CENTRE FOR ECONOMIC HISTORY
THE AUSTRALIAN NATIONAL UNIVERSITY
DISCUSSION PAPER SERIES

INTERNAL MIGRATION, EDUCATION AND UPWARD RANK MOBILITY: EVIDENCE FROM AMERICAN HISTORY

ZACHARY WARD (BAYLOR UNIVERSITY)

DISCUSSION PAPER NO. 2019-04

MARCH 2019

THE AUSTRALIAN NATIONAL UNIVERSITY
ACTON ACT 0200 AUSTRALIA
T 61 2 6125 3590
F 61 2 6125 5124
E enquiries.rse@anu.edu.au
http://rse.anu.edu.au/CEH

Internal Migration, Education and Upward Rank Mobility: Evidence from American History*

Zachary Ward

Baylor University

March 2019

Abstract: To what extent does internal migration lead to upward mobility? Using within-brother variation and a new linked dataset from 1910 to 1940, I estimate that internal migrants were more likely to improve on their father's percentile rank than non-migrants. On average, the effect of migration was nearly four times the effect of one year of education; for those raised in poorer households, migration's effect was about nine times that of education. The evidence suggests that internal migration was a key strategy for intergenerational progress in a context of rapid industrialization, high rates of rural-to-urban migration and large interregional income gaps.

JEL Codes: J61, J62, N31, N32

Keywords: internal migration, intergenerational mobility, urbanization

[*]Department of Economics, Baylor University, One Bear Place 98001, Waco, TX 76798. Email: Zach.A.Ward@gmail.com. Thanks to those at the Australian National University for helpful comments, including Bruce Chapman, Nathan Deutscher, Tim Hatton and Tue Gorgens. Thanks to Lee Alston who helped me to gain access to the full-count census.

I. Introduction

Education has long been viewed as the primary means to escape poverty, but another strategy is to simply move to better opportunities. Such investments into internal migration could theoretically lead to higher rates of intergenerational mobility (Becker and Tomes, 1986; Schultz, 1961); however, evidence for the importance of internal migration is mixed. While some argue that internal migration is a key driver of intergenerational mobility rates throughout American history (Blau and Duncan, 1967; Long and Ferrie, 2013), others do not find a strong association (Olivetti and Paserman, 2015). Moreover, there is an ongoing debate over whether any estimated gain from migration reflects a causal effect or selection into migration (e.g., Bryan et al., 2014; Collins and Wanamaker, 2014; Hicks et al., 2018; Young, 2013). To what extent does internal migration allow children to escape poverty and improve on their parents' outcomes?

In this paper, I estimate how intercounty migration was associated with intergenerational mobility with a new dataset of a million father-son pairs tracked between 1910 and 1940. This early 20th century context was a period of rapid industrialization and urbanization, and therefore this study sheds light on the importance of migration – especially rural-to-urban migration – during a key stage of economic development. For example, the 1920 Census was the first one to record that more people lived in urban areas than rural areas (Ferrie, 1999a).[1] Not only were there rural-to-urban flows, but there were also important interregional flows such as the Great Migration out of the South and the exodus from the Dust Bowl (Black et al., 2015; Collins and Wanamaker, 2014; Long and Siu, 2018).[2]

---

[1] Urban areas are defined as those with more than 2,500 people.

[2] There is evidence that the early 20th century had low interstate migration rates (Rosenbloom and Sundstrom, 2004), but the rate of intercounty migration (which captures rural to urban moves) remains unclear since this information is not available in the censuses prior to 1940.

I primarily estimate how internal migration helped a son to improve on his father's percentile rank, or "upward rank" mobility. To account for selection into migration, I compare migrant brothers to non-migrant brothers, a strategy common in the literature (e.g., Abramitzky et al., 2012; Collins and Wanamaker, 2014). A key benefit of this method over other empirical strategies, such as using natural disasters that force people to move, is that within-brother variation allows me to estimate the effect for the entire population rather than for a specific area affected by a flood, hurricane or volcanic eruption (e.g., Deryugina et al., 2018; Nakamura et al., 2017). While household-invariant unobservables are controlled for with household fixed effects, unobservables that vary across brothers are not controlled for, so I refer to estimates in this paper as the "within-brother" effect of migration on upward rank mobility.

When comparing brothers, I find that internal migration was strongly and positively associated with upward rank mobility. On average, brothers who ever migrated across county lines were 11.8 percentage points more likely to improve on their father's percentile rank than brothers who persisted in the same county.[3] The baseline rate of upward rank mobility was 45 percent; therefore, migration is associated with a 26.2 percent increase in the likelihood of upward rank mobility. Migration's effect on upward rank mobility was 3.8 times larger than education's effect – also estimated using within-brother variation – implying that intercounty migration was worth about four years of education. Therefore, migration appears to have been a key investment for moving upward in the economic distribution in the early 20th century.

[3] Ever migration is defined as being observed in a different county in either 1920, 1930 or 1940 than the source county in 1910. Those who report being in a different county in 1935 in the 1940 census are also ever migrants. Note that an "ever migrant" may be in his source county by 1940 if he migrated elsewhere and returned home. I do not observe migrations that occur between censuses if one returns to the same source county.

Note that an important limitation of these results is that the measure of economic status is not based on *actual* income (which is unavailable in historical censuses) but is instead based on *imputed* income. Following Collins and Wanamaker (2017), I impute income by occupation, race and region of residence, and then rank people based on imputed income. Therefore, any income variation that occurs within occupation, race and region is unobserved; consequently, the results are not directly comparable to modern-day rank-rank estimates based on actual income (e.g., Chetty et al., 2014). Yet, when I use outcomes not based on income scores or ranks, such as wage income for wage workers in the 1940 census, I continue to find large effects from internal migration (19.0 percent).

Internal migration was especially effective for those raised in the poorest of households. Migrant brothers raised in the bottom 10 percent of households were 18.7 percentage points more likely to improve on their father's percentile rank than non-migrant brothers – nearly double the estimate for the average migrant, and 9 times the effect of education for the poorest decile. At the same time, intercounty migration was less effective for sons raised in the richest deciles; instead of migration being worth 9 years of school, migration was equally effective as only one year of school. I find similarly large effects when using non-ranked outcomes such as wage income or income scores. Therefore, it appears that internal migration was useful for escaping poverty, but not as useful for those raised in affluent households.

The results in this paper contribute to the intergenerational mobility literature in American history, a literature which often relies on linked samples across censuses (e.g., Long and Ferrie, 2013; Feigenbaum, 2018).[4] While most of the literature describes how the intergenerational

[4] Also see Olivetti and Paserman (2015) and Clark (2014) for mobility estimates with non-linked father-son data.

relationship varied across place, race, sex or time, my paper is more related to those who move past description and estimate what affects intergenerational mobility, such as education expansion, the quality of the local labor market or economic shocks (e.g., Feigenbaum, 2015; Olivetti and Paserman, 2015; Parman, 2011; Tan, 2018). My paper shows that internal migration was one factor that led to substantial upward rank mobility during the early 20th century United States, especially for children raised in poorer households. The results are consistent with evidence from 19th century Britain and 19th century Argentina that rural-to-urban moves were important for upward mobility in the past (Long, 2005; Pérez, 2018). At the same time, I continue to find that the father's economic status persisted to the son's status for the group of migrants, suggesting that migration did not completely break the economic link between father and son.

While my focus is on internal migration, I often benchmark its effect against that of education. Therefore, another contribution is that I provide novel evidence on the value of education for upward rank mobility in the early 20th century, a period which has been studied extensively due to the sharp rise in educational attainment (e.g., Card et al., 2018; Goldin and Katz, 2008; Rauscher, 2016). Rather than instrumenting educational attainment with compulsory schooling laws, a method that primarily influences those on the lower end of the education distribution, I continue to use within-brother variation, which allows me to estimate the education premium across the entire distribution. I find that that one year of education was associated with a 3.1 percentage point increased likelihood of upward rank mobility. If I instead use wage income as the outcome, then the within-brother education premium was 5.5 percent for wage workers, which is lower than other estimates that use compulsory schooling laws (6.4 to 7.9 percent from Clay et al. (2016)). Considering these new estimates for the education premium, the results suggest that internal migration was more effective for allowing children to escape poverty since the migration wage premium was eight times larger than the education wage premium for the poorest decile. I also show that, conditional on education, the father's rank still persisted to the son's rank (a rank-rank slope of 0.37), suggesting that education did not completely equalize outcomes across households in the early 20th century.

Finally, this paper also contributes to the literature on the return to internal migration, much of which focuses on rural-to-urban migration in developing economies. Estimates in this literature are mixed, varying from highly positive to close to zero (e.g., Bryan et al., 2014; Young, 2013; Hicks et al. 2018). The size of the migration premium depends on the context, so out of this literature my paper is most closely related to research on specific migrations in the early 20th century United States, such as the Great Migration or the Dust Bowl refugees (Boustan, 2016; Collins and Wanamaker, 2014; Long and Siu, 2018). One contribution to this literature is to show that the migration premium varied widely for different people and places. For instance, the migration wage premium was particularly high for African Americans leaving the South – about four times the national estimate. Further, I provide new estimates of the premium for rural-to-urban migration.[5] I estimate that rural-to-urban moves are associated with a 30.2 percentage point increased likelihood of upward rank mobility – one of the highest estimates in the dataset. Therefore, while the average migration premium is estimated to be high, it was particularly effective for those raised in the poorer households who tended to be rural, southern, or black.

[5] There is a large literature on earnings gaps between rural and urban areas (e.g., Alston and Hatton, 1991; Boustan et al., 2014; Hatton and Williamson 1991; Hatton and Williamson 1992); a persistent question is how much of the rural-urban earnings gap is due to market imperfection or unobserved productivity differences. See Ferrie (1999b, Chapter 7), Salisbury (2014) and Stewart (2006) on internal migration with linked data in the 19th century; my article is different by using within-brother variation and focusing on intergenerational outcomes.

II. Data

Census Data Linked between 1910 and 1940

To estimate the relationship between internal migration and intergenerational mobility, I need data that observe internal migration and the adult economic outcomes of both the father and son. These data are not directly available, so I create the data by linking historical censuses. First, I take 0-14-year-old sons in the 1910 census when they are observed with their 30-55-year-old fathers (Ruggles et al., 2018); this 1910 census is when the father's outcome is observed.[6] The sons are then linked forward to their adult outcomes in the 1940 census when they are between 30 and 44 years old. Thus, the estimate of economic persistence from father to son is from a comparison of the son's adult outcome in 1940 to his father's adult outcome in 1910.

The 1910-1940 data observe both intergenerational outcomes and intercounty migration, but it does not capture all intercounty moves between 1910 and 1940. I am particularly concerned about sons who moved to another county and then returned to the source county by 1910, perhaps due to an unsuccessful trip between censuses. To address this problem, I further link the sons to the 1920 and 1930 censuses to observe intercounty migration between all decades. The final dataset tracks sons for every census between 1910 and 1940. The father's outcome is observed in 1910. Of course, I only capture location at these 10-year intervals, so I do not observe temporary moves that occurred in between censuses when the son moved back to the source county.

[6] The father's age range is limited since intergenerational relationships are best estimated in the middle of the life-cycle. The 1910-1940 dataset is the same one described in Kosack and Ward's (2018) study on mobility gaps across African, Anglo and Mexican Americans. The dataset in this paper differs from the dataset in Kosack and Ward (2018) since I additionally link the 1910-1940 dataset to the 1930 and 1920 censuses to observe intercounty migration in all decades.

Given the data structure, I define migration as whether someone ever migrated across county lines. The ever-migrant variable is an indicator variable for if the son is ever observed in a different county (in 1920, 1930, or 1940) than his county in 1910.[7] Ever migrants are also those who claimed to be in a different county in 1935 according to the five-year migration question in the 1940 Census.

To build the data, I link individuals across censuses using the machine-learning approach proposed by Feigenbaum (2016). This method creates a linking score for each person based on closeness in first name, last name, age and state of birth (see Appendix B for full details). Since last name is a key linking variable, females are dropped due to name changes after marriage. The linking scores are separately predicted for black and white sons to account for differences in naming conventions and age reports. Importantly, I set the precision of the linking algorithm such that black links are expected to be of equal quality as white links. While matching across censuses is not perfect since there are no unique identifiers across censuses, Bailey et al. (2017) show that this method accurately estimates intergenerational mobility when compared to a hand-linked dataset. Yet note that the fuzzy matching process leads to some false links and therefore the migration rates in this dataset likely overstate the true rate of intercounty migration between 1910 and 1940.

The final linked dataset is of 949,333 sons tracked across all censuses between 1910 and 1940. While large, the triple-linked dataset is only 9.1 percent of the original population of linkable sons. This low linking rate raises concerns over whether the sample is representative of the population since each link loses people non-randomly. To address this issue, I weight the data to

[7] I use the County Longitudinal Template (ICPSR 6576) to set county borders at 1910 lines. Therefore, a migrant is one who is ever observed in a different county according to the 1910 lines.

be representative of the population with inverse proportional weights, where the weights are created after predicting the characteristics that are associated with a successful link (see Appendix Table B7). While the unweighted sample is overrepresented on sons from the Midwest and sons of farmers, the weighted sample is not.

Imputing economic status with income scores

The primary outcome of interest in this paper is whether the son strictly improved on his father's percentile rank, or upward rank mobility (Bhattacharya and Mazumder, 2011). In modern-day data, a person is typically ranked by his location in the *actual* income distribution; however, this is not possible in most US historical data.[8] Therefore, I rank people based on their location in the *imputed* income distribution. The most common method of imputing income is to use the median earnings by 3-digit occupation from the 1950 Census ("occupational earnings" or *occscore* from IPUMS). Instead, I impute income based on an individual's 3-digit occupation, race, and region, where the income score is primarily estimated with the wage income data from the 1940 census (see Appendix C for full details). This imputation procedure follows Collins and Wanamaker (2017); importantly, the method also imputes self-employed earnings.

There are four benefits of using this income imputation over the more commonly used 1950 occupational score. First, the income score captures the benefit from moving across regions, a key interest in this paper. Second, it captures black-white differences in actual income more accurately than occupational earnings, which is important since black and white sons differed in their migration behavior and rate of upward rank mobility (Collins and Wanamaker, 2015; Collins and Wanamaker, 2017; Margo, 2016). Third, the income score differentiates the earnings of farm

[8] For exceptions, see Feigenbaum (2015), Feigenbaum (2018) and Parman (2011).

owners and farm tenants, where owners and tenants are delineated by whether they owned a home. Fourth, the method also imputes perquisites for farmers and farm laborers, which are important for understanding rural-urban income gaps (Alston and Hatton, 1991).

After imputing income, I rank fathers and sons based on their location in the income score distribution.[9] Therefore, upward rank mobility is defined as whether the son strictly improved on his father's percentile rank. Ranked measures are useful relative to non-ranked measures of mobility since they are less biased from measurement error (Nybom and Stuhler, 2017).

To provide an understanding of the outcomes in the dataset, Table A1 shows the mean income scores and ranks for 1-digit occupational categories (in 2018 dollars).[10] For example, farmers are estimated to earn 14,003 dollars (25th percentile) on average and professionals are estimated to earn 39,364 dollars on average (84th percentile). Note that most other categories, such as operatives (44th percentile), craftsmen (57th) and sales workers (72nd), have higher average percentile ranks than farmers, but farm laborers (8th) and general laborers (19th) do not. These scores are key to understanding the results since they suggest that a farm-to-city move will generally result in a higher rank if the son holds at least a semi-skilled occupation. At the same time, the farm-to-city move will not generally result in upward rank mobility if the son ends up as a laborer. It is vital to note that I do not observe variation within occupation, race and region, so the upward rank measures based on income scores are not directly comparable to the modern-day rank-rank measures in Chetty et al. (2014).

The baseline results do not account for cost of living differences across areas since using nominal income is more common in modern-day data studies on intergenerational mobility.

[9] Ranking is done within the dataset using weights. Those with equal income scores are given the same rank.
[10] The conversion to 2018 dollars is made using the CPI adjustment from measuringworth.com

However, others have noted that nominal earnings gaps across rural and urban areas may mostly reflect real income differences (e.g., Hatton and Williamson, 1991). I will also present results for moving up the real income score distribution. These adjustments follow Collins and Wanamaker (2014), who use rural-urban cost differences from Koffsky (1949) and city cost differences from Stecker (1937).[11] For example, if one corrects for nominal income, farmers' percentile rank increases from the 25th to the 30th percentile on average (see Table A2). In general, the cost-of-living adjustments lower the estimated effect of internal migration by about 10-25 percent.

III. Rank-Rank and Upward Rank Mobility

A first look at internal migration between 1910 and 1940

According to the linked data, intercounty migration was common in the early 20th century: 57.2 percent had ever migrated across counties, while 49.6 percent were still intercounty migrants by 1940. The difference between the intercounty and ever migration rate suggests that about 13.3 percent of ever migrants returned to the source county by 1940. A 30-year intercounty migration rate of 49.6 is less than the Long and Ferrie's (2013) estimate of 64 percent between 1850 and 1880, which is consistent with other estimates of a low interstate migration rate between 1910 and 1940 (Rosenbloom and Sundstrom, 2004).

Many of the 1910-1940 moves in the dataset were from rural to urban areas (see Table 1), showing that rural-to-urban migration was common (note that rural is defined as areas with less

[11] Stecker (1937) reports cost of living adjustments for 59 different cities for March 1935. I use this information for those living in these cities. If one is living in an urban area of over 25,000 people but not in one of the 59 cities, then I use the population-weighted average cost of living adjustment for the state. If none of the 59 cities are in the state, then I use the population-weighted average for the region. To adjust for rural-urban differences, I use the "Koffsky adjustment" reported by Collins and Wanamaker (2014, Appendix A3) and Williamson and Lindert (1980) where I scale down urban income by 1.205.

than 2,500 residents).[12] However, other types of moves were also common; for example, rural to rural moves were 27 percent of the intercounty moves between 1910 and 1940. Therefore, it appears that there was a fair amount of geographic mobility in all areas of the country. Many migrations were in fact short. However, many also moved long distances: the average 1910-1940 distance for ever migrants was 429 miles (with a median of 126 miles – see Figure A1 for a histogram of migration distances). The long-distance moves reflect the more well-known flows of the Great Migration and the Dust Bowl, which appear when plotting migration rates by 1910 county (see Figure 1).

