

Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data

By WILLIAM J. COLLINS AND MARIANNE H. WANAMAKER

ONLINE APPENDIX

1. Sensitivity to Match Quality

Table 1 of the main text suggests that the matching process did not introduce substantial bias relative to the IPUMS cross-sectional sample. That is, the average characteristics of the matched sample are quite similar to those of the full IPUMS original sample. But a remaining concern is with the quality of the matches. Incorrectly matched individuals in the data will possess observable (and unobservable) characteristics in 1910 that are uncorrelated with their subsequent labor market outcomes, including migration status and 1930 income. This will tend to bias our evaluation of selection into migration and our measures of conditional returns to migration.

To evaluate whether there is bias from incorrect matches between census years, we re-estimate the main results in Tables 4 and 7 using only the portion of the sample that was matched between 1910 and 1930 using *exact* first and last names (i.e., omitting those matched using the SOUNDEX option from footnote 14). This stricter matching criterion should generate more accurate matches. It also reduces the sample by approximately 36 percent.

Supplemental Appendix Tables S1 and S2 repeat the regression specifications from Tables 4 and 7 for this sample subset. The results are generally similar. In columns 4a and 4b of Table S2, where the samples are small, there are some non-trivial differences compared to the original estimates, but the coefficients move in different directions for 1928-based and 1960-based estimates. We conclude that the overall results are not sensitive to this restriction and unlikely to be biased from mismatches.

2. Results with 1910 Literacy as a Control Variable

Literacy is only available for those aged 10 and over in the 1910 census, and therefore we would lose a substantial fraction of the sample by including this as a covariate in all regressions. Since it does not appear to have a substantial influence on our estimates of gains from migration, we have omitted it from our baseline regressions in the main text.

Supplemental Appendix Tables S3 and S4 replicate specifications from Tables 4 and 7 of the main text, but include individual-level literacy in 1910 as an additional control variable. The differences are negligible between Tables 4 and S3. The universe of individuals included in Table 4 of the main text is all those 20 and over in 1910, and so the universe does not change with the addition of the literacy variable which is almost always reported for those age 10 and over.

Measuring the returns to migration (presented in Table 7 of the main text) with controls for literacy (Table S4 of the Supplement) requires restricting the sample to individuals aged 10 and over in 1910 – a restriction applied across all columns for consistent samples. That reduces the sample size from 5,055 individuals in the first three columns of Table 7 to 3,412 in Table S4. We see very little difference in the unadjusted returns to migration in column 1 of both Tables 7 and S4 where the difference is only the restricted sample. In column 2, we include literacy in the vector of individual characteristics in Table S4. We see a small reduction in the estimated returns to migration between columns 1 and 2, similar to the reduction we observe in Table 7 of the main text. The estimates based on the smallest sample (brothers) in Column 4b of Table S4 diverge from those in the main text (compare to column 4b of Table 7), but remain within one standard error. Overall, we conclude that the results are not sensitive to the inclusion of 1910 literacy as a control variable.

3. Alternative Cost of Living Adjustments

As discussed in Appendix A of the main text, Williamson and Lindert (1980) provide state-level estimates of cost-of-living circa 1929. They essentially create a weighted average of cost-of-living at the state level by adjusting the Stecker data according to the share of the labor force in agriculture in each state. The adjustment reflects Koffsky's (1949) estimate of the difference between farm and city price levels in 1941. Our preferred approach, followed in the main text of the paper, uses the detailed city-level data from Stecker to assign cost-of-living indices that are more sensitive to the location of the black population. Nonetheless, it is informative to see how sensitive the results are to this adjustment.

Table S5 reports results from replicating the specifications for “real earnings scores” in Table 7, but substituting the Williamson and Lindert COL. We present the results with both COLs for easier comparison. As expected, the estimated returns to migration are somewhat larger with the Williamson and Lindert COL index than in our baseline results, probably because the Williamson and Lindert state-level index tends to understate the cost-of-living for blacks in northern cities.

