

Jackson, Osborne

**Working Paper**

## The impact of migration on earnings inequality

Working Papers, No. 19-5

**Provided in Cooperation with:**

Federal Reserve Bank of Boston

*Suggested Citation:* Jackson, Osborne (2019) : The impact of migration on earnings inequality, Working Papers, No. 19-5, Federal Reserve Bank of Boston, Boston, MA

This Version is available at:

<https://hdl.handle.net/10419/202933>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Impact of Migration on Earnings Inequality

Osborne Jackson

**Abstract:**

This paper examines the impact of migration on earnings inequality using 1940–2015 data from the U.S. census and American Community Survey. Despite measurement challenges, I successfully replicate existing findings regarding national trends in earnings inequality and migration, and subsequently analyze regional and state patterns. Using 1940 birthplace information to instrument for migration, I find that recent immigration mildly increases the top decile earnings share, while recent in-migration and out-migration have no significant effects on such inequality. I estimate that immigration contributed 5.8 percent to the observed rise in U.S. earnings inequality from 1950 to 2015, primarily through a non-migrant channel.

**JEL Classifications:** R23, F22, D31

**Keywords:** migration, earnings inequality

---

Osborne Jackson is a senior economist in the research department at the Federal Reserve Bank of Boston. His e-mail address is [osborne.jackson@bos.frb.org](mailto:osborne.jackson@bos.frb.org).

The author thanks Jeff Thompson, Bob Triest, and seminar participants at the Federal Reserve Bank of Boston. Excellent research assistance was provided by Kevin Behan and Thu Tran.

This paper presents preliminary analysis and results intended to stimulate discussion and critical comment. The views expressed herein are those of the author and do not indicate concurrence by the Federal Reserve Bank of Boston, the principals of the Board of Governors, or the Federal Reserve System. The author is responsible for any errors.

This paper, which may be revised, is available on the website of the Federal Reserve Bank of Boston at <http://www.bostonfed.org/economic/wp/index.htm>.

**This version: August 2018**

# 1 Introduction

In the United States, inequality in income, earnings or wages, wealth, and consumption are areas of enormous discussion among academics, policymakers, and the public. Several studies have now determined that such inequality has generally been rising in the country for several decades (for example, Bricker, Henriques, Krimmel, and Sabelhaus 2015; Fisher, Johnson, Smeeding, and Thompson 2018; Kopczuk, Saez, and Song 2010; Piketty and Saez 2003; Saez and Zucman 2016). Nevertheless, much of the dialogue surrounding this issue remains centered on (1) how much inequality exists and how it has changed over time, as well as (2) how such inequality is best addressed, if at all.

Regarding topic (1), the current literature generally focuses on documenting national trends in inequality of various forms. However, there is much less discussion regarding movements in inequality at lower levels of geography, such as the state level (an exception being Sommeiller, Price, and Wazeter 2016, which examines income inequality at the state, county, and metro area levels). Regarding topic (2), effective policy to mitigate inequality, if desired, relies on correctly assessing the amount and causes of such disparities.

This paper therefore focuses on one type of inequality and one potential determinant: the impact of migration on U.S. earnings inequality. I focus on earnings inequality not only because it is directly tied to labor markets and how they function, which is of substantive interest, but also because such inequality may be easier to measure than overall income, wealth, or consumption, especially with the household survey data that I utilize.<sup>1</sup> Additionally, I isolate migration as a key determinant to study due to its clear mechanisms. Specifically, the extent of earnings inequality in a labor market depends on participants in that market and how their respective earnings affect the distribution. Thus, the impact of migration on the earnings distribution can be partitioned into contributions from (1) migrants, whose movements alter the population and associated earnings with which inequality is measured

---

<sup>1</sup>I also focus on earnings rather than wages in part due to the absence of hours data throughout my sample period in the survey data used.

(“migrant channel”), and (2) non-migrants, whose earnings and possibly related traits may be affected by migration (“non-migrant channel”).<sup>2</sup> Migration flows into or out of labor markets could affect earnings inequality through either channel or both.

This study will therefore examine the causal impact of migration on earnings inequality. Using data from the U.S. decennial census and American Community Survey from 1940 to 2015, I first document national, regional, and state-level trends in earnings inequality and migration, discussing measurement considerations and adding to the literature in both areas. Even absent administrative data and in the presence of top coding, I am able to closely replicate current findings that demonstrate increasing earnings inequality from 1950 to 2015. I am similarly able to reproduce migration trends documented in other studies, including a recent decline in interstate migration over the past few decades (for example, see Kaplan and Schulhofer-Wohl 2017; Molloy, Smith, and Wozniak 2011; Molloy, Smith, and Wozniak 2017).

I then examine the impact of state-level, mutually exclusive flows of immigration, in-migration, and out-migration on state-level earnings inequality.<sup>3</sup> Using 1940 historical birth-place information to instrument for observed migration flows from 1950 to 2015, I find that recent immigration—inflows from abroad of foreign-born persons 1 to 11 years earlier—increases earnings inequality, as measured by the top decile share of earnings. Meanwhile, recent in-migration and out-migration—internal inflows and outflows, respectively, of domestic residents 1 to 11 years earlier—have no significant effects on inequality. The observed immigration effect operates primarily through the non-migrant channel. I estimate that immigration accounted for 5.8 percent of the observed rise in U.S. earnings inequality from

---

<sup>2</sup>The non-migrant channel will depend on theories of labor demand, such as the degree of substitutability or complementarity of workers in production. For instance, if high-skilled non-migrant labor is a complement to high-skilled migrant labor, one might observe that inflows of high-skilled immigrant workers increase earnings inequality by raising the productivity (and thereby earnings as well) of high-skilled non-migrant labor. Alternatively, demand for labor that is affected by employer discrimination may alter non-migrant earnings in the presence of immigration, even absent any change in observed or unobserved non-migrant characteristics. Thus, a migration-related change in earnings for non-migrants may occur with or without a change in non-migrant traits.

<sup>3</sup>These flows are mutually exclusive for a given state. However, by definition, out-migration from one state corresponds to in-migration for another state.

1950 to 2015, and that 94 percent of that immigration contribution (that is, 5.4 percent of the observed rise in inequality) occurs via changes to the earnings of non-migrants in the labor force.

This paper thus adds to a small, mixed literature on this topic. For instance, Card (2009) finds modest, positive effects of immigration on the wage inequality of natives. When incorporating the wages of immigrants themselves too, he finds a larger but still small effect of immigration on overall wage inequality. Conversely, Ganong and Shoag (2017), in studying declining regional income convergence, find that reduced migration arising from supply-side housing constraints helps contribute to greater wage inequality. Thus, there is no consensus on the sign or magnitude of the impact of migration on inequality, and the effects may differ by the type of inequality examined, the type of migration flow, or both.

The remainder of the paper is organized as follows. Section 2 discusses the data used for the study. Section 3 details how earnings inequality and migration are measured, and provides national and local trends concerning both. Section 4 presents the main findings regarding a causal effect of migration on inequality and explores the mechanisms for the estimated effects. Section 5 concludes.

## 2 Data on Earnings and Migration

I use U.S. decennial census data from 1940 to 2000 (1 percent samples in 1940, 1950, and 1970 [the “Form 2 state sample,” also known as the “15 percent state sample”], and 5 percent samples in 1960, 1980, 1990, and 2000), and American Community Survey (ACS) data from 2010 and 2015, which are 1 percent samples (Ruggles, Genadek, Goeken, Grover, and Sobek 2017).<sup>4</sup> Census data before 1940 do not contain migration information, and I exclude other annual ACS data due to limited year-to-year variation in migration or earnings inequality. Aggregated across years, these micro samples contain more than 58.4 million individuals.

Thus, one advantage of using census and ACS data is the large sample count across sur-

---

<sup>4</sup>I also utilize 2000 ACS data to adjust migration flows, as discussed in section 3 and the Appendix.

veys. Additionally, the availability of earnings and migration information within the same sample benefits analysis and ensures that subsamples reflect the identical underlying population. Lastly, the access to numerous individual-level characteristics allows for the exploration of migration and more covariates than one might be afforded by administrative data. Likewise, other sources of data on migration may be more limited in scope or flexibility to define migrants (for example, the Current Population Survey [CPS], which before 1981, as Molloy, Smith, and Wozniak [2017] explain, asked migration questions only in 1964 through 1971 and 1975). Or they may be unable to measure earnings inequality (for instance, the Internal Revenue Service [IRS] migration data, which also only partially capture the migrating population because not all individuals file tax returns).

However, such advantages of these household survey data are not costless. For instance, one disadvantage is the existence of top coding on earnings. Another drawback is heterogeneous information across years on migration flows. Both of these issues, and how I address them, are discussed in depth in section 3 and the Appendix.

I restrict the sample along several dimensions. Due to the focus on earnings inequality, I examine individuals with positive real earnings (constant 1999 U.S. dollars).<sup>5</sup> I also drop individuals who are listed as not being in the labor force (a restriction I relax later when analyzing effect mechanisms in section 4), and further ensure this constraint by restricting the sample to those who are aged 16 to 64. The lower bound age threshold is motivated by state compulsory schooling laws across the country, while the upper bound age threshold is driven by the likely age of retirement of many persons in the data.<sup>6</sup> Finally, I also drop persons living in group quarters, if they have not already been eliminated by the previous restrictions. The resulting data sample contains more than 22.2 million persons.

---

<sup>5</sup>Nominal earnings are adjusted to real terms using Consumer Price Index (CPI-U) data from the U.S. Bureau of Labor Statistics.

<sup>6</sup>In census years 1940 to 1970, persons 14 and 15 years old are also in the universe of those allowed to be in the labor force, according to online documentation (Ruggles, Genadek, Goeken, Grover, and Sobek 2017). However, I exclude these individuals in order to keep the age range in the sample consistent over time.

## 3 Measurement of Inequality and Migration

### 3.1 Measuring Earnings Inequality

#### 3.1.1 Measurement Considerations

As noted earlier, I focus on earnings inequality because it is a measure of disparity directly tied to labor markets and because it is easier to measure, especially with household survey data. However, as mentioned in section 2, a potential concern regarding measuring earnings inequality using census and ACS data is that the top coding of earnings censors true labor income for individuals highest in the earnings distribution. Top coding occurs at various values throughout the sample years.<sup>7</sup>

Because many of the top coding thresholds are expressed in nominal terms that do not differ by state, I determine the state-specific percentile at which the top coding value occurs in each sample year. Across states and sample years, the lowest state-specific percentile is 97.7, corresponding to Connecticut in 2000. Thus, as a sensitivity check on the potential impact of top coding, some of the inequality measures I construct exclude earnings above the state-specific 97th percentile. However, as will be shown, top coding ultimately proves not to be a prohibitive issue for capturing trends in earnings inequality established in other studies where top coding is absent due to the use of administrative data.

#### 3.1.2 Approaches and Comparison to Existing Estimates

Even when one focuses on a particular type of inequality—namely, earnings—the question remains of how to measure it, and different approaches exist. For instance, one could measure how much of total earnings is held by persons in a subset of the top of the earnings distribution, such as the top decile. Alternatively, one could examine the ratio of earnings

---

<sup>7</sup>Specifically, top coding occurs at \$5,001 in 1940, \$10,000 in 1950, \$25,000 in 1960, \$50,000 in 1970, \$75,000 in 1980, \$140,000 in 1990 (with higher values expressed as the state medians of values above \$140,000), \$175,000 in 2000 (with higher values expressed as the state means of values above \$175,000), and at the state-specific 99.5th percentile of earnings in 2010 and 2015. Note that all monetary thresholds are in nominal terms.

between a person at one point in the earnings distribution (for instance, the 90th percentile) compared with a person at another point in the distribution (for instance, the median).

Across the various chosen measures of earnings inequality, I focus on the upper portion of the earnings distribution as the segment of interest. As Piketty and Saez (2003) and other studies have established, much of the observed growth in income inequality that has occurred in the United States has been driven by individuals at the top of the income distribution, suggesting that this is the subset of the population to focus on (and thus also motivating the aforementioned sensitivity check regarding top coding). Appendix Figure A1 similarly shows that from 1940 to 2015, there was not much of a change in the U.S. ratio of earnings at the 50th percentile compared with earnings at the 10th percentile, whether one examines the nation as a whole or the nine census divisions of the country.<sup>8</sup> Given this observed pattern, combined with absolute growth in real earnings throughout the distribution, Appendix Figure A1 suggests that any measured growth in earnings inequality during this period is primarily driven by faster growth at the top of the earnings distribution. For this reason, I also do not include the Gini coefficient as one of my measures of earnings inequality. The Gini coefficient, by reflecting the overall dispersion in earnings, would capture changes in inequality that correspond to the full earnings distribution, rather than focusing on the top of the earnings distribution.

Figure 1 therefore examines U.S. earnings-distribution trends from 1940 to 2015 using four measures that concentrate on the upper portion of the distribution while also, in some cases, circumventing potential top coding issues as a robustness check: (1) the top decile share of total earnings, (2) the share of earnings from those at or below the 97th percentile being held by those in the 90th to 97th percentiles, (3) the ratio of earnings at the 97th percentile compared with earnings at the 50th percentile, and (4) the ratio of earnings at the 90th percentile compared with earnings at the 50th percentile. As the figure shows,

---

<sup>8</sup>The nine census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC).



while the interpretation and corresponding level of each measure differs, the overall pattern is consistent across the measures. Namely, inequality fell starkly from 1940 to 1950, but then rose thereafter, exceeding the 1940 level by 2015. I plot each of these measures against the data, extended, from Piketty and Saez (2003), which reflects the top decile share of total earnings from 1927 to 2011 using annual tax returns from the IRS. As Figure 1 shows, all four of the inequality measures that I calculate correspond to the trends observed in the Piketty and Saez data. Moreover, even in the presence of top coding, the top decile share of earnings that I calculate from household survey data is remarkably close in magnitude to the analogous measure from Piketty and Saez, at all available points in time for comparison.

The only slightly puzzling result to resolve is that in Figure 1, as compared with Piketty and Saez (2003), I observe greater inequality in 1970 and, to a lesser degree, 1960. Such a pattern is unexpected given the presence of top coding in my household survey data and the absence of top coding in the Piketty and Saez administrative tax return data. This finding is likely due to differences in the underlying samples examined in my analysis versus Piketty and Saez (2003). For instance, while I study persons, the authors study tax units. Additionally, their sample excludes self-employed workers.

In Appendix Figure A2, I further adjust my sample to exclude self-employed workers as well as persons below age 25, because the latter are more likely to be dependents who do not file their own tax returns. As the figure shows, after I make these additional adjustments, measured inequality in my study always lies below inequality in the Piketty and Saez data, as expected. This result provides some assurance that top coding in my household survey data has anticipated but not prohibitive effects on observed inequality trends. Thus, throughout the subsequent analysis, I focus on the top decile share of earnings as my preferred measure of earnings inequality and use the two terms synonymously.

While Figure 1 focuses on changes in national earnings inequality, it is also informative to examine changes in the top decile earnings share across and within regions. Figure 2 shows that, across U.S. census divisions, the national pattern regarding earnings inequality

from 1940 to 2015 likewise holds for each of the nine census divisions displayed, albeit it with some differences. For all regions other than East South Central (comprising Alabama, Kentucky, Mississippi, and Tennessee), inequality in 2015 exceeds inequality in 1940, as it does in the nation. And in four regions (New England, Middle Atlantic, West South Central, and Pacific), inequality in 2015 is greater than national inequality.

Figure 3 displays trends in the top decile earnings share within each of the nine census divisions. Once again, the national and regional patterns are borne out within each region, both qualitatively and quantitatively. It's also worth noting a distinction regarding the variance of earnings inequality across regions versus the variance within each region, and how those variances have changed over time. Specifically, the variance across regions of the top decile share of earnings is nearly 10 times as great in 1940 as in 2015.<sup>9</sup> In contrast, within each region, the variance across states of the top decile share of earnings is smaller in 1940 than 2015 in five regions (New England, Middle Atlantic, East North Central, West North Central, and Pacific), and larger in 1940 than 2015 in the remaining four regions (South Atlantic, East South Central, West South Central, and Mountain).