The summary statistics in Table 1 provide a first indication about both selection into migration and the treatment effect of migration. First, there was negative selection into migration based on the father's characteristics: fathers of ever migrants were 4.8 percentiles lower in the income score distribution than fathers of persisters. Fathers of migrants were also less likely to own a home (6 percentage points). While there was negative selection into internal migration on average, migration was still common across the economic distribution where 56 percent of children from the richest decile moved by adulthood (see Figure 2).

Table 1 also shows suggestive evidence of a positive treatment effect of migration: while migrant sons came from poorer households, they ended up at higher ranks than non-migrants by 3.3 percentiles. Therefore, it appears that intercounty migrants overcame their poorer backgrounds. The raw means show that migrants were also 16 percentage points (or 44 percent) more likely to improve on their father's percentile rank than non-migrants. I will later explore this apparent

[12] The rural-to-urban moves are based on location in 1910 and 1940 and whether one is still a migrant. Since return migrants are not included in these groups, the sum of rural-rural, rural-urban, urban-rural and urban-urban moves do not add to 100 percent.

positive effect of migration in detail by examining outcomes across the distribution of fathers' ranks and by comparing brothers who migrated to brothers who never migrated.

The association between internal migration and upward rank mobility.

Before estimating the within-brother effect of migration, I first plot how upward rank mobility varied across the father's percentile rank for migrants and non-migrants in Figure 3. While I have three observations of the son's outcomes between 1920 and 1940, I focus on the outcome in 1940 when he is in the middle of the lifecycle (30 to 44 years old) to reduce life-cycle bias (Grawe, 2006). In addition to upward rank mobility, I also plot the rank-rank associations separately by migration status in Figure 3 (Panel B). Note that these figures show descriptive relationships without controls.[13]

Panel A of Figure 3 shows that, conditional on the decile of the father, intercounty migration was associated with a higher likelihood of improving on the father's rank. On average, migrants were 12.9 percentage points more likely to improve on their father's rank than non-migrants (conditional on the father's decile). The figure suggests that the benefit from internal migration was higher at the bottom end rather than at the top end of the distribution: while the increase in upward rank mobility was 12.9 percentage points on average, it was 23 percentage points for children raised in the lowest decile (i.e., 0-10th percentile) and 4 percentage points for the richest decile. A smaller gap between migrants and non-migrants for higher ranks is not just a mechanical result due to less room for upward rank mobility at the upper end; if one instead uses non-ranked measures such as wages or actual income score, the same pattern holds, as I will show later.

---

[13] For those interested in occupational category results, the underlying transition matrices between father and son are shown in Tables A3 and A4.

The rank-rank plots show that economic status was transmitted across generations less strongly for the group of migrants than for the group of non-migrants. This pattern can be measured by the slope of the line, where a flatter slope indicates higher relative mobility: indeed, the slope of the rank-rank graph is flatter for the group of migrants (0.50) than for the group of non-migrants (0.58). However, the difference in slopes is not large, suggesting that while migration was associated with both greater upward rank and relative mobility, the link between father and son was not completely broken such that slope of the migrant line is flat. In fact, persistence from father to son appears to be strong relative to modern-day rank-rank measures (e.g., Chetty et al., 2014). However, recall that historical rank-rank measures are not directly comparable to modern-day ones since I rank people on imputed income rather than actual income.

Internal migration was positively associated with upward rank mobility, but how did it compare with the more well-known strategy of increasing educational attainment? Figure 4 plots the rank-rank associations for two different education groups: those who acquired exactly 8 years of education and those who acquired exactly 12 years of education. The figure shows that those with 12 years of education were more likely to improve on the father's rank than those with 8 years, confirming that education was associated with upward rank mobility. Interestingly, the rank-rank slopes are not flat for the 8-year group (0.39) or 12-year group (0.30), showing that education did not completely equalize outcomes for children across the economic distribution.

The positive association between education and upward rank mobility is unsurprising, so a more interesting result is the size of the effect. If one treats education as a continuous variable and uses the entire dataset rather than just those with 8 or 12 years of education, then one year of education is associated with a gain of 3.1 percentiles, which is about half of the intercounty migration premium of 5.8 percentiles. Also, there is suggestive evidence that the effect of education on upward rank mobility is larger in the middle of the distribution than at the bottom or top end. However, these positive "effects" of education and internal migration may be mostly driven by selection bias rather than a true treatment effect. Therefore, I will use a regression framework to control for unobservable selection into migration or education.

IV. The within-brother migration premium

Empirical specification.

The visual evidence from Figures 3 and 4 suggests that migration and education were effective for improving on the father's percentile rank; however, these estimates do not account for selection into migration or education. One could more accurately estimate the treatment effects by controlling for observables from childhood (such as the son's observed birth order and race) with the following specification:

$$y_{i,h,g} = \alpha_0 + \alpha_1 EverMigrant_{i,h,g} + \alpha_2 Educ_{i,h,g} + \alpha_3 y_{h,g-1} + \gamma' X_{i,h,g} + \varepsilon_{i,h,g} \quad (1)$$

where $y_{i,h,g}$ is an indicator for whether son i from household h and generation g strictly improved on his father's percentile rank. Alternatively, $y_{i,h,g}$ may be the son's percentile rank. The son's outcome is regressed on an indicator variable for whether he ever moved across counties between 1910 and 1940 $(EverMigrant_{i,h,g})$ and years of educational attainment $(Educ_{i,h,g})$. The key addition to the equation over the standard Mincer equation is that I include the percentile rank of the father in 1910 $(y_{i,h,g-1})$. I also include other observable characteristics from Table 1 in $X_{i,h,g}$ such as the son's age, race and observed birth order and the father's literacy and ownership status. Sometimes I refer to α_1 as a "naïve" estimate of the migration premium (Abramitzky et al., 2012).

While this regression is an improvement over other estimates of the migration premium since it controls for childhood characteristics, it is possible to improve it further. Following Abramitzky et al. (2012), I estimate the association between migration, education and upward rank mobility using variation within the household – that is, across brothers. This methodology controls for any unobserved variable that is constant within the household such as neighborhood quality, ethnic background or exposure to local labor markets. Therefore, I estimate:

$$y_{i,h,g} = \beta_0 + \beta_1 EverMigrant_{i,h,g} + \beta_2 Educ_{i,h,g} + \gamma' X_{i,h,g} + \theta_{h,g} + \varepsilon_{i,h,g} \tag{2}$$

where $\theta_{h,g}$ is the household fixed effect. Once household fixed effects are included in the regression, all controls for the father's observables, such as rank, ownership status and literacy, are absorbed by the fixed effect. I still include controls that vary across brothers such as observed birth order and age. I also cluster standard errors at the household level. Note that this empirical method can only estimate the migration premium for households where one brother moved and one brother stayed. While the sample includes 379,429 brothers, 167,168 have variation in migration status.

Besides using upward rank mobility as the dependent variable, another way to incorporate the intergenerational component in the equation is to interact migration status and the father's rank:

$$y_{i,h,g} = \beta_0 + \beta_1 Migrant_{i,h,g} + \beta_2 Educ_{i,h,g} + \beta_3 (Migrant_{i,h,g} \times y_{h,g-1}) \tag{3}$$

$$+ \beta_4 (Educ_{i,h,g} \times y_{h,g-1}) + \gamma' X_{i,h,g} + \theta_{h,g} + \varepsilon_{i,h,g}$$

In essence, the specification captures the different intercepts and slopes between migrants and non-migrants plotted in Figure 3. Recall that Figure 3 suggests that migration was less effective for upward rank mobility for sons raised at the top of the income score distribution, which would imply that the interaction term (β_3) is negative. However, the associations in Figure 3 do not account for household-invariant unobservables. Finally, note that the specification in Equation (3) estimates an additional novel parameter of interest on how the intergenerational relationship varied across years of education (β_4).

Results.

A naïve estimate of the migration premium using OLS suggests that internal migration increased the likelihood of improving on the father's rank by about 11.4 percentage points (see Table 2). However, the naïve estimate is based on across-household variation. The within-brother migration premium is higher than the across-brother migration premium (11.8 v 11.4 percentage points), implying that there was negative selection into migration on unobservable characteristics. Negative selection into internal migration on unobservable characteristics is consistent with direct evidence that migrants were more likely to come from poorer households (Table 1 and Figure 2).

Internal migration is estimated to be much more effective for upward rank mobility than one year of education. The within-brother migration premium was 3.8 times larger than the within-brother education premium (11.8 v 3.1 percentage points). Note that, in contrast with the migration estimates, the education premium falls when going from across-brother to within-brother variation. This result is consistent with the well-known pattern of positive selection into education. The size of the migration to education premium when using within-brother variation (3.8 times) is larger than when using across-brother variation (2.9).

To provide a more complete picture of these within-brother effects, I also use log wage income as the dependent variable in Panel C. (Recall that wage income does not include self-employed income.) The naïve migration wage premium was 15.1 log points (i.e., 16.3 percent), while the return to education was 8.2 log points (8.5 percent). (In the following text, I often translate log points to percent.) When one uses household fixed effects, the wage premium increases from 16.3 percent to 19.0 percent, while the education premium decreases from 8.5 percent to 5.5 percent. This pattern suggests that intercounty migration was worth about three and a half years of education.

The nominal migration premium may overstate the actual increase in living standards since migrants moved to areas with higher prices. When one adjusts for cost of living differentials (see Columns IV-VI of Table 2), the migration premium for upward rank mobility drops by 14 percent from 11.8 percentage points to 10.1 percentage points. For the other measures, the real migration premium was about 10 to 25 percent less than the nominal premium; for example, the wage premium dropped from 19.0 percent to 17.0 percent. Therefore, while migration was associated with a large jump in income, living standards did not increase by the same amount. Nevertheless, the real migration premium was still high and 2.5-3.5 times the real education premium.

The within-brother strategy confirms the results from Figure 3 that internal migration was most effective for children raised in the poorest households. This result is shown in Table 3, which reports the regression results from interacting migration with the father's economic status (Equation (3)). For example, for upward rank mobility, the migration premium is estimated at 23.1 percentage points for children raised at the 0^{th} percentile – about twice the average premium from Table 2. The interacted effect is negative, which shows that the migration premium was smaller for children from richer households: for example, the estimated migration premium for children raised at the 75^{th} percentile was about 6.8 percentage points. Note that these results use within-brother variation, suggesting that they are not driven by changes to unobservable selection into migration across the economic distribution.

On the other hand, the interaction between years of education and the father's percentile rank is precisely estimated at zero. This result provides novel evidence that the education premium did not vary strongly across the economic distribution but was equally beneficial for children from richer and poorer households. Since the education premium was similar across the distribution but the migration premium was not, the ratio of the migration to education premium was largest for poorer households and smallest for richer households. This implies that migration was a particularly effective investment to escape poverty in the early 20th century.

A more flexible way to estimate how migration status interacted with the father's rank is to separately estimate the effect by father's decile. This removes the linearity assumption and allows the migration or education premium to be highest in the middle of the distribution compared with the top or bottom end. When one does this, there is partial evidence that the migration premium does not decrease linearly across the father's rank (see Figure 5, Panel A and B).[14] For example, the decrease in the migration premium appears to occur after the third decile, suggesting that the internal migration premium was roughly similar for the bottom 30 percent of the income score distribution. This result could be because agricultural occupations dominated the bottom 30 percent, and because migration was particularly effective for sons who grew up in rural areas. The migration premium also appears to flatten for the top 30 percent, which suggests that there was little benefit from moving to higher income places if one was already raised in one.

If one turns away from the imputed income measures and instead uses wage income for wage workers (available for sons but not fathers), then the migration premium was also very high for children raised in poorer households; subsequently, the premium decreased roughly linearly

[14] I also show the effect of migration on absolute upward mobility, or whether the son strictly improved on the father's income score in Appendix Figure A2. There is also evidence that education's effect on upward rank mobility was larger in the middle of the distribution than for children raised in the richest or poorest households. For example, for those in the sixth decile the effect was 4.1 percentage points, while it was 2.1 for the poorest decile and 2.3 for the richest decile (See Figure 5a). However, the education premium is mostly flat across the distribution when using log wage income (see Figure 6a).

across the distribution (see Figure 6a). The within-brother migration premium for wage income was 38.5 log points for children raised in the poorest decile, which translates to 47.0 percent. This result is nearly ten times larger than the education premium for the same group (5.0 percent). However, the migration premium for the poorest decile about two-thirds the size after correcting for the cost of living (29.7 percent, see Table 4 Panel C). Therefore, while the cost of living adjustment lowered the average migration premium by about 10-25 percent, it lowered the migration premium for the lowest decile by more.

One concern about these results is that children may have moved during childhood rather than on their own in adulthood such that the return to migration is mostly from the parent's decision. This concern is addressed by the empirical strategy of comparing brothers: since both brothers would move with the family, then there would be no variation in migration within the household and thus these brothers do not identify the effect of migration. Alternatively, if I limit the sample to those between ages 6 and 14 at first observation, or those who were between 16 and 24 by the next census, the same qualitative results hold (see Figure A3, Panel A). Another way to address this problem is to only consider migrants who moved between 1920-1930 or between 1930-1940 rather than between 1910-1920; if one does this, then the same patterns hold (See Figure A3, Panel B).

Of course, a significant event that may affect the results is the Great Depression. For example, there is evidence that both the economic downturn and the New Deal response influenced both intercounty migration and intergenerational mobility (Boustan et al., 2010; Feigenbaum, 2015; Fishback et al., 2006). It could be that the effect of migration between 1910 and 1940 was particularly high because those who persisted in the source county did not migrate elsewhere to avoid a local downturn. However, if one controls for Great Depression severity with the fall in retail sales between 1929 and 1933 by county, and also controls for log relief spending by county, then the results hold (see Figure A4 for results with upward rank mobility and log wage income measures).[15]

The migration premium by type of move, region and race

There was clearly heterogeneity in the migration premium across the income score distribution such that migration was most effective for children raised in poorer households. Rather than split the sample by decile, which masks the characteristics of sons, I estimate the migration premium for different types of moves or subsets of the sample, such as for interstate or interregional migration. Perhaps most importantly, I also separately estimate the premium for rural-to-urban moves, which are a dominant interest in the literature. I primarily report migration's effect on upward rank mobility rather than ranks, income scores or wages due to space considerations, but see Table A5 for other outcomes and Table A6 and for cost-of-living adjustments.

The largest effect of migration on upward rank mobility in the sample was for African Americans during the Great Migration (see Figure 7). Moving from the South to the North, once again estimated with brothers, was associated with a 38.9 percentage point increased likelihood of improving on the father's rank.[16] While others have found a large return for this group (e.g., Collins

---

[15] Relief spending is based on the sum of 1933-1939 per capita relief spending and public works spending at the county level (Fishback et al., 2006). These values are assigned to individuals based on their 1930 county. If I had assigned these values to the 1910 county, then they would not vary across brothers within the household.

[16] The sample is limited to African Americans in the South in 1910. The migration effect compares brothers who lived in the Northeast or Midwest in 1940 to brothers who lived in the South in 1940. If one instead uses wages, then the premium was 76.1 percent (see Table A5, Column 3). This wage premium is smaller than that of others who estimate the within-brother migration premium on wage or income scores at around 120-130 percent (Boustan, 2016; Collins and Wanamaker, 2014). One reason for differences between our estimates is that I additionally control for education; if one does not control for education in my data, then the migration premium was 88.0 percent. My estimates are also noisy because they are for black brothers who are triple-linked, so others' point estimates near 120-130 percent are within my estimate's error bounds

and Wanamaker, 2014; Boustan, 2016), a novel result is that Figure 7 shows that the Great Migration stands out as a unique event with an outsized effect on upward rank mobility. Otherwise intercounty migration was associated with an 11.8 percentage point increase in upward rank mobility, which is only 30 percent of the premium for the Great Migration. However, the estimated effect for the Great Migration is noisy due to having few black sons in the dataset. In general, I find a larger effect of intercounty migration for black sons than for white sons (20.6 v 11.4 percentage points).

One reason for the large Great Migration effect is the wide North-South income gap, but another reason is that many black sons moved from poorer rural areas to richer urban areas. Now I turn to isolate the effect of a rural-to-urban move (recall that urban areas are those with more than 2,500 residents). When using within-brother variation, rural-to-urban migration had the second largest effect on upward rank mobility at 30.2 percentage points. Other types of migrations, such as rural-to-rural or urban-to-urban, were less beneficial; interestingly, urban-to-rural migration was associated with zero improvement. However, the high benefit of rural-to-urban migration does not account for cost of living differentials, which are key for understanding urban and rural wage gaps (Hatton and Williamson, 1991). If one uses real ranks instead, then the rural-to-urban premium drops from 30.2 to 18.5 percentage points (Table A6).

Rural to urban moves came in many types: from migrations either to small cities of fewer than 25,000 people or to large cities of greater than 250,000 people. However, the rural-to-urban migration premium did not vary widely by the population of the destination city. Table 5 shows that the rural to urban migration premium was similar when moving to either a small town (2,500-25,000 people), a large town (25,000-50,000), a small city (50,000-100,000) or a medium-sized city (100,000-250,000).[17] The estimated premium is larger when moving to a large city with over 250,000 people (33.4 percentage points), but not much larger than the estimate for moving to a small town (28.8 percentage points). Therefore, it appears that leaving rural areas was a primary reason for upward rank mobility, rather than the size of the destination.

