boll weevil" states. They were cotton intensive and received the boll weevil prior to our window of observation (before 1910). We coded Alabama, Georgia, South Carolina, North Carolina, and Oklahoma as "late boll weevil" states. They were cotton intensive and received the boll

weevil after 1910, within our window of observation. The third group is comprised of Tennessee, Kentucky, Virginia, West Virginia, and Florida, which were not cotton intensive.[16]

If the boll weevil imparted a long-lasting negative productivity shock that is not captured by control variables and was not neutralized by cotton price changes or the redistribution of economic activity to other crops or sectors, then we would expect to see less migration to cotton-intensive places in our sample, ceteris paribus. We find that relative to states where cotton was less important (the omitted category), blacks and whites were both less likely to move to cotton-intensive states (with the exception of blacks in specification B), consistent with a negative and persistent productivity shock across cotton-producing regions.[17] Coefficients are more negative for the "late" boll weevil states than for the "early" boll weevil states, but the differences between "late" and "early" state groups are not statistically significant.

In Table 4, the coefficients on the average income variable, which varies by destination state and race, are positive but somewhat weaker in specifications that include controls for industrial composition. The coefficients in columns 1 and 4 imply that a one-standard-deviation increase in income in Pennsylvania would have raised its share of white migrants by about 2.6 percentage points and its share of black migrants by about 1.7 percentage points, relative to base shares of 3 and 8 percent, respectively. Columns 7 and 8 suggest that white migrants were more responsive than blacks to variation in pre-war income. It is possible that whites, perhaps because they had higher levels of education or access to more information, were better informed than blacks about wages in distant states and, therefore, more responsive to the existing variation. But again, there is no evidence that literate blacks were more responsive than illiterate blacks to variation in average income (Online Appendix Table 8). Another possibility is that the estimates of race-specific income circa 1910 are a better proxy for whites' expected earnings opportunities after 1910 than for blacks.

For both white and black men in columns 1, 2, 4, and 5 of Table 4, the coefficient on the Non-South variable is negative. Taken at face value,

---

[16] Results change only slightly if we shift Tennessee (where cotton was 15 percent of crop value) into the "late boll weevil" category. Note that states were not neatly divided according to the boll weevil's presence in 1910. Our coding is approximate and based on visual inspection of Hunter and Coad's map (1923).

[17] Results are reported in the Online Appendix Table 5. When the data are pooled, black-white differences in attraction to the early and late boll weevil states are statistically insignificant. Results are qualitatively similar when samples are restricted to 1910 farm residents (not shown).

neither white nor black southern migrants were especially attracted to the Non-South, conditional on other X variables.[18] There is some concern that controlling for pre-existing migrant stocks may absorb some of the attraction of regional amenities to the extent that previous migrants responded to those same amenities. However, omitting the migrant-stock variable from the specifications in columns 4 and 5 has little effect on the Non-South coefficient for black migrants (not shown). In columns 7 and 8, the coefficients on Non-South interacted with race (black=1) are not statistically significant, which again runs counter to some of the literature's emphasis on the idea that black migrants were especially motivated to escape the South (beyond that which is embedded in the other X variables).[19]

A more nuanced picture emerges by separately estimating coefficients for Northeast, Midwest, and West (expressed relative to South), instead of just "Non-South." Estimates of these coefficients are located in Table 5. Black migrants were significantly more attracted to the Northeast and Midwest than whites conditional on the other X variables, but they were significantly less likely to move to the West. Thus, the "Non-South" variable in Table 4 tends to mask substantial unexplained differences in blacks' and whites' choices of sub-region outside the South. These residual differences in migration might reflect differences in labor market discrimination, civil liberties, or social norms across northern and western locations after 1910, a hypothesis that merits further research. In this interpretation, white migrants serve as a control group to capture the influence of unobserved, race-neutral subregional characteristics, so that black-white *differences* might then be interpreted as evidence of race-specific factors that vary over subregions. Note that for blacks (columns 3 and 4) the coefficients for Midwest and Northeast are still negative (expressed relative to the South), but they are smaller in magnitude than the negative coefficients for whites.

