A Framework for Decomposing Wage Dispersion with an Application to Foreign and Internal Migration in the US from 2001-2019

Bent E. Sørensen, Fan Wang, and Adelia Zytek*

April 7, 2025

Abstract

We propose a new framework for decomposing how the composition of the work force affects the wage distribution. We consider how wage fractiles are affected by i) the total presence of a sub-group of workers such as college-educated immigrants, ii) the mean income of each sub-group, and iii) the marginal effect of adding a member from each sub-group. We apply this framework to document changes to inequality in U.S. wage income within demographic groups defined by educational attainment and migration status and show how group composition affects aggregate U.S. wage inequality over time. Every education-migration group has contributed to changing inequality over time, but college-educated migrants, both foreign and domestic, have contributed the most, as migrants' wages have become more dissimilar to the wages of non-migrants. Within-group inequality trends strongly affect the trend in overall inequality.

Keywords: Inequality, wages, immigration, domestic migration

*Bent E. Sørensen: Department of Economics, University of Houston, Houston, Texas, USA and Centre for Economic Policy Research, London, UK (email: besorensen@uh.edu); Fan Wang: Department of Economics, University of Houston, Houston, Texas, USA (fwang26@uh.edu); Adelia Zytek: Department of Economics, Rice University, Houston, Texas, USA (az29@rice.edu).

1 Introduction

We study the distribution of wages in the United States from 2001 to 2022 and decompose trends in the wage distribution by education and migration status, considering domestic non-movers, domestic movers, and immigrants. We study the impact of composition on wage inequality, taking wages as given, in the spirit of Autor, Katz, and Kearney (2008).

We propose a framework which considers the impact a migration-education group has on aggregate inequality in three ways: the impact of the difference in mean wage of a group to average wages ("mean effect"), the impact of a small increase in the relative proportion of population in a group ("marginal effect"), and the impact of adding or removing a whole group ("total effect"). In line with recent research, we begin by performing a variance decomposition on inflation adjusted log-wages (Section 2.1). We then connect this method to our proposed framework and introduce a way to consider these effects in the context of log-wage quantiles (Section 2.2). We apply this framework to the study of U.S. wage inequality and decompose changes in inequality by demographic and distributional changes in six education-migration groups (Section 3.3). We finish with a discussion of the implications of these findings on U.S. immigration policy (Section 3.4).

We find, for example, that college educated immigrants are becoming more concentrated in high wage categories and that the observed increase in wage inequality is driven by increasing upper quantiles of the wage distribution. We also find that the effect size of non-college educated immigrants decreased during the time of study.

Literature Review: The relationship between US income inequality and education, immigration, and domestic migration have been widely studied in the past two decades. Recent research has pointed to a positive association between immigration and income inequality on both a national and regional level. Hibbs and Hong (2015) argue, using instrumental variable regression on the Gini coefficient, that up to one fourth of the change in local metropolitan wage inequality in the U.S. from 1990 to 2000 could be explained by immigration alone. Similarly, Xu, Garand, and Zhu (2016) find, using data from 1996 to 2008, a positive correlation between foreign immigration into a state and greater inequality, primarily driven by low-skill immigrants. They also find that high skilled immigration is associated with lower inequality in some parts of the income distribution. A 2019 paper from Lin and Weiss find that, contrary to popular belief, low skill immigration is not associated with declines in the lower quantiles of the wage distribution.

There is an extensive work that attempts to quantify the effect of immigrants on the wages of natives; see e.g. Borjas (2003) and Card (2009). Lin and Weiss (2019) find that both low and high skill immigration are associated with wage gains for high wage natives, resulting in a widening of the income distribution at the top end.[1]

The relationship between educational trends and wage inequality has also been extensively explored. Hoxby and Terry (1999) find that within-group inequality for those with a college education rose in the latter half of the 20th century, while Autor, Katz, and Kearney

1. Llull (2017) found that workers who are most similar in skill to immigrants are the most affected by immigration. These migrants tend to adjust to immigration trends by either increasing or decreasing the level of education they attain.

(2008) find that wage dispersion in the upper half of the wage distribution grew faster than that of the lower half at the end of the 20th century. Part of this trend could be explained by a rising wage premium for holders of college and graduate degrees. In contrast, Hershbein, Kearney, and Pardue (2020) found, by simulating an increase in college degree attainment, that upper to lower quantile inequality would decrease, assuming bachelor degree attainment were raised to at least fifty percent. Fernández and Rogerson (2004) proposed that sorting could be one potential mechanism for education's impact on inequality.

We expand upon the existing literature by examining the impact of trends in the interaction between education and migration in the United States specifically, considering the effects of domestic migration across the four Census regions, as well as by making a methodological contribution by examining for each sub-group the total effect on inequality, as well as the effect of the mean wages of each group and the marginal effect on quantiles, capturing within group dispersion.

2 Methods

2.1 Variance Decomposition

Historically, demographic inequality analysis has relied on variance decompositions, in which the effect of a particular group on overall inequality is decomposed into the contribution of the within-group variance and the addition to across-group variance. Overall variance can then be decomposed into the within and between group effects of each group in the population. Therefore, we begin by decomposing inequality as measured by a variance decomposition.

We denote the set of education and migration statuses that an individual i in group t can take as

$$\begin{gathered} m_{i,t} \in M = \{\text{Domestic-born-non-mover}, \text{Domestic-born-mover}, \text{Foreign-born}\} \\ \text{and } e_{i,t} \in E = \{\text{Non-college}, \text{College}\} \ . \end{gathered} \tag{2.1}$$

In our setting, variance can then be decomposed as

$$\sigma^2 = \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot \sigma^2_{e,m} + \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot (\mu_{e,m} - \mu^2) \ , \tag{2.2}$$

where the first term represents the total within group variance of each education-migration group $\{e, m\}$ and the second term is the total across group variance. More information on the interpretation of the above equation can be found in B.1.

Variance decompositions can be highly effective in understanding relative contributions to inequality in contexts where within-group log-wages are approximately normally distributed. In reality, however, the log-wage distributions of U.S. subgroups are both skewed and contain more mass in the tails than that of a normal distribution. To account for these differences, we introduce a new framework that examines quantiles and quantile ranges in place of variance.

2.2 Framework—Total, Mean, and Marginal Effects

We introduce a new framework for considering the effect that a demographic group has on inequality. In principle, there are three ways that a demographic group can contribute to changes in overall inequality over time. The first is an increase in their within-group inequality, which may impact overall inequality (the "within group inequality" mechanism). The second mechanism is a change in the group's mean income relative to the whole sample (the "between group inequality" mechanism). For example, if the gap between college educated and non-college educated individuals' average wage were to widen, we would expect national inequality to increase as well. Finally, if a group is currently positively contributing to inequality and their relative proportion in the population rises, we would expect overall inequality to rise (the "demographic" mechanism). We explore these three mechanisms using a combination of three methods: *total effect*, *mean effect*, and *marginal effect*.

We define the "total effect" as the impact on inequality (or quantiles) of adding a particular group to the population. For group $\{e, m\}$ with relative contribution share $\omega_{e,m}$, this is equivalent to changing the group's relative share from 0 to $\omega_{e,m}$. The "mean effect" denotes the effect of a group's mean wage on the level of inequality and can be observed by modifying the mean of group $\{e, m\}$ to be equal to the national mean. If we denote the mean of group $\{e, m\}$ as $\mu_{e,m}$ then the mean effect will be proportional to $|\mu - \mu_{e,m}|$. Finally, the "marginal effect" is the effect on inequality of marginally increasing the share of a particular group by a small percent of the population.

2.2.1 Relationship to Variance Decomposition

We first consider the total, mean, and marginal effects in the context of a variance decomposition. Recall that for a particular year, the national variance of wage-income, σ^2, can be decomposed as:

$$\sigma^2 = \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot \sigma^2_{e,m} + \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot (\mu_{e,m} - \mu)^2 \ , \tag{2.3}$$

where $\sigma^2_{e,m}$ is the within group variance of group $\{e, m\}$ and $\mu_{e,m}$ is the within group mean of group $\{e, m\}$.