Another major flow during this period was of refugees fleeing the Dust Bowl (see Long and Siu, 2018 for an in-depth exploration). I define a Dust Bowl migrant as someone who started in a Dust Bowl county in 1930 and was not in one by 1940; a Dust Bowl county is identified as an area that experienced more than 25 percent of topsoil erosion when using data from Hornbeck (2012).[18] When using within-brother variation, I estimate that Dust Bowl migrants were 19.1 percentage points more likely to improve on their father's rank, which is higher than the average intercounty return of 11.8 percentage points but similar to the interregional return of 19.3 percentage points.[19] Besides the upward rank measure, leaving during the Dust Bowl was associated with a 12.8 percent wage return for wage workers. Therefore, the Dust Bowl was associated with economic improvements for brothers who left, not to mention the potential health consequences for brothers who remained behind (Arthi, 2018). Yet the estimated income premium was not uniquely large for Dust Bowl migrants relative to other interregional migrants.

This evidence from Dust Bowl migrants highlights that the migration premium was larger the farther someone moved. One can explicitly show the relationship between the migration premium and distance after calculating the straight-line distance based on county centroids. A short

[17] The distribution of rural-to-urban moves between 1910 and 1940 is 33.5 percent to small towns, 10.3 percent to large towns, 10.8 percent to small cities, 12 percent to medium-sized cities, and 33.3 percent to large cities.
[18] I also limit it to the states shown in Hornbeck (2012, Figure 2). Long and Siu (2018, Figure 1) define a Dust Bowl county as the ten counties that were most wind eroded (located in Oklahoma, Texas, Colorado and Kansas). I expand this definition since there are too few brothers from these ten counties across all decades between 1910 and 1940. There are 5,985 Dust Bowl migrants in the brother sample, and 48,855 in a Dust Bowl county in 1930.
[19] The comparison group are brothers who remained in a Dust Bowl county in 1930 and 1940.

move of between 0 and 50 miles was associated with a 7.2 increased likelihood of improving on the father's rank, while the effect from a 1000+ mile move was nearly four times larger at 26.6 percentage points. This result is consistent with income gaps being larger across longer distances, such as between the South and non-South, than for shorter distances.

Note that the benchmark migration premium is for those who *ever* moved across counties (as observed in the 1920, 1930 and 1940 censuses), not for those who were migrants in 1940. According to the linked data, about 13 percent of those who ever migrated returned to the source county by 1940. Temporary migration may have been a strategy to exploit higher wages in urban areas to build savings and return to a buy a farm; alternatively, it may be that temporary migration was the result of unemployment in the destination county (Hatton and Williamson, 1992). If one estimates the within-brother premium for temporary migration by comparing temporary migrants to persisters, then the effect of temporary migration on upward rank mobility was effectively zero. The migration effect for permanent migrants, or those who continued to live in a different county by 1940, was 20 to 40 percent higher than the baseline estimate for those who ever migrated.

The lack of economic benefit for temporary migration suggests that the migration was not universally beneficial. Rather, the effect of migration was low enough for some that many decided not to stay permanently in the destination county. Unfortunately, another possibility is that temporary migration is observed due to noisy data: if I correctly link someone between 1910-1940 but not between 1920-1930 or 1930-1940, then I will mischaracterize one as a temporary migrant when he was in fact a persister. Therefore, the estimated null effect of temporary migration should be interpreted with caution because because it may be that they are actually persisters. Yet if false links are driving this result, then it also implies that I am underestimating the migration premium in general since I include non-migrants in the ever-migrant group. This interpretation reinforces my main argument that migration's effect on upward rank mobility was, on average, large and positive during the early 20^{th} century.

V. Discussion and Conclusions

Sons can improve on their fathers' outcomes by moving to better opportunities elsewhere. In this paper, I directly show that brothers who did move across counties between 1910 and 1940 ended up with higher-paying occupations and wages than brothers who remained home. Since migrants ended up at a higher rank, they were more likely to improve on their father's rank. When comparing the effect of internal migration on upward rank mobility to the more well-studied effect of education, I find that the migration premium was three to four times the education premium on average. The migration premium was highest for those at the bottom end of the income score distribution – that is, for sons with fathers who worked in low-paying jobs in low-paying regions.

The results leave open a natural question: if the migration premium was so high, why didn't more people migrate? As Duncan and Blau note, "Men do not flow from places of poor to places of good opportunities with the ease of water" (Duncan and Blau, 1967, pg. 244). There are several potential reasons why more did not migrate, such as a lack of information about migration benefits (Bryan et al., 2014) or because the source community provided insurance against income risk while the destination community did not (Munshi and Rosenzweig, 2016). On the other hand, there may have been substantial disutility from leaving the source county or from living in poor conditions in the destination county despite higher wages. Moving to poor conditions was certainly the case for some in the early to mid-20^{th} century; for example, those who migrated during the Great Migration had higher mortality rates and infant mortality rates due to entering unsanitary northern cities; moreover, those who left during the Great Migration were more likely to enter jail (Black et al., 2015; Eriksson and Niemesh, 2016; Eriksson, 2018). Therefore, the large migration wage premium found in this study masks the overall welfare benefits of internal migration (Lagakos et al., 2018).

Nevertheless, sons who did move elsewhere ended up higher in the economic distribution and were more likely to improve on their father's economic rank. Therefore, migration was key for improving wage and occupational outcomes across generations in the early 20th century. This result suggests that variation in migration rates across time may indeed contribute to variation in intergenerational mobility rates (Long and Ferrie, 2013). However, when considering the results, it is important to note that they are specific to the early 20th century when there were wide income gaps across regions and rural to urban areas. After this period there was regional convergence in income, suggesting that the importance of migration for upward rank mobility may have grown smaller in the United States since the mid-20th century.

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* 102.5 (2012): 1832-56.

Alston, Lee J., and Timothy J. Hatton. "The earnings gap between agricultural and manufacturing laborers, 1925–1941." *The Journal of Economic History* 51.1 (1991): 83-99.

Becker, Gary S., and Nigel Tomes. "Human capital and the rise and fall of families." *Journal of labor economics* 4.3, Part 2 (1986): S1-S39.

Black, Dan A., et al. "The impact of the Great Migration on mortality of African Americans: Evidence from the Deep South." *American Economic Review* 105.2 (2015): 477-503.

Blau, P. Michael., Duncan, O. Dudley. (1967). *The American occupational structure.* New York: Wiley.

Bhattacharya, Debopam, and Bhashkar Mazumder. "A nonparametric analysis of black–white differences in intergenerational income mobility in the United States." *Quantitative Economics* 2.3 (2011): 335-379.

Boustan, Leah Platt. *Competition in the promised land: Black migrants in northern cities and labor markets*. Princeton University Press, 2016.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh." *Econometrica* 82.5 (2014): 1671-1748.

Card, David, Ciprian Domnisoru, and Lowell Taylor. "The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility." (2018).

Chetty, Raj, et al. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics* 129.4 (2014): 1553-1623.

Clark, Gregory. *The son also rises: Surnames and the history of social mobility*. Vol. 49. Princeton University Press, 2014.

Clay, Karen, Jeff Lingwall, and Melvin Stephens Jr. *Laws, Educational Outcomes, and Returns to Schooling: Evidence from the Full Count 1940 Census*. No. w22855. National Bureau of Economic Research, 2016.

Collins, William J., and Marianne H. Wanamaker. "Selection and economic gains in the great migration of African Americans: new evidence from linked census data." *American Economic Journal: Applied Economics* 6.1 (2014): 220-52.

Collins, William J., and Marianne H. Wanamaker. "The great migration in black and white: New evidence on the selection and sorting of southern migrants." *The Journal of Economic History* 75.4 (2015): 947-992.

Collins, William J., and Marianne H. Wanamaker. *Up from Slavery? African American Intergenerational Economic Mobility Since 1880*. No. w23395. National Bureau of Economic Research, 2017.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. "The economic impact of Hurricane Katrina on its victims: evidence from individual tax returns." *American Economic Journal: Applied Economics* 10.2 (2018): 202-33.

Eriksson, Katherine. "Moving North and into jail? The great migration and black incarceration." *Journal of Economic Behavior & Organization* (2018).

Eriksson, Katherine, and Gregory Niemesh. "Death in the Promised Land: The Great Migration and Black Infant Mortality." (2016).

Feigenbaum, James J. "Intergenerational mobility during the great depression." (2015).

Feigenbaum, James J. "Automated census record linking: A machine learning approach." (2016).

Feigenbaum, James J. "Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940." *The Economic Journal* 128.612 (2018): F446-F481.

Ferrie, Joseph P. "How Ya Gonna Keep 'Em Down on the Farm [When They've Seen Schenectady]? Rural-to-Urban Migration in 19th Century America, 1850-1870." Mimeo. (1999a)

Ferrie, Joseph P. *Yankeys now: Immigrants in the antebellum US 1840-1860*. Oxford University Press on Demand, 1999b.

Fishback, Price V., William C. Horrace, and Shawn Kantor. "The impact of New Deal expenditures on mobility during the Great Depression." *Explorations in Economic History* 43.2 (2006): 179-222.

Grawe, Nathan D. "Lifecycle bias in estimates of intergenerational earnings persistence." *Labour economics* 13.5 (2006): 551-570.

Goldin, Claudia Dale, and Lawrence F. Katz. *The race between education and technology*. Harvard University Press, 2008.

Hatton, Timothy J., and Jeffrey G. Williamson. "Wage gaps between farm and city: Michigan in the 1890s." *Explorations in Economic History* 28.4 (1991): 381-408.

Hatton, Timothy J., and Jeffrey G. Williamson. "What explains wage gaps between farm and city? Exploring the Todaro model with American evidence, 1890-1941." *Economic Development and Cultural Change* 40.2 (1992): 267-294.

Hicks, Joan Hamory, et al. *Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata*. No. w23253. National Bureau of Economic Research, 2017.

Hornbeck, Richard. "The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe." *American Economic Review* 102.4 (2012): 1477-1507.

Katz, Lawrence F., and Robert A. Margo. "Technical change and the relative demand for skilled labor: The united states in historical perspective." *Human capital in history: The American record*. University of Chicago Press, 2014. 15-57.

Koffsky, Nathan. "Farm and urban purchasing power." *Studies in Income and Wealth*. NBER, 1949. 151-220.

Kosack, Edward, and Zachary Ward "El Sueño Americano? The generational progress of Mexican Americans prior to World War II". Working Paper, 2018.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh. *The Welfare Effects of Encouraging Rural-Urban Migration*. No. w24193. National Bureau of Economic Research, 2018.

Long, Jason, and Henry E. Siu. *Refugees from dust and shrinking land: Tracking the Dust Bowl migrants*. No. w22108. National Bureau of Economic Research, 2016.

Long, Jason, and Joseph Ferrie. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *American Economic Review* 103.4 (2013): 1109-37.

Long, Jason. "Rural-urban migration and socioeconomic mobility in Victorian Britain." *The Journal of Economic History* 65.1 (2005): 1-35.

Margo, Robert A. "Obama, Katrina, and the persistence of racial inequality." *The Journal of Economic History* 76.2 (2016): 301-341.

Munshi, Kaivan, and Mark Rosenzweig. "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." *American Economic Review* 106.1 (2016): 46-98.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. *The Gift of Moving: Intergenerational Consequences of a Mobility Shock*. No. w22392. National Bureau of Economic Research, 2016.

Olivetti, Claudia, and M. Daniele Paserman. "In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940." *American Economic Review* 105.8 (2015): 2695-2724.

Parman, John. "American mobility and the expansion of public education." *The Journal of Economic History* 71.1 (2011): 105-132.

Pérez, Santiago. *Railroads and the Rural to Urban Transition: Evidence from 19th-Century Argentina*. Technical Report, University of California, Davis, 2018.

Rauscher, Emily. "Does educational equality increase mobility? Exploiting nineteenth-century US compulsory schooling laws." *American Journal of Sociology* 121.6 (2016): 1697-1761.

Rosenbloom, Joshua L., and William A. Sundstrom. "The decline and rise of interstate migration in the United States: Evidence from the IPUMS, 1850–1990." *Research in economic history*. Emerald Group Publishing Limited, 2004. 289-325.

Salisbury, Laura. "Selective migration, wages, and occupational mobility in nineteenth century America." *Explorations in Economic History* 53 (2014): 40-63.

Schultz, Theodore W. "Investment in human capital." *The American economic review* 51.1 (1961): 1-17.

Stecker, Margaret Loomis. *Intercity differences in costs of living in March, 1935, 59 cities*. No. 12. US Govt. Print. Off., 1937.

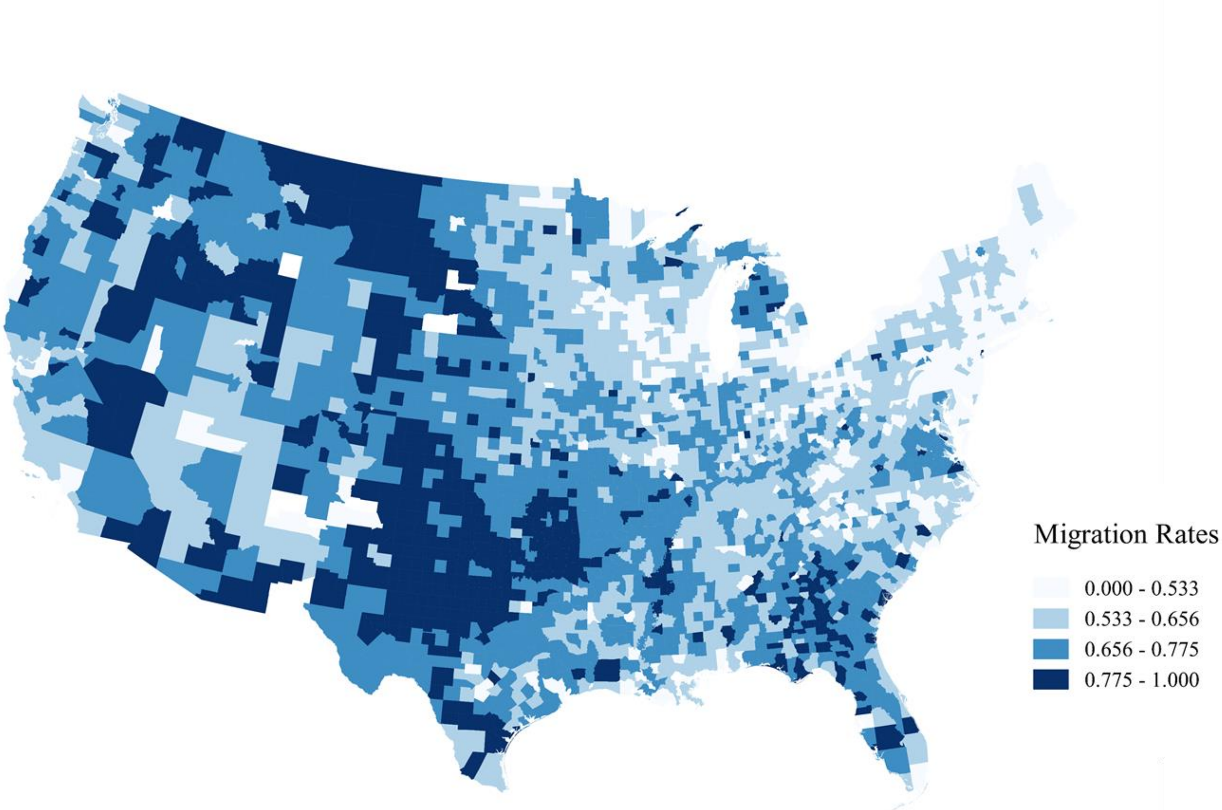
Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 7.0* [dataset]. Minneapolis: University of Minnesota, 2017. https://doi.org/10.18128/D010.V7.0.

Tan, Hui Ren. *A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early 20th Century US.* Boston University Working Paper, 2018.

Williamson, Jeffrey G., and Peter H. Lindert. *American inequality: A macroeconomic history*. Academic Pr, 1980.

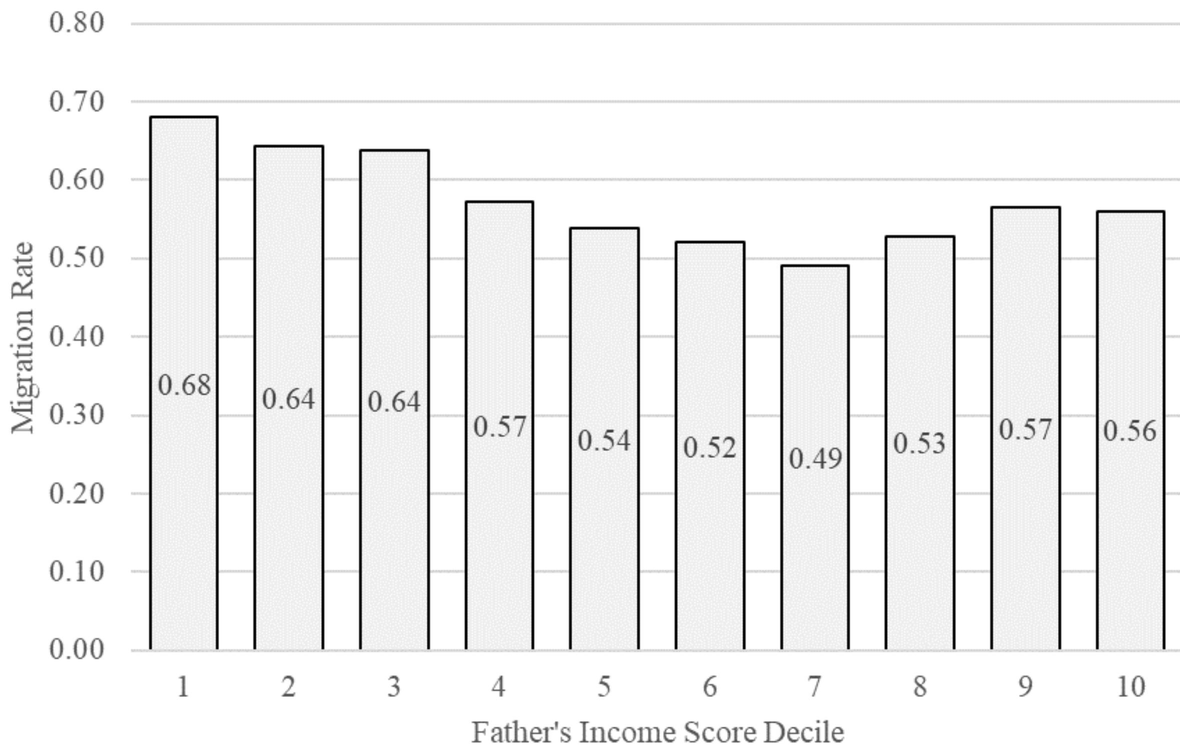
Young, Alwyn. "Inequality, the urban-rural gap, and migration." *The Quarterly Journal of Economics* 128.4 (2013): 1727-1785.