Returning to Table 4's results, in columns 1 and 4, the coefficient on "urban" is negative for white men but positive for black men. The coefficients suggest small effects on migration patterns. Raising Mississippi's

---

[18] Vigdor (2002) finds that black migrants to cities were more likely to choose non-southern locations in the 1940 census cross-section. His conditional logit framework differs from ours in several ways. He uses census cross-sections in 1940 and 1970, a different set of covariates, cities as destination choices, and a sample of interstate migrants who resided in cities in 1940 (omitting interstate migrants who do not reside in cities).

[19] In a specification that expresses variables in levels rather than logs, the coefficients on "Non-South" are consistent with the hypothesis that blacks were more attracted to the North than whites, but the coefficient for blacks is still non-positive (the white coefficient is larger in magnitude and negative). See the Online Appendix Table 4.

TABLE 5
MIGRANT SORTING, CONDITIONAL LOGIT COEFFICIENTS ON REGIONAL
INDICATORS

	(1)	(2)	(3)	(4)	(5) Pooled, Coefficients on Race Interaction	(6) Pooled, Coefficients on Race Interaction
	White (A)	White (B)	Black (A)	Black (B)	(A)	(B)
Regional Indicators (South is omitted)						
Northeast	−1.70*** (0.100)	−1.87*** (0.105)	−1.02*** (0.172)	−1.14*** (0.177)	0.684*** (0.200)	0.733*** (0.208)
Midwest	−0.830*** (0.0573)	−0.928*** (0.0622)	−0.513*** (0.130)	−0.456*** (0.136)	0.318** (0.142)	0.472*** (0.151)
West	1.054*** (0.0807)	1.224*** (0.0863)	−0.0580 (0.215)	0.381* (0.219)	−1.11*** (0.227)	−0.843*** (0.233)
Pseudo R^2	0.24	0.24	0.26	0.26	0.25	0.25
N	7,498	7,498	2,114	2,114	9,612	9,612

Notes: All regressions also include corresponding covariates reported in Table 4, with the exception of the "Non-South" indicator. Regional definitions follow standard census regions with the exception of Delaware, Maryland, and Washington DC, which we classify as "Northeast" rather than "South." See notes to Table 4 for more details.
Sources: Data are from the linked sample of census records, as described in the text and Appendix.

urban share by one standard deviation (22 percentage points) would raise its share of black migrants by only 0.5 percentage points, relative to a base of 3.5 percent. Because urbanization is highly correlated with manufacturing and agricultural employment, we do not emphasize the coefficient on urban in columns 2 and 5.

In sum, the article's earlier analysis found that it was not possible to account for differences in white and black migration patterns with the individuals' background characteristics. Instead, the different patterns must have resulted from differences in perceived opportunities and migration costs across space (X) and differences in how men responded to those opportunities and costs (γ). To gauge how much of the black-white differences in destination choice can be accounted for by distance and migrant stock variables, we apply the parameters of the black conditional logit model with destination fixed effects to all men in the sample (black and white), and then apply the parameters of the white model to all

men.[20] This yields two counterfactual distributions of migrants with the same underlying X values. The dissimilarity index of the two counterfactual distributions is approximately 0.20, substantially less than the unadjusted index value of 0.28. From this perspective, equalizing the migrant-stock and distance variables across black and white migrants reduces the dissimilarity index by about one-fourth. The remaining differences in migration patterns reflect differences in race-specific wages across states (absorbed by race-specific destination fixed effects) and model coefficients (the γ's), which are difficult to interpret definitively. Nonetheless, we believe that additional research could isolate and test specific hypotheses and interpretations (e.g., whether the pattern of black-white differences in destination fixed effects correlates with a measure of local labor market discrimination, perhaps following Sundstrom 2007).

CONCLUSIONS

To study migration patterns during the first decades of the Great Migration, we built a new dataset of more than 26,000 southern men that we linked from the 1910 to the 1930 census. The dataset is novel in that it observes the same men before and after the start of the Great Migration. The scope of analysis is novel in that it includes both white and black southerners and it incorporates information about those who moved within the South as well as those who left the region.