If we assume that the distribution of sub-group log-income is perfectly normal, then we can calculate the total, mean, and marginal effects of a given group analytically, knowing only their variance and mean and the national variance and mean in a particular year. In these terms, the mean effect for education-migration group $\{e, m\} = \{e^*, m^*\}$ can be expressed as

$$R^{mean}_{e^*,m^*} = \omega_{e^*,m^*} \cdot (\mu_{e^*,m^*} - \mu)^2 \ . \tag{2.4}$$

The larger the mean difference a group has relative to the overall mean, the larger its mean effect will be. A group's total effect is a composition of its within group and across group variance and can be expressed as

$$R^{total}_{e^*,m^*} = \omega_{e^*,m^*} \cdot (\sigma^2_{e^*,m^*} - \overline{\sigma}^2) + \omega_{e^*,m^*} \cdot (\mu_{e^*,m^*} - \mu)^2 \,, \ . \tag{2.5}$$

where $\overline{\sigma}^2 = \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot \sigma^2_{e,m}$. That is, the effect of the within variance comes from the deviation from the average variance, not from the within variance not being zero.

Finally, under conditions of normality, the marginal effect on a particular quantile of income is equivalent to the derivative of the inverse mixed-normal CDF, often denoted the "quantile function," with respect to a group's weight in the population:

$$R^{marg}_{e^*,m^*} = \frac{dF^{-1}(\cdot)}{d\omega_{e^*,m^*}} \ . \tag{2.6}$$

In order to derive an expression for this derivative, consider a distribution made up of two subgroups with distribution function F_1 and F_2 and relative population weights ω and $1-\omega$. We can consider F_1 to be the distribution function of our subgroup of interest and F_2 to be a mixture of normals of the other five subgroups.The population distribution function F is then $F = \omega F_1 + (1-\omega)F_2$. We are interested in how the log income value of a particular quantile, denoted q_τ, changes with respect to ω. First, note that $q_\tau = F^{-1}(\tau)$, where $F^{-1}(\cdot)$ is the inverse distribution function of log income. We can therefore find $\frac{dq_\tau}{d\omega}$ by noticing that $\frac{d\tau}{d\omega} = 0$ and that $\tau = F(q_\tau) = \omega F_1(q_\tau) + (1-\omega)F_2(q_\tau)$. We perform implicit differentiation to obtain:

$$R^{marg}_{e^*,m^*} = \frac{d\,q_\tau}{d\,\omega} = \frac{F_2(q_\tau) - F_1(q_\tau)}{\omega f_1(q_\tau) + (1-\omega) f_2(q_\tau)} \tag{2.7}$$

where $f(\cdot)$ is the probability density function of income.

Figure 1 shows the results of applying this formula to normalized distributions of the six education-migration groups of interest in 2001 and 2022. We assume that each education-migration group $\{e, m\}$ has a log income distribution that is normal with mean $\mu_{e,m}$ and variance $\sigma^2_{e,m}$. As expected, groups with means below the national average, such as the three non-college educated groups, generally have a negative effect on each quantile, whereas groups with means above the national average have positive effects on every quantile.

Figure 1 shows the differences between this normal approximation of the marginal effect and the marginal effect calculated using a non-parametric "RIF-regression" which will be explained shortly. While the normal approximation is good for most quantiles and groups, the approximation is inaccurate for the two college-educated mover groups at the highest quantiles of log income. The observed distributions of income for each group is displayed for 2001 and for 2022 in Figures 8 and 9. We note that for all six groups, the actual log wage distribution has become less normal overtime, particularly at the high end of the wage distribution.

This normal derivation is useful in situations where the entire distribution of income for all groups is not known, but the mean and variance of income by group is known. However, for our subsequent analysis,h we choose to use the RIF method as the inter-decile range of log income is sensitive to these differences for some of the education-migration groups, particularly the two college educated migrant groups.

2.2.2 Precise Numerical Implementation on Generic Distribution

In our analysis, we primarily use the 90/10 inter-quantile range (the "inter-decile range") to track inequality over time. For individual i in year t, we denote log wage as denoted $Y_{i,t}$. In this generic distribution setting, we consider the group-specific *total effect* of an education-migration group to be the effect of excluding that group entirely from our sample when calculating inequality. For an education-migration group $\{e^*, m^*\}$, we identify this effect, which we denote $F^{total}_{e^*,m^*,t^*}$, by calculating the raw inter-decile range for the entire sample year-by-year and then recalculating this range for all individuals in the sample that do not belong to group $\{e^*, m^*\}$. We therefore calculate the total effect as:

$$R^{e^*,m^*,t}_{total,90/10} = IPR_{90/10}(\{Y^{e,m}_{i,t}\}) - IPR_{90/10}(\{Y^{e,m}_{i,t} | \{e, m\} \neq \{e^*, m^*\}\}) \ , \tag{2.8}$$

where $IPR_{q/r}(\cdot)$ is a function for the weighted inter-percentile range between quantiles $q > r$ of a set. This total effect measure is particularly sensitive to changes in the relative proportion of each group over time.

The group-specific *mean effect* of an education-migration group is the impact of this group's average wages—for example, the average wage of foreign-born college graduates—on wage inequality. We ask the question of how wage inequality changes if we equalize the mean wage of a particular education-migration group to the mean wage of all other groups, without shifting within group wage distributions. We calculate this effect by residualizing out the mean difference in wage of this particular group compared to all other groups, while controlling for differences in age distributions. If we denote $Y^{age}_{i,t}$ to be the age adjusted log wage of individual i in year t and $Y^{e^*,m^*,mean}_{i,t}$ to be the wage of individual i in year t once one normalizes the wages of individuals in group $\{e, m\}$ to the mean national wage, then the mean effect can be expressed as

$$R^{e^*,m^*,t}_{mean,90/10} = IPR_{90/10}(\{Y^{age}_{i,t}\}) - IPR_{90/10}(\{Y^{e^*,m^*,mean}_{i,t}\}) \ . \tag{2.9}$$

The derivation for this equation can be found in Appendix B.2.1.

Finally, we calculate the *marginal effect* of an education-migration group to be the effect on the inter-decile range of increasing the relative proportion of a group in the population by 1% (i.e. 35% to 36% or 98% to 99%), while proportionally decreasing the relative proportion of all other individuals by 1%. This metric is a function of the"within group inequality" and "between group inequality" mechanisms, but independent of the relative size of each group. In order to easily implement age controls, we calculate this effect using recentered influence function regressions on the inter-decile range of income.

Recentered Influence Functions and Unconditional Quantile Regressions We employ unconditional quantile regressions (UQR), as described in Firpo, Fortin, and Lemieux (2009), to the inter-decile range. UQR regressions can be used to estimate the marginal change to the distribution that a dummy variable, for example, college-education status, or the average value of a numerical variable, such as age, would have on a particular quantile or on a measure such as the Gini coefficient or inter-quartile range, given controls placed on other variables.

We employ this method in order to estimate the changes to the national income distribution that arise from marginal changes to the distribution of various demographic variables, such as education-migration interaction groups.

Intuition of RIF Regression In order to better understand the mathematical intuition of the RIF Regression and our implementation, suppose we divide the total population in a given year between individuals who are foreign born and college educated and those who are not. We would like to estimate the effect on a quantile τ of a marginal (1%) increase in the relative proportion of foreign born college educated individuals. This effect will vary depending on the actual proportion of foreign born college educated individuals. A change from 45% to 46% will have a different impact than a change from 99% to 100%.