Figure 1. Intercounty Migration Rates 1910-1940 by 1910 County



Notes: Data are from a linked sample of males between childhood (ages 0-14) in 1910 and adulthood in 1940. Migration rate, calculated from the linked sample, is the ratio of those who lived in a different county in 1940 to total number in the county in 1910.

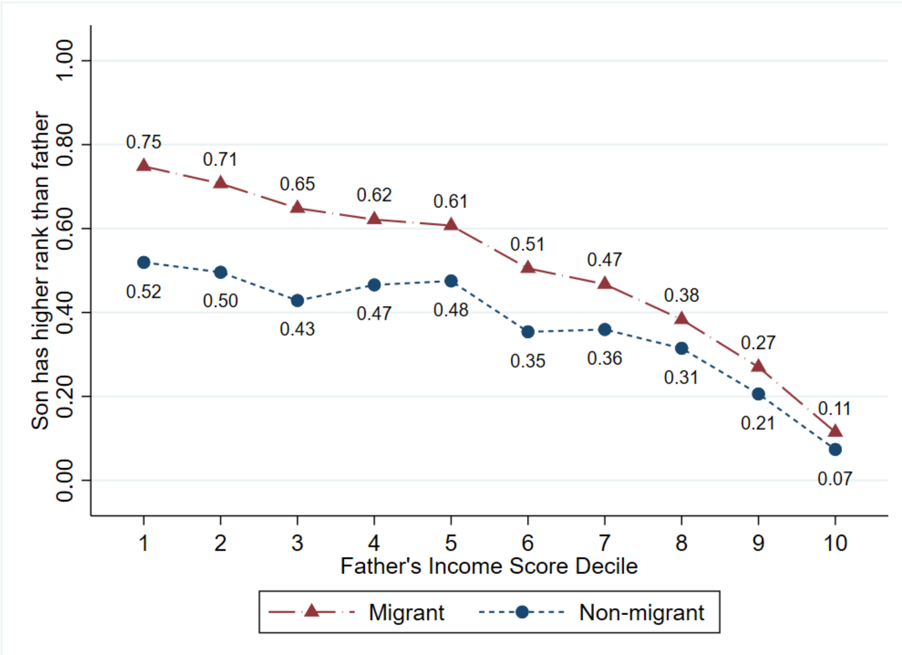
Figure 2. Migration rate by father's income score decile



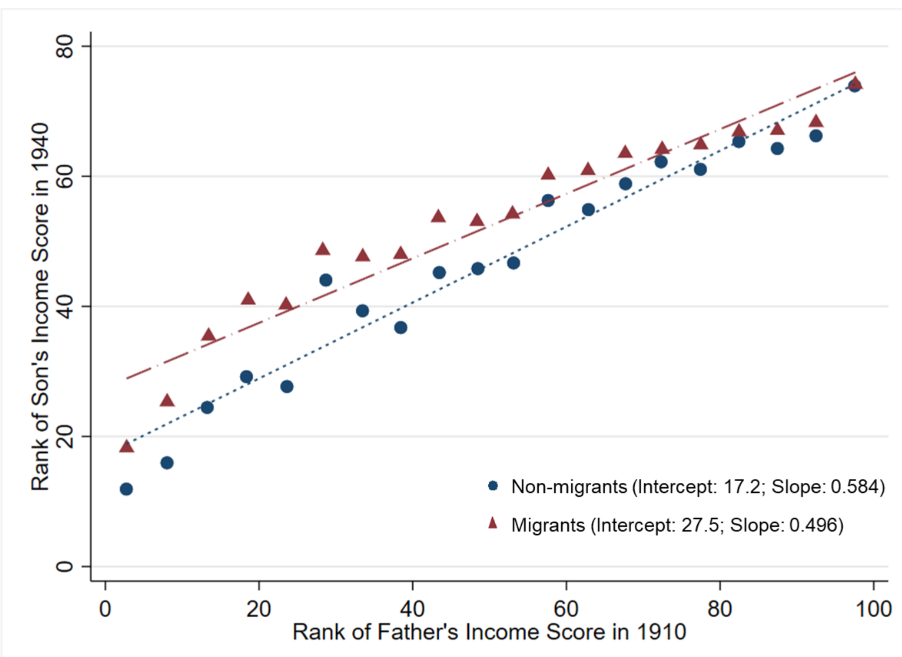
Notes: Data are from the linked sample of individuals between the 1910, 1920, 1930 and 1940 censuses. Income scores are based on the father's occupation, race and region (see Appendix C). The migration rate is for those who ever migrated across counties, as observed in the censuses.

Figure 3. Mobility measures for ever migrants and non-migrants

A. Upward Rank Mobility



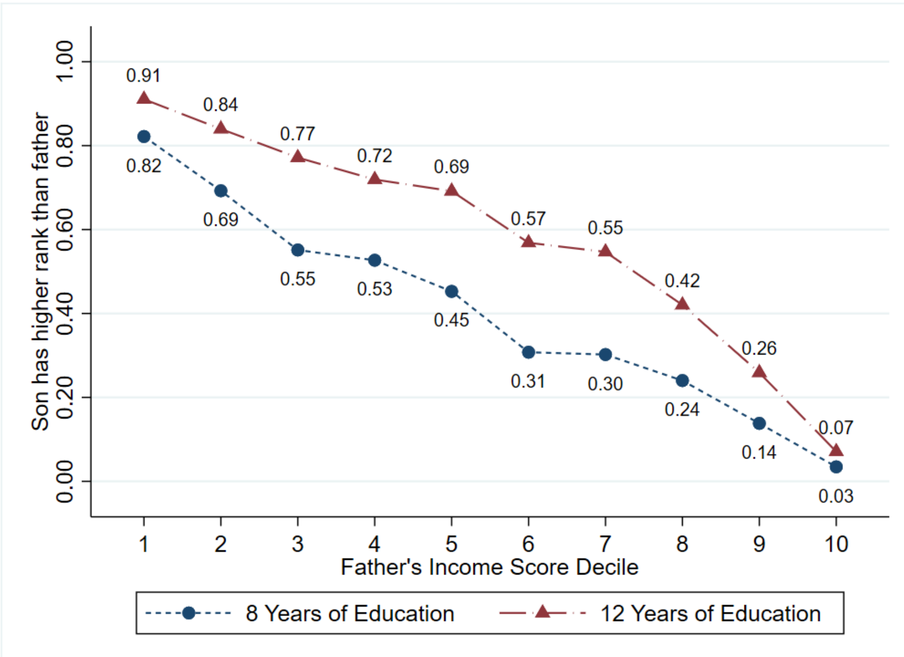
B. Rank-Rank



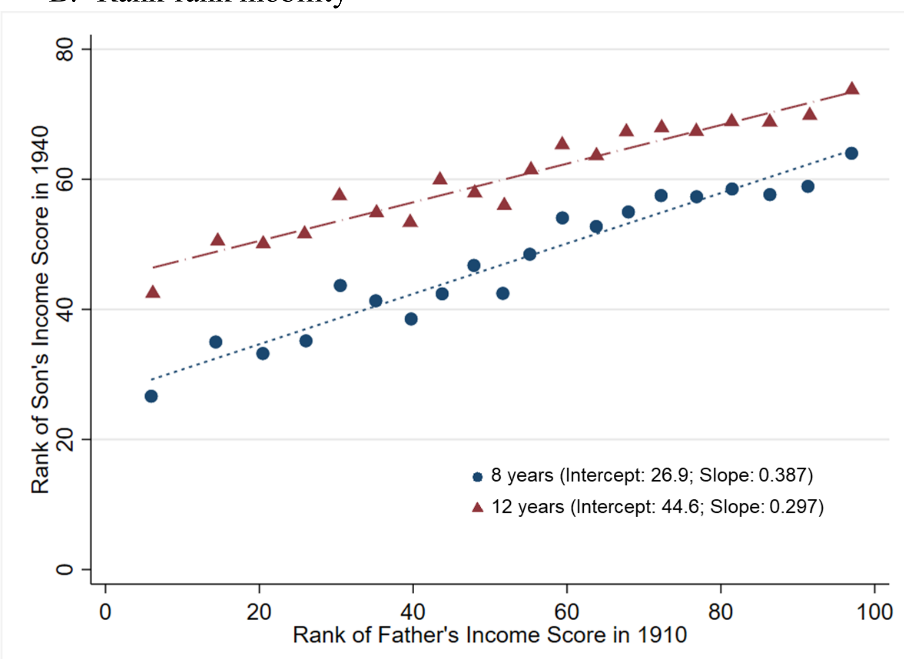
Notes: Data are from the linked sample of individuals between the 1910, 1920, 1930 and 1940 censuses. Son's outcomes are observed in the 1940 census. Migration is defined as having ever migrated across counties according observation at the 1920, 1930 and 1940 censuses. Upward rank mobility is whether the son strictly improved on the father's percentile rank. Income scores are based on the father's occupation, race and region (see Appendix C). No controls are included in the above relationships.

Figure 4. Rank-rank relationship by educational attainment

A. Upward rank mobility

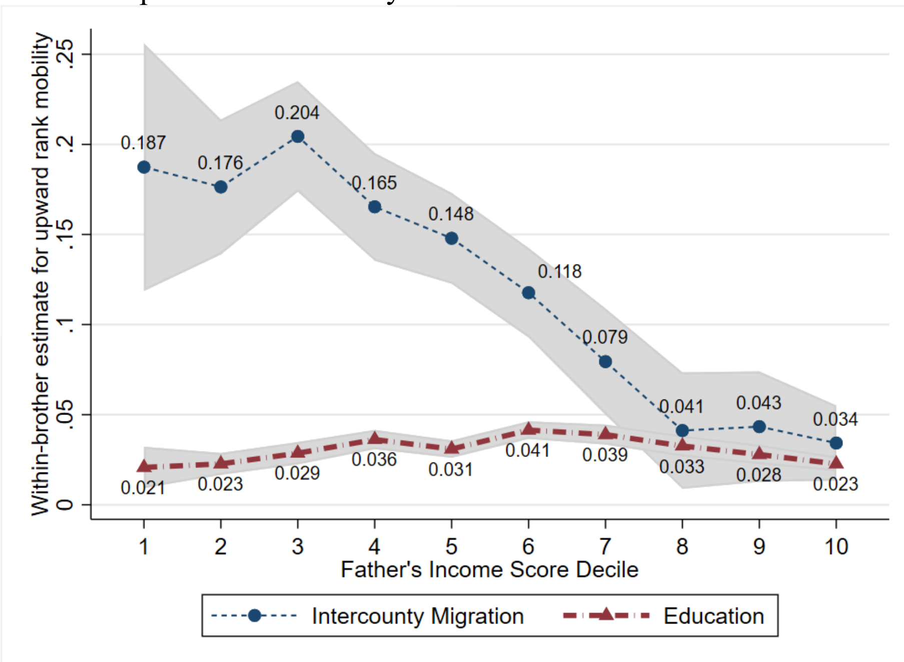


B. Rank-rank mobility

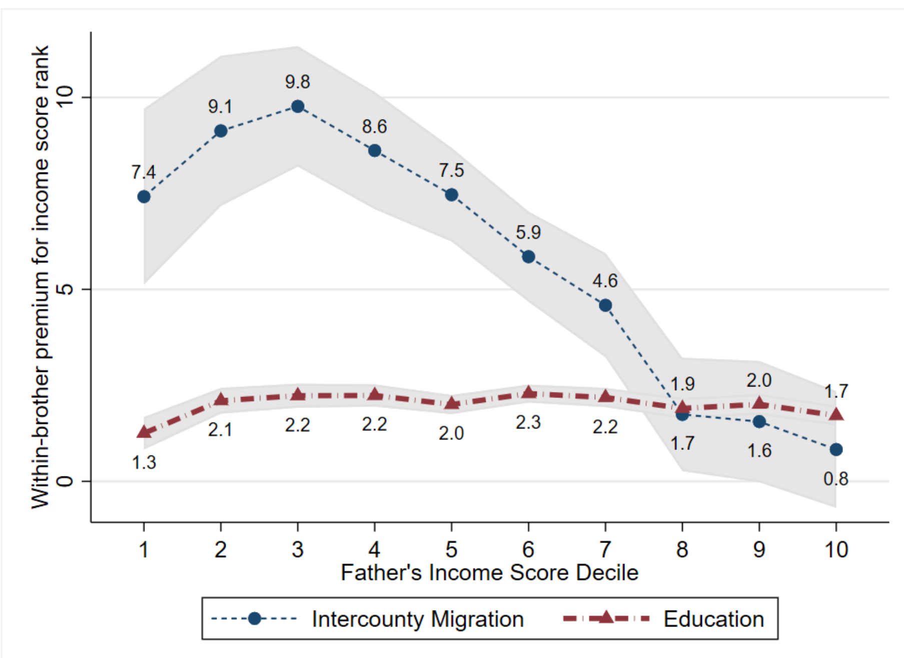


Notes: Data are from the linked sample of individuals between the 1910, 1920, 1930 and 1940 censuses. Son's outcomes are observed in the 1940 Census. Education is taken from the 1940 Census; only those with exactly 8 years or 12 years are included. Upward rank mobility is whether the son strictly improved on the father's percentile rank. Income scores are based on the father's occupation, race and region (see Appendix C). No controls are included in the above relationships.

Figure 5. The within-brother association between intercounty migration and upward rank mobility
Panel A. Upward rank mobility



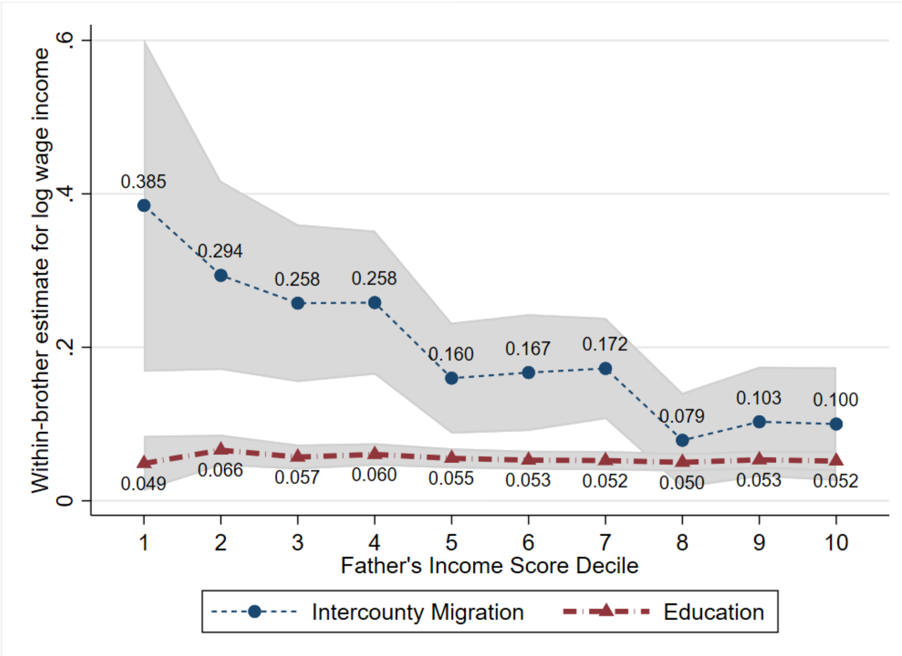
Panel B. Rank of income score



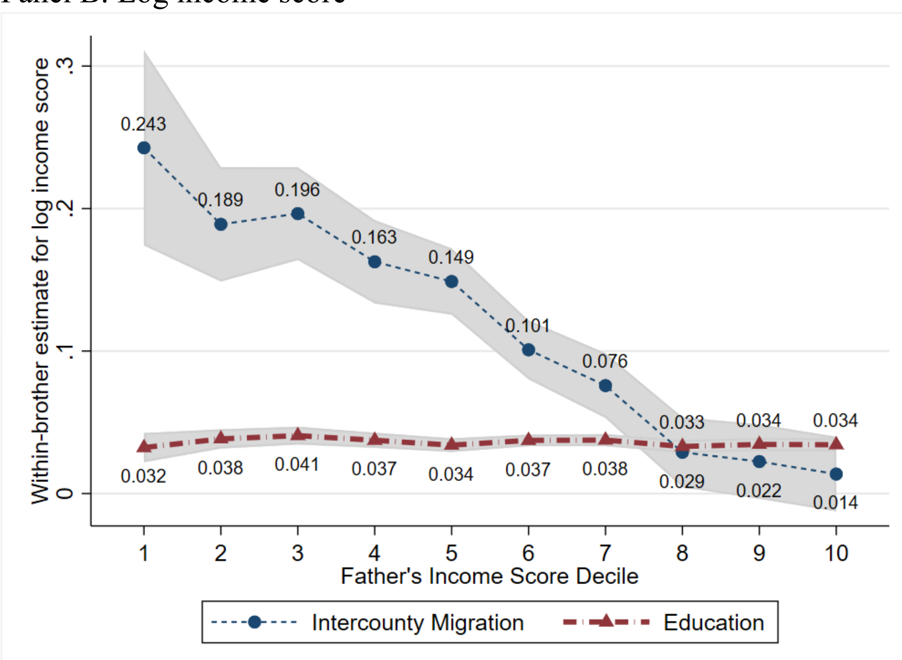
Notes: Data are from the linked sample of individuals between the 1910, 1920, 1930 and 1940 censuses. Estimates are created separately for each decile. Upward rank mobility is whether the son strictly improved on the father's economic rank. Father's income is split into deciles based on position in 1910 nominal income score distribution. Income scores are based on the father's occupation, race and region (see Appendix C). Sample is weighted to be representative and standard errors are clustered at the 1910 household level.