First, we study selection into inter-state and inter-regional migration. There is some evidence of positive selection into migration among both whites and blacks in terms of occupational status, and it is clear that non-farm residents (in 1910) were more likely to move across state and regional lines than farm residents. Overall, however, the differences between migrants and non-migrants were small within race categories, even for those moving inter-regionally. In this sense, participation in internal migration by southern men was remarkably widespread after 1910.

Second, we examine migration patterns between origins and destinations and ask whether individual and local background characteristics account for differences in black and white migration choices. Although

---

[20] Note that distance and migrant stock variables differ by race across destinations, and their coefficients are identified in the conditional logit model with destination fixed effects. The "all white" and "all black" counterfactuals are estimated with a pooled sample, thus the same underlying distributions of distance and migrant stock variables.

there was substantial overlap in black and white migrants' choices of destinations, there were also notable differences. In OLS and multinomial logit frameworks, we find that on net a surprisingly small amount of the black-white differences in migration patterns can be accounted for by differences in background characteristics. In other words, to the extent that black and white migration patterns differed, those differences were not, in general, a simple reflection of where the men started geographically or their personal circumstances circa 1910. Rather, observationally similar men made different location choices depending on their race.

Third, given the findings above, we measure black and white migrants' responsiveness to variation in the characteristics of potential destinations. Black and white men were similarly responsive to pre-existing migrant stocks, but black men were more deterred by distance than whites, more attracted to manufacturing centers, and more responsive to variation in labor demand growth. Conditional on the potential destination states' characteristics, we find only mixed evidence that black migrants were more likely than whites to leave the South. There is, however, interesting variation across areas outside the South, with blacks sorting more strongly than whites into the Midwest and Northeast and whites sorting more strongly into the West, conditional on state characteristics. Variation in the characteristics of potential destinations, such as distance and pre-existing migrant stocks, can account for a non-trivial portion of the black-white dissimilarity index in destinations, but a larger portion is associated with racial differences in responsiveness to the destinations' characteristics and in destination fixed effects. These differences invite further research.

The article's findings rely heavily on the linked dataset of census records for southern males from 1910 to 1930, and it is our hope that other scholars will find this dataset to be useful. We have endeavored to keep the analyses focused on a set of fundamental questions about the nature of internal migration in this period, but it is not difficult to imagine extensions that look at finer levels of geographic detail, exploit more information on the migrants' outcomes, or link to other historical or administrative datasets. More generally, it is clear that this kind of dataset holds great promise for historical research on migration (e.g., Ferrie 2005; Abramitzky, Boustan, and Eriksson 2012), and new technologies and initiatives are likely to make them far more common and accessible in the future. Lying just beyond the scope of this article, there is much that remains to be learned about migration during the Great Depression, the growth of black ghettos and suburbanization of whites, the interaction

of migration and intergenerational mobility, and much more. All these topics are ripe for reassessment as new datasets that follow individuals over time are brought to light.

Appendix

For each male, age 0 to 40, in the 1910 IPUMS sample (Ruggles et al. 2010), we ran two searches in the 1930 census manuscripts: one with exact last name and one with a SOUNDEX version of the last name. In both cases, we conditioned on the first three letters of the first name (or less, if shorter than three letters in length), exact state of birth, age within two years, race, and gender. We counted any individual with a unique match in the exact last name or SOUNDEX search as a successful match. We then eliminated all duplicate matches (i.e., two different individuals in 1910 linked to the same individual in the 1930 census).

From the 1930 census, we extracted detailed location of residence as well as information on literacy, employment, occupation, industry, and veteran status. Individual and household-level variables available in the linked sample are detailed in Appendix Table 1. The availability of some variables depends on an individual's age at the time of enumeration. And although some variables are technically "available" for all age groups, they are practically missing (e.g., *occupation* for 0–9 year olds).