First, we consider the quantiles of log wage-income for different relative proportions of college-educated foreigners. Suppose we consider only two groups in the population, group A, for which $\{e, m\} = \{e^*, m^*\}$ and group B, for which $\{e, m\} \neq \{e^*, m^*\}$. As an input, we use two vectors of quantiles of log income for individuals in groups A and B. We denote log income as Y and we denote y a particular value of log income of interest. Therefore, the conditional income distributions is $F(Y \leq y \mid x = A)$ and $F(Y \leq y \mid x = B)$, where F is the cumulative distribution function of log income. Additionally, we have for each year the relative proportion of individuals in each group, $p_A = \mathbb{P}(x = A)$ and $p_B = \mathbb{P}(x = B)$. Using these values, we can construct the unconditional log wage distribution in terms of p_A:

$$F\left(Y \leq y\right) = p_A \cdot F\left(Y \leq y \mid x = A\right) + (1 - p_A) \cdot \left(Y \leq y \mid x = B\right) \ . \tag{2.10}$$

The log wage distribution is discrete, and we separate the distribution into tenths of percentiles. For the nth percentile in a group G, we denote its weight $\nu_{n,G} = \frac{1}{1000}$. We can therefore rewrite the conditional wage distribution as

$$F(Y \leq y) = \sum_{n=1}^{N_A} \nu_{n,A} \cdot \mathbf{1}\{Y_{n,A} \leq y\} \tag{2.11}$$

and the unconditional wage distribution as

$$F(Y \leq y) = p_A \cdot \left(\sum_{n=1}^{N_A} \nu_{n,A} \cdot \mathbf{1}\{Y_{n,A} \leq y\} \right) + (1 - p_A) \cdot \left(\sum_{n=1}^{N_B} \nu_{n,B} \cdot \mathbf{1}\{Y_{n,B} \leq y\} \right) \ . \tag{2.12}$$

We are interested in estimating the value of the log wage that corresponds to a particular quantile of interest τ, given a particular value of p_A, which we denote $Q_\tau(p_A)$. This value is equal to $F^{-1}(\tau)$.

Denote Ω the full set of all possible values of Y. Given the τ of interest, we observe the values of Y for which

$$\begin{aligned} \underline{y}(\tau, p_A) &= argmax_{Y \in \Omega}(F(Y \leq y; p_A) - \tau) \cdot \mathbf{1}\{F(Y \leq y) \leq \tau\} \\ \overline{y}(\tau, p_A) &= argmax_{Y \in \Omega}(F(Y \leq y; p_A) - \tau) \cdot \mathbf{1}\{F(Y \geq y) \leq \tau\} \end{aligned} \tag{2.13}$$

Finally, we compute the value of $Q_\tau(p_A)$ as:

$$Q_\tau(p_A) = \frac{\overline{y}(\tau, p_A) - \underline{y}(\tau, p_A)}{F(Y \leq \overline{y}(\tau, p_A); p_A) - F(Y \leq \underline{y}(\tau, p_A); p_A)} \cdot (\tau - F(Y \leq \underline{y}(\tau; p_A); p_A)) + \underline{y}(\tau; p_A) \tag{2.14}$$

Figure 10 shows the results of this analytical exercise for college educated foreigners as the base group on the 10th, 50th, and 90th quantiles of log wage. As a simplification, we use tenths of percentiles for each group as our input. As expected, as the proportion of college-educated foreigners rises, the expected value of both the highest and lowest decile of wage increase. From these estimates, we can also calculate the derivative of the unconditional quantile of log wage-income on the relative share. These results are shown in Figure 11.

While these results give us a reliable estimate of the marginal effect of each group on each quantile of income, a RIF regression allows for other factors, such as age, to be controlled for in a straight forward manner, as well as for the entire distribution of log wages to be considered.

For quantile τ and individual log wages $Y_{i,t}$, the re-centered influence function is equal to

$$RIF(Y_i, q_\tau) = q_\tau + \frac{\tau - \mathbf{1}\{Y_i \leq q_\tau\}}{f(q_\tau)} \ , \tag{2.15}$$

where $\mathbf{1}\{y \leq q_Y(\tau)\}$ is an indicator function that takes the value 1 if the particular log wage $Y_{i,t}$ falls below the value of the τth quantile, $q_Y(\tau)$ and $f_Y(q_Y(\tau))$ is the density function of the τth quantile of income. A RIF regression is simply a linear regression with this set of RIF values as the dependent variable.

NEW: A regression coefficient is $(X'X)X'Y$ on a dummy D_i with 0-1 variables for group (e, m) gives $\frac{\Sigma_{i\in(e,m)} D_i Y_i}{N^2_{e,m}}$. In a large sample, the converges to $\frac{E(Y_i|(e,m)}{N_{e,m}}$. In the RIF case, $E(\mathbf{1}\{Y_i \leq q_\tau | i \in (e, m))\}$ is $F_{e,m}(q_\tau)$. However, in a regression that also includes constant, the constant terms will be absorbed in the constant and by Frisch-Waugh, we can find the regression coefficient by regressing $Y - \bar{Y}$ on the regressor. In our case (omitting the constants) we will regress $\mathbf{1}\{Y_i \leq q_\tau\} - F(q_\tau)$ on the dummy. Asymptotically, the constant term just stays constant, so the regression will return $F_{e,m}(q_\tau) - F(q_\tau)$ still divided by the overall density. This differs from our exact derivative which subtracts the CDF of population minus the group in question. (The normalization with $N_{e,m}$ is immaterial, this is only reflects whether we change the share in population or, in the OLS above, one person.) So the general RIF method, which is valid not just for changing shares, is an approximation (valid around the share being zero) while ours is exact, and of course much less data dependent if functional forms for the CDFs are available.

Implementation of RIF Regression We employ several versions of this regression, using dummy variables for migration-education groups as regressors. We also employ age controls. Each regression takes the following form and is performed year-by-year, where $RIF^{90/10range}_{i,t}$ is the re-centered influence function value for individual i in year t for the inter-decile range of log wages, α is a regression constant, $D^{e,m}_{i,t}$ is a dummy variable for a particular migration group m

and education group e, and $Age_{i,t}$ is a discrete variable for Age:

$$\text{RIF}_{i,t}^{IPR^{90/10},e,m} = \alpha + \gamma^{e,m} D_{i,t}^{e,m} + \beta Age_{i,t} \ . \tag{2.16}$$

We run the regressions for one group at at time considering the rest of the population to be the base group. More information on dummy variables and RIF regressions can be found in Appendix B. The marginal effect, denoted $R_{marg}^{e,m,t}$, of group $\{e^*, m^*\}$ in year t is then equivalent to:

$$R_{marg,90/10}^{e^*,m^*,t} = \gamma^{e^*,m^*} \ . \tag{2.17}$$

For our interpretation, we divide this marginal effect by one-hundred and interpret the result as the impact on the inter-decile range of a one percent increase in the relative proportion of a particular education-migration group.

3 Application of Empirical Analysis to Inequality

We apply the framework outlined to a decomposition of changes in log-wage inequality from 2001 to 2022 by six education-migration groups.

3.1 Data

3.1.1 ACS Census Data

The wage income and demographic data used in this paper is taken from the publicly available annual American Community Survey (ACS) taken by the Census Bureau. Data is collected on an individual and household level. Those surveyed were all documented U.S. residents at the time of surveying living within one of the 50 U.S. states or the District of Columbia.[2]

We focus our analysis on those who are between the ages of twenty-five and sixty-five at the time of the survey, who consider themselves fully employed, and who work at least thirty-six hours in a typical week. This is consistent with the recent inequality literature. We study wage income, calculated as pre-tax income from wages, salaries, commissions, cash bonuses, tips, or any other source of income coming from an employer in the year prior to being surveyed.[3]

Consistent with much of the contemporary wage inequality literature, we focus our analysis on the 90/10 inter-percentile range (the "inter-decile range"). We also consider inequality in the upper and lower halves of the wage distribution (the 90/50 and 50/10 inter-percentile ranges respectively). Because of top-coding on the wage-income data, we focus our analysis only on metrics that are not sensitive to income concentrations in the highest percentiles.

Income values in the ACS survey are originally reported in contemporary dollars so we adjust wages for inflation using CPI data from the St. Louis Federal Reserve Bank's FRED

2. Between the years of 2001 and 2004, approximately 1 in every 250 U.S. residents were surveyed. This number increased to 1 in every 100 from 2005 onwards.

3. For incomes below $1,000, values are rounded to the nearest 10. For incomes between $1,000 and $50,000, values are rounded to the nearest thousand. Finally, above $50,000, incomes are rounded to the nearest thousand. Additionally, there is a top code on the income values, equivalent to $200,000 for 2001-2002 and to the 99.5 quantile by state from 2003 onwards

database. All inflation adjusted income values are in 2022 dollars.