Figure 6. The within-brother migration and education premium for wages and income score

Panel A. Log wage income for wage workers

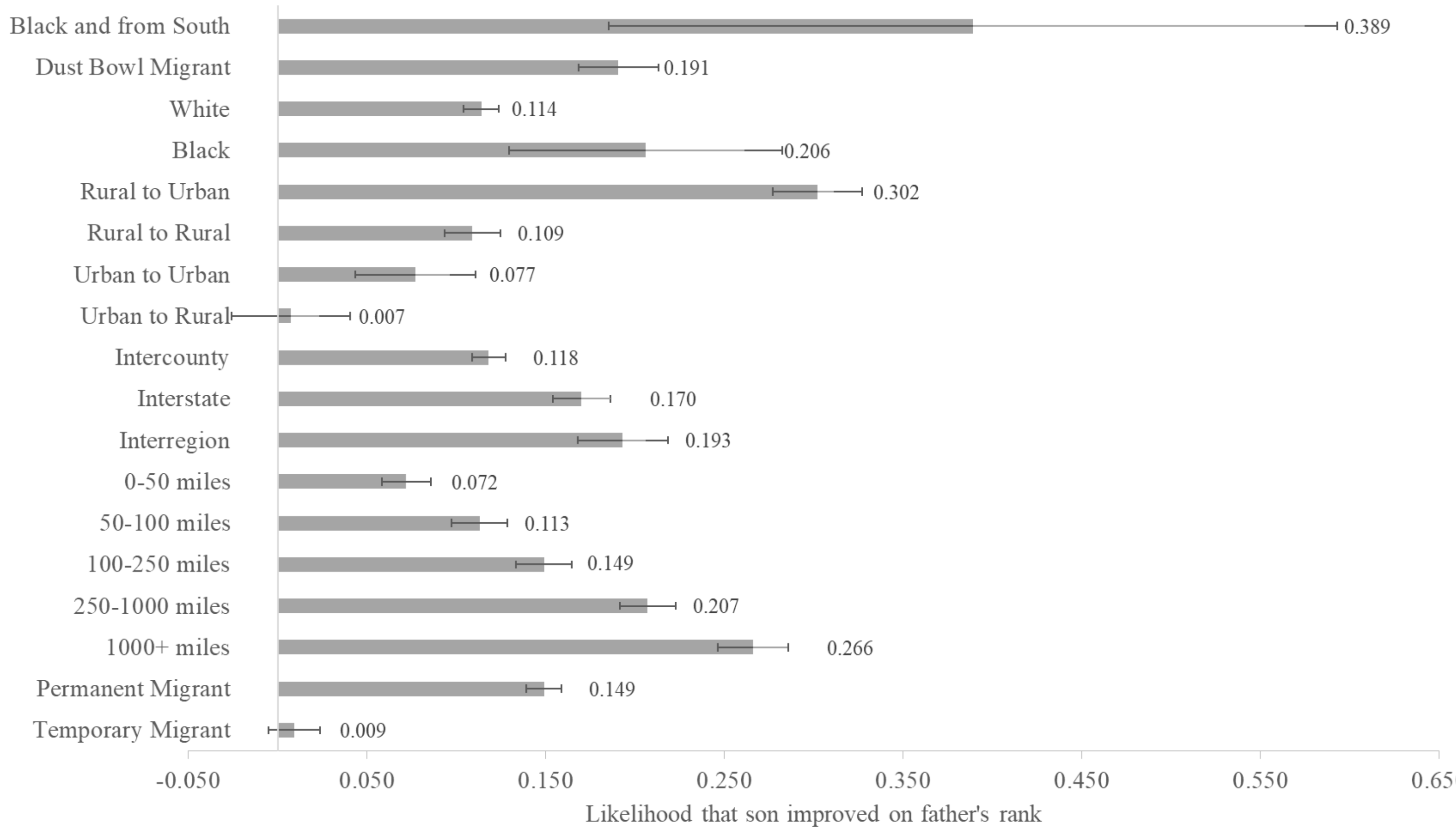


Panel B. Log income score



Notes: Data are from the linked sample of individuals between the 1910, 1920, 1930 and 1940 censuses. Estimates are based on within-brother variation and is estimated separately for each decile. Upward rank mobility is whether the son strictly improved on the father's economic rank. Father's income is split into deciles based on position in 1910 nominal income score distribution. Income scores are based on the father's occupation, race and region (see Appendix C). Wages are only for wage workers and does not include business of farmer income. Sample is weighted to be representative and standard errors are clustered at the 1910 household level.

Figure 7. The within-brother effect on upward rank mobility, by type of move



Notes: Data are from 1910-1940 linked sample. Each row is a separate regression, always using within-brother variation. See Table A5 for other outcomes besides nominal upward rank mobility. Table A6 reports effects with cost of living adjustments. The first row estimates the migration effect for black sons who are in the North or Midwest census regions in 1940 but were in the South census region in 1910; the comparison group are those who remained in the South. Dust Bowl migrants are those who were in a Dust Bowl county in 1930 but not in one by 1940; non-migrants are those in Dust Bowl counties in both 1930 and 1940. Dust Bowl counties are those where any part of the county had more than 25 percent topsoil erosion (see Hornbeck (2012, Figure 2)). Urban is defined as an area with over 2,500 residents; moves are separated by location in 1910 and 1940. Only those who are still intercounty migrants by 1940 are in the rural-urban, rural-rural, urban-urban, urban-rural moves. Miles are measured based on straight-line distances between county centroids in 1910 and 1940. Permanent migrants are those who are in a different county in 1940 than in 1910. Temporary migrants are those who are in the same county in 1940 as they were in 1910 but are observed in a different county in either 1920, 1930, or according to the 1935 migration question in the 1940 Census.

Table 1. Descriptive Statistics across persisters and ever migrants

	I All	II Persisters	III Ever Migrants	IV Difference (III-II)
1910 Census Outcomes				
Age of son	7.31	7.30	7.32	0.03***
	(4.06)	(4.07)	(4.05)	(0.01)
Father's percentile rank	50.30	53.03	48.25	-4.78***
	(28.58)	(27.35)	(29.31)	(0.08)
Father's age	40.59	40.80	40.43	-0.37***
	(6.53)	(6.49)	(6.55)	(0.02)
Father was literate	0.93	0.94	0.93	-0.00***
	(0.25)	(0.25)	(0.25)	(0.00)
Father owned home	0.52	0.56	0.49	-0.06***
	(0.50)	(0.50)	(0.50)	(0.00)
Black	0.09	0.07	0.11	0.03***
	(0.29)	(0.26)	(0.31)	(0.00)
Observed Birth Order	2.61	2.65	2.58	-0.07***
	(1.18)	(1.18)	(1.18)	(0.00)
Lived in City (>25,000 pop)	0.30	0.38	0.24	-0.14***
	(0.46)	(0.48)	(0.43)	(0.00)
1940 Census Outcomes				
Son's percentile rank	50.08	48.20	51.49	3.29***
	(28.88)	(28.47)	(29.09)	(0.08)
Upward Rank Mobility	0.45	0.36	0.52	0.16***
	(0.50)	(0.48)	(0.50)	(0.00)
Years of Education	9.41	9.08	9.65	0.58***
	(3.39)	(3.07)	(3.58)	(0.01)
Lived in City (>25,000 pop)	0.38	0.37	0.39	0.02***
	(0.49)	(0.48)	(0.49)	(0.00)
Owned a home	0.45	0.53	0.39	-0.14***
	(0.50)	(0.50)	(0.49)	(0.00)
Types of 1910-1940 moves:				
Ever moved across states	0.30	0.00	0.52	
	(0.46)	(0.00)	(0.50)	
Ever moved across regions	0.15	0.00	0.26	
	(0.36)	(0.00)	(0.44)	
Rural to urban	0.17	0.00	0.30	
	(0.38)	(0.00)	(0.46)	
Rural to rural	0.16	0.00	0.27	
	(0.36)	(0.00)	(0.45)	
Urban to rural	0.05	0.00	0.08	
	(0.21)	(0.00)	(0.28)	
Urban to urban	0.12	0.00	0.21	
	(0.32)	(0.00)	(0.40)	
Observations	949,333	408,137	541,196	

Notes: Data are from linked sample between 1910 and 1940. Ranks are based on income scores, which are imputed based on the occupation, race and region (see Appendix C). Persisters are those in the same county according to the 1910, 1920, 1930 and 1940 censuses and ever migrants are those who ever switched counties. Urban is defined as a place with over 2,500 residents in 1910 or 1940.

Table 2. The within-brother premium for internal migration and education

	Nominal Value			Real Value		
	I	II	III	IV	V	VI
Panel A. Upward Rank Mobility						
Intercounty Migrant	0.113	0.114	0.118	0.093	0.094	0.101
	(0.001)	(0.002)	(0.005)	(0.001)	(0.002)	(0.005)
Education	0.039	0.039	0.031	0.037	0.037	0.029
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Father's Rank	-0.010	-0.010		-0.010	-0.009	
	(0.000)	(0.000)		(0.000)	(0.000)	
Panel B. Percentile Rank of Income Score						
Intercounty Migrant	4.826	5.108	5.693	3.728	3.797	4.658
	(0.049)	(0.112)	(0.243)	(0.051)	(0.116)	(0.255)
Education	2.715	2.683	2.030	2.705	2.677	2.016
	(0.009)	(0.020)	(0.045)	(0.009)	(0.021)	(0.048)
Father's Rank	0.217	0.232		0.183	0.201	
	(0.001)	(0.003)		(0.001)	(0.003)	
Panel C: Log Income Wage for wage workers						
Intercounty Migrant	0.151	0.151	0.174	0.131	0.127	0.157
	(0.002)	(0.005)	(0.014)	(0.002)	(0.005)	(0.014)
Education	0.084	0.082	0.054	0.081	0.079	0.053
	(0.000)	(0.001)	(0.002)	(0.000)	(0.001)	(0.002)
Father's Income Score	0.167	0.179		0.145	0.158	
	(0.003)	(0.006)		(0.003)	(0.006)	
Panel D: Log Income Score						
Intercounty Migrant	0.091	0.097	0.112	0.065	0.068	0.086
	(0.001)	(0.002)	(0.005)	(0.001)	(0.002)	(0.005)
Education	0.049	0.048	0.036	0.045	0.044	0.033
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Father's Income Score	0.200	0.222		0.170	0.193	
	(0.001)	(0.003)		(0.001)	(0.003)	
1910 Controls	Y	Y	Y	Y	Y	Y
Brother Sample	N	Y	Y	N	Y	Y
HH Fixed Effect	N	N	Y	N	N	Y

Notes: Data are from linked sample between 1910 and 1940. Ranks are based on income scores, which are imputed based on the occupation, race and region (see Appendix C). Cost of living adjustments are made across cities and rural-urban areas following Collins and Wanamaker (2014). Controls include son's observed birth order and age fixed effects, father's age (and square), ownership, literacy and number of children; indicator for race, and indicators for 1910 population (<2,500; 10,000-50,000; 50,000-100,000; 100,000+). Observations for Panel A, B and D are 949,333 total individuals and 211,558 for the brother sample; for Panel C it is 642,543 individuals and 136,604 for the brother sample. Regressions are weighted for representativeness. Standard errors are clustered at the 1910 household level.

Table 3. Migration, Education and Intergenerational Mobility

	Nominal Value		Real Value	
Panel A. Upward Rank Mobility				
Migrant	0.219	0.231	0.170	0.186
	(0.006)	(0.013)	(0.006)	(0.013)
Migrant x Rank of Father (divided by 100)	-0.202	-0.218	-0.144	-0.161
	(0.009)	(0.020)	(0.009)	(0.020)
Education	0.046	0.030	0.042	0.027
	(0.001)	(0.002)	(0.001)	(0.002)
Education x Rank of Father (divided by 100)	-0.013	0.002	-0.010	0.005
	(0.002)	(0.003)	(0.002)	(0.003)
Panel B. Percentile Rank of Income Score				
Migrant	10.399	11.380	7.738	9.055
	(0.263)	(0.578)	(0.274)	(0.609)
Migrant × Rank of Father	-0.102	-0.109	-0.075	-0.083
	(0.004)	(0.010)	(0.004)	(0.010)
Education	2.957	2.090	2.822	1.986
	(0.047)	(0.106)	(0.049)	(0.112)
Education × Rank of Father	-0.005	-0.001	-0.003	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
Panel C. Log Income Wage for wage workers				
Migrant	0.936	1.287	0.584	0.840
	(0.089)	(0.260)	(0.092)	(0.265)
Migrant × Rank of Father	-0.113	-0.160	-0.067	-0.100
	(0.013)	(0.037)	(0.013)	(0.038)
Education	0.167	0.093	0.116	0.057
	(0.014)	(0.045)	(0.014)	(0.046)
Education × Rank of Father	-0.012	-0.006	-0.005	-0.001
	(0.002)	(0.006)	(0.002)	(0.007)
Panel D. Log Income Score				
Migrant	1.005	1.116	0.730	0.816
	(0.043)	(0.090)	(0.043)	(0.089)
Migrant × Rank of Father	-0.132	-0.146	-0.097	-0.107
	(0.006)	(0.013)	(0.006)	(0.013)
Education	0.109	0.046	0.066	0.020
	(0.007)	(0.014)	(0.007)	(0.015)
Education × Rank of Father	-0.009	-0.002	-0.003	0.002
	(0.001)	(0.002)	(0.001)	(0.002)
Household Fixed Effects	N	Y	N	Y

Notes: Data are from linked sample between 1910 and 1940. Ranks are based on income scores, which are imputed based on the occupation, race and region (see Appendix C). Cost of living adjustments are made across cities and rural-urban areas following Collins and Wanamaker (2014). Controls include son's observed birth order and age fixed effects. Observations for Panel A, B and D are 211,558; for Panel C it is 136,604. Regressions are weighted for representativeness. Standard errors are clustered at the 1910 household level. Migrant is an indicator for whether one ever migrated across counties.

Table 4. Within-brother migration and education premium after accounting for cost of living

	Father's Decile in Income Score Distribution									
	1	2	3	4	5	6	7	8	9	10
Panel A: Upward rank mobility, real ranks										
Intercounty Migrant	0.124	0.125	0.147	0.190	0.150	0.039	0.098	0.034	0.052	0.024
	(0.034)	(0.020)	(0.016)	(0.015)	(0.013)	(0.019)	(0.012)	(0.015)	(0.015)	(0.011)
Education	0.018	0.020	0.026	0.034	0.032	0.031	0.039	0.034	0.029	0.020
	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Panel B: Percentile rank of real income score										
Intercounty Migrant	5.779	6.371	7.791	7.836	7.322	2.707	3.819	1.958	1.821	0.726
	(1.204)	(1.024)	(0.886)	(0.800)	(0.635)	(0.876)	(0.598)	(0.764)	(0.803)	(0.823)
Education	1.270	1.967	2.137	2.413	1.980	2.017	2.135	2.120	1.901	1.891
	(0.218)	(0.184)	(0.165)	(0.160)	(0.127)	(0.158)	(0.116)	(0.141)	(0.137)	(0.139)
Panel C: Log of real wage income										
Intercounty Migrant	0.260	0.211	0.204	0.192	0.206	0.110	0.122	0.144	0.114	0.100
	(0.092)	(0.057)	(0.047)	(0.047)	(0.041)	(0.035)	(0.037)	(0.033)	(0.037)	(0.039)
Education	0.038	0.066	0.057	0.057	0.049	0.050	0.047	0.058	0.054	0.052
	(0.016)	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Panel D: Log of real income score										
Intercounty Migrant	0.175	0.120	0.135	0.136	0.138	0.047	0.066	0.031	0.030	0.011
	(0.031)	(0.019)	(0.016)	(0.015)	(0.012)	(0.014)	(0.010)	(0.012)	(0.013)	(0.014)
Education	0.027	0.033	0.035	0.038	0.031	0.033	0.033	0.035	0.031	0.035
	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: Data are from 1910 to 1940 linked sample. Table is split by father's position in the 1910 real income score distribution. See Equation (2) for estimating equation; brother fixed effects, observed birth order and age in 1940 are included in the control variables. Regressions are weighted for representativeness. Standard errors are clustered at the 1910 household level. Migrant is an indicator for whether one ever migrated across counties.