The characteristics of males in the matched sample are compared to those of the entire 1910 IPUMS sample in Table 1 of the main text. There, we concluded that the differences are not economically significant and are unlikely to bias the study of migration patterns. Here, we also provide estimates of a linear probability model of being found in 1930 (out of those in the original IPUMS sample), conditional on observable 1910 characteristics. The results are located in Appendix Table 2 and largely echo the conclusions from Table 1 in the main text. Specifically, we estimate the probability of being located in the 1930 manuscripts separately by race and three age categories: 0–9, 10–18, and 19–40. The age groupings correspond to the availability of school attendance, literacy, and farming occupation information in the 1910 data. We observe a slightly increased probability of being found for literate individuals, farmers, and residents of West Virginia, conclusions highlighted in discussion of Table 1, as well. On the other hand, homeownership is a statistically insignificant predictor of sample inclusion conditional on other factors observable in 1910. In each case, the enhanced probability of being located in the 1930 manuscripts is small: 0.78–2.4 percentage points in the case of literacy and 2.0–2.9 percentage points for farming occupations. Residence in West Virginia in 1910 raises the probability of being located in 1930 by as much as 5.5 percentage points, but for a relatively small number of men (6.2 percent of white males in the sample lived in West Virginia in 1910).

In Appendix Table 3, we compare the 1930 characteristics of inter-regional migrants in the linked sample to those in the 1930 IPUMS sample. The matched sample's inter-regional migrants are individuals with a southern state of residence in 1910 and a non-southern residence in 1930. IPUMS migrants, on the other hand, are individuals with a southern *place of birth* and a non-southern residence in 1930. Thus, individuals who

APPENDIX TABLE 1
VARIABLE AVAILABILITY, 1910–1930 LINKED SAMPLE OF CENSUS RECORDS

	1910			1930
	Ages 0–9	Ages 10–18*	Ages 19–40	All Ages
Race	x	x	x	x
Owner-occupied housing status	x	x	x	x
Value of home/rent paid	x	x	x	x
Farm residence status	x	x	x	x
School attendance	x	x	x	x
Literacy		x	x	x
Marital status		x	x	x
City population	x	x	x	x
Industry of employment	x	x	x	x
Father's industry of employment	x	x		
Occupation	x	x	x	x
Father's occupation	x	x		
Employment status	x	x	x	x
Father's employment status	x	x		
State of residence	x	x	x	x
County of residence	x	x	x	x
State of birth	x	x	x	x
Age	x	x	x	x
Veteran status				x

Notes: The table reports data availability for 1910 and 1930 observations in the linked sample of census records. Some variables (e.g., father's industry of employment in 1930) are occasionally available for unchecked age groups, but usually such information is missing. (Observing father's industry would require co-residence with father in 1930.) In general, variable definitions are consistent across years. We use the IPUMS variable *occ1950* and *ind1950* for 1910 and coded the 1930 variables into the same categories. *For marital status in this column, availability pertains to ages 12–18, rather than 10–18.
Source: Linked sample of census records, as described in the text.

migrated out of the South prior to 1910 will not be included in our sample but will be included in the IPUMS sample. We cannot correct for these fundamental sample differences.[21] Nonetheless, the differences across samples are small even when they are statistically significant. The most notable exception is the share of white migrants in owner-occupied housing in the linked sample (38 percent) relative to the 1930 cross-sectional sample (33 percent). As mentioned in the main text, when comparing destination shares in the linked and cross-sectional datasets (separately for black and white inter-state migrants), the correlation in migration patterns is high. With inter-regional destination states grouped as in Appendix Table 3, the correlation is more than 0.98 for both whites and blacks.

[21] We focus on inter-regional migrants for comparison because the relatively high rate of within-South migration prior to 1910 is likely to confound comparisons for within-South migrants across the datasets.