3.1.2 Data Demographic Manipulations

We study the impacts of education and migration status on national inequality and we define the following subgroups for this analysis. We group individuals by college attainment status and define a college educated individual as someone who as attained a bachelor's degree and/or a graduate degree. For our migration analysis, we use the Census Bureau's division of four U.S. regions to categorize each individual's region of residence. These regions are as follows: Northeast, Midwest, South, and West. We group individuals into three categories of migration status depending on the relationship between where they were born and which Census region they reside in at the time of survey. Domestic non-movers are defined as individuals who were born in the same Census Bureau region as they report residing in at the time of survey. Domestic movers are defined as those who were born within the United States, excluding U.S. territories, and now live in a different Census Bureau region than that of their birth. Finally, Foreign-born individuals are defined as those born outside the United States or in a U.S. territory. We refer to domestic movers and foreign-born individuals collectively as "migrants."

3.2 Summary Statistics of Data

3.2.1 Demographics of Sample

Gender and Age Across all twenty-two years of our sample, the share of woman is between forty-one and forty-four percent. The average age of those included in our sample is between forty-one and forty-four depending on the year.[4]

Education and Migration In order to understand the total and mean effects of migration groups and education status on national once inequality, we report the demographic shifts in these groups over time. Table 1 presents the proportion of individuals in the sample belonging to each education-migration group nationally and by Census region, in 2001 and 2019. We note that within each migration group, the proportion of college educated individuals has increased. Most notably, the proportion of college educated foreign born individuals has almost doubled from 4.8 percent to 8.5 percent. By migration group, the proportion of U.S. non-movers has modestly decreased, whereas the proportion of domestic movers has remained nearly unchanged and the proportion of foreign-born individuals has grown.

3.2.2 Income

Table 2 shows the mean and median incomes in 2022 dollars in 2001 and 2019 for the entire sample and for each education-migration group. We note that across the entire population, median real wages grew by 7.6% but these returns were not distributed equally. Median wages grew 13.6% and 11.5% for college and non-college educated foreign born individuals respectively, whereas wages for non-college educated domestic non-movers fell by 1.7%.

4. Unweighted, the smallest sample, in 2001, was 329,915, and the largest sample, in 2022, was 1,007,995. The weighted average of hours worked in the sample is between 43.86 and 44.60.

3.2.3 Wage inequality

As documented in the inequality literature, inequality has been on the rise for several decades. This change is reflected across various inequality measures, including the inter-decile range, inter-ventile range, and the Gini coefficient of income. Based on the ACS sample, between 2001 and 2022, the inter-decile range of log-income rose by 0.066.

Within group inequality has also risen within almost every education-migration group between 2001 and 2019. The rise has been most pronounced in college educated domestic non-movers, for which the inter-decile range rose by 0.18. Across all years, within group inequality has been between 1.42 and 1.96 for the three college educated groups and between 1.43 and 1.61 for the three non-college educated groups. Within group inequality has consistently been lowest for non-college educated foreign-born individuals, for whom the inter-decile range actually decreased between 2001 and 2022 by 0.11. Within group inequality has consistently been highest among college educated foreign-born individuals. Additionally, these results are consistent with the findings of Autor, Katz, and Kearney (2008) in that the 90/50 range rose between 2001 and 2022 while the 50/10 range actually decreased.

3.3 Wage Inequality Analysis

3.3.1 Variance Decomposition

We begin our analysis with a variance decomposition to calculate the relative contribution of each demographic group to national variance in log income. Figure 13 shows the total within-group and across-group variances for our six education migration groups, as well as the across group variation attributable to variance across just education and just migration groups. We first observe that across all years in our sample, within group variance is over four times larger than between group variance. Additionally, across group variance is relatively stable over time, whereas within group variance is increasing. This difference seems to imply that within group inequality is a larger contributor to national inequality than inequality between groups, as it relates to education-migration groups. We can also break down this within and between group inequality by each education-migration group.

Figure 14a shows the within group variance of each group over time. We note that for all three college educated group, within group inequality is rising. Additionally, both migrant groups have greater within group variance than their non-migrant equally educated counterparts. In particular, college educated foreign-born migrants have the greatest within group variance. In contrast, this same group's mean income is not furthest from the national mean income (see Figure 14c). Therefore, it appears that migrant groups may contribute most significantly to inequality via their within group inequality rather than through their mean income difference relative to the population.

We also consider the weighted contribution of each group's within group variance to the overall variance in the national wage distribution. This contribution is a function of both the group's relative representation in the population and their within group variance. Figure 14b presents these values over time. We note that the two largest group, domestic non-movers of each education status make the largest contribution. For college-educated non-movers, their contribution is growing, whereas for their non-college educated counter parts, their contribution is decreasing. For both groups, within group variance has marginally increased over time so these trends appear to be driven by the respective increase and decrease of these groups representation in the population over time. Since 2001, the relative contrition of the within group variance of college educated foreign-born migrants has increased from 0.0281 to 0.585, which represents a 108% increase. This increase appears to be a result of both an increase in the group's representation in the population and an increase in their within group variance. Both the absolute variance and the contribution to variance for non-college educated foreign-born individuals have been stable over time.

3.3.2 Total and mean effect

We begin our analysis of inequality trends by documenting changes in the magnitude and direction of total and mean effects by education-migration group. Figure 15 shows the total and mean effects for each of the six education-migration groups on the inter-decile range of log income, as well as the "raw" (unadjusted) inequality across the entire population, and each group's proportion of the population in each year.

In five out of the six education-migration groups, the mean effect is smaller than the total effect, implying that the majority of the group's impact on inequality is attributable to variations in their entire distribution, not simply how their mean compares to the national mean. We also observe that this mean effect is disproportionately large for foreign born individuals without a college degree relative to their proportion of the national population.

With the exception of non-mover non-college educated individuals, the total effect of all groups is positive, meaning that removing all members of the group lowers inequality. Both the mover- and college-educated groups have a higher total effect on inequality than their non-mover and/or non-college-educated counterparts, even considering that non-college-educated domestic non-movers represent the largest proportion of the population in all years of the sample. Foreign college educated individuals have the highest total group effect of any group on the inter-decile range of log income, despite representing the smallest proportion of the population in each year. The size of this effect has remained stable over time. Domestic movers with college educations have a very similar effect but represent a slightly higher proportion of the population, especially in the earlier years of the sample. Interestingly, despite representing the largest share of the population, non-college educated domestic non-movers have the smallest absolute total effect on inequality in most years of the sample.

We also decompose the inter-decile range effect into 90/50 and 50/10 range effects in order to examine if these effects on inequality occur due to widening in the upper or lower quantile distributions–see Figures 16 and 17. From 2001 to 2022, the "raw"" unadjusted 90/50 and 50/10 ranges have changed little, with the 90/50 range slightly increasing over time and the 50/10 range remaining stable.

Among both migrant non-college educated groups the total effects are more pronounced on the 50/10 quantile range than 90/50 quantile range. In contrast, the mean effect of non-college educated domestic migrants is lower on the 50/10 range than the 90/50 range. Among college educated migrants of both types, the total effect has generally been larger for the 90/50 range than the 50/10 while the mean effect was larger for the 90/50 range in every year of the sample. We also note that when removing both college educated domestic and foreign migrants, the raw 90/50 range becomes almost stable over time and is significantly lower in absolute value.

3.3.3 Marginal Effects

We examine marginal effects that are independent of relative population size. Figure 18 shows the marginal effects on the inter-decile range and three quantiles calculated from a regression on one cross migration-education group at a time. In every year of the sample, college educated migrants of both types have a higher marginal effect on inequality than all other groups, implying that a randomly selected college educated migrant is more likely to be found in the tails of the income distribution than a college-educated non-migrant. Similarly, in every year, all four migrant groups have a greater marginal effect on inequality than their equally educated domestic counterparts.

Additionally, since 2008, foreign-born college educated individuals have had a higher marginal effect on inequality than college educated domestic movers. Between 2011 and 2018, the marginal effect of every group except non-college educated foreigners rose. In contrast, the effect of non-college educated foreign-born individuals has declined consistently since 2012 and has become negative since the 2019. This is largely driven by the decrease in the magnitude of their effect on the lower quantiles (see Figure 18d), as their marginal effect on the 50th and 90th quantiles has been stable over time.