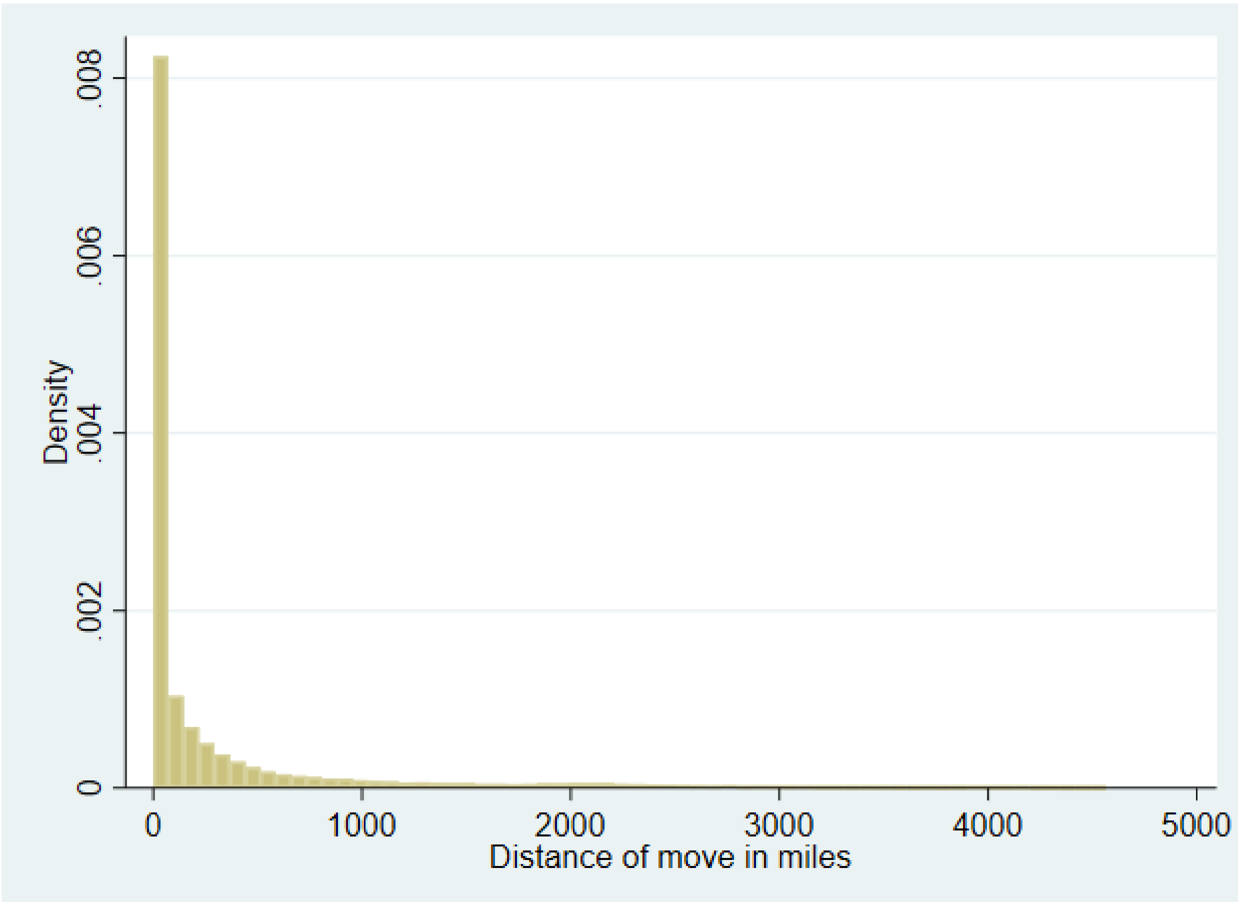
Table 5. The within-brother rural-urban migration premium does not widely vary by population of destination

	Nominal				Real			
	Upward Rank	Rank of Score	Log(Wage Inc.)	Log(Inc. Score)	Upward Rank	Rank of Score	Log(Wage Inc.)	Log(Inc. Score)
Population in 1940:								
2,500-25,000	0.288	15.163	0.423	0.282	0.170	6.852	0.281	0.123
	(0.015)	(0.733)	(0.042)	(0.014)	(0.015)	(0.766)	(0.041)	(0.013)
25,000-50,000	0.285	16.355	0.551	0.308	0.179	8.023	0.408	0.149
	(0.023)	(1.105)	(0.060)	(0.020)	(0.023)	(1.158)	(0.060)	(0.019)
50,000-100,000	0.289	15.295	0.427	0.287	0.166	7.483	0.292	0.134
	(0.024)	(1.150)	(0.058)	(0.021)	(0.025)	(1.197)	(0.057)	(0.021)
100,000-250,000	0.294	16.251	0.498	0.292	0.169	8.349	0.362	0.139
	(0.023)	(1.150)	(0.055)	(0.022)	(0.023)	(1.201)	(0.054)	(0.022)
250,000+	0.334	17.737	0.467	0.346	0.218	8.728	0.313	0.174
	(0.017)	(0.796)	(0.044)	(0.017)	(0.018)	(0.826)	(0.043)	(0.016)
Observations	96,883	96,883	56,669	96,883	96,883	96,883	56,669	96,883
R-squared	0.782	0.812	0.837	0.805	0.765	0.787	0.824	0.779

Notes: Data are from 1910 to 1940 linked sample. The sample is only of rural-urban movers between 1910 and 1940, who are compared to persisters who remained in a rural county. Wage income is only for wage workers. Real income is adjusted for across state and rural-urban cost differences. Income scores are imputed based on occupation, race and region (see Appendix C).

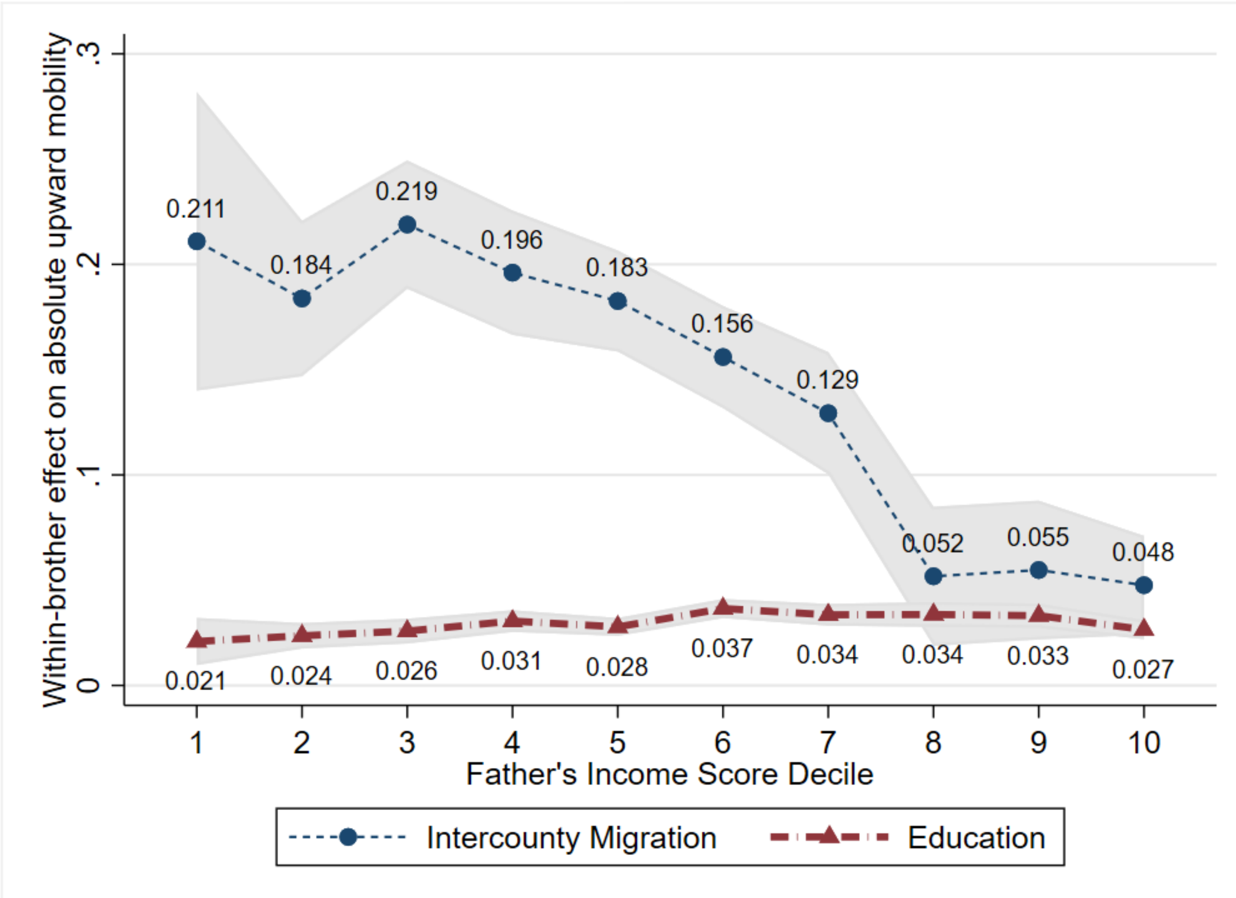
Online Appendix, not for publication

Figure A1. Histogram of migration distances



Notes: Distance between the 1910 and 1940 county centroids as measured by a straight line.

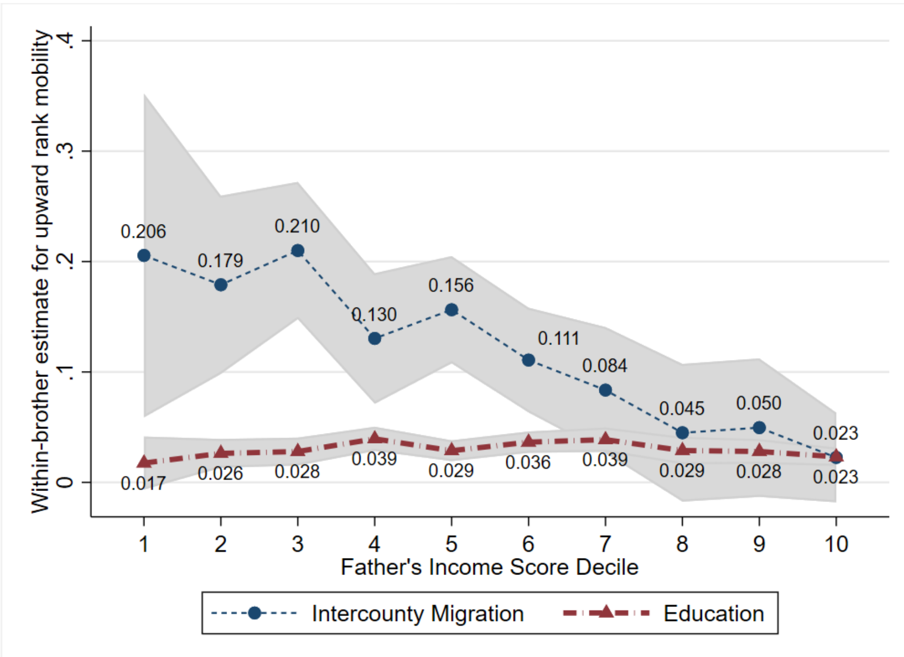
Figure A2. Within-brother effect on absolute upward mobility, or strictly improving on the father's income score



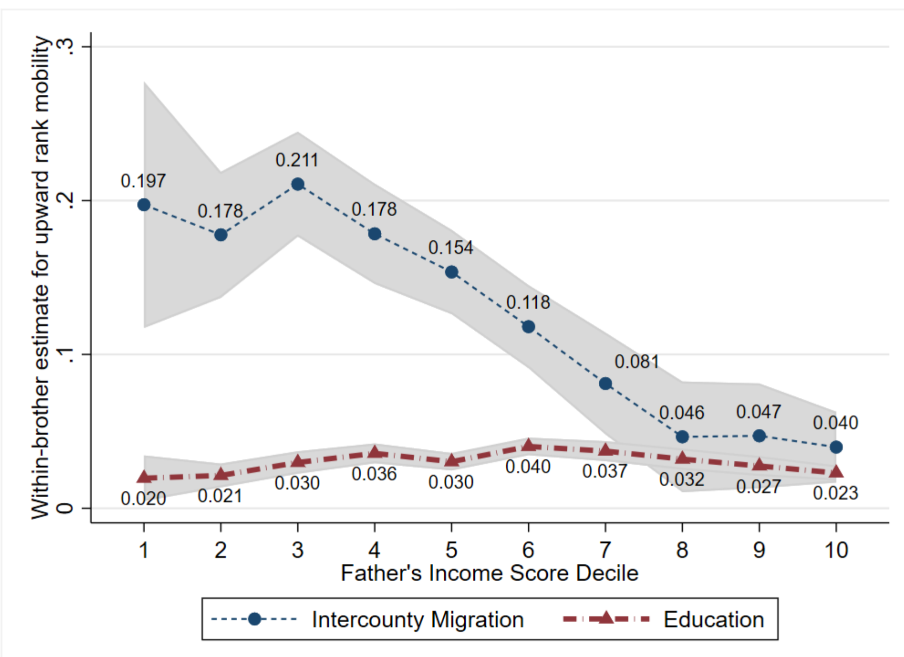
Notes: This figure recreates the results from Figure 5(A) but uses whether the son have a strictly higher income score than the father as the dependent variable.

Figure A3. Upward Rank Measures, robustness check

Panel A. Limit to ages 6 to 14 at first observation



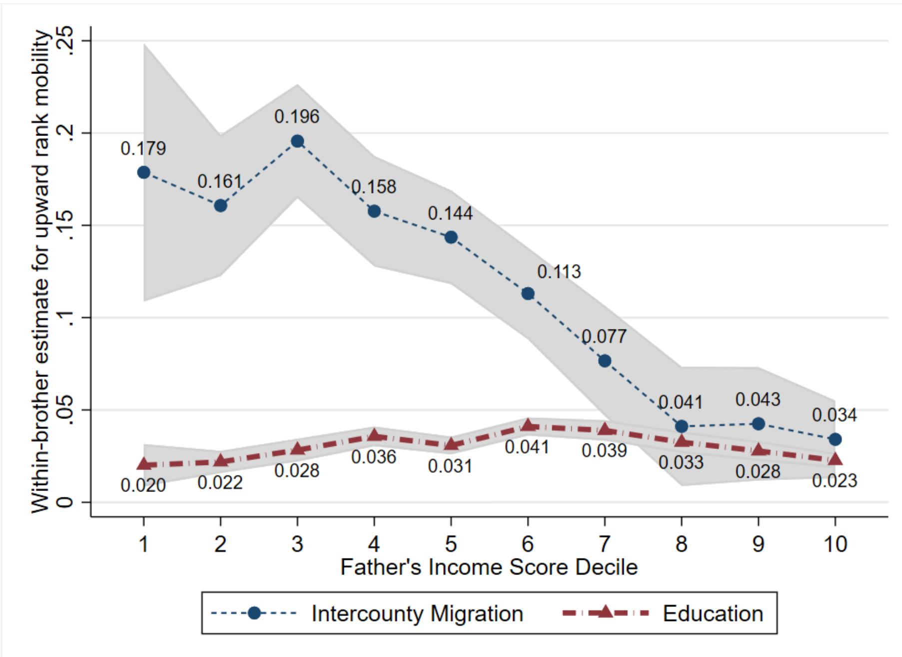
Panel B. Limit migration measure to those migrating between 1920-1930 or 1930-1940 but not 1910-1920



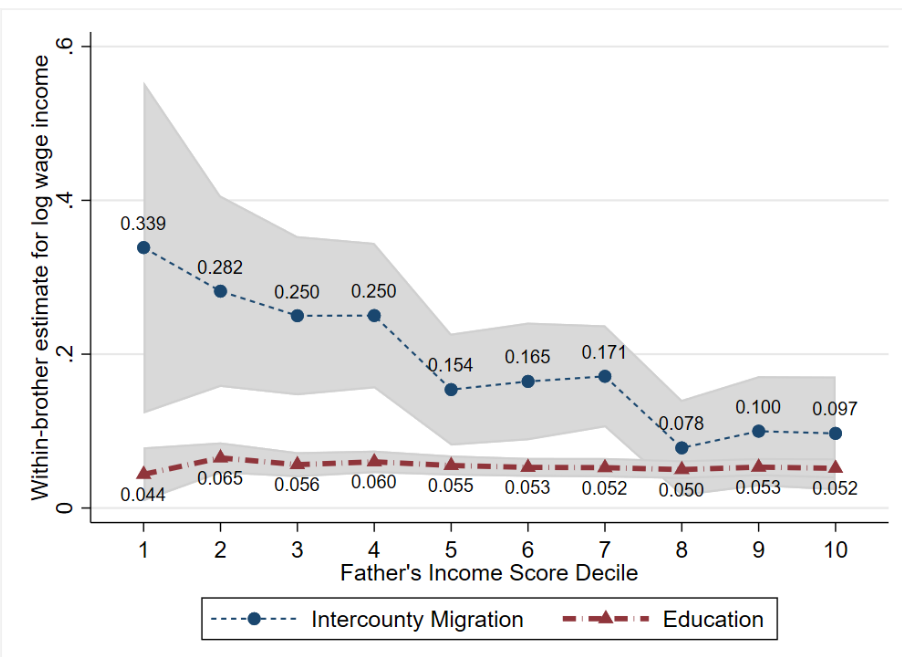
Notes: These figures check whether the results are sensitive to whether the child may have moved with the family or on his own. The first panel limits the sample to those who are older at first observation to ensure that the move was more likely to occur when the child was on his own. The second panel limits the sample to those who moved in either 1920-1930 or 1930-1940 but not 1910-1920 to ensure the child is older at time of move.

Figure A4. Controlling for Great Depression and New Deal Response does not alter estimates

Panel A. Upward Rank Mobility



Panel B. Log Wage Income



Notes: These figures recreate the results from Figures 5(A) and 6(A) from the main text. I additionally control for the log fall in per capita retail sales by county between 1929 and 1933, which controls for the downturn from the Great Depression. I also control for the log per capita relief spending at the county level. These are created after merging data from Fishback et al. (2006) to the 1930 county.