APPENDIX TABLE 2
ESTIMATED PROBABILITY OF INCLUSION IN MATCHED SAMPLE,
BY AGE GROUP AND RACE

	White Males			Black Males		
	Age 0 to 9	10 to 18	19 to 40	Age 0 to 9	10 to 18	19 to 40
Homeownership	−0.0091	−0.017	−0.0014	0.0012	0.0044	0.001
	(0.005)	(0.007)	(0.004)	(0.010)	(0.010)	(0.007)
School attendance	—	0.0143	—	—	−0.0061	—
	—	(0.009)	—	—	(0.011)	—
Literacy	—	0.024***	0.017***	—	0.0078*	0.0081***
	—	(0.004)	(0.003)	—	(0.004)	(0.003)
City population						
City pop. <=25,000	0.0094	−0.016	0.0028	0.0056	−0.016	0.0041
	(0.008)	(0.010)	(0.007)	(0.012)	(0.014)	(0.010)
City pop. >25,000	0.0077	0.0019	0.0050	0.018	0.016	0.0067
	(0.012)	(0.013)	(0.008)	(0.018)	(0.020)	(0.012)
Father is farmer/	0.020***	0.010	0.029***	0.0044	−0.0041	0.029***
individual is farmer†	(0.007)	(0.009)	(0.006)	(0.010)	(0.013)	(0.009)
State of residence (1910)						
Alabama	0.011	0.014	−0.003	−0.025	0.0063	−0.038
	(0.015)	(0.018)	(0.014)	(0.021)	(0.023)	(0.018)
Arkansas	0.015	0.036*	0.0017	−0.0033	−0.014	−0.016
	(0.015)	(0.018)	(0.014)	(0.025)	(0.028)	(0.022)
Florida	−0.014	0.028	−0.042**	−0.017	−0.0050	−0.0053
	(0.021)	(0.025)	(0.018)	(0.028)	(0.031)	(0.022)
Georgia	−0.0021	0.017	−0.0012	−0.013	−0.012	−0.027
	(0.014)	(0.017)	(0.013)	(0.020)	(0.022)	(0.017)
Kentucky	−0.035*	−0.0098	−0.014	−0.0088	0.0025	−0.036
	(0.016)	(0.019)	(0.014)	(0.022)	(0.024)	(0.019)
Louisiana	−0.00010	0.025	−0.0058	0.016	0.005	−0.017
	(0.017)	(0.020)	(0.016)	(0.020)	(0.023)	(0.018)
Mississippi	0.016	0.021	0.014	0.031	0.037	−0.0090
	(0.014)	(0.017)	(0.013)	(0.022)	(0.024)	(0.02)
North Carolina	−0.036*	0.0011	0.010	0.0019	0.0063	−0.012
	(0.018)	(0.021)	(0.016)	(0.020)	(0.023)	(0.019)
Oklahoma	−0.012	0.011	−0.018	−0.035	−0.00070	−0.0026
	(0.012)	(0.015)	(0.011)	(0.021)	(0.024)	(0.018)

APPENDIX TABLE 2 (CONTINUED)
ESTIMATED PROBABILITY OF INCLUSION IN MATCHED SAMPLE,
BY AGE GROUP AND RACE

	White Males			Black Males		
	Age 0 to 9	10 to 18	19 to 40	Age 0 to 9	10 to 18	19 to 40
State of residence (1910)						
South Carolina	0.011	0.0072	0.016	0.040	0.029	–0.0044
	(0.013)	(0.016)	(0.012)	(0.032)	(0.034)	(0.025)
Tennessee	0.019	0.017	–0.025	–0.027	0.016	–0.026
	(0.014)	(0.017)	(0.013)	(0.038)	(0.044)	(0.032)
Texas	0.026	0.0046	–0.016	–0.0061	0.016	–0.022
	(0.014)	(0.016)	(0.013)	(0.025)	(0.027)	(0.021)
West Virginia	0.028	0.055***	0.041***	0.068	–0.11	–0.037
	(0.015)	(0.018)	(0.013)	(0.060)	(0.069)	(0.037)
N	25,791	19,565	32,716	9,214	7,331	11,665
Adjusted R^2	0.0060	0.0032	0.0031	0.0017	0.00050	0.0025

* = Significant at the 95 percent level.
** = Significant at the 97.5 percent level.
*** = Significant at the 99 percent level.
Notes: Omitted categories in the regressions are No City (City Population) and Virginia (State of Residence). The model also includes age fixed effects (not shown).
†Indicator for father's farming occupation (ages 0–18) or self (ages 19+).
Sources: The linked sample is created by taking the 1910 IPUMS sample of white and black men, age 0–40, who reside in the South and searching for these men in the 1930 census manuscripts. The text contains details on sample construction. The IPUMS data are from Ruggles et al. (2010).