3.4 Discussion

First, differences in mean income across groups only partly explain inequality. This is particularly true with regards to the four migrant groups, whose income distributions tend to be much wider than the non-migrant distributions. Second, removing a group from the population measures the impact of this group. In either case, a larger group will have a larger total effect than a comparable smaller group with the same wage distribution. However, we note that even despite this, the smallest group, college educated foreigners, have the largest total effect on inequality, with a comparably small mean effect. This total effect is particularly pronounced on the 90/50 quantile range. This reflects substantial within group variation in the wages of foreign-born college graduates.

Third, the marginal effect of each group provides insight into the effect of each group irrespective of their relative size. This effect may be of particular relevance for policy as it measures the impact of an small increase in, say, immigration of college graduates. The trends in marginal effects are largely similar to the size sensitive total and mean effects, with college educated foreigners having the largest marginal impact.

Our results suggest that a relative increase in the size of any college-educated group or in numbers of non-college educated foreigners increases wage inequality. However, our results do not suggest that an increase in the proportion of any one group makes another group worse off in terms of wages, as we do not attempt to measure the impact of any group on the wages of another. The income distribution of non-college educated non-migrants could remain completely unchanged and wage inequality would still rise if another 1% of the population were to become college educated foreigners. Additionally, the marginal effect of each group affects different sections of the income distribution. College educated movers increase inequality largely by widening the range between the median and upper quantiles more than the range between the median and lower quantiles. In contrast, non-college educated foreigners increase inequality by increasing the lower quantile range more than the upper quantile range, though this effect has weakened over time. Non-college educated domestic born individuals have a negative marginal effect on inequality despite having neutral or positive mean and total effects on inequality. This demonstrates the importance of considering all three measures, as a mass exodus of domestic born individuals could raise inequality, while a gradual and small movement towards more domestic born individuals in the population would be expected to lower inequality.

3.4.1 Implications for Education and Migration Policies

Our results have implications for the current discourse about education and immigration. First, the continuing trend of increasing educational attainment should be expected to result in increasing inequality. This change will be particularly pronounced in the upper quantiles of the wage distribution. In contrast, the positive mean return to higher education is expected to have little effect on inequality.

Second, policies targeting mass removal of immigrants may have a short term negative effect on national inequality; however, this does not mean that within group inequality would change or that any individual person would necessarily be made better off. In contrast, policies encouraging marginal increases in the relative proportion of college educated foreigners are expected to be accompanied by increases in inequality, but these increases are predominately reflected in changes to the upper quantiles of the income distribution. Additionally, should domestic migration continue to rise we should expect wage inequality to rise as well, although this marginal effect is decreasing over time. Once again, this effect is predominantly observed as a widening in the upper quantiles of income.

3.4.2 Limitations

We note the following limitations to our analysis. We take wages in each year as exogenous and hence do not make claims about the impact of demographic changes in one group to wages in another. Across years, we note trends in demographic changes and changes to inequality but do not make claims about casual relationships in these trends.

We use the inter-decile range as our main wage inequality metric. Due to the nature of our data being top-coded, we cannot make precise statements about the Gini coefficient or any other inequality measure that is sensitive to income concentration above the top 1% of income.

4 Conclusion

We propose a new framework to decomposing contributions to inequality by demographic groups over time. This framework involves three types of effect: the "total effect", which is a composition of a group's within group wage distribution, their across group variance, and their relative size in the population; the "mean effect", which is a composition of a group's mean relative to the population and their relative size; and the "marginal effect", which is a composition of a group's within group wage distribution and their across group variance, independent of the group's relative size. We use recentered influence function regressions applied to the inter-decile range to calculate the marginal effects.

We apply this framework to six education-migration groups, all of which have contributed to increases in wage inequality as measured by the inter-decile range. All three college educated have greater total, mean, and marginal effects on the upper quantile (90/50) range whereas non-college educated foreigners have had a greater impact on the lower quantile (50/10) range. We demonstrate that both non-college educated foreign migrants and college educated domestic and foreign migrants have a pronounced positive impact on wage inequality but in different ways. The distributions of college educated domestic and foreign migrants have been associated with an increase wage inequality by being positively correlated with the upper quantiles of the income distribution. This impact has increased for foreign migrants over time and has surpassed the impact of domestic migrants. Non-college educated foreign migrants have a negative impact on the lower quantiles of the income distribution but this effect has decreased over time. Domestic non-college migrants have close to zero correlation with inequality. For all groups, total effects are significantly higher than mean effects.

These findings also have policy implications. With educational attainment and foreign immigration continuing to rise, we should expect inequality to rise as well, irrespective of each group's effect on the wages of other groups. Policies which discourage immigration may have a negative effect on inequality simply due to decreasing the heterogeneity in the population. In contrast, policies which encourage domestic migration and international immigration are likely to be accompanied by increases in wage inequality.

Tables and Figures

4.1 Migration Matrix

	US Non-Mover		US Mover		Foreign Born	
Migration Status	No Coll	Coll	No Coll	Coll	No Coll	Coll
2001						
All	46.3%	19.0%	11.3%	7.8%	10.9%	4.8%
Midwest	55.7%	22.5%	8.8%	4.9%	5.2%	2.9%
Northeast	46.8%	24.7%	4.5%	4.4%	13.0%	6.5%
South	47.6%	16.6%	13.8%	9.2%	9.0%	3.8%
West	33.6%	14.1%	15.4%	11.4%	18.5%	6.9%
2019						
All	38.4%	24.3%	8.7%	8.3%	12.3%	7.9%
Midwest	47.1%	29.7%	5.5%	6.7%	6.4%	4.6%
Northeast	35.5%	31.1%	6.0%	3.5%	13.8%	10.2%
South	38.3%	21.0%	10.4%	11.4%	11.8%	7.2%
West	33.2%	19.9%	10.8%	8.4%	17.5%	10.3%

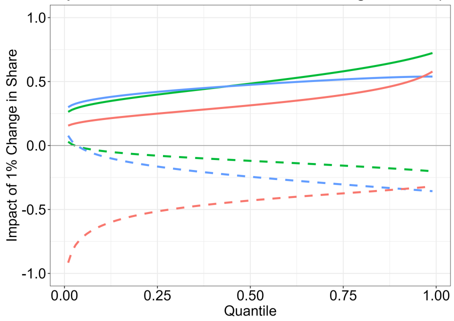
Table 1: Education-Migration Matrix (2001-2019)

	All	US Non-Mover		US Mover		Foreign-Born	
Measure	All	No Coll	Coll	No Coll	Coll	No Coll	Coll
2001							
Median	56.0	49.4	79.0	51.0	85.6	37.9	79.0
Mean	71.4	56.8	102.5	60.4	115.9	45.7	101.5
2019							
Median	59.9	47.5	81.1	50.2	90.4	40.6	88.1
Mean	81.0	58.2	104.4	62.4	122.6	49.5	116.1

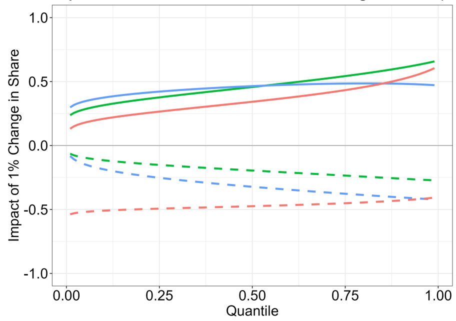
Table 2: Raw Average Measures of Income by Education-Migration Group (in thousands, 2022 dollars)

4.2 RIF Regression and Normal Simulation

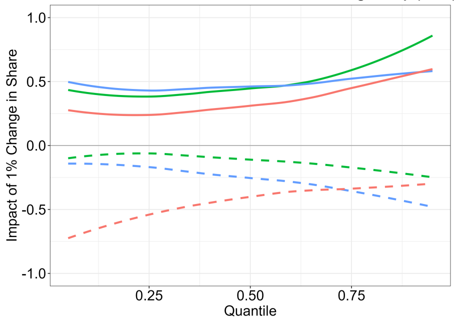
(a) Normal Estimation of Impact of 1% Change on Log Income across Quantiles (2001)