Table A1. Overview of Nominal Income Scores

	National	White	Black	Northeast	Midwest	South	West
Income Scores							
Professional	39,364	40,065	17,749	42,232	37,526	37,941	39,129
Farmers	14,003	14,905	7,744	23,942	15,247	10,655	20,441
Managers and Officials	37,934	38,188	13,999	41,609	37,052	35,113	38,360
Clerical and Kindred	27,008	27,140	22,145	27,436	26,850	26,249	27,922
Sales Workers	28,450	28,606	12,742	30,693	27,739	25,735	28,690
Craftsmen	23,359	23,700	12,507	24,529	23,999	20,519	24,705
Operatives	19,262	19,815	12,494	20,035	20,435	15,888	22,027
Service Workers	19,301	22,315	11,316	22,719	19,745	14,609	22,002
Farm laborers	7,959	8,686	5,296	13,180	7,854	5,783	12,048
Laborers	12,158	13,080	9,427	14,362	13,098	9,070	14,904
Percentile Ranks							
Professional	84	85	35	88	83	79	85
Farmers	25	28	5	59	30	14	48
Managers and Officials	89	90	25	92	88	85	91
Clerical and Kindred	69	70	54	71	69	67	72
Sales Workers	72	73	21	77	71	65	74
Craftsmen	57	58	20	61	59	48	61
Operatives	44	46	20	47	48	32	54
Service Workers	41	51	16	51	43	27	50
Farm laborers	8	10	2	22	7	3	18
Laborers	19	22	10	26	21	9	29

Notes: Data shows the mean income scores within broad occupational categories. The means are calculated using the 1940 outcomes of the sons in the linked sample. The categories are separated by the first-digit of the *occ1950* codes from IPUMS. See Appendix C for description of income scores; generally, they are estimated by one's 3-digit occupation, race and region.

Table A2. Overview of Real Income Scores

	National	White	Black	Northeast	Midwest	South	West
Income Scores							
Professional	33,472	34,064	15,209	35,301	31,445	34,013	32,809
Farmers	13,943	14,799	8,005	22,698	14,928	11,031	20,049
Managers and Officials	32,931	33,154	11,942	35,185	31,753	32,260	32,639
Clerical and Kindred	22,881	23,020	17,763	22,794	22,477	23,309	23,328
Sales Workers	24,207	24,342	10,585	25,679	23,288	23,170	23,971
Craftsmen	20,269	20,572	10,652	20,822	20,509	18,853	21,226
Operatives	16,844	17,325	10,955	17,043	17,496	14,913	19,196
Service Workers	16,252	18,823	9,440	18,620	16,481	12,987	18,365
Farm laborers	7,794	8,470	5,324	12,393	7,618	5,913	11,490
Laborers	10,922	11,818	8,270	12,464	11,558	8,633	13,430
Percentile Ranks							
Professional	83	85	32	86	82	81	84
Farmers	30	34	7	66	34	18	57
Managers and Officials	89	90	21	91	88	88	90
Clerical and Kindred	67	68	47	67	66	68	69
Sales Workers	71	72	16	75	69	68	71
Craftsmen	56	58	16	59	57	50	60
Operatives	43	45	17	44	46	34	53
Service Workers	38	49	12	46	39	27	47
Farm laborers	9	11	2	24	8	3	20
Laborers	17	21	7	23	19	9	28

Notes: Data shows the mean income scores within broad occupational categories. The means are calculated using the 1940 outcomes of the sons in the linked sample. The categories are separated by the first-digit of the *occ1950* codes from IPUMS. See Appendix C for description of income scores; generally, they are estimated by one's 3-digit occupation, race and region.

Table A3. Occupation Transition Matrix for non-migrants

Father in 1910	Son in 1940					
	White Collar	Semi-skilled	Unskilled	Farmer, Owner	Farmer, Tenant	Total
White Collar	36,562	16,943	6,596	1,318	675	62,093
	(58.88)	*(27.29)*	*(10.62)*	*(2.12)*	*(1.09)*	*(100.00)*
Semi-skilled	44,214	57,037	19,457	1,469	875	123,051
	(35.93)	*(46.35)*	*(15.81)*	*(1.19)*	*(0.71)*	*(100.00)*
Unskilled	18,740	27,768	20,667	2,269	2,240	71,684
	(26.14)	*(38.74)*	*(28.83)*	*(3.16)*	*(3.12)*	*(100.00)*
Farmer, Owner	14,366	18,132	22,855	32,657	20,605	108,615
	(13.23)	*(16.69)*	*(21.04)*	*(30.07)*	*(18.97)*	*(100.00)*
Farmer, Tenant	4,495	8,416	12,122	5,683	11,883	42,599
	(10.55)	*(19.76)*	*(28.46)*	*(13.34)*	*(27.89)*	*(100.00)*
Total	118,376	128,297	81,696	43,396	36,277	408,041
	(29.01)	*(31.44)*	*(20.02)*	*(10.64)*	*(8.89)*	*(100.00)*

Notes: Data are from the 1910 and 1940 linked sample. Table only shows results for those who never moved across counties (according to the 1910, 1920, 1930 and 1940 Censuses). Sample is split into occupational categories based on the *occ1950* variable. White collar are professionals, managers, sales and clerical. Semi-skilled are craftsmen and operatives. Unskilled are service workers, farm laborers and laborers. Farmers are separated by owners and tenants based on whether claimed to own a home. Farmers without a farm ownership variable are excluded from the matrix.

Table A4. Occupation Transition Matrix for Ever Migrants

Father in 1910	Son in 1940					
	White Collar	Semi-skilled	Unskilled	Farmer, Owner	Farmer, Tenant	Total
White Collar	57,943	22,985	9,234	1,889	1,461	93,512
	(61.96)	*(24.58)*	*(9.88)*	*(2.02)*	*(1.56)*	*(100.00)*
Semi-skilled	55,091	56,221	20,222	2,478	2,145	136,156
	(40.46)	*(41.29)*	*(14.85)*	*(1.82)*	*(1.58)*	*(100.00)*
Unskilled	25,848	33,293	25,901	2,541	3,669	91,251
	(28.33)	*(36.48)*	*(28.38)*	*(2.79)*	*(4.02)*	*(100.00)*
Farmer, Owner	35,169	40,716	29,442	12,627	14,407	132,362
	(26.57)	*(30.76)*	*(22.24)*	*(9.54)*	*(10.88)*	*(100.00)*
Farmer, Tenant	15,219	26,146	28,228	5,079	13,062	87,735
	(17.35)	*(29.80)*	*(32.17)*	*(5.79)*	*(14.89)*	*(100.00)*
Total	189,269	179,360	113,027	24,615	34,745	541,016
	(34.98)	*(33.15)*	*(20.89)*	*(4.55)*	*(6.42)*	*(100.00)*

Notes: Data are from the 1910 and 1940 linked sample. Table only shows results for those who ever moved across counties (according to the 1910, 1920, 1930 and 1940 Censuses). Sample is split into occupational categories based on the *occ1950* variable. White collar are professionals, managers, sales and clerical. Semi-skilled are craftsmen and operatives. Unskilled are service workers, farm laborers and laborers. Farmers are separated by owners and tenants based on whether claimed to own a home. Farmers without an ownership variable are excluded from the matrix.

Table A5. The within-brother *nominal* migration premium, alternative moves

	Upward Rank	Percentile Rank	Log Wage Income	Log Income Score
Intercounty	0.118	5.693	0.174	0.112
	(0.00504)	(0.243)	(0.0138)	(0.00473)
Interstate	0.170	7.996	0.211	0.154
	(0.00859)	(0.412)	(0.0222)	(0.00804)
Interregion	0.193	9.263	0.227	0.184
	(0.0136)	(0.652)	(0.0383)	(0.0132)
Rural to urban	0.302	16.21	0.463	0.306
	(0.0118)	(0.567)	(0.0325)	(0.0114)
Rural to rural	0.109	4.602	0.208	0.104
	(0.0102)	(0.491)	(0.0395)	(0.0105)
Urban to urban	0.0771	4.000	0.169	0.0641
	(0.0129)	(0.599)	(0.0259)	(0.00981)
Urban to rural	0.00722	-2.133	-0.00203	-0.0382
	(0.0201)	(1.016)	(0.0443)	(0.0171)
White	0.114	5.709	0.166	0.107
	(0.005)	(0.248)	(0.013)	(0.005)
Black	0.206	5.068	0.329	0.219
	(0.042)	(1.176)	(0.112)	(0.040)
0-50 miles	0.072	3.300	0.131	0.066
	(0.008)	(0.358)	(0.020)	(0.007)
50-100 miles	0.113	5.128	0.198	0.100
	(0.009)	(0.439)	(0.024)	(0.008)
100-250 miles	0.149	7.147	0.276	0.134
	(0.009)	(0.414)	(0.022)	(0.008)
250-1000 miles	0.207	10.467	0.367	0.197
	(0.008)	(0.400)	(0.021)	(0.008)
1000+ miles	0.266	14.095	0.304	0.264
	(0.010)	(0.507)	(0.027)	(0.009)
Dust Bowl Migrant	0.191	10.231	0.115	0.191
	(0.036)	(0.777)	(0.048)	(0.014)
Black and from South	0.389	10.88	0.566	0.396
	(0.114)	(3.550)	(0.305)	(0.104)
Permanent	0.149	7.235	0.234	0.140
	(0.00542)	(0.260)	(0.0146)	(0.00505)
Temporary	0.00949	0.223	-0.0504	0.0134
	(0.00780)	(0.377)	(0.0221)	(0.00719)

Notes: Data are from 1910-1940 linked sample. Each cell is a separate regression, always based on within-brother variation. The first row estimates the migration effect for those who are in the North or Midwest census regions in 1940 but were in the South census region in 1910. Dust Bowl migrants are defined based on whether they were in a Dust Bowl county in 1930 and not in one in 1940; non-migrants are those in Dust Bowl counties in both 1930 and 1940. Dust Bowl counties are those where any part of the county had more than 25 percent topsoil erosion (see Hornbeck (2012), Figure 2). Urban is defined as an area with over 2,5000 residents. Miles are measured based on straight-line distances between county centroids. Permanent migrants are those who are in a different county in 1940 as they were in 1910. Temporary migrants are those who are in the same county in 1940 as they were in 1910, but are observed in a different county in either 1920, 1930, or according to the 1935 migration question the 1940 Census.

Table A6. The within-brother *real* migration premium, alternative moves

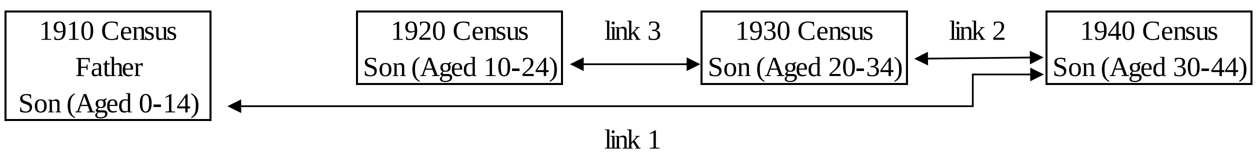
	Upward Rank	Percentile Rank	Log Wage Income	Log Income Score
Intercounty	0.101	4.658	0.157	0.0859
	(0.00502)	(0.255)	(0.0136)	(0.00457)
Interstate	0.139	6.221	0.182	0.112
	(0.00860)	(0.428)	(0.0219)	(0.00769)
Interregion	0.164	7.157	0.189	0.132
	(0.0135)	(0.674)	(0.0374)	(0.0124)
Rural to urban	0.185	7.731	0.318	0.143
	(0.0124)	(0.593)	(0.0321)	(0.0110)
Rural to rural	0.133	6.283	0.240	0.120
	(0.0102)	(0.534)	(0.0394)	(0.0104)
Urban to urban	0.0399	2.329	0.139	0.0332
	(0.0128)	(0.642)	(0.0259)	(0.00984)
Urban to rural	0.142	7.456	0.159	0.123
	(0.0198)	(1.052)	(0.0443)	(0.0170)
White	0.100	4.731	0.151	0.083
	(0.005)	(0.260)	(0.013)	(0.004)
Black	0.128	2.890	0.265	0.146
	(0.043)	(1.159)	(0.109)	(0.037)
0-50 miles	0.071	3.375	0.135	0.062
	(0.008)	(0.379)	(0.020)	(0.007)
50-100 miles	0.100	4.063	0.181	0.074
	(0.009)	(0.461)	(0.024)	(0.008)
100-250 miles	0.119	5.498	0.247	0.095
	(0.009)	(0.438)	(0.022)	(0.008)
250-1000 miles	0.164	7.922	0.320	0.139
	(0.008)	(0.416)	(0.021)	(0.007)
1000+ miles	0.225	11.333	0.259	0.197
	(0.010)	(0.521)	(0.026)	(0.009)
Dust Bowl Migrant	0.166	8.769	0.0929	0.152
	(0.033)	(0.813)	(0.0472)	(0.0138)
Black and from South	0.290	5.391	0.437	0.243
	(0.119)	(3.660)	(0.302)	(0.0992)
Permanent	0.126	5.870	0.211	0.106
	(0.00542)	(0.272)	(0.0143)	(0.00487)
Temporary	0.0117	0.357	-0.0470	0.0148
	(0.00770)	(0.401)	(0.0221)	(0.00705)

Notes: Data are from 1910-1940 linked sample. Each cell is a separate regression, always based on within-brother variation. The first row estimates the migration effect for those who are in the North or Midwest census regions in 1940 but were in the South census region in 1910. Dust Bowl migrants are defined based on whether they were in a Dust Bowl county in 1930 and not in one in 1940; non-migrants are those in Dust Bowl counties in both 1930 and 1940. Dust Bowl counties are those where any part of the county had more than 25 percent topsoil erosion (see Hornbeck (2012), Figure 2). Urban is defined as an area with over 2,5000 residents. Miles are measured based on straight-line distances between county centroids. Permanent migrants are those who are in a different county in 1940 as they were in 1910. Temporary migrants are those who are in the same county in 1940 as they were in 1910, but are observed in a different county in either 1920, 1930, or according to the 1935 migration question the 1940 Census.

Appendix B. Details on linking data

I combine three different linked datasets in this paper: 1910-1940 (sons from childhood to adulthood); 1940-1930 (sons to another observation); and 1930-1920 (sons to another observation) – see Figure B1. The first link (1910-1940) allows me to observe the adult outcomes of both the father and son since I observe fathers and sons at the same time in the household in 1910, and then the son's adult outcome in 1940. The second link takes sons in 1940 and finds them in the 1930 censuses; the third link finds them in the 1920 census. I will describe each of the three links (1910-1940; 1920-1930; and 1930-1940) links in detail.[20] Note that all links are made in the same basic way (that is, based on the method described in Feigenbaum (2016)).

Figure B1. Linking Process to build dataset



Building the set of potential matches.

I build new datasets of US-born whites and US-born blacks by linking the 1910-1940, 1920-1930 and 1930-1940 censuses. I use the same broad strategy as in Feigenbaum (2016) where I build a set of potential links, handlink a subset of them, and then train a probit to pick the best link.

I first extract the entire set of US-born white and black males who are over 10 and under 40 years of age in both 1920 and 1930. For the 1910-1940 link, I extract US-born sons who are 0-14 years old from the 1910 census. After dropping those with the exact same combinations of first name string, last name string, race, state of birth and year of birth, I then search for all possible combinations in the second census that meet the following criteria:

1) First letter of first name match
2) First letter of last name match
3) Jaro-Winkler distance of first name is less than 0.20
4) Jaro-Winkler distance of last name is less than 0.20
5) Year of birth is less than three years in difference

[20] Note that the 1910-1940 link is the same one created in Kosack and Ward (2018, Appendix B).

6) State of birth and race match exactly

The first two criteria differ from Feigenbaum (2016), who does not block on first letters of last or first name; I keep these criteria to reduce computing costs and keep the matching process managable when matching complete to complete-count censuses. The race match requirement also misses some matches because race identification may change between censuses; therefore, the US-born black results only apply to fathers and sons listed as black in all censuses. Finally, I do not block on mother or father's state of birth because there appears to be some error in how these variables are recorded, perhaps because another person of the household was answering the enumerator for the entire household. (Further, mother and father's state of birth available in the 1940 census). However, mother and father's state of birth is useful for choosing the best matches for the 1920-1930 and 1930-1940 links, so I will incorporate it into the probit model.

Based on these linking criteria, not everyone in the starting census has a potential match in the second census. For the white population, about 70 to 80 percent of the starting sample has a possible match in the second census; for the black poulation, only about 60 percent of the starting sample has a possible match. The different rates for the black and white population may reflect differential mortality between the two groups, or that true matches in the black population are less likely to meet the above criteria. Either way, the results suggest that the maximum linking rate is not near 100 percent even if I could find a true link among the set of potential matches. However, I first need to determine which of the potential matches is the true link.

Choosing the best link.

After creating the set of potential matches, I draw a sample of 2,000 black and 2,000 white individuals and all of their potential matches in the later census. I do this each for the 1920-1930 match and the 1930-1940 match, where the 2,000 are drawn from the starting census. These three datasets will form the basis of the training data, but first I need to handlink the people in the dataset.

From the dataset of potential matches, I handpick which is the best match. If there are two close potential links that look similar to the original link, then I do not pick a match since I am not confident which one is the true link. The matching rates for the training data are given in Table B1. After going through this handlinking process, I am able to find a true link for about 60 percent of the white population with at least one potential match, and about 50 percent of the black population with at least one potential match. Part of the reason why I fail to find a link for all of the training data is because none of the potential links are close in names or year of birth; part of the reason is because there are multiple good matches.