Finally, Table 1 reports mean values of earnings inequality, averaged over the entire 1940–2015 sample period for areas (that is, states plus Washington, D.C., or alternatively, regions and the United States), or averaged across states (including Washington, D.C.) in each year. The table shows that, for instance, among the states, average earnings inequality in the sample period is highest in Georgia and lowest in Wyoming. Among regions, the inequality is highest in the West South Central region and lowest in the East North Central region. Lastly, in terms of time periods, and as expected given the figures, inequality is highest in 2015 and lowest in 1950, rising consistently since the latter date. Moreover, on average across states, earnings inequality from 2000 onward is at levels not seen since 1940.

---

<sup>9</sup>The variance is 0.000992 in 1940 and 0.000105 in 2015.

## 3.2 Measuring Migration Flows

### 3.2.1 Measurement Considerations

I now turn to measuring migration, focusing on recent flows in each included census or ACS sample for three, mutually exclusive types of migration: immigration, internal in-migration, and internal out-migration.<sup>10</sup> I start by defining those flows. I classify immigration as inflows from abroad of persons who are either noncitizens or naturalized citizens.<sup>11</sup> I classify inflows and outflows within the country by domestic residents as internal in-migration and out-migration, respectively. These three flows are distinct phenomena, and internal migration accounts for only current residence, not birthplace. For instance, a foreign-born person immigrating to the United States from Germany will contribute to immigration inflows. If this same individual relocates within the United States several years later, he or she would then also contribute to internal in-migration for the location moved to, and internal out-migration for the location moved from.

To further specify migrant flows, as Molloy, Smith, and Wozniak (2011) discuss, migration researchers typically make two decisions to define migrants: (1) “[T]hey choose geographic units to define potential origin and destination locations”; and (2) “they define the time period in which individuals must move between origins and destinations.” These decisions will thus define “internal” and “recent” in my earlier description of the migration flows. Regarding both terms, I focus on migration flows from the prior year that are going to U.S. states (immigration, interstate in-migration) or coming from U.S. states (interstate

---

<sup>10</sup>Emigration is necessarily omitted, because capturing this flow would require analyzing household survey data from all foreign destinations of former U.S. residents.

<sup>11</sup>Regarding further details and exceptions, I treat all countries, territories, and U.S. possessions whose inhabitants are statutory U.S. citizens as non-foreign (for example, Puerto Rico). Also, if individuals may have been born abroad of U.S. parents, I do not allow them to contribute to immigration inflows. Specifically, for sample years 1940, 1950, and 1980 to 2015, and using citizenship and birthplace information, I recode a foreign-born indicator as 0 if a person was born abroad to American parents, or if they are residents of a country where “automatic U.S. citizenship” applies and are not already coded as native-born, leading to a recoding of persons from the Northern Mariana Islands. For sample years 1960 and 1970, because citizenship information is not available, I use parental nativity as a rough proxy for who would be a citizen at birth, reclassifying the individual as native-born if either parent is native-born. This adjustment is motivated by domestic rules regarding citizenship. For more details, see <https://travel.state.gov/content/travel/en/legal-considerations/us-citizenship-laws-policies/citizenship-child-born-abroad.html>.

out-migration).

The choice of “prior year” as the time horizon for the migration flows stems from data availability in the census and ACS being limited to one-year or five-year flows. I opt to focus on the former in part because longer time horizons for migration flows are more likely to miss intermittent relocations. For instance, suppose a current Massachusetts resident has lived in the state for six months, but had lived in New York for the five preceding years. Such an individual would thus be recorded as an in-migrant to Massachusetts from New York according to both a one-year in-migration measure and a five-year in-migration measure. Now consider a current Massachusetts resident who has lived in the state for six months, but lived in New York for only four years preceding Massachusetts residency and lived in Massachusetts for a year before New York residency. This individual would be recorded as an in-migrant to Massachusetts from New York according to a one-year migration measure, but listed as a non-migrant according to a five-year migration measure. Thus, although five-year migration is meant to capture individuals who migrated in the prior five years, certain short-term relocations may be missed. Thus, I concentrate on one-year migration because this shorter time horizon likely captures a larger share of the relevant movements within the time period. Additionally, given this time horizon, it is worth noting that “non-migrants” reflect persons who have migrated less recently (for instance, two years earlier or five years earlier), as well as persons who have never migrated.<sup>12</sup> Meanwhile, the selection of states as the relevant local labor market helps maintain geographically consistent market boundaries throughout the sample, as well as ensuring full coverage of both rural and urban areas (Molloy, Smith, and Wozniak 2011).<sup>13</sup> I rescale each type of migration flow by the size of the labor force in the local labor market.

With those two choices made, the primary difficulty in using census and ACS data spanning 1940 to 2015 to identify prior year migration is that the time horizon of the recorded

---

<sup>12</sup>Although some results may be heterogeneous across these two groups, for the purposes of this study, I do not distinguish them.

<sup>13</sup>Additionally, for some descriptive analysis, I examine regions or the nation as a whole as the relevant labor markets.

flows is heterogeneous across sample years, as mentioned. In 1950, 2010, and 2015, I observe the location (state or country) in which a person resided one year earlier, for all individuals age 1 or older. In contrast, in 1940 and 1960 to 2000, I observe the location (again, state or country, excluding country in 1960 because previous location abroad is not available in that year) in which a person resided five years earlier, for all individuals age 5 or older.<sup>14</sup> Thus, I need to convert five-year migration flows into one-year migration flows so that all of the flows represent prior year migration and are comparable across years. Before determining a reasonable conversion process (detailed in the Appendix), I turn to existing literature documenting U.S. migration flows. Such established migration patterns help serve as a basis of comparison to validate the converted one-year migration flows.

### 3.2.2 Features of Existing Estimates and Approaches

Regarding documented prior year migration trends in the literature, I focus on U.S. interstate migration. Figure 4 displays patterns in prior year interstate migration arising from three recent papers: Molloy, Smith, and Wozniak (2011) (top); Molloy, Smith, and Wozniak (2017) (middle); and Kaplan and Schulhofer-Wohl (2017) (bottom). When one analyzes these studies, at least two stylized facts emerge that are useful to bear in mind for validating this paper’s estimates of prior year migration. First, depending on the data source and underlying sample of the population, the *level* of interstate migration as a share of the population from 1948 to 2013 ranges between 1 percent and 4 percent. Second, the size of the observed *change* in interstate migration between 1980 and 2010 appears to be roughly in the  $-0.5$  to  $-1.5$  percentage point range, again varying by the data source and underlying population.<sup>15</sup>

Thus, in my own migration estimates, a useful check on any conversion procedure to generate

---

<sup>14</sup>Regarding the exclusion of immigration information in the 1960 5 percent sample, IPUMS notes that this is a “restoration” sample that “includes restored data originally missing from 1960 PUMS,” and as such, may be subject to more missing or inconsistent information than non-restoration samples (Ruggles, Schroeder, Rivers, Alexander, and Gardner 2011).

<sup>15</sup>The long-run decline in interstate migration documented by these studies arguably begins as early as the 1970s. However, given the presence of some missing data in the 1970s (in the CPS, as discussed earlier), as well as a positive spike in migration rates in 1990, I opt to focus on the 1980–2010 period for data validation purposes.

one-year migration flows throughout the 1940–2015 sample is to confirm that I can replicate the aforementioned results regarding the level and change in interstate migration rates.

In the Appendix, I detail various possible conversion approaches to generate one-year migration flows throughout the estimation sample. Using my preferred approach, Figure 5 shows rates from 1940 to 2015 of prior year U.S. interstate migration and immigration as a share of the area labor force. In the figure, one-year migration flows in 1940 and 1960 to 2000 are adjusted from five-year flows by dividing the latter by a scaling factor.<sup>16</sup> Focusing on interstate migration, the figure depicts patterns in such migration that are fairly close to the trends in Figure 4. In both Figures 4 and 5, the level of interstate migration is between 1 percent and 4 percent (specifically, from 1.8 percent to 3.1 percent in Figure 5), and the change in such migration between 1980 and 2010 is in the desired  $-0.5$  to  $-1.5$  percentage point range (specifically,  $-0.9$  percentage point in Figure 5). Thus, my preferred conversion approach appears to do a reasonable job of converting five-year migration flows to one-year flows.

Additionally, Figure 6 displays examples of region-level (New England) and state-level (Massachusetts) estimated prior-year migration rates, separately for immigration, in-migration and out-migration. Figure 6 highlights differences in migration patterns at the regional or state level compared with national trends, including distinguishing between in-migration and out-migration versus interstate migration only. Table 1 reports mean values of each of the three migration flows, averaged over the 1940–2015 sample period for states and regions, or averaged across states in each year. For instance, the table shows that average immigration inflows are highest in Washington, D.C., at 0.8 percent, and lowest in Arkansas at 0.0 percent (0.046 percent, rounded).

---

<sup>16</sup>Specifically, scaling factor 3 in the Appendix,  $\psi_{ijt,2000}$ , where  $i = U.S.$ ,  $j = \{\textit{interstate migration, immigration}\}$ , and  $t = \{1940, 1960, \dots, 2000\}$ .

### 3.3 Counterfactual “No Migration” Earnings Inequality

Before turning to regressions analyzing the impact of migration on earnings inequality, I delve a bit deeper into the mechanisms that might drive such an effect. As discussed earlier, any observed effect would stem from either (1) how migrants alter the population and associated earnings with which inequality is measured (“migrant channel”) or (2) changes in the earnings of those who do not migrate and possibly related observed or unobserved traits (“non-migrant channel”).

Regarding the migrant channel, for instance, an internal outflow of very high-earning individuals from New York to California might reduce inequality in New York and increase inequality in California due solely to the migrants themselves and their relative locations in the earnings distributions of those two states. Regarding the non-migrant channel, the aforementioned in-migration to California might further increase inequality in the state by increasing the earnings of non-migrants. Such an increase in non-migrant earnings might occur absent changes in non-migrant traits (for instance, due to employer discrimination), or rather, due to changes in non-migrant traits that may be observed (for example, adjustments in non-migrant educational attainment) or unobserved (for instance, production complementarities between migrants and non-migrants that increase non-migrant productivity).

To separate the migrant and non-migrant channels, I would ideally be able to measure earnings inequality in a state of the world without migration. If I could then examine the impact of migration on such counterfactual “no migration” inequality, any causal effect estimated would necessarily arise due to the non-migrant channel only. Although such counterfactual inequality is not observable, a reasonable proxy can be obtained by recalculating earnings inequality for a state-year population that “reverts” prior year migration—that is, reassigns estimated prior year immigrants, in-migrants, and out-migrants back to their source locations. This proxy thus assumes that only the location of migrants is affected by their moves and not their earnings amount, or that earnings across locations are highly

correlated even if not perfectly correlated.<sup>17</sup>

Upon calculating such counterfactual earnings inequality and comparing it with actual inequality, I find only small differences between the two measures.<sup>18</sup> Figure 7 shows actual versus counterfactual earnings inequality for two sample states, Maryland and Connecticut. In Maryland, migrant channel effects from migration tend to decrease inequality because of where migrants are drawn from in the earnings distribution, whereas in Connecticut, migrant channel effects from migration tend to increase inequality. However, the size of the migrant effect is quite small, both in absolute (top subfigures) and proportional (bottom subfigures) terms, falling between  $-0.51$  percent and  $0.17$  percent in Maryland and between  $-0.17$  percent and  $1.15$  percent in Connecticut. Such small migrant effects are not surprising, though, given the small share of the labor force that prior year migrants tend to comprise, as Figures 5 and 6 indicate.

Table 2 lists average percent differences between actual and counterfactual earnings inequality, ranked separately by state, region, and year. Maryland, as shown in Figure 7, is one of the states where migrant channel effects of migration tend to decrease earnings inequality during the 1940–2015 sample period; it ranked fourth lowest, above only Alaska, Delaware, and Wyoming. Conversely, Connecticut, as shown in Figure 7, is one of the states where migrant channel effects tend to increase earnings inequality; it ranked third highest, behind South Dakota and Hawaii. For most regions and the nation as a whole, the earnings distribution of one-year migrant flows tends to increase earnings inequality, although the proportional effects are quite small. New England, with a  $0.17$  percent difference on average between actual and counterfactual inequality, displays the largest migrant channel effect. East North Central, with a  $-0.03$  percent difference, shows the smallest migrant effect. East

---

<sup>17</sup>This assumption could perhaps be tested by using longitudinal data and examining the earnings correlation across locations in narrow periods that stratify interstate moves.

<sup>18</sup>I use scaling factor version 3, as described in the Appendix, in order to determine a random share of observed five-year migrants ( $1/\psi_{ijt,2000}$ ) in relevant sample years to reflect estimated prior year migrants. I then reassign these individuals, along with observed one-year migrants in relevant sample years, to their source locations. The remaining five-year migrants ( $1 - [1/\psi_{ijt,2000}]$ ) in relevant sample years I designate as non-migrants (that is, those estimated not to have migrated in the prior year), along with observed non-migrants in relevant sample years containing one-year migration flows.



North Central is also one of only three regions with a lower migrant effect than that of the nation, and it is the only region with a negative effect, reflecting that the earnings distribution of migrant flows tends to decrease earnings inequality on average. Finally, Table 2 also shows year-specific averages (across the 50 states and Washington, D.C.) in actual and counterfactual inequality, as well as their proportional difference. There does not appear to be any particular pattern over time in the migrant channel effect of migration, and 1950 is the only year in the sample when this effect tends to decrease inequality.

## 4 The Impact of Migration on Earnings Inequality

### 4.1 Estimation Strategy

Previous sections show, descriptively, that interstate migration in the country was declining for much of the same period that earnings inequality was increasing. The question remains, however, whether there is a causal link between changes in interstate migration rates or immigration rates, at the state level, and state-level earnings inequality. The paper now turns to implementing such causal estimation. For state  $i$ , year  $t$ , and migration flow type  $j$ , the basic ordinary least squares (OLS) estimating equation is:

$$Y_{1it} = \mu + \sum_{j=1}^3 \beta_j M_{ijt} + \phi_i + \lambda_t + \varepsilon_{it}, \quad (1)$$

where  $Y_{1it}$  is the top decile's share of total earnings;  $M_{ijt}$  is the share of persons in the labor force migrating in the prior year and distinguishes immigration ( $j = 1$ ), in-migration ( $j = 2$ ), and out-migration ( $j = 3$ );  $\phi_i$  and  $\lambda_t$  capture state and year fixed effects, respectively; and  $\varepsilon_{it}$  is an error term, with standard errors clustered at the state level to allow for possible within-state correlation of errors. In more stringent specifications, I include division-year fixed effects,  $\kappa_{rt}$ , in lieu of year effects, to account for region-level macroeconomic shocks that vary over time and where  $r$  indicates the nine census divisions discussed in section 3.

Due to this array of fixed effects, additional controls are omitted, which also is helpful in the presence of limited degrees of freedom. However, I will explore possible effect mechanisms as alternative outcomes in later estimation, to be discussed. Parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  reflect the short-run effects of immigration, in-migration, and out-migration on earnings inequality, respectively. These effects operate through migrant and non-migrant channels, as discussed in section 3, with both mechanisms reflected in the  $\beta$  coefficients.

To separately identify the aforementioned migrant and non-migrant channels, I can instead estimate an alternative version of equation (1) as follows:

$$Y_{2it} = \zeta + \sum_{j=1}^3 \alpha_j M_{ijt} + \chi_i + \tau_t + \nu_{it}, \quad (2)$$

where  $Y_{2it}$  is the counterfactual, “no migration” version of the top decile’s share of total earnings. Thus, parameters  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  reflect the short-run effects of immigration, in-migration, and out-migration on earnings inequality, respectively, that operate solely through the non-migrant channel. Additionally,  $(\beta_j - \alpha_j)$  for any given migration flow type  $j$  captures the short-run, migrant channel effects of migration on inequality.