4. Robustness to Using Alternative Adjustment Factors for Agricultural Workers

The first of the two income estimation methods used in Tables 4 and 7 assigns a 1928-based average income level, by industry, to each individual in the dataset. We then calculate an industry/region/employment status-specific “adjustment factor,” a ratio to scale down southern-born black earnings relative to the industry average, based on income data in the 1940 IPUMS data. For example, the adjustment factor for an employed, black, southern-resident male in agriculture would be the ratio of mean income for all black, southern-residing (and southern-born) males employed in agriculture in 1940 divided by the mean income of all agriculture employees observed in 1940.

Because agriculture was a large sector of employment for southern blacks in this period, this adjustment is important, but it also very difficult to pin down with the “wage and salary” income measure provided by the 1940 census. Farmers are not wage and salary workers, and so the ratio, which is based on relative earnings among wage and salary workers, might not be an accurate proxy for the relative earnings of all blacks in agriculture.¹ Even among wage workers, the prevalence of payment in-kind in agriculture complicates the issue.

To test the sensitivity of our migration returns estimates to changes in the agricultural worker adjustment factor, we undertake two separate robustness checks. In each case, we replace the adjustment factor for agricultural workers with a ratio based on different but reasonable alternatives. Of course, if the 1928-based average income level for the industry as a whole is misstated, our robustness checks will not address this bias.²

- (a) First, we simply move away from agriculture and calculate adjustment factors based on all non-agricultural workers, by region and employment status. So, for example, if southern-born, southern-resident, employed black workers in non-agricultural industries earned 50 percent of what all workers earned in non-agricultural jobs, then the adjustment factor is 0.5.
- (b) Second, we follow the original process for calculating the adjustment factors with agricultural wage workers from the 1940 IPUMS, but we limit the sample to those who reported receiving less than \$50 of in-kind income in 1940.³ This may insulate the ratio from bias due to differential rates of in-kind pay.

¹ These difficulties could bias the estimate of average black earnings in agriculture in either direction. For instance, the ratio of black/all earnings among agricultural wage workers could be higher than the black/all ratio among all those who work in agriculture (e.g., there are more relatively rich white farmers than relatively rich white farm laborers), which would set the ratio too high. On the other hand, if black farm workers received a larger share of income in-kind than others, then the ratio based on observed wages would be set too low.

² This issue is addressed in the text of the paper, footnote 20.

³ The specific question is: “Did this person receive income of \$50 or more from sources other than money wages or salary?” and instructions to enumerators indicate that “Yes is to be entered for each person who had an income of \$50 or more from any one or any combination of the following sources: Roomers or boarders; business profits; fees; rents; . . . income in kind paid as wages, etc.”

The results of these alternative specifications of adjustment factors are contained in Supplemental Tables S6 and S7, respectively. We replicate the results in Table 7 of the main text for both of these robustness checks. (Only the top panel of these tables will differ relative to Table 7 as only the 1928-based income assignment method is affected.) Table S6 shows that using an adjustment ratio based on *non-agricultural* workers leads to higher estimated returns to migration (i.e., the non-agricultural adjustment factor is lower than the original baseline adjustment factor), a reflection of higher relative wages in agriculture for southern-born, southern-residing blacks than in other industries.

Table S7 reports the results of the second robustness test. Despite the fact that 20 percent of agricultural workers in the 1940 IPUMS sample report receiving in-kind income of more than \$50, limiting the 1940 data to those without in-kind income does not substantially affect our estimates. We see no major change in the estimated returns to migration in Table S7 relative to those reported in Table 7 of the main text.

We see no upward bias in our estimates of migration returns based on these two alternative methods for estimating wages in the agricultural industry.