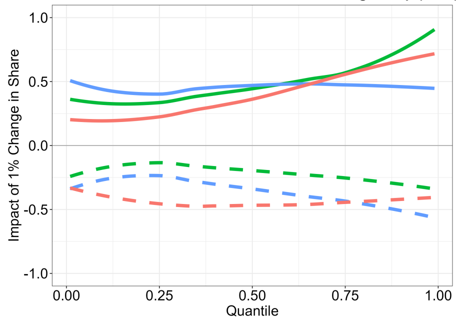
(b) Normal Estimation of Impact of 1% Change on Log Income across Quantiles (2019)



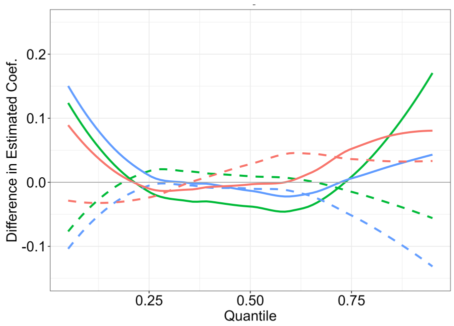
(c) Non-Parametric Estimation of Impact of 1% Change on Log Income Across Quantiles (2001)



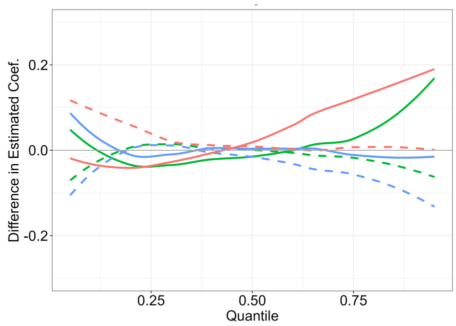
(d) Non-Parametric Estimation of Impact of 1% Change on Log Income Across Quantiles (2019)



(e) Difference Between Actual vs Normal RIF Coefficients by Edu-Mig Group (2001)



(f) Difference Between Actual vs Normal RIF Coefficients by Edu-Mig Group (2019)



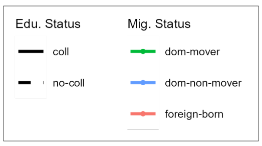
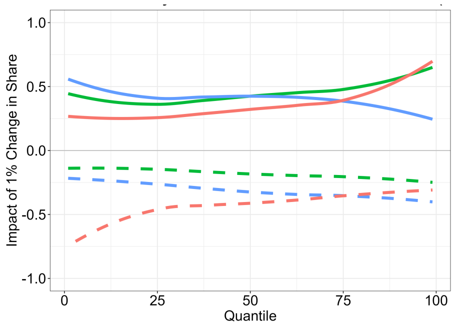


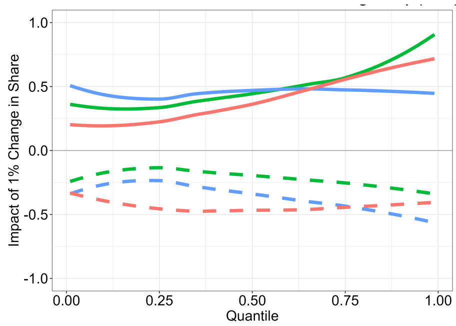
Figure 1: Normal and Non-Parametric Estimation of Impact of 1% Change on Log Income across Quantiles (2019)

Note: Figure reflects impact of 1% change in share multiplied by 100

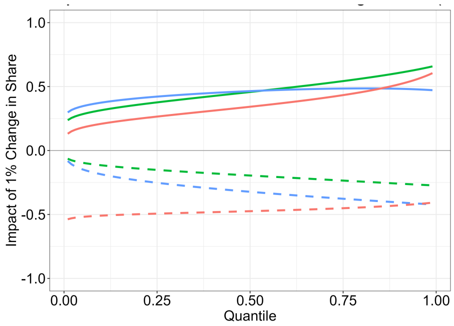
(a) Our Estimation from Quantiles of Impact of 1% Change on Log Income across Quantiles (2019)



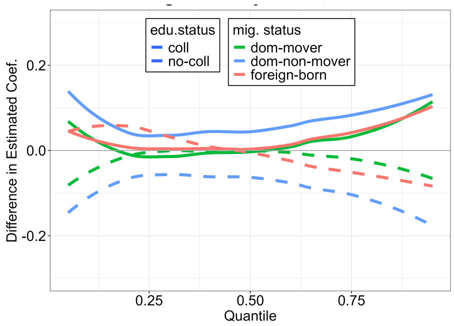
(b) Non-Parametric Firpo Estimation of Impact of 1% Change on Log Income Across Quantiles (2019)



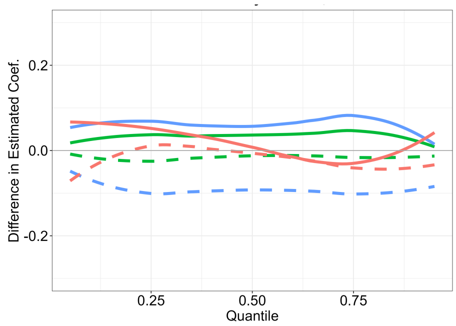
(c) Normal Estimation of Impact of 1% Change on Log Income Across Quantiles (2019)



(d) Difference Between Firpo RIF vs Our RIF Coefficients by Edu-Mig Group (2019)



(e) Difference Between Normal vs Quantile RIF Coefficients by Edu-Mig Group (2019)



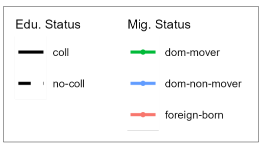
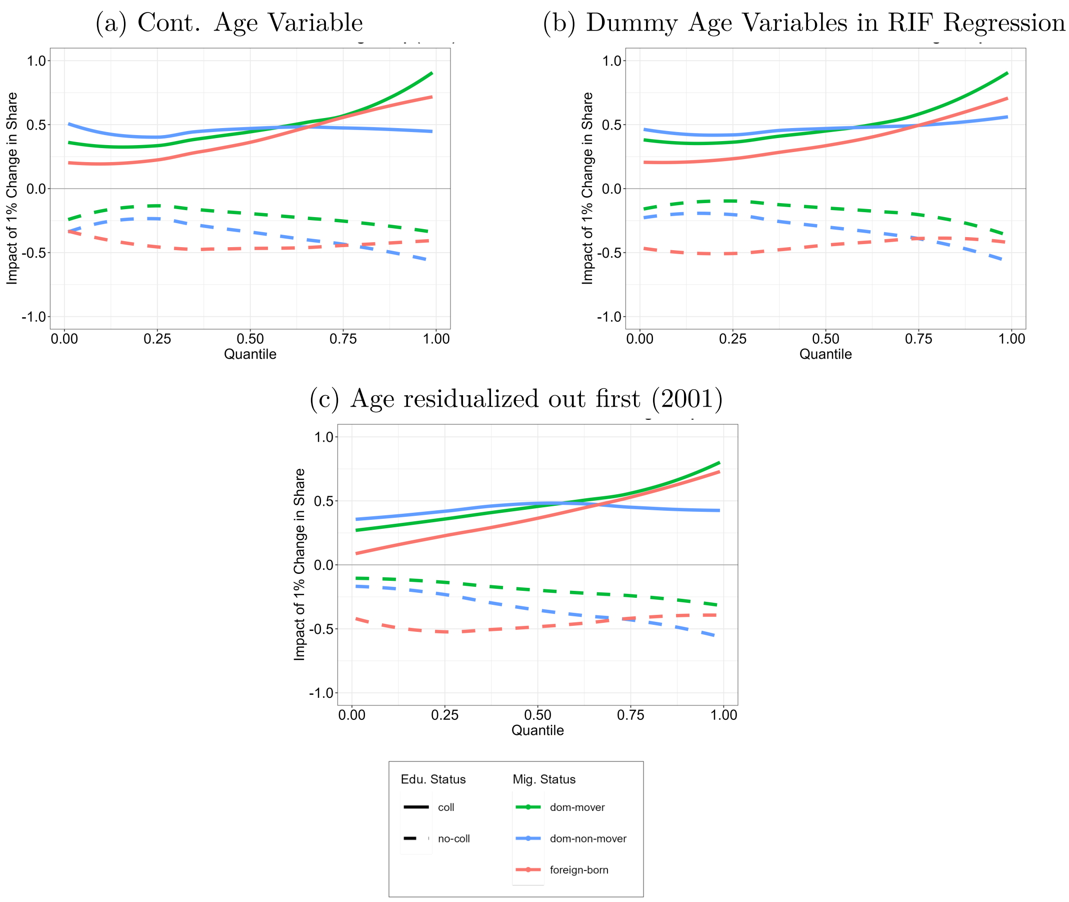