Other variants of equations (1) and (2) that I also estimate are specifications including a one-period lag of the migration regressors, as follows:

$$Y_{1it} = \mu + \sum_{j=1}^3 \beta_j M_{ijt} + \sum_{j=1}^3 \beta_{j+3} M_{ijt-1} + \phi_i + \lambda_t + \varepsilon_{it}, \quad (3)$$

$$Y_{2it} = \zeta + \sum_{j=1}^3 \alpha_j M_{ijt} + \sum_{j=1}^3 \alpha_{j+3} M_{ijt-1} + \chi_i + \tau_t + \nu_{it}. \quad (4)$$

Because my estimation sample uses decadal data from 1940 to 2010 plus data from 2015, the lagged migration in equations (3) and (4) corresponds to migration flows 11 years earlier (for data from 1950 to 2010) or 6 years earlier (for data from 2015), rather than the prior year migration captured by the migration regressors without lags. Such estimation allows for the

exploration of long-run effects of each of the migration flow types  $j$  on earnings inequality. Parameters  $\beta_1$  through  $\beta_6$  of equation (3) capture the long-run effects of immigration, in-migration, and out-migration on earnings inequality, while parameters  $\alpha_1$  through  $\alpha_6$  of equation (4) capture analogous long-run effects of migration on earnings inequality in the absence of migrant channel effects.

Lastly, to explore potential changes in observed non-migrant traits that could partly underlie effects of migration on inequality that operate through the non-migrant mechanism, I estimate versions of earlier specifications with certain non-migrant variables,  $\mathbf{Y}_{kit}$ , as alternative outcomes. The  $k$  variables in the vector  $\mathbf{Y}$ , which pertains solely to non-migrants in the estimation sample, are the share employed, as well as the shares with fewer than four years of high school, four years of high school, three years or fewer of college (which, for naming ease, I will call “some college,” noting that this does not reflect degree completion status), and four years of college or more. Examining these auxiliary regressions should yield insights into whether changes in observed non-migrant traits, primarily those related to skills (via education), play a notable role in the non-migrant channel, or rather if alternative factors matter—namely, changes in non-migrant earnings conditional on stable non-migrant traits (for instance, through employer discrimination), or changes in non-migrant unobserved traits that could affect non-migrant earnings (for instance, productivity changes due to potential production complementarities with migrants).

#### 4.1.1 Instrumental Variables

OLS estimation of equations (1) to (4) and related analysis will be biased if migration flows are endogenous. For instance, building on a model by Roy (1951), Borjas (1987) suggests that immigrants might self-select based on their skill level and the distribution of earnings in potential destinations. He proposes that high-skilled immigrants will choose destinations with a larger variance of earnings where they can reap the return of being high-skilled, while low-skilled immigrants will choose destinations with a smaller variance of earnings in order to

insulate themselves from being low-skilled. However, if the skill composition of immigrants is balanced, or if such selection along the skill dimension or other dimensions doesn't occur, then it's not clear that immigrant flows or other migrant flows will be endogenous in the context of earnings inequality.

Nevertheless, to protect against the possibility of such OLS bias, I propose alternative estimation by instrumental variables (IV). The instruments use the 1940 U.S. distribution of foreign-born and native persons with different birthplaces to form predictions about the flow of migrants over the sample period, 1950 to 2015. The instruments are motivated by the immigration literature, which proposes that the existence of immigrant networks is an important determinant of the location choices of prospective immigrants because they increase cultural benefits (for example, due to common preferences) and reduce information costs of migration into local markets (for example, Bartel 1989 and Card 2001). I use this reasoning regarding immigrant networks and apply it more broadly to generate three instruments for all three types of migration flows in this study, discussed in further detail in the Appendix.<sup>19</sup>

## 4.2 Main Results

Table 3 displays results from estimation of equations (1) and (2). OLS results for equation (1) reveal large differences in coefficients and explanatory power between specification 1 with no effects and specifications 2 and 3 with various fixed effects. For instance, Model 1 explains only 16 percent of the spatial and temporal variation in the top decile share of earnings, while model 3 explains 96 percent of such variation. Therefore, I focus on the most stringent specification that includes state and division-year fixed effects. Doing so, I observe significantly positive relationships between the top decile share and immigration rates as well as out-migration rates. A 1.0 percentage point increase in the immigration rate is associated

---

<sup>19</sup>Boustan, Fishback, and Kantor (2010) similarly mention this approach in their study of internal migration. However, they opt for another approach due to their examination of U.S. history preceding 1940, which was the first year in which the Census Bureau asked residents about internal mobility.

with a 1.2 percentage point increase in the top decile's share of total earnings, while a 1.0 percentage point increase in the out-migration rate is associated with a 0.14 percentage point increase in the top decile share. In-migration similarly has a positive but not statistically significant association with earnings inequality.

While the immigration coefficient is an order of magnitude larger than the out-migration coefficient, it's worth noting that the mean immigration share of the labor force in the data is 0.2 percent, while the mean in-migration share is 3.5 percent and the mean out-migration share is 3.7 percent. Thus, a 1.0 percentage point increase in immigration is a huge, 500 percent increase, while a 1.0 percentage point increase in out-migration is a much smaller, 27 percent increase. Focusing on 10 percent increases for both migration rates thus results in associated increases in the top decile share of 0.02 percentage point for immigration and 0.05 percentage point for out-migration. Thus, when the migrant flows have comparable proportional increases, the resulting relationship with earnings inequality is similar. To give the migration flow rates some context in absolute terms, the mean state-year labor force size in my sample is 2,135,447 persons. Thus, a mean immigration share of 0.2 percent is equivalent to roughly 4,271 prior year immigrants, with a proportional increase of 10 percent therefore equivalent to 427 persons.

Because a causal interpretation of the OLS coefficients may be problematic, as discussed earlier, I also estimate equations (1) and (2) by IV. Appendix Table A1 shows reasonable F-statistics for the first stage regressions examining the impact of the three predicted migrant rate instruments on each of the rates of immigration, in-migration, and out-migration, for the most stringent specifications with state and division-year fixed effects included. Additionally, the significant coefficients are generally of the expected, positive sign. Thus, I reasonably suspect that weak instruments are not a large problem in this estimation.<sup>20</sup>

---

<sup>20</sup>Stock and Yogo (2005) do not provide critical values in their paper for the just-identified case of three endogenous variables and three instruments. However, for the case of two endogenous variables and two instruments, regarding the actual size of hypothesis testing when the nominal size is 5 percent, the critical values for actual test sizes of 10 percent and 15 percent are 7.03 and 4.58, respectively. The F-statistics I obtain in specifications 7 to 9 of Appendix Table A1, ranging from 6.25 to 10.84, all exceed at least one of those critical values if not both (and noting that appropriate corresponding critical values for three

IV results in column 6 of Table 3 show that in-migration and out-migration have no significant causal impact on earnings inequality. However, immigration does have such an impact, with a 1.0 percentage point increase in the immigration share of the labor force causing a significant, 3.83 percentage point increase in the top decile earnings share.<sup>21</sup> Figure 5 shows that during the estimation period, the immigration rate rose from 0.14 percent in 1950 to 0.31 percent in 2015, an increase of 0.17 percentage point. Thus, given the IV coefficients, this rise in immigration would cause an increase of  $3.83 \times 0.17 = 0.65$  percentage point in the top decile share of earnings from 1950 to 2015. My measure of the U.S. top decile earnings share rises from 24.01 percent in 1950 to 35.24 percent in 2015, an increase of 11.23 percentage points. Thus, the IV analysis suggests that immigration contributed to  $(0.65/11.23) \times 100 = 5.79$  percent of the observed rise in national earnings inequality.

As discussed previously, such causal effects of migration on inequality could operate through either migrant channel effects regarding the location of the earnings distribution from where migrants are drawn or non-migrant channel effects due to migration-induced changes in non-migrant earnings. To disentangle these channels and their relative importance, I re-run the estimation on the counterfactual earnings inequality measure developed in section 3. Upon doing so, I obtain coefficients that are typically smaller in magnitude but remain very similar, which is not surprising given the small size of prior year migration flows. As column 9 shows, when I remove migrant channel effects, I still find that a 1.0 percentage point increase in immigration causes a significant 3.58 percentage point increase in the top decile share of earnings via the non-migrant channel alone. Therefore, the observed rise in immigration from 1950 to 2015 would cause a  $3.58 \times 0.17 = 0.61$  percentage point increase in the top decile share, suggesting immigration contributed  $(0.61/11.23) \times 100 = 5.43$  percent of the observed rise in U.S. earnings inequality through changes in non-migrant earnings. In

---

endogenous variables and three instruments, if available, would be lower).

<sup>21</sup>For a diagnostic of model misspecification, I also examine estimation by weighted least squares and weighted IV, where the weights are the state-year specific population counts (Solon, Haider, and Wooldridge 2015). I obtain coefficients that are quantitatively similar, suggesting that the model with state and division-year fixed effects is specified correctly.

other words,  $(5.43/5.79) \times 100 = 94$  percent of the contribution of immigration to the rise in national earnings inequality occurred through the non-migrant channel.

As mentioned, this non-migrant channel reflects migration-induced—and now, more specifically, immigration-induced—changes in non-migrant earnings. Such earnings changes could occur absent changes in non-migrant characteristics (for instance, due to employer discrimination) or alternatively, as a result of changes in observed or unobserved non-migrant characteristics. For instance, focusing on potential changes in observed non-migrant traits, perhaps non-migrants respond to immigration by adjusting their skills in a manner that exacerbates inequality. I will explore such potential responses of non-migrant observed characteristics in the next section of the paper.

The positive causal relationship between immigration and earnings inequality that I observe here, as well as a larger but still relatively small effect I observe when considering the contributions of both non-migrants and migrants rather than non-migrants alone, is fairly consistent with the findings of Card (2009). Focusing on wages, not earnings, he finds an approximate contribution of immigration to wage inequality (as measured by the variance of wages) from 1980 to 2005/2006 of 4 percent to 6 percent, which is similar to this paper’s 5.79 percent contribution of immigration to earnings inequality from 1950 to 2015. However, unlike this paper, Card (2009) finds that most of the observed impact of immigration on wage inequality operates through the distribution of immigrant wages, rather than through the distribution of native wages.

In Table 4, I use IV to also estimate long-run effects (that is, 6 to 11 years) of migration on earnings inequality, reflecting equations (3) and (4) and now instrumenting for six endogenous migration flows per specification. Here, again focusing on the most stringent specifications 3 and 6 that examine effects on inequality and counterfactual inequality, respectively, I observe qualitatively similar results. Namely, the only significant effects of migration on the top decile earnings share occur via immigration, with a similar coefficient of 2.9, and a comparatively smaller coefficient of 2.7 for the counterfactual decile share specification that

removes migrant channel effects. However, Table 4 shows that the effect appears to operate through the long-run, previous period migration flow rather than the prior year flow, which now has a negative effect on earnings inequality that is not significantly different from zero. In light of this finding, I reinterpret the main IV results in Table 3 to actually reflect a longer, 6- to 11-year time horizon regarding the impact of migration on earnings inequality, rather than just a 1-year time horizon. I opt for such reinterpretation of the results in Table 3, instead of using the results in Table 4 as my preferred estimates, because given limited degrees of freedom, I am more confident in credible IV identification in the presence of three endogenous variables rather than six endogenous variables.<sup>22</sup>

### 4.3 Non-Migrant Channel: Observed Non-Migrant Traits

I now turn to further exploration of the non-migrant channel that accounts for the majority (94 percent) of the effect of migration—namely, immigration—on earnings inequality. I focus here on whether migration flows affect certain observed non-migrant characteristics, thus either contributing to or mitigating observed inequality. As discussed, I return to the specification presented in Table 3, now reinterpreting the results from such estimation to reflect a broader time horizon for the migration flows (namely, flows within the previous 6 to 11 years). Although I focus on non-migrant observed traits here, it remains possible, as stated earlier, that unobserved non-migrant traits also respond to higher rates of immigration to help explain the non-migrant channel, or that immigration-induced changes in non-migrant earnings occur even absent changes in non-migrant traits.<sup>23</sup>

---

<sup>22</sup>Beyond these results, one might still wonder whether the observed IV effects of migration on inequality actually reflect a cumulative effect resulting from less-recent migrants who, along with those who never migrate, comprise the non-migrants in the sample. This would effectively amount to a reverse causality concern. If less-recent migrants, who partially account for the non-migrants comprising the majority of the inequality effect, influence immigration inflow rates, then this might reflect the true channel through which variation in immigration affects earnings inequality. However, while such a concern would bias OLS estimation, it should not affect valid IV estimation, and so such cumulative effects are not a concern.

<sup>23</sup>In additional analysis, I further explore the overall non-migrant channel (which may operate through any of the aforementioned three mechanisms) by examining changes in non-migrant earnings across deciles of the earnings distribution as well as educational attainment categories. Although largely inconclusive, this analysis provides some suggestive results, indicating positive welfare implications for all non-migrants from



Table 5 shows that increases in the immigration rate do not appear to have a significant effect on the non-migrant employment share. However, immigration has a large, significantly positive effect on the share of non-migrants in the labor force with fewer than four years of high school education. A 1.0 percentage point increase in the immigration rate causes a 9.0 percentage point increase in the share of this “high school dropout” proxy, although this response may mitigate observed earnings inequality rather than contribute to it.<sup>24</sup>

I also observe in Table 5 that immigration causes a large, 7.3 percentage point decrease in the share of non-migrants with four years of high school. It is possible that the two significant effects in the table are causally linked, with some individuals who would have attained four years of high school education in the absence of immigration attaining less schooling in the presence of immigration. Some existing studies suggest that there is indeed such a link between immigration and native educational attainment (for example, Jackson 2015 and McHenry 2015). However, another possibility is that no educational attainment is counterfactually changed. Rather, it could be that there is an adjustment regarding the non-migrants in the population who are in the labor force, thereby affecting the distribution of education in the labor force if the aforementioned labor force participation change is heterogeneous across educational attainment groups.

Given the different qualitative implications of these two potential explanations, I take further steps to determine which one is more relevant for explaining my findings. Table 6 reexamines the same effects as Table 5, but now for the entire non-migrant population rather than only non-migrants in the labor force. If the effects from Table 5 were due solely or primarily to labor force participation effects, they would disappear in Table 6 or at least greatly diminish. However, to the contrary, the observed non-migrant education responses among those with fewer than four years of high school and those with four years of high

---

immigration-driven changes in earnings inequality. Such positive welfare results would align with studies that have found beneficial effects of immigration on innovation, suggesting a positive impact on economic output (for example, Kerr and Lincoln 2010).

<sup>24</sup>Because degree completion is not available throughout the sample period, I cannot rule out the possibility that some individuals in this group may have obtained a General Equivalency Diploma. Therefore, I refer to this characterization as a proxy.

school are actually slightly larger in magnitude in Table 6.

Consistent with this finding, Table 7 shows that non-migrant labor force participation rates due to immigration are actually negative. A 1.0 percentage point increase in the immigration rate decreases non-migrant labor force participation by 5.2 percentage points overall, with this effect predominantly occurring among those with fewer than four years of high school and monotonically decreasing in magnitude as educational attainment increases. Therefore, the non-migrant education response in Table 5 is actually in spite of, rather than due to, changes in labor force participation. This suggests that this change in non-migrant education is indeed consistent with a counterfactual change due to increased immigration, seemingly operating in part through some individuals opting for this level of education rather than four years of high school. Additionally, although the effects are not significant, I observe in Table 5 that immigration also decreases the share of non-migrants with some college education in the labor force, and increases the share of non-migrants with four years of college or more in the labor force. Taken together, the non-migrant education results are consistent with a “hollowing out” of the education distribution of non-migrants due to increased immigration.<sup>25</sup>

One question that still remains is why non-migrant education adjusts in response to immigration but not in response to in-migration or out-migration. Although this inquiry is challenging to answer causally given the difficulty in credibly identifying education-specific immigration inflows in IV estimation, I can nevertheless explore some suggestive evidence via OLS estimation. Table 8 first establishes that the OLS results regarding the association between migration flows and non-migrant educational attainment are reasonably similar to the IV results in Table 5, in terms of signs as well as magnitudes. Then, in each specification, I interact the immigration rate coefficient with three of four immigrant-specific education shares, thus leaving the baseline immigration rate to reflect the omitted education share.<sup>26</sup> Although

---

<sup>25</sup>Mandelman and Zlate (2016) similarly find a reduction in the middle portion of the occupational distribution due to low-skilled immigration and offshoring.