5. Sensitivity to Alternative Methods for Assigning Earnings Scores to the Unemployed

As noted in Table 5 of the main text, unemployment rates in 1930 are higher for northern migrants (16.0%) than for non-migrants (5.9%). We account for this difference in the main text by assigning earnings scores to unemployed and employed individuals separately. Each earnings score is based not only on a region, race, and occupation or industry, but also on employment status. As an example, for industry-based assignments, we assign an adjustment factor for manufacturing of 0.6715 for southern-born, northern-residing, employed black males and a factor of 0.3999 for southern-born, northern-residing, unemployed black males, where the adjustment factors depend on annual earnings observed in the 1940 census data. When using the 1960 microdata, the calculation of annual earnings conditional on employment status is straightforward. In this sense, our approach to assigning annual earnings scores to the unemployed is fully consistent with our approach to assigning scores to the employed. If the relative earnings of unemployed workers are similar in 1930, 1940 and 1960, then the assigned scores will adequately reflect differences in earnings associated with differences in employment status at the time of census enumeration. Note that this does not require that the unemployment rates be similar in 1930, 1940, or 1960, but rather that the relative annual earnings of unemployed and employed workers at the time of census enumeration is similar.

To gauge the sensitivity of our results on this issue, we have taken two alternative approaches to assigning earnings scores for unemployed workers. First, we re-estimated the baseline model from Table 7 in dollar levels and then replaced the scores with \$0 for the unemployed and re-estimated the model (compare columns 3 and 4 of Table S8 below). The estimated returns to migration fall by approximately 10 to 16 percent depending on the score assignment method (1928/1940- or 1960-based). Second, we simply set the earnings scores for the unemployed at 25 percent of the earnings for employed workers in the same reported industry or occupation and region. We think this is too low based on what we can directly observe about the annual earnings of those unemployed at the time of the 1940 and 1960 census, but it does provide additional perspective on the sensitivity of the results to this issue. We observe a reduction in the estimated real gains of approximately 5 to 8 log points relative to the baseline model. See Table S8, column 2.

For a different perspective, we gathered direct evidence on differences in weeks worked in the previous year between men who were employed or unemployed at the time of the census enumeration. The difference in weeks worked per year is relevant because the scores are based on estimates of *annual* earnings, and being unemployed at the time of census enumeration does not necessarily imply large differences in employment or earnings during the previous year. In 1910, the census inquired about weeks unemployed in the previous year. From this basis, and using the IPUMS data, it appears that black men in the North who were unemployed at the time of enumeration worked approximately 85 percent as many weeks in the previous year as those who were employed at the time of enumeration. Information on weeks worked is not available again until 1940, at the end of the Great Depression, when the ratio was approximately 65 percent.

Ultimately, it is clear that those unemployed at the time of census enumeration had lower annual earnings than those who were employed (within region and industry or occupation categories). We think that assigning the unemployed earnings scores by the same methods used to assign scores to the employed provides a reasonable basis for inference. All the same, the key results are not very sensitive to substituting score assignments for the unemployed that are deliberately too low.

6. *Decomposing Gains*

As described in Section 4 of the main text, a potential source of bias in our estimates is the possibility that northern migrants were positively selected within industry or occupation cells and that the returns to migration based on wage assignments utilizing these individuals are overstated. We have implemented a method of assigning earnings scores to the men in our sample that does not

rely directly on the observed earnings of black migrants in 1940 or 1960. We proceeded in two steps. First, we assigned scores to the black migrants (in North) that are equal to those of southern blacks in the same industry or occupation, conditional on employment status. This isolates the role of industry/occupation upgrades by migrants while omitting gains associated with the North-South difference in wage levels and avoiding reliance on the observed earnings of migrants. Next, we adjust the scores upward for black migrants according to the regional wage premium found among whites within industry/occupation and employment status cells (based on 1940 or 1960 microdata). This “adds on” a regional wage premium that is insulated from unobserved selection of black migrants, though it also omits any black-specific northern wage premium within job categories (e.g., from entering a less discriminatory market). The results are reported in Supplemental Tables S9a and S9b. The results in Table S9a indicate that job changes alone generate earnings gains of between 34 and 48 log points (about half of the total gains suggested in the baseline results). The results in Table S9b, which include a northern wage premium derived from comparisons among white workers, are comparable in magnitude to our baseline results in the main text, but typically about 20 percent lower.