Figure 2: Quantile and Non-Parametric Estimation of Impact of 1% Change on Log Income across Quantiles (2019)

Note: Figure reflects impact of 1% change in share multiplied by 100



(a) Cont. Age Variable

(b) Dummy Age Variables in RIF Regression

(c) Age residualized out first (2001)

Figure 3: Comparison of Three RIFs in Stata: cont. age variable, dummy age variables, age residualized out first via dummies (2019)

Note: Figure reflects impact of 1% change in share multiplied by 100

(a) Difference Between Cont. Age and Age Residualized

(b) Difference Between Dummy Age and Age Residualized

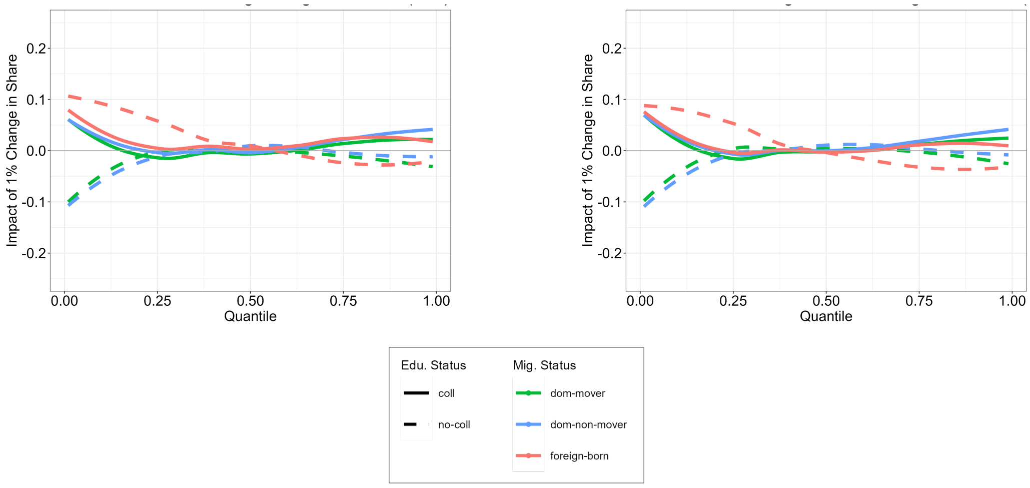


Figure 4: Comparison of Three RIFs in Stata (Differences): cont. age variable, dummy age variables, age residualized out first via dummies (2019)

Note: Figure reflects impact of 1% change in share multiplied by 100

(a) Normal Estimation of Impact of 1% Change on Log Income across Quantiles (2001)

(b) Non-Parametric Estimation of Impact of 1% Change on Log Income across Quantiles (2019)

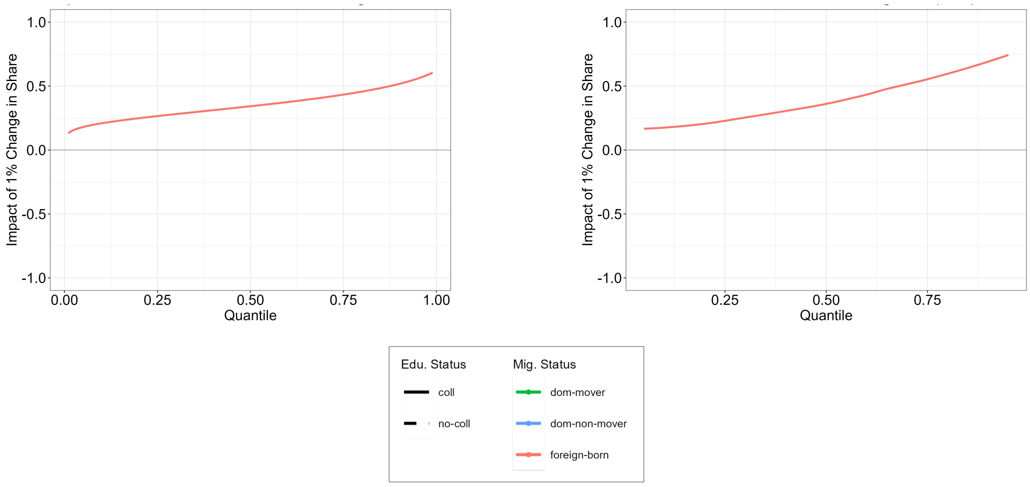
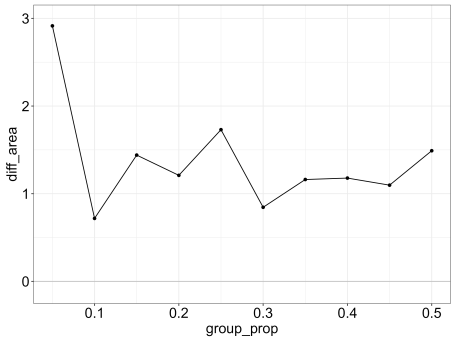


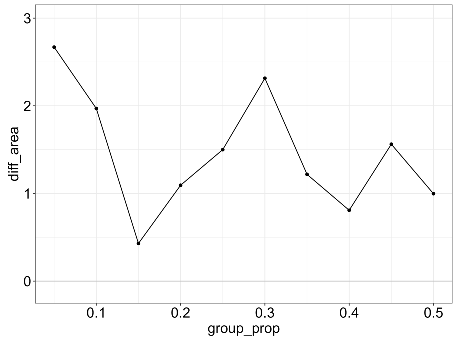
Figure 5: Estimates of RIF Coefficients for Log Income by Quantile for One Group

Note: Figure reflects impact of 1% change in share multiplied by 100

(a) Abs. Difference by Proportion for variances 1,1



(b) Abs. Difference by Proportion for variances 1,1.4



(c) Abs. Difference by Proportion for variances 1,1.8

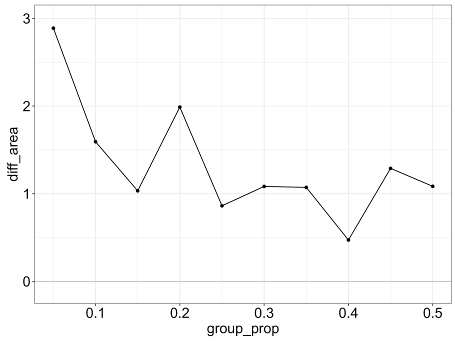
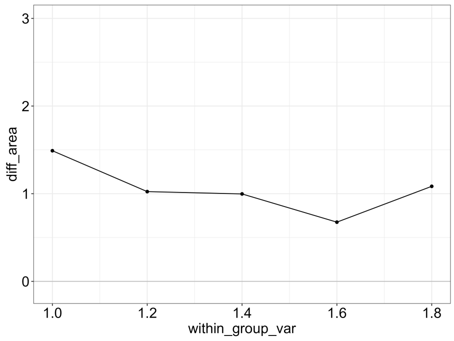
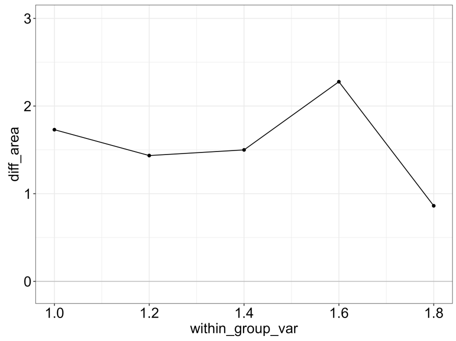