<sup>26</sup>I omit the highest (lowest) immigration education share in the two specifications with lower (higher) non-migrant education levels.

these results are descriptive in nature and thus should not be interpreted causally, they seem to suggest that highly educated immigrant workers may be complementary to non-migrant workers who have the highest and lowest levels of education while perhaps substitutable for non-migrant workers in the middle of the education distribution.<sup>27</sup> Additionally, immigrants during the estimation period became increasingly highly educated (Appendix Figure A4), with more than half of prior year immigrants in 2015 having four years of college education or more, which exceeds the analogous share for prior year internal migrants (Figure 8). In combination, these findings suggest that the increasingly highly educated immigrant labor force may be what has driven significant share increases at the lower end of the education distribution among non-migrants in the labor force, and perhaps non-migrant share increases (albeit, not significant) at the upper end of the education distribution too.

## 5 Conclusion

This paper examines the extent of a causal effect of migration on earnings inequality. Such an effect could operate through the earnings of migrants or non-migrants in the labor force. Using U.S. census and ACS data, I find that state-level immigration has a small, significantly positive effect on the top decile’s share of earnings in the state, but I find no significant effects of in-migration or out-migration on such earnings inequality. I estimate that immigration accounted for 5.8 percent of the observed rise in U.S. earnings inequality from 1950 to 2015, and that an overwhelming 94 percent of that immigration contribution operates through changes in the earnings of non-migrants in the labor force. Education appears to be a key observed non-migrant trait affected by immigration, with non-migrants in the labor force becoming significantly more likely to have fewer than four years of high school in the presence of higher immigration rates.

---

<sup>27</sup>This result, if also causal, may be partly related to imperfect labor market transferability of immigrant skills from abroad into the United States. For instance, employers in domestic labor markets may perceive some immigrants with four years or more of college education to most closely correspond to non-migrants with three or fewer years of college (“some college”).

These findings suggest that discussion designating immigration as a substantial reason for earnings inequality growth is misguided. That said, because immigration does play a small role in earnings inequality trends, this link may be of interest to state policymakers. Because immigration is regulated at the federal level, state governments have limited ability to influence the level or composition of immigrant inflows into their jurisdiction, other than indirect effects via changes in local economic conditions. However, states experiencing large inflows of immigration may wish to be aware of possible resultant effects on local earnings inequality and respond accordingly, to the extent that such inequality is not desirable. Such a state-level response may take the form of minimum wage increases, greater state income tax progressivity, or other redistributive policies. Additionally, while observed non-migrant responses to increased immigration, such as lower amounts of education, may be optimal given constraints, the state might also play a role in helping alleviate any existent barriers (for example, borrowing constraints) potentially preventing alternative responses that could increase non-migrant welfare by a larger amount than current responses.

## References

- Bartel, Ann P. 1989. “Where Do the New Immigrants Live?” *Journal of Labor Economics* 7(4): 371–391.
- Borjas, George J. 1987. “Self-Selection and the Earnings of Immigrants.” *American Economic Review* 77(4): 531–553.
- Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor. 2010. “The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression.” *Journal of Labor Economics* 28(4): 719–746.
- Bricker, Jesse, Alice M. Henriques, Jake A. Krimmel, and John E. Sabelhaus. 2015. “Measuring Income and Wealth at the Top Using Administrative and Survey Data.” Finance and Economics Discussion Series 2015-030. Washington, D.C.: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.030>.
- Card, David. 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” *Journal of Labor Economics* 19(1): 22–64.
- Card, David. 2009. “Immigration and Inequality.” *American Economic Review: Papers & Proceedings* 99(2): 1–21.
- Fisher, Jonathan, David Johnson, Timothy Smeeding, and Jeffrey Thompson. 2018. “Inequality in 3D: Income, Consumption, and Wealth.” Finance and Economics Discussion Series 2018-001. Washington, D.C.: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2018.001>.
- Ganong, Peter, and Daniel Shoag. 2017. “Why Has Regional Income Convergence in the U.S. Declined?” *Journal of Urban Economics* 102(1): 76–90.
- Jackson, Osborne. 2015. “Does Immigration Crowd Natives Into or Out of Higher Education?” Federal Reserve Bank of Boston Research Department Working Papers No. 15-18.
- Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. “Understanding the Long-Run Decline in Interstate Migration.” *International Economic Review* 58(1): 57–94.
- Kerr, William R., and William F. Lincoln. 2010. “The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention.” *Journal of Labor Economics* 28(3): 473–508.
- Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. “Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937.” *Quarterly Journal of Economics* 125(1): 91–128.
- Mandelman, Federico, and Andrei Zlate. 2016. “Offshoring, Low-Skilled Immigration, and Labor Market Polarization.” Federal Reserve Bank of Boston RPA Working Papers No. 16-3.
- McHenry, Peter. 2015. “Immigration and the Human Capital of Natives.” *Journal of Human Resources* 50(1): 34–71.

- Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. “Internal Migration in the United States.” *Journal of Economic Perspectives* 25(3): 173–196.
- Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2017. “Job Changing and the Decline in Long-Distance Migration in the US.” *Demography* 54(2): 631–653.
- Piketty, Thomas, and Emmanuel Saez. 2003. “Income Inequality in the United States, 1913–1998.” *Quarterly Journal of Economics* 118(1): 1–39.
- Roy, A.D. 1951. “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers* 3(2): 135–146.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. “Integrated Public Use Microdata Series: Version 7.0 [dataset].” Minneapolis: University of Minnesota. <https://doi.org/10.18128/D010.V7.0>.
- Ruggles, Steven, Matthew Schroeder, Natasha Rivers, J. Trent Alexander, and Todd K. Gardner. 2011. “Frozen Film and FOSDIC Forms: Restoring the 1960 Census of Population.” *Historical Methods* 44(2): 69–78.
- Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data.” *Quarterly Journal of Economics* 131(2): 519–578.
- Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. “What Are We Weighting For?” *Journal of Human Resources* 50(2): 301–316.
- Sommeiller, Estelle, Mark Price, and Ellis Wazeter. 2016. “Income Inequality in the U.S. by State, Metropolitan Area, and County.” Report. Washington, D.C.: Economic Policy Institute.
- Stock, James, and Motohiro Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. James H. Stock and Donald W. K. Andrews, 80–108. Cambridge, UK: Cambridge University Press.

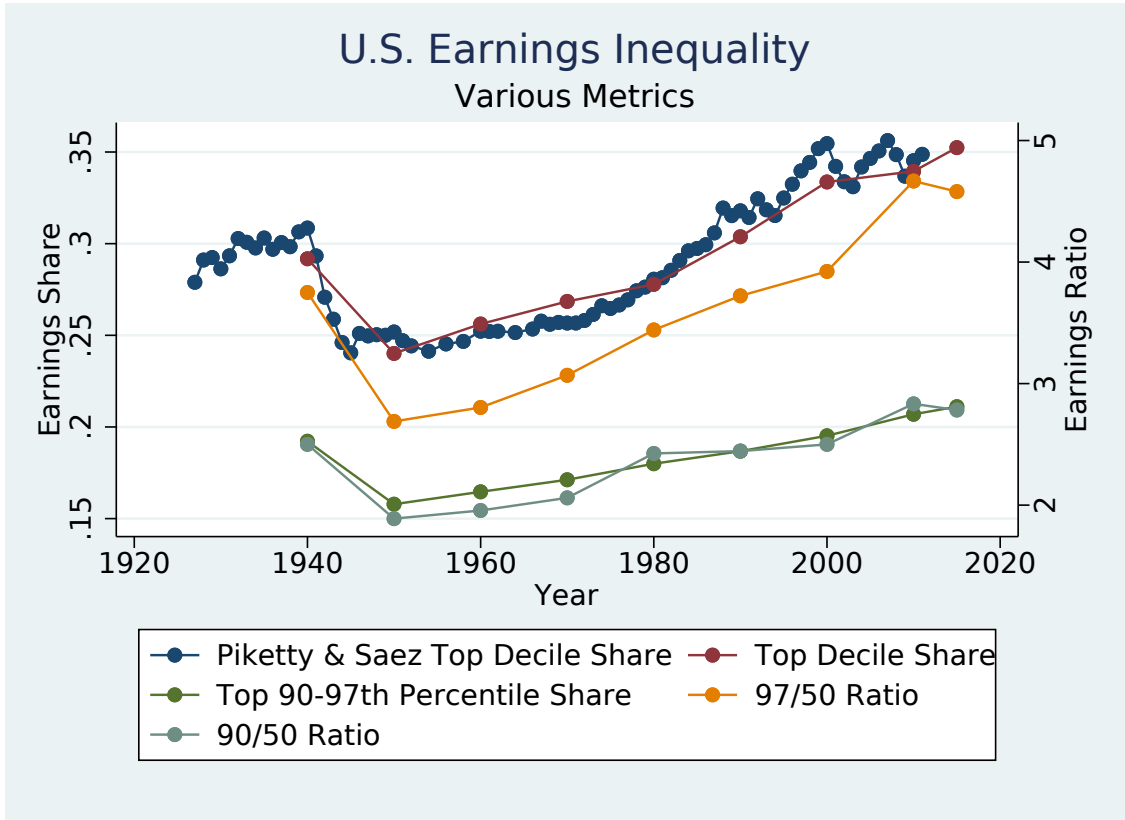


Figure 1: Comparing Measures of Earnings Inequality

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

Notes: This study’s sample includes self-employed workers and persons 16 to 64 years old. Sample excludes Alaska and Hawaii in 1940 and 1950.

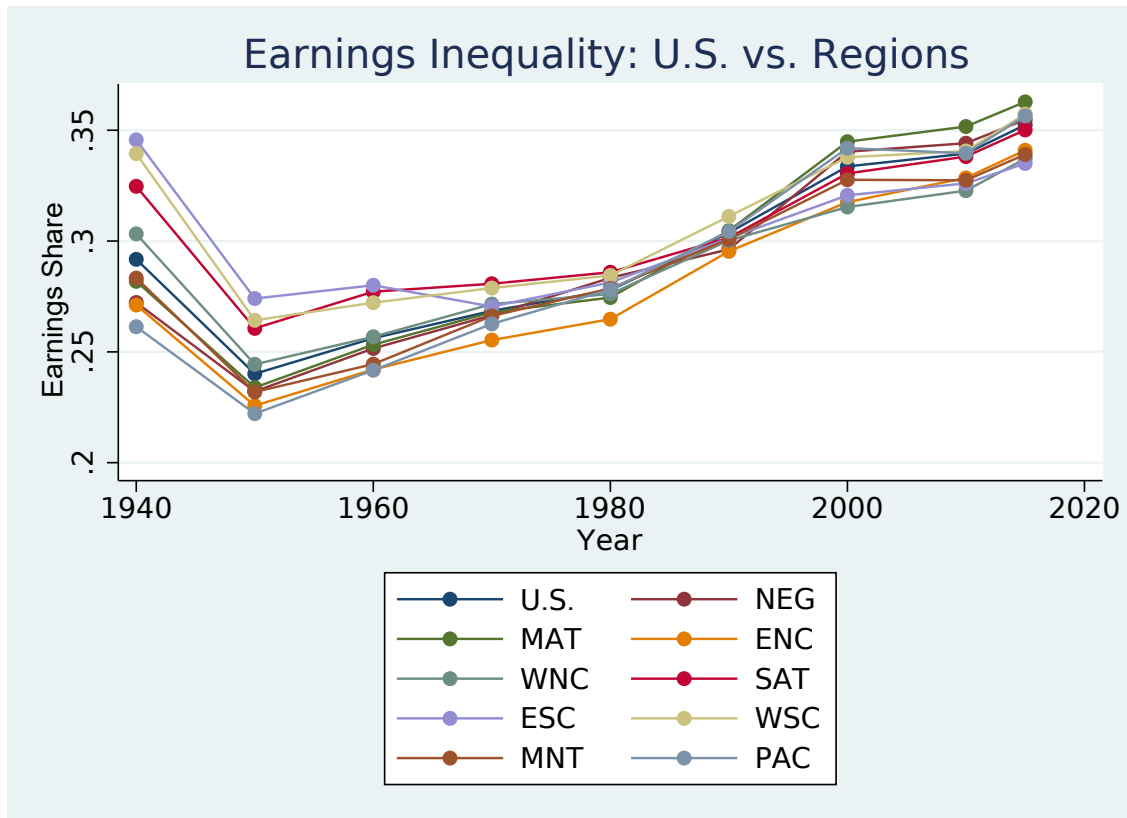


Figure 2: Top Decile Earnings Share Across Regions

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

Notes: Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC).



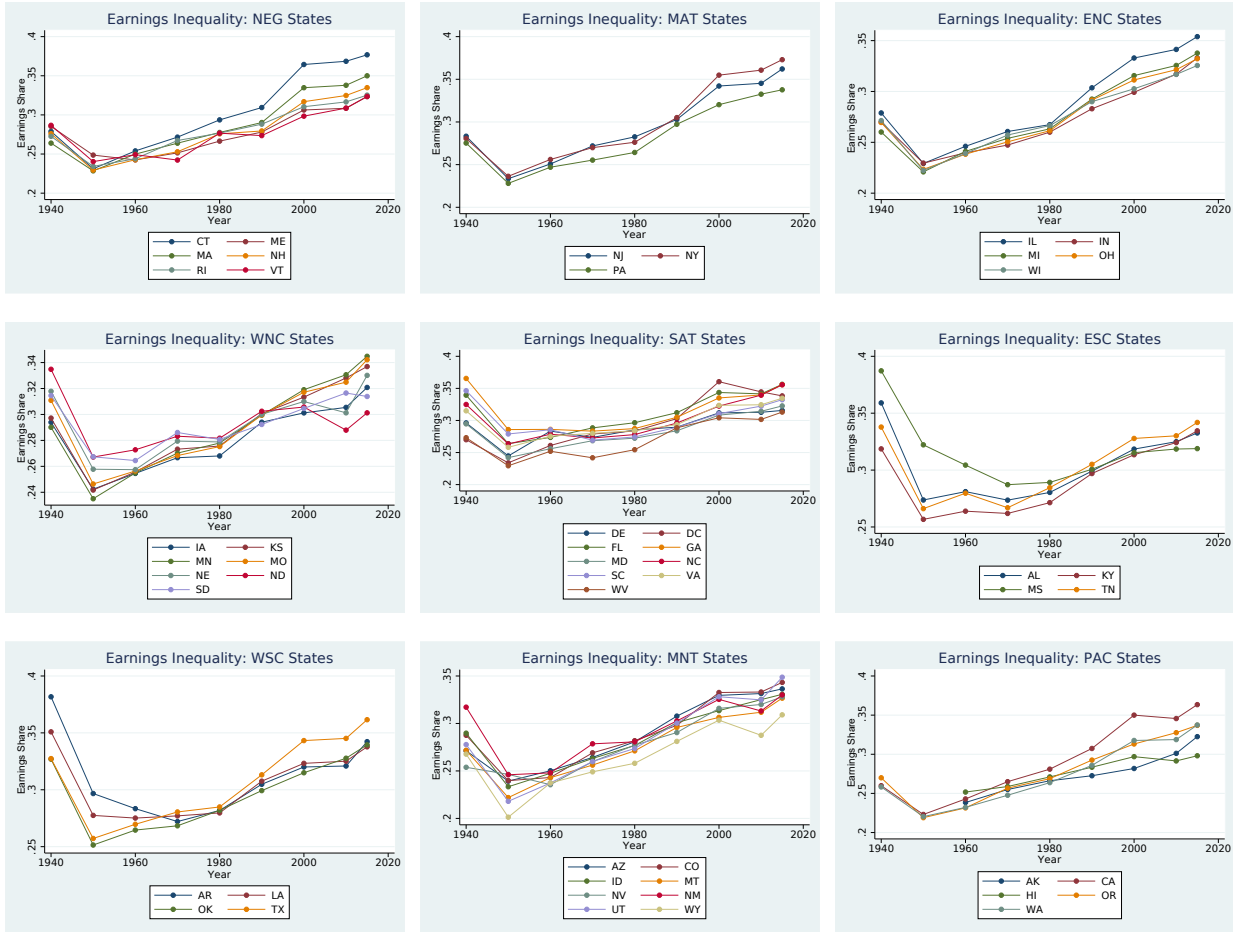


Figure 3: Top Decile Earnings Share Within Regions

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.  
 Notes: Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC). States, which include the District of Columbia, are listed according to standard U.S. Postal Service abbreviations.