APPENDIX TABLE 3
COMPARISON OF LINKED AND FULL SAMPLES, SOUTHERN MIGRANTS
TO THE NON-SOUTH, 1930

	Black Migrants			White Migrants		
	Matched Sample of Migrants (Resided in South in 1910)	IPUMS Sample of Migrants (Born in South)	p-value of Difference	Matched Sample of Migrants (Resided in South in 1910)	IPUMS Sample of Migrants (born in South)	p-value of Difference
Distribution of state of residence (1930)						
Arizona	0.01	<0.01	0.77	0.02	0.03	0.12
California	0.03	0.04	0.81	0.14	0.15	0.36
Colorado	0.00	<0.01	0.05	0.02	0.02	0.38
Connecticut/Rhode Island	0.01	0.01	0.57	<0.01	<0.01	0.79
Delaware	<0.01	<0.01	0.93	<0.01	<0.01	0.56
D.C./Maryland	0.07	0.08	0.33	0.05	0.06	0.08
Idaho/Oregon/Washington	<0.01	<0.01	0.42	0.05	0.05	0.23
Illinois	0.14	0.15	0.75	0.09	0.08	0.71
Indiana	0.04	0.04	0.63	0.06	0.07	0.00
Iowa	0.01	0.00	0.13	0.01	0.01	0.60
Kansas	0.02	0.02	0.88	0.05	0.03	0.00
Maine/NH/Vermont	0.00	0.00	—	<0.01	<0.01	0.03
Massachusetts	<0.01	<0.01	0.08	0.01	0.01	0.51
Michigan	0.10	0.08	0.10	0.06	0.06	0.30
Minnesota	<0.01	<0.01	0.40	0.01	<0.01	0.04
Missouri	0.08	0.06	0.03	0.08	0.07	0.07
Montana/Wyoming	0.00	<0.01	0.53	0.01	0.01	0.96
Nebraska	<0.01	<0.01	0.78	0.01	0.01	0.31
Nevada/Utah	<0.01	0.00	0.02	<0.01	0.01	0.33

APPENDIX TABLE 3 (CONTINUED)
COMPARISON OF LINKED AND FULL SAMPLES, SOUTHERN MIGRANTS
TO THE NON-SOUTH, 1930

	Black Migrants			White Migrants		
	Linked Sample of Migrants (Resided in South in 1910)	IPUMS Sample of Migrants (Born in South)	p-value of Difference	Linked Sample of Migrants (Resided in South in 1910)	IPUMS Sample of Migrants (born in South)	p-value of Difference
Distribution of state of residence (1930)						
New Jersey	0.06	0.08	0.10	0.02	0.02	0.69
New Mexico	<0.01	<0.01	0.32	0.03	0.03	0.41
New York	0.12	0.13	0.45	0.05	0.04	0.24
North/South Dakota	0.00	<0.01	0.66	<0.01	<0.01	0.60
Ohio	0.14	0.12	0.14	0.17	0.17	0.95
Pennsylvania	0.15	0.17	0.08	0.06	0.05	0.54
Wisconsin	0.01	<0.01	0.25	0.01	0.01	0.12
Personal characteristics (1930)						
Mean age	35.7	36.3	0.05	36.4	36.8	0.06
In owner-occupied housing	0.18	0.18	0.71	0.38	0.33	0.00
Literate	0.95	0.95	0.94	0.99	0.99	0.17
Veteran	0.13	0.11	0.12	0.18	0.18	0.44
Job characteristics (1930)						
Farmer	0.01	0.01	0.06	0.09	0.09	0.10
Employed	0.84	0.82	0.06	0.91	0.90	0.10

Notes: The IPUMS sample includes all southern-born black males observed outside of the South in the 1930 census. The text contains more details on the construction of the linked sample of census records. Values of "0.00" in the above table represent true zeroes. Values between 0.00 and 0.05 are represented as "<0.01". Sample sizes for black migrants in our matched sample and in the IPUMS sample are 1,103 and 5,675, respectively. For white migrants, sample sizes are 3,654 and 8,828, respectively.

Sources: Data are from Ruggles et al. (2010) and the linked dataset of census records described in the text.