7. Additional Evidence of Selection into the Migrant Stream

As an additional metric of selection into the migrant stream, we calculate the *ex ante* “return” to eventual migration status for other variables beyond earnings score as measured in 1910. The variables of interest are literacy, homeownership, big city status, and employment status, and we calculate the return to migration for each binary variable with controls for age (fixed effects) and either industry or occupation of employment in 1910. If there is unobserved selection within industry or occupation categories, it will not be observable to us using earnings scores based on industry or occupation. But if, conditional on these observable employment variables, we observe differences in other outcomes of interest, it would be evidence of within-job-category selection.

We report these “returns” to migration in Table S10. None of the four outcomes differ at a statistically significant level across eventual migrant status, conditional on age and either an occupation (column 1) or industry (column 2) fixed effect. We take this as further evidence of the limited role of unobservables to explain the large returns to inter-regional migration reported in the main text.

Supplement Table S1: Log Earnings Score Differentials in 1910 between Subsequent Migrants and Non-Migrants, Sample Limited to Those without SOUNDEX Variation

	1	2	3
<i>Panel A: Earnings score based on Lebergott, 1928</i>			
Nominal	0.167 (0.0318)	0.0531 (0.0259)	0.0213 (0.0343)
Real	0.152 (0.0305)	0.0507 (0.0262)	0.0231 (0.0345)
<i>Panel B: Earnings score based on IPUMS, 1960</i>			
Nominal	0.216 (0.0359)	0.0716 (0.0299)	0.0193 (0.0410)
Real	0.201 (0.0347)	0.0692 (0.0302)	0.0210 (0.0412)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes
1910 County fixed effects	No	No	Yes
Observations	1,305	1,305	1,305

Notes: Corresponds to Table 4 in the main text. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant), sample limited to those individuals matched between years without a SOUNDEX variation of their last name (i.e., more exact matches). Earnings are assigned according to the industry or occupation held in 1910, as described in the text. The control variables differ across the columns. Standard errors are adjusted for clustering at the household level. Column 1 has no control variables. Column 2 controls for age fixed effects, veteran status, a binary variable for blank veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All variables pertain to 1910 status except veteran status. Column 3 includes county fixed effects.

Supplement Table S2: Log Earnings Score Differentials in 1930 by Migrant Status,
Sample Limited to Those without SOUNDEX Variation

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Nominal	0.889 (0.0123)	0.866 (0.0126)	0.856 (0.0172)	0.851 (0.157)	0.804 (0.183)	0.854 (0.0234)	0.801 (0.0351)
Real	0.684 (0.0119)	0.667 (0.0122)	0.659 (0.0163)	0.665 (0.148)	0.626 (0.168)	0.661 (0.0222)	0.611 (0.0346)
<i>Panel B: Earnings score based on IPUMS, 1960</i>							
Nominal	0.897 (0.0170)	0.871 (0.0174)	0.854 (0.0226)	0.819 (0.195)	0.743 (0.216)	0.877 (0.0319)	0.796 (0.0447)
Real	0.692 (0.0167)	0.671 (0.0171)	0.656 (0.0217)	0.632 (0.191)	0.565 (0.204)	0.685 (0.0310)	0.606 (0.0443)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	3235	3235	3235	254	254	1212	1212

Notes and sources: Corresponds to Table 7 in the main text. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant), sample limited to those individuals matched between years without a SOUNDEX variation of their last name (i.e., more exact matches). All are statistically significant at the 5 percent level. Controls variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930. Cost-of-living adjustments are applied to the "real earnings" rows, based on Stecker (1937) and Koffsky (1949). See the main text and data appendix for more discussion of the cost-of-living adjustments.