B: Interstate

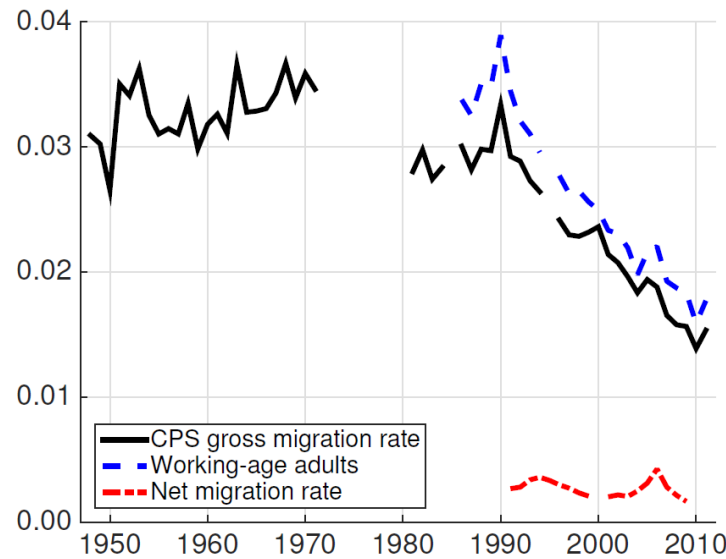
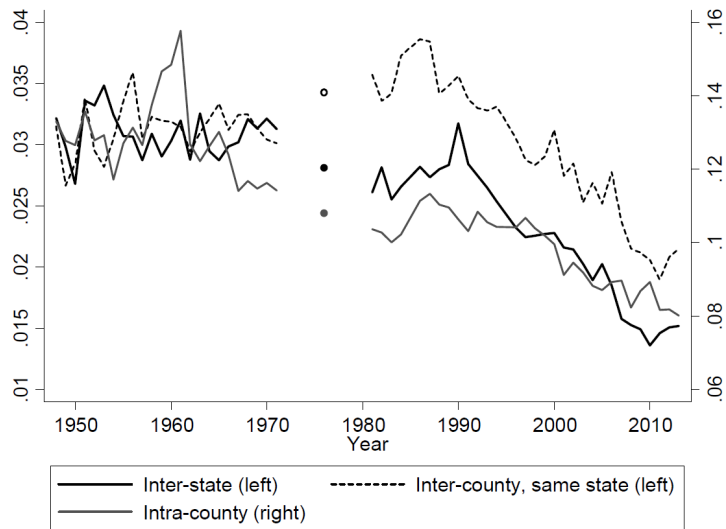
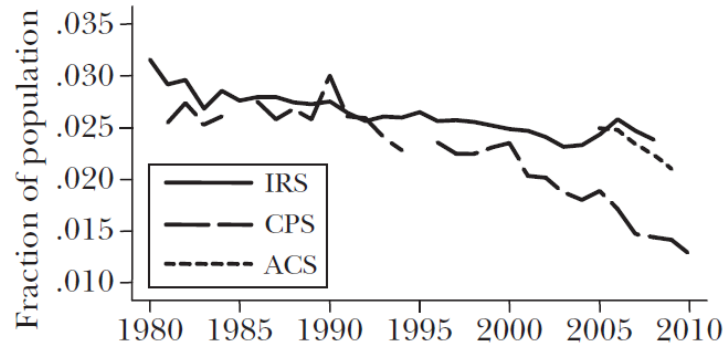


Figure 4: Some Existing Estimates of U.S. Interstate Migration  
*Sources:* Molloy, Smith, and Wozniak (2011) (top); Molloy, Smith, and Wozniak (2017) (middle); Kaplan and Schulhofer-Wohl (2017) (bottom).

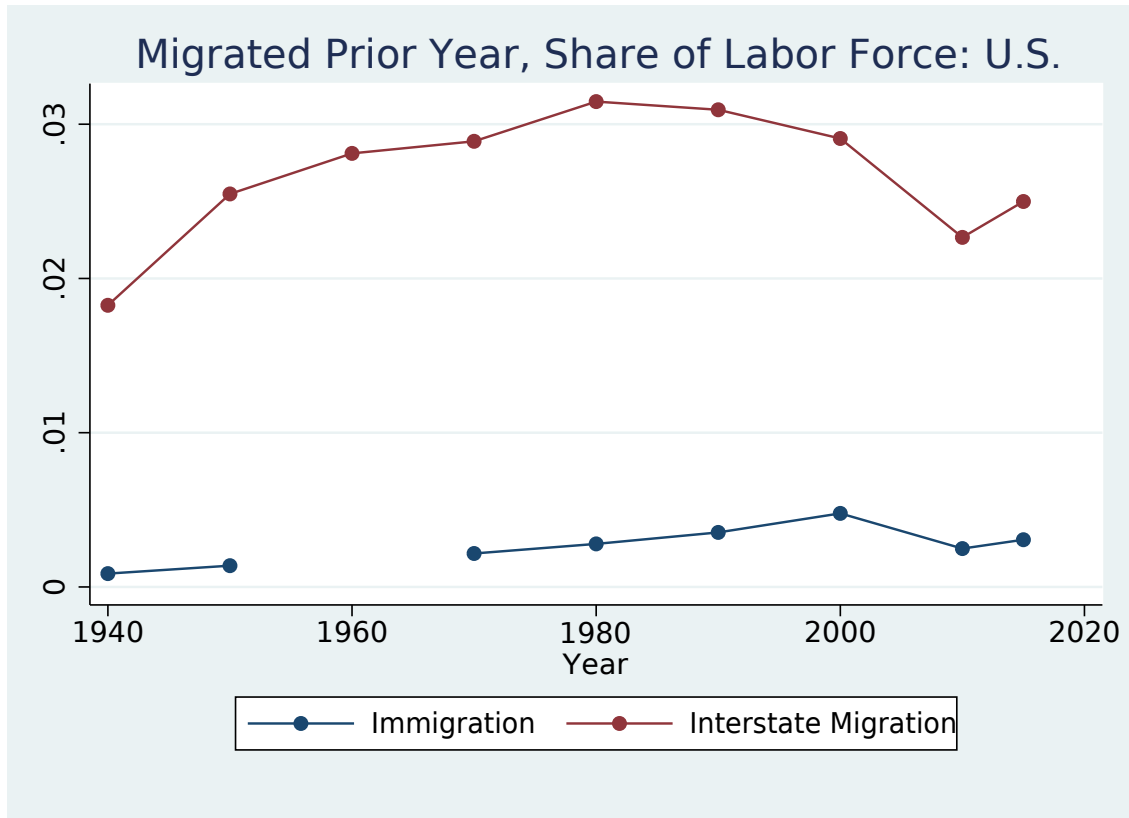


Figure 5: Estimating U.S. Interstate Migration and Immigration: Valid Approach  
 Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

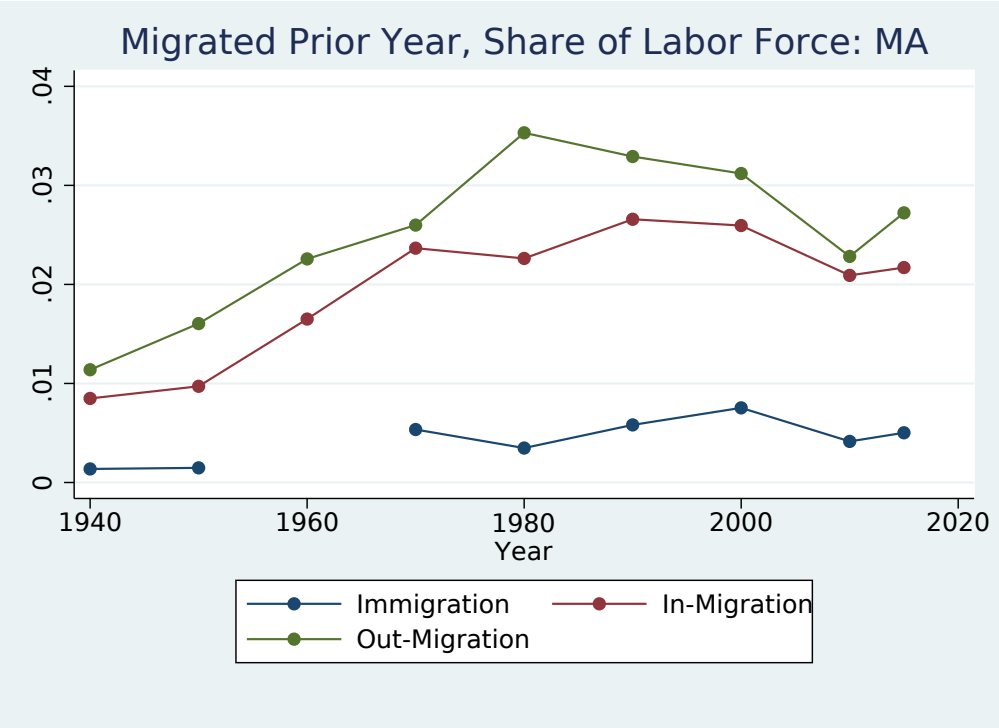
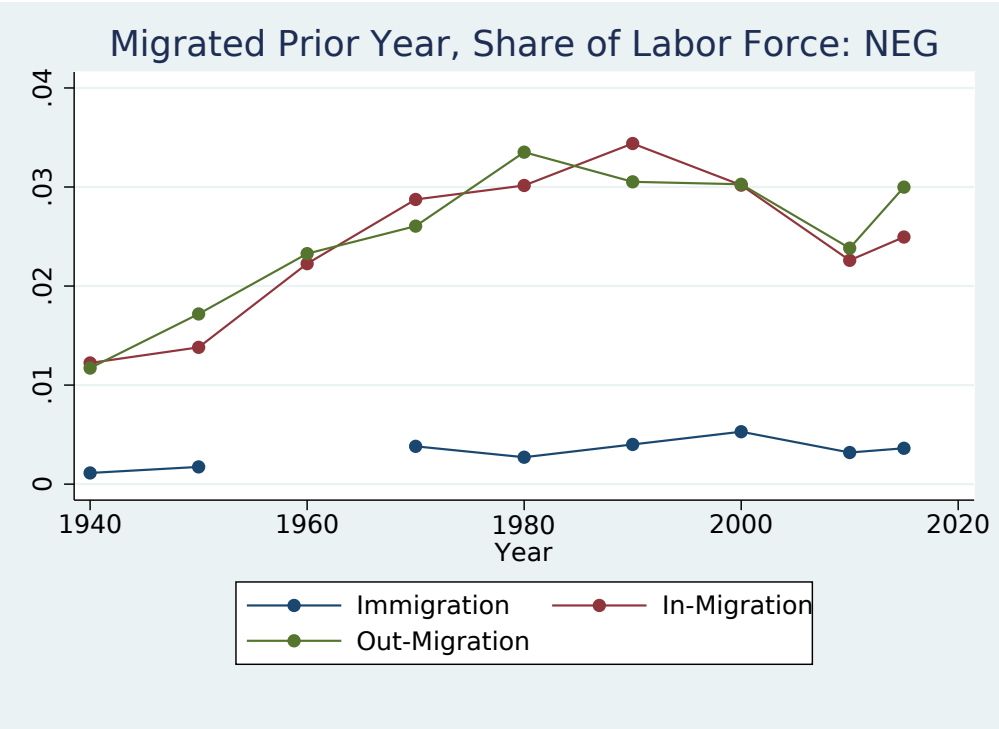


Figure 6: Region and State Immigration, In-Migration, and Out-Migration

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

Notes: Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC). States, which include the District of Columbia, are listed according to standard U.S. Postal Service abbreviations.

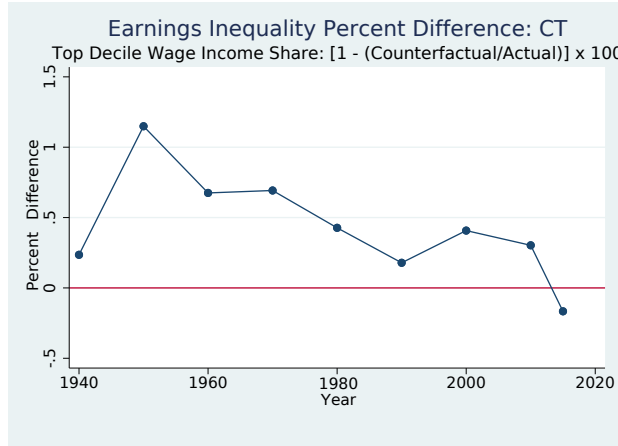
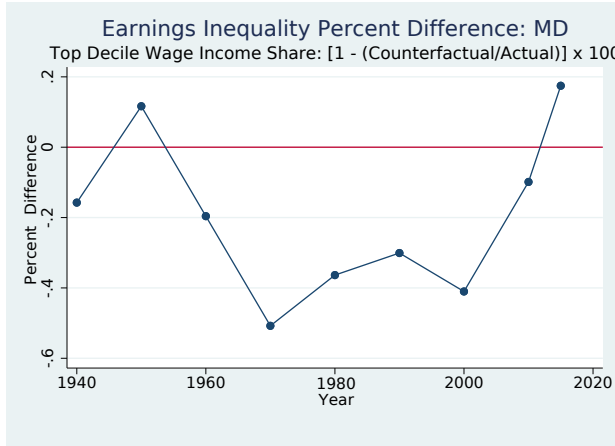
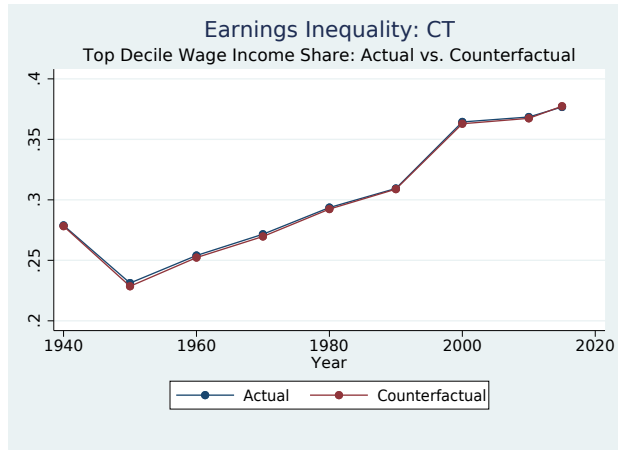
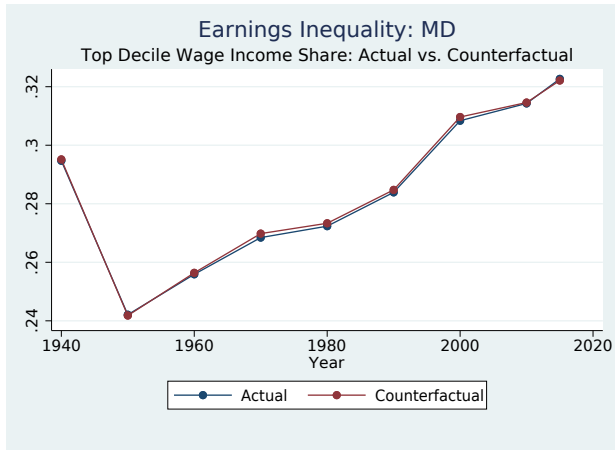