REFERENCES

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, no. 5 (2012): 1832–56.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Have the Poor Always Been Less Likely to Migrate? Evidence from Inheritance Practices during the Age of Mass Migration." *Journal of Development Economics* 102 (2013): 2–14.

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

Atack, Jeremy. "On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited." *Journal of Economic History* 73, no. 2 (2013): 313–38.

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

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

Baker, Richard B. "From the Field to the Classroom: The Boll Weevil's Impact on Education in Rural Georgia." Unpublished Working Paper, 2015.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, et al. "The Impact of the Great Migration on Mortality of African-Americans: Evidence from the Deep South." *American Economic Review* 105, no. 2 (2015): 477–503.

Bodnar, John, Roger Simon, and Michael P. Weber. *Lives of Their Own: Black, Italians, and Poles in Pittsburgh, 1900–1960*. Champaign, IL: University of Illinois Press, 1982.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. "Self Selection and Internal Migration in the United States." *Journal of Urban Economics* 32, no. 2 (1992): 159–85.

Boustan, Leah Platt. "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970." *Journal of Economic History* 69, no. 3 (2009): 755–82.

Boustan, Leah Platt, Matthew E. Kahn, and Paul W. Rhode. "Moving to Higher Ground: Migration Responses to Natural Disasters in the Early Twentieth Century." *American Economic Review: Papers and Proceedings* 102, no. 3 (2012): 238–44.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. "Migration with Endogenous Moving Costs." *American Economic Review* 86, no. 4 (1996): 909–30.

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

Collins, William J. "When the Tide Turned: Immigration and the Delay of the Great Black Migration." *Journal of Economic History* 57, no. 3 (1997): 607–32.

Collins, William J. "African-American Economic Mobility in the 1940s: A Portrait from the Palmer Survey." *Journal of Economic History* 60, no. 3 (2000): 756–81.

Collins, William J., and Robert A. Margo. "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, 107–54. New York, NY: North-Holland, 2006.

Collins, William J., and Marianne H. Wanamaker. "Selection and Income Gains in the Great Migration of African Americans." *American Economic Journal: Applied* 6, no. 1 (2014): 220–52.

Davies, Paul S., Michael J. Greenwood, and Haizheng Li. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science* 41, no. 2 (2001): 337–60.

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

Ferrie, Joseph P. "History Lessons: The End of American Exceptionalism? Mobility in the United States since 1850." *Journal of Economic Perspectives* 19, no. 3 (2005): 199–215.

Foote, Christopher L., Warren C. Whatley, and Gavin Wright. "Arbitraging a Discriminatory Labor Market: Black Workers at the Ford Motor Company, 1918–1947." *Journal of Labor Economics* 21, no. 3 (2003): 493–532.

Gregory, James N. *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*. Chapel Hill: University of North Carolina Press, 2005.

Grogger, Jeffrey, and Gordon H. Hanson. "Income Maximization and the Selection and Sorting of International Migrants." *Journal of Development Economics* 95, no. 1 (2011): 42–57.

Grossman, James R. *Land of Hope: Chicago, Black Southerners, and the Great Migration*. Chicago: University of Chicago Press, 1989.

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

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

Hatton, Timothy J., and Jeffrey G. Williamson. "What Fundamentals Drive World Migration?" NBER Working Paper 9159. Cambridge, MA, 2002.

Higgs, Robert. "The Boll Weevil, the Cotton Economy, and Black Migration 1910–1930." *Agricultural History* 50, no. 2 (1976): 335–50.

———. "Accumulation of Property by Southern Blacks Before World War I." *American Economic Review* 72, no. 4 (1982): 725–37.

Hornbeck, Richard, and Suresh Naidu. "When the Levee Breaks: Black Migration and Economic Development in the South." *American Economic Review* 104, no. 3 (2014): 963–90.

Hunter Walter David, and B. R. Coad. *The Boll Weevil Problem: Farmers' Bulletin No. 1329*. Washington, D.C.: U.S. Department of Agriculture, 1923.

Killian, Lewis M. "The Adjustment of Southern White Migrants to Northern Urban Norms." *Social Forces* 32, no. 1 (1953): 66–69.