Supplement Table S3: Log Earnings Score Differentials in 1910 between Subsequent Migrants and Non-Migrants, including Literacy as a Control Variable

	1	2	3
<i>Panel A: Earnings score based on Lebergott, 1928</i>			
Nominal	0.130 (0.0249)	0.0462 (0.0198)	0.0216 (0.0226)
Real	0.121 (0.0238)	0.0437 (0.0200)	0.0225 (0.0227)
<i>Panel B: Earnings score based on IPUMS, 1960</i>			
Nominal	0.159 (0.0287)	0.0513 (0.0228)	0.0149 (0.0264)
Real	0.150 (0.0277)	0.0488 (0.0230)	0.0157 (0.0265)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes
1910 County fixed effects	No	No	Yes
Observations	2,079	2,079	2,079

Notes: Corresponds to Table 4 of the main text, limited to those individuals who report literacy in 1910. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). Earnings are assigned according to the industry or occupation held in 1910, as described in the text. The control variables differ across the columns. Standard errors are adjusted for clustering at the household level. Column 1 has no control variables and is identical to Column 1 in Table 4 of the main text. Column 2 controls for age fixed effects, literacy, veteran status, a binary variable for blank veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All variables pertain to 1910 status except veteran status. Column 3 includes for county fixed effects.

Supplement Table S4: Log Earnings Score Differentials in 1930 by Migrant Status, including Literacy as a Control Variable

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Nominal	0.898 (0.0119)	0.867 (0.0122)	0.859 (0.0162)	0.677 (0.159)	0.698 (0.205)	0.877 (0.0177)	0.832 (0.0273)
Real	0.694 (0.0114)	0.669 (0.0117)	0.663 (0.0153)	0.510 (0.149)	0.515 (0.193)	0.679 (0.0167)	0.636 (0.0268)
<i>Panel B: Earnings score based on IPUMS, 1960</i>							
Nominal	0.915 (0.0165)	0.878 (0.0170)	0.869 (0.0217)	0.672 (0.225)	0.777 (0.282)	0.888 (0.0250)	0.829 (0.0346)
Real	0.711 (0.0162)	0.680 (0.0167)	0.673 (0.0208)	0.505 (0.224)	0.594 (0.283)	0.691 (0.0244)	0.633 (0.0343)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	3412	3412	3412	209	209	1935	1935

Notes and sources: Corresponds to Table 7 of the main text, restricted to those who report a literacy value in 1910. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, literacy, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930. Cost-of-living adjustments are applied to the "real earnings" rows, based on Stecker (1937) and Koffsky (1949). See the main text and data appendix for more discussion of the cost-of-living adjustments.

Supplement Table S5: Log Earnings Score Differentials in 1930 by Migrant Status, Alternative COL Adjustments

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Real – Baseline COL	0.685 (0.00950)	0.667 (0.00968)	0.661 (0.0119)	0.604 (0.0759)	0.595 (0.0935)	0.680 (0.0167)	0.636 (0.0268)
Real – Alternative COL	0.739 (0.00977)	0.718 (0.00997)	0.711 (0.0124)	0.643 (0.079)	0.643 (0.0981)	0.732 (0.0176)	0.691 (0.0274)
<i>Panel B: Earnings score based on IPUMS, 1960</i>							
Real – Baseline COL	0.694 (0.0133)	0.671 (0.0136)	0.661 (0.0161)	0.604 (0.0993)	0.592 (0.121)	0.691 (0.0243)	0.633 (0.0342)
Real – Alternative COL	0.747 (0.0135)	0.722 (0.0137)	0.712 (0.0165)	0.642 (0.0980)	0.639 (0.119)	0.743 (0.0248)	0.689 (0.0345)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	5055	5055	5055	403	403	1935	1935

Notes and sources: Corresponds to Table 7 in the main text. Cost-of-living adjustments under the baseline are based on Stecker (1937) and Kofsky (1949) as described in the data appendix to the main text. COL adjustments in the second row are based on Williamson and Lindert (1980), also described in the data appendix. See the text and data appendix for more discussion of the cost-of-living adjustments. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930.