Figure 7: Counterfactual (No “Migrant Channel”) Top Decile Earnings Share  
*Sources:* 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.  
*Notes:* States, which include the District of Columbia, are listed according to standard U.S. Postal Service abbreviations.  
 Counterfactual sample reverts immigration, in-migration, and out-migration from prior year. Bottom panel depicts proportional difference in each year between actual and counterfactual inequality shown in top panel  

$$\left( \frac{Actual - Counterfactual}{Actual} \times 100\% \right)$$

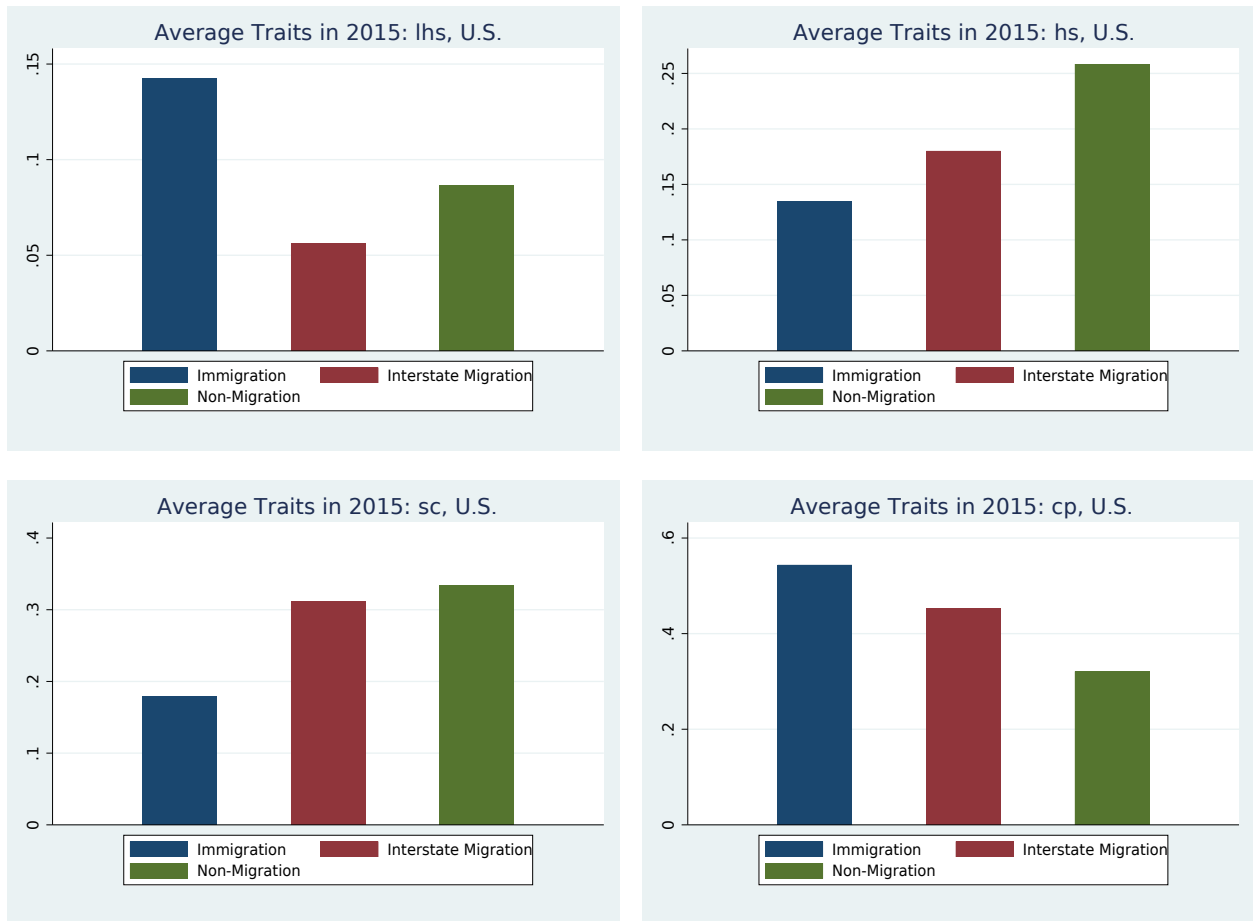


Figure 8: 2015 U.S. Migrant and Non-Migrant Education Shares

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

Notes: Average traits in 2015 of estimated prior year migration groups and non-migrating group. All three groups are mutually exclusive and reflect the full labor force sample. “LHS” is fewer than four years of high school; “HS” is four years of high school; “SC” is three years or fewer of college; and “CP” is four years of college or more.

Table 1: Top Decile Earnings Share and Migration as Share of Labor Force, Average Values

Rank	Inequality		Immigration		In-Migration		Out-Migration		
	Area or Year	Average Share	Area or Year	Average Share	Area or Year	Average Share	Area or Year	Average Share	
<b>States</b>									
1	GA	0.316	D.C.	0.008	AK	0.098	D.C.	0.093	
2	MS	0.316	CA	0.005	NV	0.080	ND	0.077	
3	FL	0.313	NV	0.005	D.C.	0.079	WY	0.070	
4	AR	0.312	FL	0.004	AZ	0.063	AK	0.067	
5	TX	0.309	MA	0.004	WY	0.062	MT	0.063	
47	AK	0.277	MO	0.001	MA	0.020	CA	0.019	
48	WA	0.276	AL	0.001	MI	0.017	IL	0.019	
49	IN	0.275	KY	0.001	OH	0.017	WI	0.019	
50	WV	0.273	ND	0.001	PA	0.015	PA	0.018	
51	WY	0.266	AR	0.000	NY	0.012	MI	0.017	
<b>U.S. &amp; Regions</b>									
1	WSC	0.310	PAC	0.004	MNT	0.055	MNT	0.050	
2	SAT	0.305	NEG	0.003	SAT	0.034	ESC	0.036	
3	ESC	0.304	MAT	0.003	PAC	0.034	WNC	0.034	
4	MAT	0.297	U.S.	0.003	WSC	0.030	SAT	0.033	
5	U.S.	0.296	WSC	0.002	WNC	0.028	WSC	0.028	
6	NEG	0.294	SAT	0.002	U.S.	0.027	U.S.	0.027	
7	WNC	0.292	MNT	0.002	ESC	0.026	NEG	0.025	
8	PAC	0.290	ENC	0.002	NEG	0.024	PAC	0.024	
9	MNT	0.289	WNC	0.001	ENC	0.019	ENC	0.020	
10	ENC	0.282	ESC	0.001	MAT	0.015	MAT	0.019	
<b>Years</b>									
1	2015	0.336	2000	0.004	1980	0.042	1960	0.043	
2	2010	0.324	2015	0.002	1970	0.038	1990	0.043	
3	2000	0.319	1990	0.002	1960	0.038	1970	0.043	
4	1940	0.299	2010	0.002	1950	0.038	1980	0.041	
5	1990	0.296	1980	0.002	1990	0.038	2000	0.037	
6	1980	0.276	1970	0.002	2000	0.036	1950	0.036	
7	1970	0.266	1950	0.002	2015	0.032	2015	0.032	
8	1960	0.256	1940	0.001	2010	0.030	1940	0.031	
9	1950	0.245	1960	—	1940	0.026	2010	0.030	

*Notes:* Measures for “inequality” (top decile earnings share), “immigration” (prior year immigration as a share of the labor force), “in-migration” (prior year in-migration as a share of the labor force), and “out-migration” (prior year out-migration as a share of the labor force) are the mean values across years (for “States” or “U.S. & Regions”) or areas (for “Years,” focusing only the 50 states and D.C.). Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC). States, which include the District of Columbia, are listed according to standard U.S. Postal Service abbreviations.

Table 2: Actual and Counterfactual Top Decile Earnings Share, Average Values

	Rank (by Avg. Pct Diff.)	Area or Year	Average Actual (Share)	Average Counterfactual (Share)	Average Percent Difference (%)
<b>States</b>					
	1	SD	0.293	0.291	0.845
	2	HI	0.279	0.294	0.480
	3	CT	0.305	0.304	0.434
	4	AZ	0.290	0.289	0.415
	5	ID	0.287	0.286	0.380
	47	ME	0.279	0.279	-0.184
	48	MD	0.285	0.285	-0.194
	49	AK	0.277	0.271	-0.221
	50	DE	0.290	0.292	-0.660
	51	WY	0.266	0.267	-0.662
<b>U.S. &amp; Regions</b>					
	1	NEG	0.294	0.293	0.169
	2	WNC	0.292	0.292	0.138
	3	PAC	0.290	0.289	0.136
	4	MNT	0.289	0.288	0.127
	5	WSC	0.310	0.309	0.063
	6	ESC	0.304	0.304	0.062
	7	U.S.	0.296	0.296	0.042
	8	SAT	0.305	0.305	0.034
	9	MAT	0.297	0.297	0.015
	10	ENC	0.282	0.282	-0.031
<b>Years</b>					
	1	1940	0.299	0.299	0.245
	2	1990	0.296	0.295	0.164
	3	2000	0.319	0.319	0.134
	4	1980	0.276	0.276	0.091
	5	2015	0.336	0.335	0.077
	6	1970	0.266	0.266	0.077
	7	1960	0.256	0.256	0.070
	8	2010	0.324	0.324	0.031
	9	1950	0.245	0.247	-0.121

*Notes:* The counterfactual (no “migrant channel”) top decile earnings share is calculated using state-year populations that revert immigration, in-migration, and out-migration from the prior year. “Average Percent Difference” is the mean proportional difference across years (for “States” or “U.S. & Regions”) or areas (for “Years,” focusing only the 50 states and D.C.) between the actual top decile share and counterfactual top decile share (that is,  $\text{Avg}[\frac{\text{Actual}-\text{Counterfactual}}{\text{Actual}} \times 100\%]$ ; it is not the difference between the actual and counterfactual average shares in the preceding columns,  $\text{Avg}(\text{Actual}) - \text{Avg}(\text{Counterfactual})$ ). Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC). States, which include the District of Columbia, are listed according to standard U.S. Postal Service abbreviations.



Table 3: Short-Run Impact of Migration on the Top Decile Share of Earnings

Model	(1)	(2)	(3)	(4)	(5)
Estimation	OLS	OLS	OLS	IV	IV
Outcome	Decile Share	Decile Share	Decile Share	Decile Share	Decile Share
Immigration Rate	6.13*** (0.59)	0.69 (0.52)	1.19* (0.61)	5.61*** (0.65)	3.99*** (0.96)
In-Migration Rate	-0.41*** (0.09)	0.03 (0.11)	0.05 (0.12)	-0.56*** (0.17)	-1.01*** (0.27)
Out-Migration Rate	-0.06 (0.08)	0.19* (0.09)	0.14* (0.08)	0.34*** (0.12)	-0.25 (0.18)
State FEs	No	Yes	Yes	No	Yes
Year FEs	No	Yes	No	No	Yes
Div-Year FEs	No	No	Yes	No	No
$R^2$	0.16	0.93	0.96		
Observations	326	326	326	326	326

Model	(6)	(7)	(8)	(9)
Estimation	IV	IV	IV	IV
Outcome	Decile Share	Counter. Decile Share	Counter. Decile Share	Counter. Decile Share
Immigration Rate	3.83* (2.16)	5.58*** (0.65)	4.03*** (0.94)	3.58* (1.91)
In-Migration Rate	-0.67 (0.60)	-0.54*** (0.17)	-0.94*** (0.26)	-0.54 (0.53)
Out-Migration Rate	-0.08 (0.27)	0.32*** (0.12)	-0.26 (0.18)	-0.05 (0.26)
State FEs	Yes	No	Yes	Yes
Year FEs	No	No	Yes	No
Div-Year FEs	Yes	No	No	Yes
$R^2$				
Observations	326	326	326	326

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. Instruments in IV specifications are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. “Counter. Decile Share” is the counterfactual, “no migration” top decile earnings share that reverts immigration, in-migration, and out-migration. Standard errors clustered by state are in parentheses.

Table 4: Long-Run Impact of Migration on the Top Decile Share of Earnings (IV)

Model	(1)	(2)	(3)	(4)	(5)	(6)
				Counter.	Counter.	Counter.
Outcome	Decile Share	Decile Share	Decile Share	Decile Share	Decile Share	Decile Share
Immig. Rate	-0.55 (1.91)	-1.46 (1.30)	-1.94 (3.60)	-0.60 (1.96)	-1.44 (1.32)	-1.85 (3.54)
Immig. Rate [t-1]	3.59 (2.50)	3.39** (1.50)	2.92* (1.46)	3.62 (2.53)	3.24** (1.47)	2.65* (1.45)
In-Mig. Rate	-1.97** (0.80)	0.50 (1.63)	0.43 (2.32)	-1.95** (0.79)	0.59 (1.63)	0.48 (2.29)
In-Mig. Rate [t-1]	1.64* (0.84)	0.29 (0.71)	0.31 (0.48)	1.63* (0.83)	0.29 (0.71)	0.34 (0.46)
Out-Mig. Rate	-1.31*** (0.34)	0.78 (1.23)	0.77 (2.24)	-1.33*** (0.34)	0.81 (1.24)	0.79 (2.22)
Out-Mig. Rate [t-1]	1.08*** (0.26)	0.04 (0.70)	0.29 (0.39)	1.07*** (0.26)	-0.02 (0.72)	0.27 (0.39)
State FEs	No	Yes	Yes	No	Yes	Yes
Year FEs	No	Yes	No	No	Yes	No
Div-Year FEs	No	No	Yes	No	No	Yes
Observations	245	245	245	245	245	245

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author’s calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. Instruments in IV specifications are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. “Counter. Decile Share” is the counterfactual, “no migration” top decile earnings share that reverts immigration, in-migration, and out-migration. Standard errors clustered by state are in parentheses.

Table 5: Impact of Migration on Observed Outcomes of Non-Migrants in Labor Force (IV)

Model	(1)	(2)	(3)	(4)	(5)
	Employment	< 4 Years	4 Years	Some	4 Years
Outcome	Share	HS	HS	College	College +
		Share	Share	Share	Share
Immig. Rate	1.10 (1.58)	8.96*** (2.15)	-7.31* (3.78)	-4.35 (4.19)	2.70 (2.13)
In-Mig. Rate	-0.08 (0.61)	-0.58 (1.14)	-0.63 (1.36)	1.77 (1.48)	-0.56 (0.78)
Out-Mig. Rate	-0.41 (0.38)	0.22 (0.83)	-0.07 (0.80)	0.15 (0.89)	-0.30 (0.46)
Observations	326	326	326	326	326

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. All regressions include state and division-year fixed effects. Instruments are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. Standard errors clustered by state are in parentheses.

Table 6: Impact of Migration on Observed Outcomes of Non-Migrants in Population (IV)

Model	(1)	(2)	(3)	(4)	(5)
	Employment	< 4 Years	4 Years	Some	4 Years
Outcome	Share	HS	HS	College	College +
		Share	Share	Share	Share
Immig. Rate	-3.99 (2.76)	10.80*** (2.58)	-7.70** (3.32)	-4.79 (4.44)	1.69 (1.90)
In-Mig. Rate	0.78 (1.17)	-0.88 (1.21)	-0.48 (1.20)	1.87 (1.50)	-0.50 (0.71)
Out-Mig. Rate	-0.25 (0.50)	0.08 (0.78)	0.05 (0.72)	0.22 (0.88)	-0.35 (0.44)
Observations	326	326	326	326	326

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. All regressions include state and division-year fixed effects. Instruments are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. Contrary to primary analysis, non-migrant outcomes here are based on a non-migrant sample that is not restricted to persons in the labor force. Standard errors clustered by state are in parentheses.

Table 7: Impact of Migration on Non-Migrant Labor Force Participation Rates (IV)

Model	(1)	(2)	(3)	(4)	(5)
	Overall	< 4 Years HS	4 Years HS	Some College	4 Years College +
Outcome	LFPR	LFPR	LFPR	LFPR	LFPR
Immig. Rate	-5.16*	-6.57**	-5.04**	-4.21*	-0.57
	(2.90)	(3.13)	(2.47)	(2.32)	(1.32)
In-Mig. Rate	0.97	0.74	0.71	0.73	-0.05
	(1.06)	(1.38)	(0.90)	(1.13)	(0.66)
Out-Mig. Rate	0.22	0.32	-0.17	0.06	-0.13
	(0.49)	(0.69)	(0.44)	(0.53)	(0.32)
Observations	326	326	326	326	326

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. All regressions include state and division-year fixed effects. Instruments are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. Contrary to primary analysis, non-migrant outcomes here are based on a non-migrant sample that is not restricted to persons in the labor force. Standard errors clustered by state are in parentheses.