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

Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode. "The Impact of the Boll-Weevil 1892–1932." *Journal of Economic History* 69, no. 3 (2009): 685–718.

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

Logan, Trevon. "Health, Human Capital and African American Migration before 1910." *Explorations in Economic History* 46, no. 2 (2009): 169–85.

Long, Jason, and Joseph Ferrie. "Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." *American Economic Review* 103, no. 4 (2013): 1109–37.

Maloney, Thomas N. "Migration and Economic Opportunity in the 1910s: New Evidence on African-American Occupational Mobility in the North." *Explorations in Economic History* 38, no. 1 (2001): 147–65.

Margo, Robert A. "Accumulation of Property by Southern Blacks Before World War I: Comment and Further Evidence." *American Economic Review* 74, no. 4 (1984): 768–76.

———. *Race and Schooling in the South 1880–1950: An Economic History*. Chicago: University of Chicago Press, 1990.

McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–42. New York: Academic Press, 1974.

McKenzie, David, and Hillel Rapoport. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico." *Journal of Development Economics* 84, no. 1 (2007): 1–24.

Mitchener, Kris James, and Ian W. McLean. "U.S. Regional Growth and Convergence, 1880–1980." *Journal of Economic History* 59, no. 4 (1999): 1016–42.

Moretti, Enrico. "Local Labor Markets." In *Handbook of Labor Economics*, Vol. 4B, edited by Orley Ashenfelter and David Card, 1237–313. Amsterdam: North Holland, 2011.

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

Naidu, Suresh. "Recruitment Restrictions and Labor Markets: Evidence from the Postbellum U.S. South." *Journal of Labor Economics* 28, no. 2 (2010): 413–45.

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

Roback, Jennifer. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90, no. 6 (1982): 1257–78.

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

Roy, A. D. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, no. 2 (1951): 135–46.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, et al. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Scroggs, Wm. O. "Interstate Migration of Negro Population." *Journal of Political Economy* 25, no. 10 (1917): 1034–43.

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

StataCorp. "Post-Estimation Commands for Multinomial Logit." *Stata Base Reference Manual: Release 11*. College Station, TX: StataCorp LP, 2009.

Stuart, Bryan A., and Evan J. Taylor. "Social Interactions and Location Decisions: Evidence from U.S. Mass Migration." Unpublished Working Paper, 2014.

Sundstrom, William A. "The Geography of Wage Discrimination in the in the Pre-Civil Rights South." *Journal of Economic History* 67, no. 2 (2007): 410–44.

Thomas, Brinley. *Migration and Economic Growth*. New York, NY: Cambridge University Press, 1954.

Tolnay, Stewart E. "Educational Selection in the Migration of Southern Blacks, 1880–1990." *Social Forces* 77, no. 2 (1998): 487–514.

Tolnay, Stewart E., and E.M. Beck. "Black Flight: Lethal Violence and the Great Migration, 1900 to 1930." *Social Science History* 14, no. 3 (1990): 347–70.

Tolnay, Stewart E., Katherine Curtis White, Kyle D. Crowder, et al. "Distances Traveled During the 'Great Migration': An Analysis of Racial Differences Among Male Migrants." *Social Science History* 29, no. 4 (2005): 523–48.

U.S. Department of Agriculture. *Public Roads Mileage and Revenues in the Southern States, 1914*. USDA Bulletin No. 387, Office of Public Roads and Rural Engineering. Washington, DC, 1917.

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

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

Vigdor, Jacob L. "The Pursuit of Opportunity: Explaining Selective Black Migration." *Journal of Urban Economics* 51, no. 3 (2002): 391–417.

Whatley, Warren. "Getting a Foot in the Door: 'Learning,' State Dependence, and the Racial Integration of Firms." *Journal of Economic History* 50, no. 1 (1990): 43–66.

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

Woodward, C. Vann. *The Strange Career of Jim Crow*, Third Revised Edition. New York: Oxford University Press, 1974. (Original work published 1955.)

Wozniak, Abigail. "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources* 45, no. 4 (2010): 944–70.

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