Supplement Table S6: Log Earnings Score Differentials in 1930 by Migrant Status,
with Ag Adjustment Factors Calculated as per Option (a)

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Nominal	0.987 (0.0114)	0.959 (0.0117)	0.948 (0.0146)	0.857 (0.0925)	0.860 (0.114)	0.970 (0.0208)	0.913 (0.0326)
Real	0.781 (0.0110)	0.758 (0.0112)	0.749 (0.0139)	0.673 (0.0883)	0.666 (0.108)	0.772 (0.0197)	0.717 (0.0320)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	5055	5055	5055	403	403	1935	1935

Notes and sources: Corresponds to Table 7 of the main text. Agricultural earnings adjustment ratios based on all non-agriculture industries in 1940. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930. Cost-of-living adjustments are applied to the "real earnings" rows, based on Stecker (1937) and Koffsky (1949). See the text and data appendix for more discussion of the cost-of-living adjustments.

Supplement Table S7: Log Earnings Score Differentials in 1930 by Migrant Status,
with Ag Adjustment Factors Calculated as per Option (b)

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Nominal	0.888 (0.00973)	0.865 (0.00995)	0.856 (0.0123)	0.785 (0.0791)	0.786 (0.0977)	0.874 (0.0175)	0.829 (0.0271)
Real	0.682 (0.00943)	0.664 (0.0096)	0.658 (0.0118)	0.602 (0.0755)	0.592 (0.093)	0.677 (0.0166)	0.633 (0.0266)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	5055	5055	5055	403	403	1935	1935

Notes and sources: Corresponds to Table 7 of the main text. Agricultural earnings adjustment ratios determined using those with <\$50 in kind income in 1940 IPUMS. Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930. Cost-of-living adjustments are applied to the "real earnings" rows, based on Stecker (1937) and Koffsky (1949). See the text and data appendix for more discussion of the cost-of-living adjustments.

Supplement Table S8: Log Earnings Score Differentials in 1930 by Migrant Status, with Alternative Method for Assigning Earnings to Unemployed

	(1)	(2)	(3)	(4)
	Baseline Result (Table 7, Column 3)	25% Sensitivity	Baseline (In Levels)	\$0 Sensitivity – (In Levels)
<i>Panel A: Earnings score based on Lebergott, 1928</i>				
Nominal	0.860 (0.0124)	0.782 (0.0204)	499.4 (6.85)	433.9 (12.4)
Real	0.661 (0.0119)	0.584 (0.0202)	410.6 (7.59)	349.6 (12.8)
<i>Panel B: Earnings score based on IPUMS, 1960</i>				
Nominal	0.860 (0.0166)	0.811 (0.0218)	1800.2 (40.66)	1620.5 (454.5)
Real	0.661 (0.0161)	0.612 (0.0215)	1478.5 (43.11)	1312.6 (56.3)
Controls for personal, household and county characteristics in 1910	Yes	Yes	Yes	Yes
1910 County fixed effects	Yes	Yes	Yes	Yes
1910 Household fixed effects	No	No	No	No
Differenced dependent variable (1930-1910)	No	No	No	No
N	5055	5055	5055	5055

Note: Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). Standard errors are adjusted for clustering at the household-of-origin level. Results in Column (1) are replicated from Table 7, Column 3 of the main text. Results in Column (2) reflect assigning all unemployed individuals in the sample earnings equivalent to 25% of the employed in their industry/ occupation cell. Column (3) re-calculates the baseline result in a levels estimator, rather than in logs and Column (4) replaces earnings levels for unemployed individuals with \$0's.