Table 8: Relationship between Immigrant Education & Non-Migrant Education (OLS)

Model	(1)	(2)	(3)	(4)
	< 4 Years	4 Years	Some	4 Years
	HS	HS	College	College +
Outcome	Share	Share	Share	Share
Immigration Rate	6.46*** (2.21)	-7.71*** (1.52)	-2.42 (1.58)	3.68** (1.60)
In-Migration Rate	-0.10 (0.43)	-0.14 (0.42)	-0.69 (0.48)	0.93* (0.51)
Out-Migration Rate	0.16 (0.34)	-0.31 (0.34)	-0.57 (0.43)	0.72 (0.43)
Observations	326	326	326	326

Model	(6)	(7)	(8)	(9)
	< 4 Years	4 Years	Some	4 Years
	HS	HS	College	College +
Outcome	Share	Share	Share	Share
Immigration Rate	14.69*** (3.42)	-15.10*** (3.09)	3.30 (2.65)	-2.36 (4.29)
× Immig. Share, < 4 Years HS	-18.23*** (5.17)	17.70*** (4.68)		
× Immig. Share, 4 Years HS	-6.57 (14.06)	2.28 (13.81)	7.55 (8.75)	-3.79 (9.80)
× Immig. Share, Some College	-1.42 (9.57)	7.25* (4.29)	-2.80 (8.38)	-3.56 (6.26)
× Immig. Share, 4 Years College +			-23.42*** (3.48)	22.90*** (4.73)
In-Migration Rate	-0.17 (0.40)	-0.07 (0.34)	-0.42 (0.27)	0.66** (0.27)
Out-Migration Rate	0.10 (0.30)	-0.25 (0.27)	-0.18 (0.20)	0.33** (0.16)
Observations	323	323	323	323

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. All regressions include state and division-year fixed effects. Standard errors clustered by state are in parentheses.

# A Appendix

## A.1 Converting Five-Year Migration Flows to One-Year Flows

As discussed in section 3, in order to obtain consistent migration measures throughout the estimation sample, I need to convert five-year migration flows into one-year flows. One conversion approach is to divide the five-year flows by five. However, upon doing so, although the level of the resulting migration rate is reasonable (between 1 percent and 3 percent), changes in the observed rate do not align with established patterns. Specifically, as shown in Appendix Figure A3, I obtain a national interstate migration rate that falls from 1950 to 1960 and rises from 2000 to 2010, unlike trends in Figure 4.

Two potential reasons explain why the aforementioned approach does not capture the decline in interstate migration in recent decades. One reason (“demographic”) is because there are different age thresholds for the one-year migration flows versus the five-year migration flows. The other reason (“behavioral”) is because there are potentially different migration flow rates conditional on the same age threshold.

Regarding the demographic cause (focusing, without loss of generality, on interstate flows rather than immigration), the one-year flow in year  $t$  is given by  $\delta$ : persons age 1+ who lived in a different state in year  $t - 1$ . The five-year flow in year  $t$  is given by  $\gamma$ : persons age 5+ who lived in a different state in year  $t - 5$ . Object  $\gamma$  can be thought of as the sum of five distinct one-year flows: persons age 5+ in year  $t$  who lived in a different state in year  $t - 1$ , persons age 4+ in year  $t - 1$  who lived in a different state in year  $t - 2$ , . . . , persons age 1+ in year  $t - 4$  who lived in a different state in year  $t - 5$ . Thus, letting  $\gamma/\delta = \theta$ , the scaling factor, it is clear that  $\theta$  is 5 if and only if each one-year subcomponent of  $\gamma$  equals  $\delta$ , such that  $\theta = \gamma/\delta = 5\delta/\delta = 5$ . However, this need not be the case.

Besides age thresholds and demographics, another reason that simply dividing five-year flows by five may be a poor approximation for one-year flows is that, even conditional on the same age threshold, the migration flow rate in some year  $t$  compared with some other year  $t - 1$  may differ for behavioral reasons. In other words, there may be higher or lower propensities to migrate across years. Thus, to address both the “demographic” and “behavioral” considerations for converting five-year migration flows into one-year flows, I consider the following decomposition of the true scaling factor,  $\theta_{ijt}$ , for state  $i$ , migration flow type  $j$  (that is, immigration, in-migration, or out-migration), and year  $t$ :

$$\theta_{ijt} = d_{ijt}b_{ijt}, \tag{5}$$

where  $d_{ijt}$  is the demographic component, and  $b_{ijt}$  is the behavioral component. The objective is to find an alternative approach to  $\theta_{ijt} = 5 \forall i, j, t$ .

I consider four distinct but related approaches to deriving the scaling factor. The key across all four approaches is that I can actually observe the true scaling factor in the year 2000,  $\theta_{ij2000}$ , due to the availability of five-year migration flow data from the 2000 decennial census and one-year migration flow data from the 2000 ACS. As a result, I can calculate  $\theta_{ij2000}$  as follows:

$$\theta_{ij2000} = \frac{(\text{age } 5^+ \text{ five-year migration flow})_{ij2000}}{(\text{age } 1^+ \text{ one-year migration flow})_{ij2000}}. \tag{6}$$

Approach 1 is to use  $\theta_{ij2000}$  as the scaling factor for all years  $t$  given state  $i$  and migration flow  $j$ , dividing by  $\theta_{ij2000}$  in each case.

Approach 2 takes advantage of the availability of one-year migration flow data in the year 1950 to adjust the approach 1 scaling factor in the year 1940. Specifically, approach 2 applies  $\theta_{ij2000} = d_{ij2000}b_{ij2000}$  to five-year migration flows from years 1960 to 2000, and applies  $\pi_{ij1950} = d_{ij1950}b_{ij2000}$  to five-year flows from year 1940. The true scaling factor for year 1950,  $\theta_{ij1950}$ , cannot be calculated because  $b_{ij1950}$  is unobserved. However, the approximation,  $\pi_{ij1950}$ , can be calculated. First, I use one-year migration flow data and apply different age restrictions in the 2000 ACS to calculate  $d_{ij2000}$  as follows:

$$d_{ij2000} = \sum_{k=1}^5 \frac{(\text{age } k^+ \text{ one-year migration flow})_{ij2000}}{(\text{age } 1^+ \text{ one-year migration flow})_{ij2000}}, \quad (7)$$

which will be less than or equal to five.<sup>28</sup> Then, with estimates of  $\theta_{ij2000}$  and  $d_{ij2000}$ , I can recover estimates of  $b_{ij2000} = \theta_{ij2000}/d_{ij2000}$ . Finally, I can calculate  $d_{ij1950}$  following the approach outlined in equation (7) but using the 1950 census, thereby allowing me to calculate  $\pi_{ij1950}$ . It should be noted that it's unclear whether  $\pi_{ij1950}$  or  $\theta_{ij2000}$  is a better proxy for the unobserved true scaling factor in 1940,  $\theta_{ij1940}$ , which is what I would ideally use to convert the 1940 five-year migration flows to one-year flows. How closely  $\pi_{ij1950}$  versus  $\theta_{ij2000}$  compares to  $\theta_{ij1940}$  depends on  $d_{ij1940}$  and  $b_{ij1940}$ , both of which are unobserved. Thus, a preference between approach 1 and approach 2 must be based either on assumptions or some other criteria (for example, coverage across states and years).

Approach 3 recognizes that I can further decompose the demographic factor into two components: (a) the overall population at each age threshold (available in all survey years), and (b) the share of the population at each age threshold that migrates (available in one-year migration flow survey years). Thus, I can apply scaling factor  $\psi_{ijt,2000} = c_{ijt,2000}b_{ij2000}$  to the five-year migration flows for a given state  $i$ , migration flow type  $j$ , and year  $t$ , where  $c_{ijt,2000}$  is calculated as follows:

$$c_{ijt,2000} = \sum_{k=1}^5 \left[ \frac{(\text{age } k^+ \text{ population stock})_{ijt}}{(\text{age } 1^+ \text{ population stock})_{ijt}} \right] \left[ \frac{(\text{age } k^+ \text{ migration rate})_{ij2000}}{(\text{age } 1^+ \text{ migration rate})_{ij2000}} \right]. \quad (8)$$

For a given age  $k$ , the second bracketed term in equation (8)—the migration rate ratio in the year 2000—can be obtained by dividing the migration flow ratio in the year 2000 (as described in equation [7] but summed across  $k$  in that case) by the population stock ratio in the year 2000. Comparing approach 3 with approach 1, it's unclear whether  $\psi_{ijt,2000}$  or  $\theta_{ij2000}$ , respectively, is a better proxy for the unobserved true scaling factor,  $\theta_{ijt}$ . The aforementioned choice effectively depends on whether allowing for differences in the population stock over time improves the approximation to  $\theta_{ijt}$ , although it seems likely that doing so would be beneficial.

Lastly, approach 4 is similar to approach 3, but utilizes the availability of one-year migration flow data in the year 1950 to adjust the approach 3 scaling factor in the year 1940. Specifically, approach 4 applies  $\psi_{ijt,2000} = c_{ijt,2000}b_{ij2000}$  to five-year migration flows from

---

<sup>28</sup>Note that  $d_{ij2000}$  will equal 5 only if no persons migrate who are under age 5.

years 1960 to 2000, and applies  $\eta_{ij1940,1950} = c_{ij1940,1950}b_{ij2000}$  to five-year flows from year 1940. The true scaling factor for year 1950,  $\psi_{ijt,1950}$ , cannot be calculated because  $b_{ij1950}$  is unobserved. However, the approximation,  $\eta_{ij1940,1950}$ , can be determined by calculating  $c_{ijt,1950}$  following the approach outlined in equation (8) but using the 1950 census, thus allowing me to determine  $\eta_{ij1940,1950}$ . As with the previous scaling factor approaches, it is ambiguous whether  $\eta_{ij1940,1950}$  (approach 4) or  $\psi_{ijt,2000}$  (approach 3) is a better proxy for the unobserved true scaling factor in 1940,  $\theta_{ij1940}$ , because this depends on several unobserved variables. Across all four scaling factor approaches, I opt for approach 3 as the preferred factor based on methodology. However, all four scaling factor approaches are highly correlated, with significant pairwise correlation coefficients ranging from 0.9996 to 1.000 for the United States, 0.9996 to 1.000 across regions, and 0.9982 to 1.000 across states.

## A.2 Constructing Instrumental Variables

For state  $i$ , year  $t$ , and migrant type  $j = \{1, 2, 3\}$ , the three instruments for prior year migration as a share of the labor force,  $Z_{ijt}$ , take the following forms, described in turn. First, to help primarily capture exogenous variation regarding immigration ( $j = 1$ ):

$$Z_{i1t} = \frac{\sum_h \left( \frac{F_{hi,1940}}{F_{h,1940}} \right) \times M_{h1t}}{\psi_{i1t,2000} L_{it}}, \quad (9)$$

where  $h$  is birthplace continent,  $F$  is foreign-born, and  $L$  is labor force.<sup>29</sup> The numerator of  $Z_{i1t}$  uses the 1940 share of foreign-born persons who were born in continent  $h$  and chose to live in state  $i$ , and uses this share to allocate future one-year immigrants from origin continent  $h$  in year  $t$ , before then aggregating across all origin continents to generate a predicted prior year immigrant flow for residence state  $i$  and year  $t$ . The denominator of  $Z_{i1t}$  is the labor force for residence state  $i$  and year  $t$ , multiplied by scaling factor 3 (which the predicted migration flow is divided by in order to convert the flow to prior year migration; note that this scaling factor is set equal to 1 in sample years with one-year migration flows already available). Instrument  $Z_{i1t}$  thus reflects the predicted immigrant share of the labor force.

So, for instance, if 15 percent of foreign-born persons born in South America were living in Massachusetts in 1940, then the instrument numerator would allocate 15 percent of the total South American one-year immigrant inflow in 1990 to Massachusetts. The validity of this instrument and the identification strategy hinges on three assumptions, two of which are related to the two components of the instrument (numerator). First, it is assumed that any unobserved differential market shocks between states  $i$  and  $i'$  in 1940 or earlier that caused foreign-born persons to locate in state  $i$  rather than  $i'$  are uncorrelated with such relative market shocks from 1950 to 2015. In other words, suppose that, in 1920, a market shock increasing earnings inequality occurred in New York that was large relative to a similar shock in Arizona. As a result, more (high-skilled) immigrants from Europe located in New York rather than Arizona. Then, for the instrument to be valid, it cannot be the case that over the

<sup>29</sup>I aggregate foreign-born stocks to birthplace continents rather than countries in order to increase sample sizes given small counts of foreign-born persons from some countries. The continent groupings are North America (including Central America), South America, Europe, Asia, Africa, Australia, and Antarctica (including “others,” such as “abroad [unknown] or at sea” or “other not elsewhere classified”).



1950–2015 period, all such shocks increasing inequality were also large in New York relative to Arizona (that is, inequality was growing at a faster rate in New York than Arizona). If so, then the 1950–2015 allocations of European immigrants to New York and Arizona predicted by the instrument would be correlated with the 1950–2015 relative inequality-related shocks, causing the instrument to be endogenous.

Secondly, instrument validity requires that the total origin-continent immigration flow is exogenous to such unobserved relative market shocks between states from 1950 to 2015. For example, suppose that a 1990 inequality-related shock in Arizona that was large relative to a similar shock in New York caused some European immigrants to locate in Arizona rather than New York. For the instrument to be valid, such a relative inequality-related shock cannot have caused some Europeans to immigrate to the United States who otherwise would not have or, alternatively, dissuaded some Europeans from immigrating, such that the shock altered the total flow of European immigrants in 1990.

Combined, these two assumptions form the instrument exogeneity assumption, or the exclusion restriction. Here, this restriction imposes that the only channel through which the instrument-predicted immigrant inflows affect local earnings inequality is through their impact on the endogenous immigrant inflows. I can include division-year fixed effects for the nine U.S. census divisions in the estimation to try to ensure that the exclusion restriction holds (for instance, inequality in some regions may have been growing at a rate different from the growth rates of other regions since 1940, although Figure 2 suggests that this may not be a large concern).

The other necessary assumption for instrument validity and consistent IV estimation is instrument relevance, such that the immigrant flows predicted from the instruments are sufficiently related to the endogenous immigrant flows. This assumption is testable, however, via first-stage estimation, which I will examine to ensure problematic weak instrument estimation does not occur.

Moving on to the second instrument, to help primarily capture exogenous variation regarding in-migration ( $j = 2$ ):

$$Z_{i2t} = \frac{\sum_h \left( \frac{N_{hi,1940}}{N_{h,1940}} \right) \times M_{h2t}}{\psi_{i2t,2000} L_{it}}, \quad (10)$$

where  $h$  is birthplace state,  $N$  is native-born, and all other variables are as previously defined. The numerator of  $Z_{i2t}$  uses the 1940 share of native-born persons who were born in state  $h$  and live in state  $i$  (conditional on  $h \neq i$ ) to allocate future one-year in-migrants from origin state  $h$  in year  $t$ , then aggregates across all origin states to generate a predicted prior year in-migrant flow for residence state  $i$  and year  $t$ . Following  $Z_{i1t}$ , the denominator of  $Z_{i2t}$  is again the labor force multiplied by scaling factor 3. So, for instance, if 15 percent of native-born persons born in Illinois were living in Massachusetts in 1940, the instrument numerator would allocate 15 percent of the total prior year internal migrants from Illinois in 1990 to go to Massachusetts. The instrument numerator would then aggregate across all origin states until generating a predicted in-migrant flow to Massachusetts in 1990. Meanwhile, other than a change in focus from origin continents to origin states, the discussion earlier regarding identification assumptions of the instrument still applies.

Lastly, I describe the third and final instrument, to help primarily capture exogenous

variation regarding out-migration ( $j = 3$ ):

$$Z_{i3t} = \frac{\sum_h \left( \frac{N_{hi,1940}}{N_{h,1940}} \right) \times M_{h3t}}{\psi_{i3t,2000} L_{it}}, \quad (11)$$

where  $i$  now indicates the origin state of the outflow,  $h$  indicates the state of residence, and all other variables are as previously defined. The numerator of  $Z_{i3t}$  identifies the 1940 share of native-born persons who live in state  $h$  and were born in state  $i$  (conditional on  $i \neq h$ ), and uses this share to allocate future one-year out-migrants from residence state  $h$  in year  $t$  (in terms of predicting their origin state), before then aggregating across all residence states to generate a predicted prior year out-migrant flow for origin state  $i$  and year  $t$ . Once again, following  $Z_{i1t}$  and  $Z_{i2t}$ , the denominator of  $Z_{i3t}$  is similarly the labor force multiplied by scaling factor 3. So, for instance, if 15 percent of native-born persons living in California were born in New Mexico in 1940, the instrument numerator would allocate 15 percent of the total prior year internal migrants to California in 1990 to be from New Mexico. The instrument numerator would then aggregate across all residence states until generating a predicted out-migrant flow from New Mexico in 1990. Meanwhile, other than a change in focus from origin states to destination states, the earlier discussion regarding instrument identification assumptions once again still holds. Thus, given the vector of three endogenous migration rate variables,  $\mathbf{M}_{it}$ , I use the vector of exogenous migration rate instruments,  $\mathbf{Z}_{it}$ , in IV estimation.