Table B1. Details for the handlinked dataset

	1910-1940		1920-1930		1930-1940	
	White	Black	White	Black	White	Black
Random sample in base year	2,000	2,000	2,000	2,000	2,000	2,000
Potential links in second census	16,248	9,695	15,547	9,302	14,550	9,137
Successfully linked	1,121	862	1,224	992	1,198	958
Handlinking Rate for training data (given 1 potential match)	56.1	43.1	61.2	49.6	59.9	47.9

Notes: Data are from the handlinked sets from 1910-1940, 1920-1930 or 1930-1940.

With the training dataset of potential links and actual links in hand, I model the true link as a function of observable differences between matches. I include the Jaro-Winkler distance in the first and last name; absolute difference in year of birth; number of potential links and its square; mother's place of birth and father's place of birth. I also include information on whether there are unique and exact matches for either the first or last name in terms of NYSIIS codes or exact string match; this is based on the handlinking process where having the same last name that was unique (that is, no other potential links has the same last name) was a strong predictor of a link. The probit models for each of the 1920-1930 and 1930-1940 matches, separately by black and white, are shown in Table B2, and the one for the 1910-1940 link is in Table B3.

The probit models estimate a predicted match score for each potential link in the training dataset. From this information, I set two tuning parameters to determine who will be included in my linked dataset. The first parameter is the cut off for predicted probability, where a potential link needs to have a predicted probability above this level to be included in the linked dataset. The second parameter is the ratio of the 1st best probability to the 2nd best probability; this ensures that I do not keep a match that has a close alternative. I set these parameters to maximize the efficiency of the algorithm in terms of true positive rate (TPR, or the percentage of true links that I keep), as long as the positive predictive value (PPV) is at least 0.9. The positive predictive value is the ratio of true positives to total matches; viewed from the opposite direction, it sets the false positive rate to 10 percent. This false positive rate is slightly lower than Feigenbaum's (2016) training data in Iowa and thus is on the conservative end; however, one could easily change this parameter to be more or less restrictive. A consequence of the decision to limit false positives is that it reduces the matching rate for the full sample. See Table B4 for the tuning parameters and the resulting PPV and TPR.

Table B2. Predicting the handlinked match using a probit model

	1920-1930 White	1920-1930 Black	1930-1940 White	1930-1940 Black
Jaro-Winkler Distance, First name	-6.073***	-6.992***	-7.622***	-6.012***
	(0.553)	(0.530)	(0.669)	(0.542)
Jaro-Winkler Distance, Last name	-13.07***	-13.11***	-14.38***	-13.17***
	(0.869)	(1.012)	(0.969)	(0.977)
Year of Birth Difference = 1	-0.180	-0.172	-0.678***	-0.247
	(0.128)	(0.178)	(0.138)	(0.161)
Year of Birth Difference = 2	-0.544***	-0.254	-1.102***	-0.442***
	(0.144)	(0.173)	(0.165)	(0.158)
Year of Birth Difference = 3	-0.920***	-0.634***	-1.546***	-0.630***
	(0.156)	(0.179)	(0.191)	(0.164)
No. of potential links	-0.0257	-0.100***	-0.0932***	-0.126***
	(0.0192)	(0.0199)	(0.0212)	(0.0219)
No. of potential links squared	-9.29e-05	0.00253***	0.00211***	0.00301***
	(0.000668)	(0.000803)	(0.000739)	(0.000901)
Unique and Exact NYSIIS First name match	0.315**	0.266**	-0.0212	0.121
	(0.156)	(0.132)	(0.164)	(0.119)
Unique and Exact NYSIIS Last name match	-0.00698	0.797**	0.0484	0.157
	(0.304)	(0.340)	(0.307)	(0.509)
Unique and Exact NYSIIS First AND Last name match	0.616***	0.961***	1.062***	0.994***
	(0.128)	(0.119)	(0.134)	(0.115)
Unique Exact Last name String match	0.491***	0.137	0.769***	-0.242
	(0.180)	(0.219)	(0.215)	(0.226)
Middle initial match, if have one	0.718***	1.557***	1.121***	1.163***
	(0.113)	(0.387)	(0.139)	(0.271)
NYSIIS last name match AND Year of Birth Diff=0	1.539***	0.850***	0.889***	1.343***
	(0.245)	(0.285)	(0.235)	(0.453)
NYSIIS last name match AND Year of Birth Diff=1	1.172***	0.790***	0.915***	1.147**
	(0.241)	(0.274)	(0.251)	(0.452)
NYSIIS last name match AND Year of Birth Diff=2	0.914***	0.169	0.470*	0.792*
	(0.255)	(0.270)	(0.275)	(0.445)
2 Potential links with NYSIIS last name match	-0.582***	-0.498**	-0.622***	-0.811***
	(0.178)	(0.228)	(0.193)	(0.260)
>2 potential links with NYSIIS last name match	-0.952***	-0.286	-0.684***	-1.014**
	(0.227)	(0.263)	(0.228)	(0.439)
2 Potential links with last name string match	-0.798***	-0.450**	-0.935***	-0.321*
	(0.151)	(0.182)	(0.184)	(0.193)
>2 Potential links with last name string match	-1.415***	-1.168***	-1.433***	-1.501***
	(0.133)	(0.144)	(0.142)	(0.144)
One potential link	0.636***	0.645***	0.986***	0.980***
	(0.170)	(0.133)	(0.180)	(0.123)
Difference in length of last name strings	-0.383***	-0.479***	-0.619***	-0.552***
	(0.0586)	(0.0711)	(0.0738)	(0.0683)
Mother place of birth match	0.653***	0.327***		
	(0.0784)	(0.119)		
Father place of birth match	0.546***	0.0173		
	(0.0775)	(0.110)		
Constant	0.269	0.648***	1.845***	1.322***
	(0.182)	(0.229)	(0.190)	(0.200)
Observations	15,547	9,302	14,205	9,096

Notes: Data are from the handlinked sample between 1920-1930 or 1930-1940. The coefficients are from a probit model that predicts the correct link.

Table B3. Modeling the linking process with a probit, 1910-1940

	White	Black
Jaro-Winkler Distance, First name	-4.885***	-4.203***
	(0.576)	(0.635)
Jaro-Winkler Distance, Surname	-13.64***	-12.35***
	(0.853)	(1.113)
Year of Birth Difference = 1	-0.557***	0.0740
	(0.114)	(0.198)
Year of Birth Difference = 2	-0.906***	-0.199
	(0.131)	(0.202)
Year of Birth Difference = 3	-1.426***	-0.355*
	(0.157)	(0.203)
Number of Potential links	-0.114***	-0.129***
	(0.0187)	(0.0235)
Number of Potential links squared	0.00217***	0.00283***
	(0.000639)	(0.000922)
Exact surname match AND unique surname	0.619***	0.228
	(0.235)	(0.303)
Exact first and surname string match AND unique first and surname string	0.382**	0.763***
	(0.159)	(0.154)
Exact first name match AND unique first name	-0.391*	-0.0317
	(0.205)	(0.178)
Exact Soundex first name match AND unique soundex first name	0.277	0.149
	(0.279)	(0.202)
Exact Soundex surname match AND unique soundex surname	-0.238	0.489***
	(0.192)	(0.168)
Exact Soundex first and surname match AND unique soundex first and surname	0.829***	0.622***
	(0.188)	(0.158)
Exact NYSIIS first name match AND unique NYSIIS first name	0.204	0.489**
	(0.298)	(0.215)
Exact NYSIIS surname match AND unique NYSIIS surname	0.445	-2.197
	(0.320)	(126.1)
Exact NYSIIS first and surname match AND unique NYSIIS first and surname	0.0126	0.253
	(0.196)	(0.171)
Middle initial match, if have one	1.212***	1.068***
	(0.103)	(0.201)
NYSIIS last name match AND YOB Diff=0	1.131***	4.537
	(0.234)	(126.1)
NYSIIS last name match AND YOB Diff=1	1.066***	4.175
	(0.243)	(126.1)
NYSIIS last name match AND YOB Diff=2	0.795***	3.757
	(0.255)	(126.1)
2 Potential links with NYSIIS last name match	-0.308**	-0.951**
	(0.147)	(0.407)
>2 potential links with NYSIIS last name match	-0.637***	-3.921
	(0.224)	(126.1)
2 Potential links with last name string match	-1.090***	-0.682**
	(0.181)	(0.273)
>2 Potential links with last name string match	-1.575***	-1.194***
	(0.123)	(0.162)
One potential link	0.398**	0.407***
	(0.173)	(0.143)
Constant	1.370***	0.318
	(0.167)	(0.233)
Observations	16,248	9,695

Notes: This paper shows a regression of whether one is a true link on observable characteristics. Data set is the set of potential matches in 1940 for a random sample of 2,000 individuals in 1910 with one potential link.

Table B4. Tuning parameters for determining who to keep in the linked sample

Census Years	Race	Cutoff for predicted probability	Score Ratio of 1st best link to 2nd best	PPV	TPR
1910-1940	White	0.335	2.6	0.901	0.790
	Black	0.784	5.8	0.901	0.580
1920-1930	White	0.412	2.5	0.900	0.786
	Black	0.587	2.8	0.901	0.596
1930-1940	White	0.33	4.4	0.900	0.836
	Black	0.639	1.7	0.901	0.631

Notes: PPV stands for positive predictive value and gives the ratio of true positives to all links. TPR stands for true positive rate and gives the proportion of true links that would appear in the final linked dataset.

Provided these estimates from the probit model, I then predict the linking scores for the full to full count match with the probit model; afterwards, I keep only those who meet the parameters set in Table B4. See Table B5 and B6 for the linking rates when applying this process to the full-count data. I link of 32 to 36 percent of the white population, and 15 to 17 percent of the black population for the decadal links. The 1910-1940 link is 29.8 for the white population and 11.9 for the black population. These linking rates are lower than Feigenbaum's link from the 1915 Iowa Census to the 1940 Federal Census of near 60 percent. This may be due to several reasons: because Iowa is a smaller state and thus has fewer other potential matches, because the data quality is higher from Iowa, because modelling the hand-linking process is easier for Iowans versus the rest of the country, or because there are lower mortality rates for Iowans relative to the rest of the country. I also have more restrictive requirements for a potential link (requiring the first letter of the first name and first letter of the last name to match exactly). While the linking rate is somewhat low, I still have millions of individuals linked across censuses.

Table B5. Applying the probit model to the full 1920-1930 and 1930-1940 link, details

	1920-1930 Census		1930-1940 Census	
	White	Black	White	Black
Starting group in base year	21,234,490	2,819,891	25,676,888	3,193,903
Starting group in base year with a potential link ten years later	16,320,377	1,489,797	19,144,294	1,979,364
Potential links ten years later	159,810,633	7,071,600	258,313,387	9,217,794
Unique match amongst links	7,559,217	418,958	8,197,666	539,359
Overall Linking Rate	**35.6**	**14.9**	**31.9**	**16.9**
Linking Rate given Potential Match	46.3	28.1	42.8	27.2

Table B6. Applying handlinking results to full 1910-1940 link, details

	Anglo American	African Americans
Starting group in 1910	12,567,861	1,851,076
Starting group in 1910 with a potential link in 1940 based on criteria	10,180,244	1,094,394
Potential links in 1940	136,372,727	6,085,262
Unique match in 1940 amongst links	3,748,917	220,145
Overall Linking Rate	**29.8**	**11.9**
Linking Rate given Potential Match	36.8	20.1

Getting into the sample used in the main analysis

To be included into the final linked sample used in this paper, an individual must be in the 1910-1940, 1930-1940 and 1920-1930 link. That is, the son must survive being triple linked. The resulting sample is of 949,333 sons. This number is only 9.1 percent of the 1910 children that I could have possibly linked to the 1940 census. Given that the general linking rate of two censuses is around 33 percent, it would be expected that about (0.33)(0.33)(0.33) = 3.6 percent of individuals would be linked three times if linking is independent across censuses. The actual linking rate is higher than 3.6 percent, indicating that being successfully linked is not independent, perhaps because individuals have unique names. The linking rate for black sons is even lower at 2.0 percent of the original sons; this reflects that the black linking rate is lower at around 15 percent.

Weighting

Only a select group (9.1 percent) of the original population shows up in the triple-linked sample. Therefore, this group may be unrepresentative of the original population and provide misleading information on the convergence of economic gaps. I address this problem by reweighting the data to be representative of the population. To weight the data, I use the inverse probability weight approach as suggested by Bailey et al. (2017). The process is as follows: pool the linked and linkable sample together (that is, the children in 1910), run a probit to determine which observables predict being in the linked sample, and then weight each observation in the linked sample using the inverse probability weight.[21]

The representativeness of the sample (as found in the probit model) is shown in Table B7. I run representativeness checks separately by the white population and black population so that I tailor the weights to be race specific. Table B7 shows that there is selection into the linked sample, where fathers with white-collar jobs and farmers are more likely to be in the sample than unskilled or semi-skilled fathers. Further, those in the Midwest and West are more likely to be in the sample than those in the South or in the Northeast. Therefore, estimating the mobilities using the unweighted data will erroneously reflect Midwestern rural states like Iowa, rather than the full population. The weighted representative characteristics are also shown in Table B7.

The final step to the weighting process is to upweight African Americans. Since the regressions in Table B7 are done separately by black and white, the average weight for the white sample is about one and the average weight in the black sample is about one. However, since the linking rate is lower for African Americans, they only end up being 2.22 percent of linked sample, in contrast to 9.25 percent of population (of 30-44 year old adults in 1940). Therefore, I reweight the black sample up by multiplying the black weights by 9.25/2.22; I also reweight the white sample down by multiplying their weights by 90.75/97.78. Now the black weighted proportion of the sample reflects the population proportion.

[21] Let q be the share of linked records and p be the predicted probability. The weight is $[(1-p)/p] \times [q/(1-q)]$

Table B7. Representativeness of the linked sample on a probit

	White, unweighted	White, weighted	Black, unweighted	Black, weighted
Length of First name	0.0267***	0.00387***	0.0247***	0.00340**
	(0.000310)	(0.000316)	(0.00157)	(0.00153)
Length of last name	0.0203***	0.000425	0.00220	0.000912
	(0.000308)	(0.000322)	(0.00172)	(0.00180)
Urban	-0.0390***	0.00777***	0.0169**	0.00815
	(0.00141)	(0.00150)	(0.00818)	(0.00853)
Father has white-collar job	0.0879***	-0.00281	0.150***	0.0104
	(0.00177)	(0.00186)	(0.0201)	(0.0213)
Father is farmer	0.124***	-0.00610***	0.156***	0.00625
	(0.00177)	(0.00185)	(0.0145)	(0.0155)
Father has unskilled job	-0.0145***	0.000245	0.150***	-0.00289
	(0.00165)	(0.00175)	(0.0141)	(0.0152)
Age in 1910	-0.0231***	-0.00122**	-0.0185***	-0.00308
	(0.000461)	(0.000474)	(0.00229)	(0.00230)
Age in 1910 squared	0.000818***	7.73e-05**	0.000302*	0.000220
	(3.21e-05)	(3.34e-05)	(0.000160)	(0.000165)
Midwest	0.141***	0.00703***	0.0378**	6.14e-05
	(0.00140)	(0.00145)	(0.0151)	(0.0139)
South	-0.0976***	0.0177***	-0.326***	0.00157
	(0.00160)	(0.00174)	(0.0124)	(0.0118)
West	0.203***	0.0100***	0.0604*	-0.0139
	(0.00213)	(0.00215)	(0.0361)	(0.0325)
Constant	-1.543***	-1.217***	-1.874***	-1.969***
	(0.00374)	(0.00386)	(0.0247)	(0.0248)
Observations	10,462,321	10,462,321	1,070,349	1,070,349

Notes: The regression is the pooled linkable and linked sample between 1910 and 1940. The dependent variable is an indicator for being in the linked sample. The likelihood of being in the linked sample is modelled in a probit. The probit coefficients are reported in this table.

Appendix C. Income Score

The main income score used for estimating rank-rank associations follows the process of Collins and Wanamaker (2017), who provide more detail on assumptions behind the adjustments. I provide the general strategy here for the interested reader. The benefit of the score over the traditional occupational score based on *occscore* from IPUMS is that this income score further adjusts income by region of residence and race. These adjustments are key for estimating the economic benefit from internal migration, since income gaps were large across geography and black/white families.

I first use the 1940 census and limit the sample to 25 and 55-year-olds to measure occupational-based earnings at prime ages in the lifecycle. I then separate the sample into cells based on 3-digit occupational code (occ1950), race/ethnicity (black/white/Mexican), and 9 census regions. After splitting the 1940 full-count census into cells, I use the average income in the cell as the income score. While this may seem straightforward, a few further corrections need to be made before taking the average income to address that the 1940 census only includes wage income but not self-employed earnings. That is, I need to estimate how much self-employed workers earn. To do so, I take the ratio of total income to wage income by occupational code in the 1960 census. I then multiply this ratio by the average wage income in the 1940 census in the occupation/race/state cell for self-employed workers.

I also make adjustments for farmers and farm laborers to address that some compensation may be in-kind. To do so, in the 1960 census I increase farmer income by 35 percent and farm laborer income by 19 percent (Collins and Wanamaker, 2017). Then, I take the ratio of farmer to farm laborer income in the 1960 census in order to proxy for how much farmers earn over farm laborers in the 1940 census. Before multiplying this ratio by farm laborer wages in the 1940 census, I increase farm laborer wages by 26 percent in 1940 to reflect perquisies in this earlier period. Note that I use the same ratios for land-holding farmers (who are assumed to be farm owners) and non-land-holding farmers (who are assumed to be farm tenants). According to my data, this results with increasing farmer income by about 43 percent to account for perquisites.

After these adjustments, the average earnings in the occupation/race/region cell is the income score. If there are less than 30 people in the regional cell, then the income score is the national average in the occupation/race cell. Finally, if there are less than 30 people in the national cell, then the income score is the average income at the 1-digit level by race.