Supplement Table S9a: Log Earnings Score Differentials in 1930 by Migrant Status,
All Earnings Scores Based on Southern Black Earnings

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
	0.391 (0.0100)	0.367 (0.0102)	0.356 (0.0127)	0.291 (0.0818)	0.297 (0.101)	0.393 (0.0181)	0.347 (0.0277)
<i>Panel B: Earnings score based on IPUMS, 1960</i>							
	0.476 (0.0134)	0.447 (0.0138)	0.435 (0.0163)	0.340 (0.0912)	0.346 (0.112)	0.460 (0.0250)	0.401 (0.0347)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	5056	5056	5056	403	403	1935	1935

Notes and sources: Corresponds to Table 7 in the main text. Migrant estimated wages equivalent to non-migrant wages in the same industry (Panel A) and occupation (Panel B). Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930.

Supplement Table S9b: Log Earnings Score Differentials in 1930 by Migrant Status,
All Earnings Scores Based on Southern Black Earnings with a Regional Adjustment Based on White Wages

	1	2	3	4a	4b	5a	5b
<i>Panel A: Earnings score based on Lebergott, 1928</i>							
Nominal	0.662 (0.00956)	0.639 (0.00976)	0.627 (0.0121)	0.571 (0.0776)	0.575 (0.0954)	0.767 (0.0242)	0.614 (0.0264)
Real	0.456 (0.00941)	0.438 (0.00956)	0.429 (0.0118)	0.387 (0.0750)	0.381 (0.0917)	0.570 (0.0232)	0.418 (0.0260)
<i>Panel B: Earnings score based on IPUMS, 1960</i>							
Nominal	0.793 (0.0133)	0.765 (0.0136)	0.753 (0.0161)	0.661 (0.0893)	0.676 (0.109)	0.660 (0.0167)	0.708 (0.0343)
Real	0.587 (0.0129)	0.563 (0.0132)	0.555 (0.0154)	0.478 (0.0893)	0.482 (0.109)	0.462 (0.0160)	0.512 (0.0337)
Controls for personal, household and county characteristics in 1910	No	Yes	Yes	Yes	Yes	Yes	Yes
1910 County fixed effects	No	No	Yes	Yes	No	No	No
1910 Household fixed effects	No	No	No	No	Yes	No	No
Differenced dependent variable (1930-1910)	No	No	No	No	No	No	Yes
N	5056	5056	5056	403	403	1935	1935

Notes and sources: Each coefficient is from a separate regression of log earnings on migrant status (=1 if inter-regional migrant). All are statistically significant at the 5 percent level. Control variables and sample sizes vary across columns. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 has no control variables; it is simply the difference between migrants and non-migrants earnings. Column 2 controls for age fixed effects, veteran status, city status, owner-occupied housing interacted with headship status, state-level log income per capita, black percent of county population, black adult literacy rate in the county, black children's school attendance in the county, and percent of farm acres in cotton. All control variables pertain to 1910 except veteran status. Column 3 adds county fixed effects. Columns 4a and 4b are based on a subsample of brothers, and 4b includes household level fixed effects. Columns 5a and 5b are based on a subsample of men observed in the labor market in both 1910 and 1930, and the dependent variable in 5b is change in log earnings from 1910 to 1930.

Supplement Table S10: “Returns” to Migration for 1910 Observable Characteristics

	w/ Industry FEs	w/ Occupation FEs
Literacy	0.0298 (0.0273)	0.0391 (0.0281)
Home Ownership	-0.0167 (0.0237)	-0.0115 (0.0234)
Big city status	-0.0072 (0.0193)	-0.0074 (0.0192)
Employment status	0.0016 (0.0125)	0.00095 (0.0128)
1910 age fixed effects	Yes	Yes
1910 industry fixed effects	Yes	No
1910 occupation fixed effects	No	Yes
Observations	1935	1935

Notes and sources: Each coefficient is from a separate regression of the outcome variable (literacy, home ownership, big city status, employment status) on migrant status (=1 if inter-regional migrant) and an age fixed effect. Standard errors are adjusted for clustering at the household-of-origin level. Column 1 includes industry in 1910 fixed effects. Column 2 includes occupation in 1910 fixed effects.