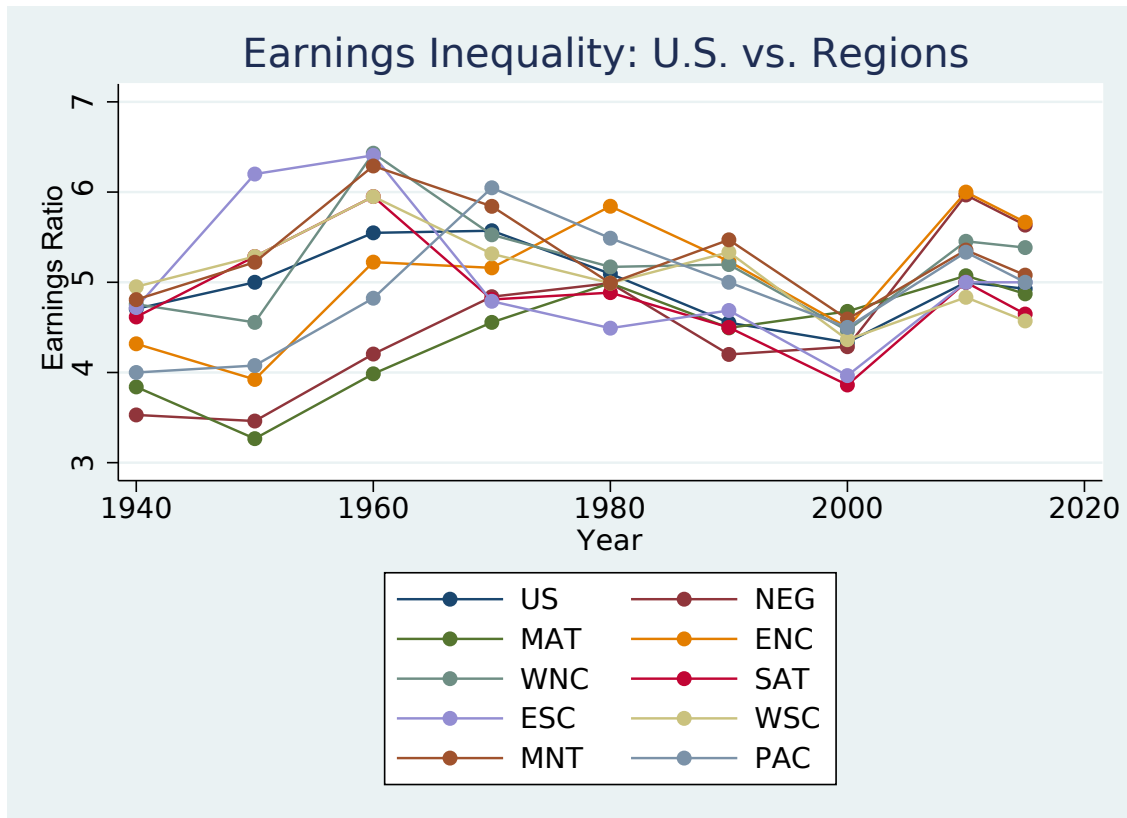


Figure A1: Earnings Ratio at the 50th/10th Percentile

Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

Notes: Census divisions are New England (NEG), Middle Atlantic (MAT), East North Central (ENC), West North Central (WNC), South Atlantic (SAT), East South Central (ESC), West South Central (WSC), Mountain (MNT), and Pacific (PAC).

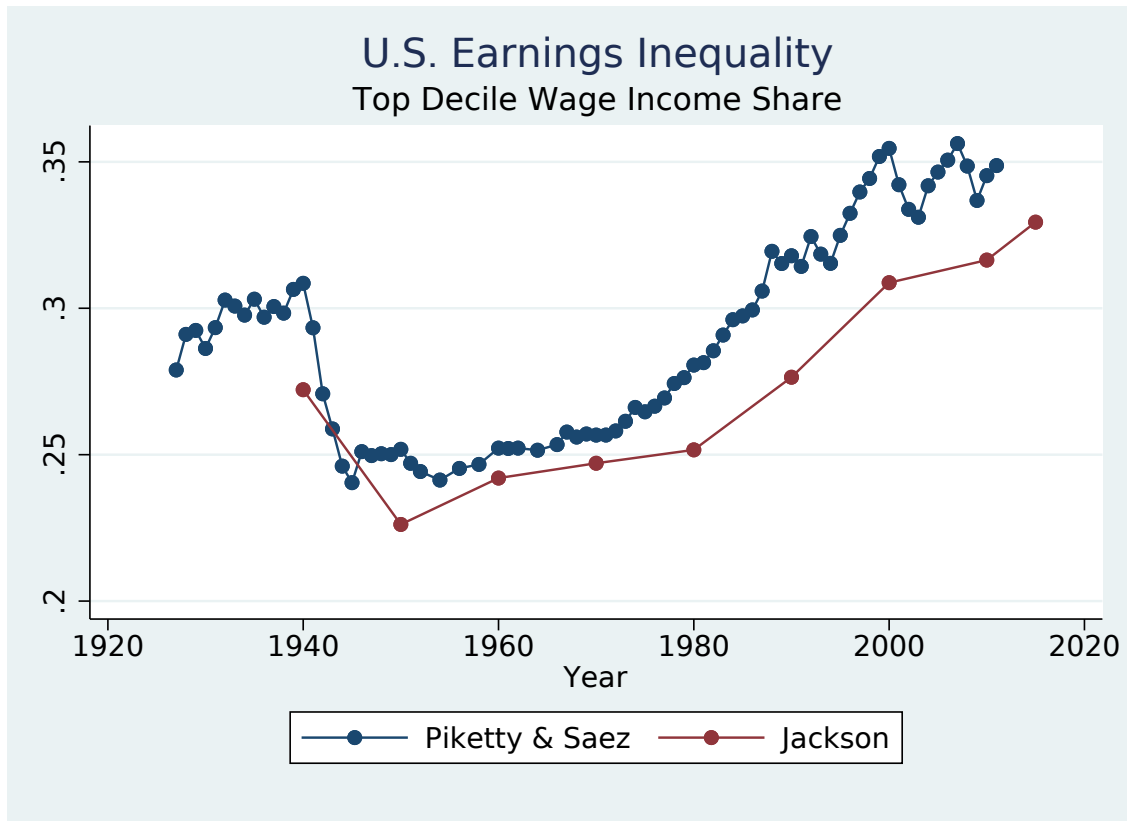


Figure A2: Piketty & Saez vs. Jackson (Adjusted)

*Sources:* 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.  
*Notes:* Jackson sample excludes self-employed workers and includes ages 25 to 64. Also excludes Alaska and Hawaii in 1940 and 1950.

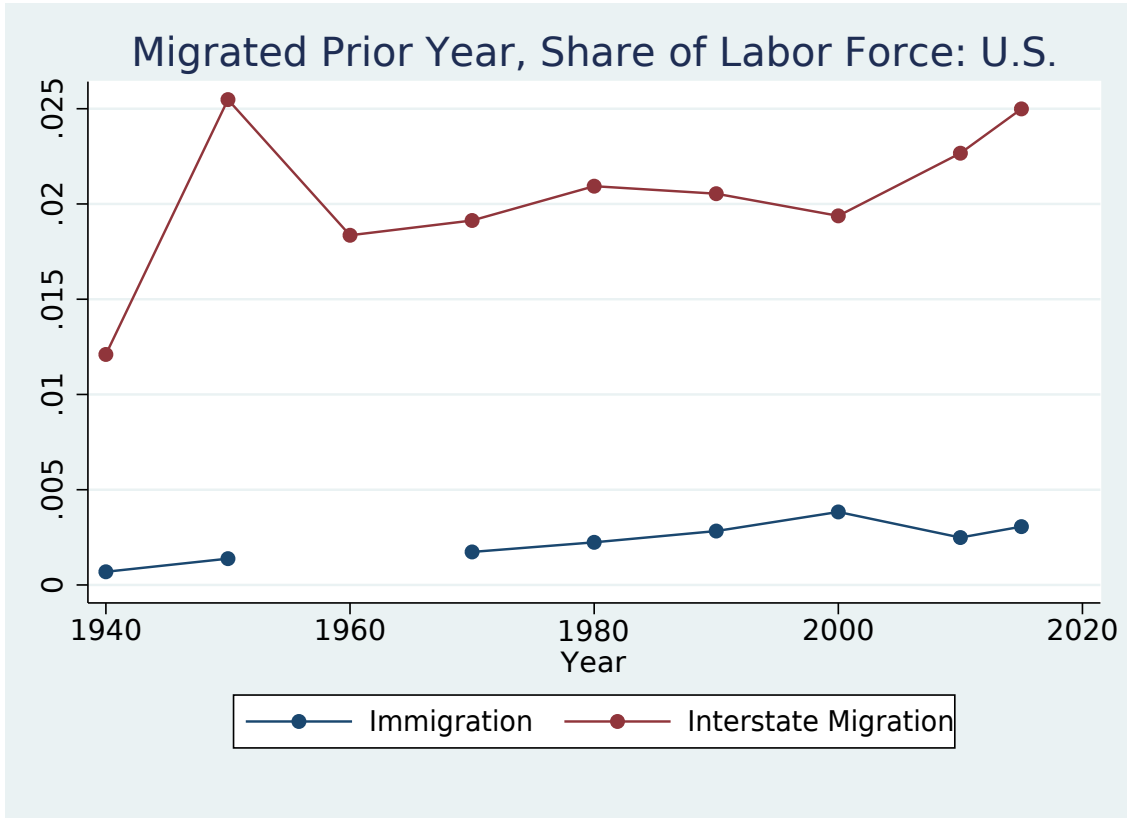


Figure A3: Estimating U.S. Interstate Migration and Immigration: Invalid Approach  
 Sources: 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.

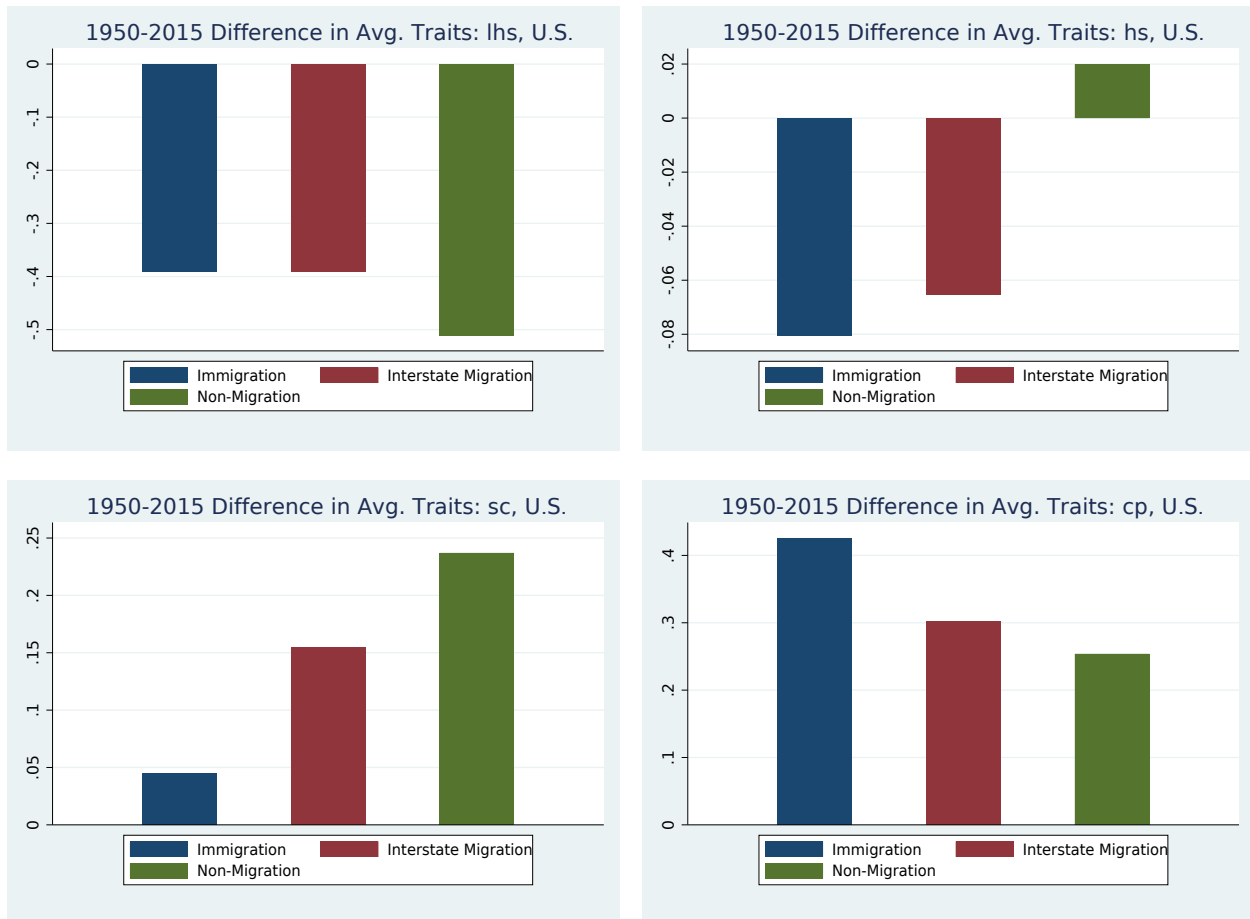


Figure A4: 2015–1950 Differences in U.S. Migrant and Non-Migrant Education Shares  
*Sources:* 1940–2000 U.S. Census, 2010 and 2015 American Community Survey, and author’s calculations.  
*Notes:* Differences (2015 – 1950) in the average traits of estimated prior year migration groups and non-migrating group. All three groups are mutually exclusive and reflect the full labor force sample. “LHS” is fewer than four years of high school; “HS” is four years of high school; “SC” is three years or fewer of college; and “CP” is four years of college or more.

Table A1: First Stage of Migrant Rates on Predicted Migrant Rates (OLS)

Model	(1)	(2)	(3)
Outcome	Immigration Rate	In-Migration Rate	Out-Migration Rate
Predicted Immig. Rate	0.00010** (0.00004)	0.00006*** (0.00002)	0.00008*** (0.00003)
Predicted In-Mig. Rate	0.00008*** (0.00003)	0.00006** (0.00002)	0.00006*** (0.00002)
Predicted Out-Mig. Rate	-0.00001 (0.00001)	0.00003 (0.00002)	0.00003 (0.00002)
Any FEs	No	No	No
$R^2$	0.33	0.77	0.83
F-stat on instruments = 0	9.86	8.42	8.20
Observations	326	326	326

Model	(4)	(5)	(6)
Outcome	Immigration Rate	In-Migration Rate	Out-Migration Rate
Predicted Immig. Rate	-0.00051** (0.00019)	-0.00013** (0.00006)	-0.00002 (0.00008)
Predicted In-Mig. Rate	0.00035* (0.00020)	0.00024*** (0.00009)	0.00024*** (0.00008)
Predicted Out-Mig. Rate	0.00006 (0.00009)	-0.00009 (0.00009)	-0.00015 (0.00010)
State and Year FEs	Yes	Yes	Yes
$R^2$	0.05	0.79	0.87
F-stat on instruments = 0	3.87	2.65	3.44
Observations	326	326	326

Model	(7)	(8)	(9)
Outcome	Immigration Rate	In-Migration Rate	Out-Migration Rate
Predicted Immig. Rate	-0.00027* (0.00014)	0.00009** (0.00004)	0.00019 (0.00014)
Predicted In-Mig. Rate	-0.00014 (0.00013)	-0.00006 (0.00010)	0.00000 (0.00014)
Predicted Out-Mig. Rate	0.00065*** (0.00013)	0.00055*** (0.00012)	0.00056*** (0.00013)
State and Div-Year FEs	Yes	Yes	Yes
$R^2$	0.17	0.77	0.85
F-stat on instruments = 0	10.84	7.70	6.25
Observations	326	326	326

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* Author's calculations using 1940–2000 U.S. Census, 2010 and 2015 American Community Survey. Instruments are predicted immigrant, in-migrant, and out-migrant flows as a share of the labor force using historical (1940) birthplace information of foreign-born and native persons. Standard errors clustered by state are in parentheses.