Figure 6: Area of Absolute Difference Between Exact and Firpo RIF Estimates by Group Proportion for Different Variances

(a) Abs. Difference by Variances for proportions 0.5, 0.5



(b) Abs. Difference by Variances for proportions 0.25, 0.75



(c) Abs. Difference by Variances for proportions 0.05, 0.95

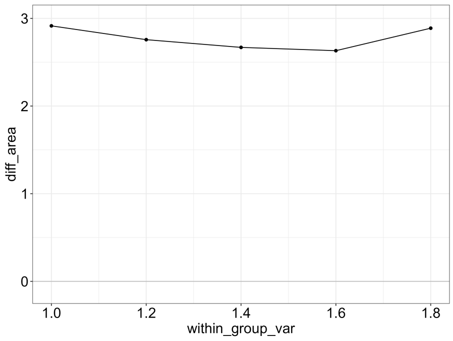


Figure 7: Area of Absolute Difference Between Exact and Firpo RIF Estimates by Group Variances for Different Proportions

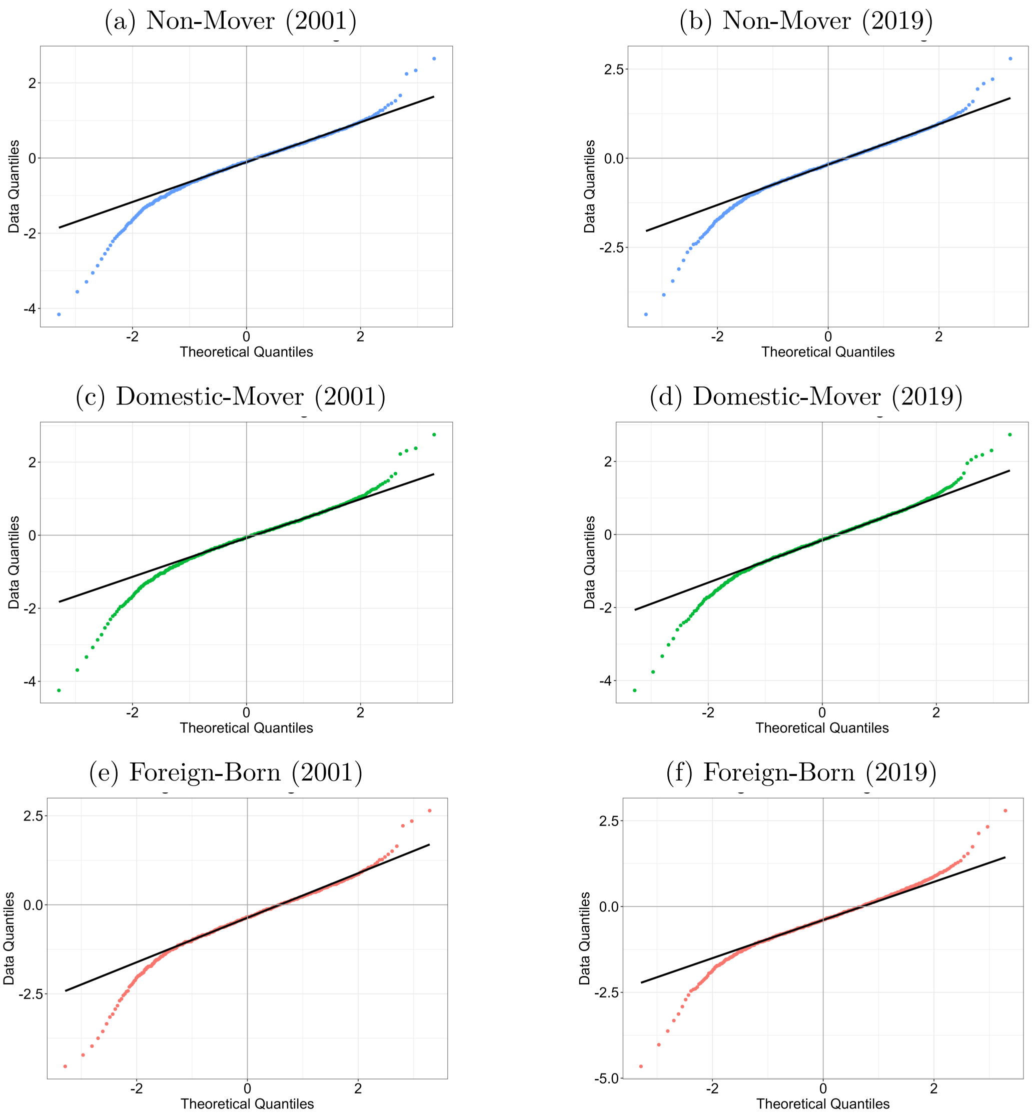
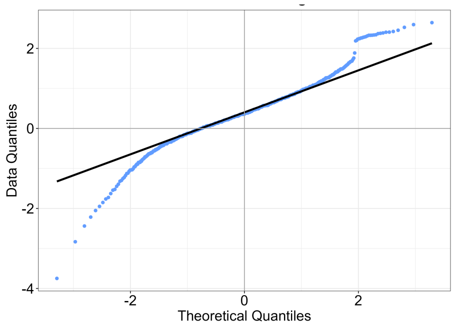
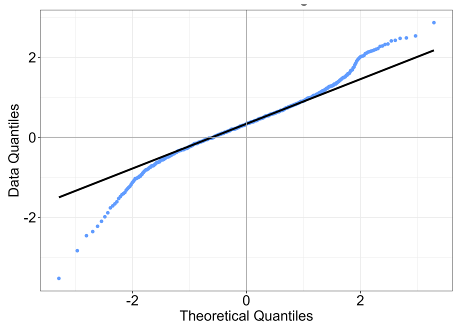


Figure 8: QQ Plots of Log Wage Income for Non-College Educated Migration Groups (2001-2022)

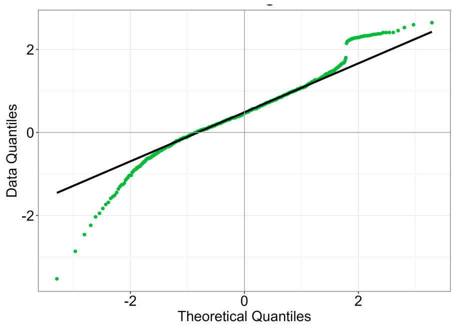
(a) Non-Mover (2001)



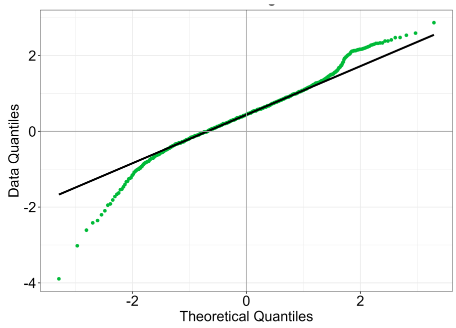
(b) Non-Mover (2019)



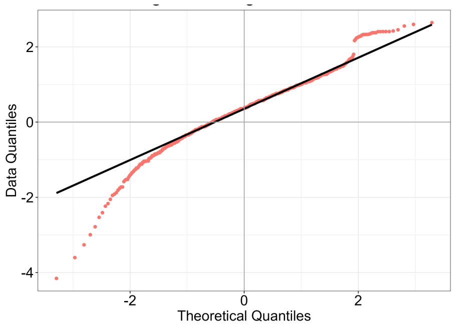
(c) Domestic-Mover (2001)



(d) Domestic-Mover (2019)



(e) Foreign-Born (2001)



(f) Foreign-Born (2019)

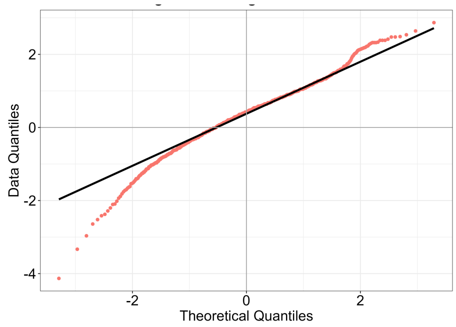


Figure 9: QQ Plots of Log Wage Income for College-Educated Migration Groups (2001-2022)

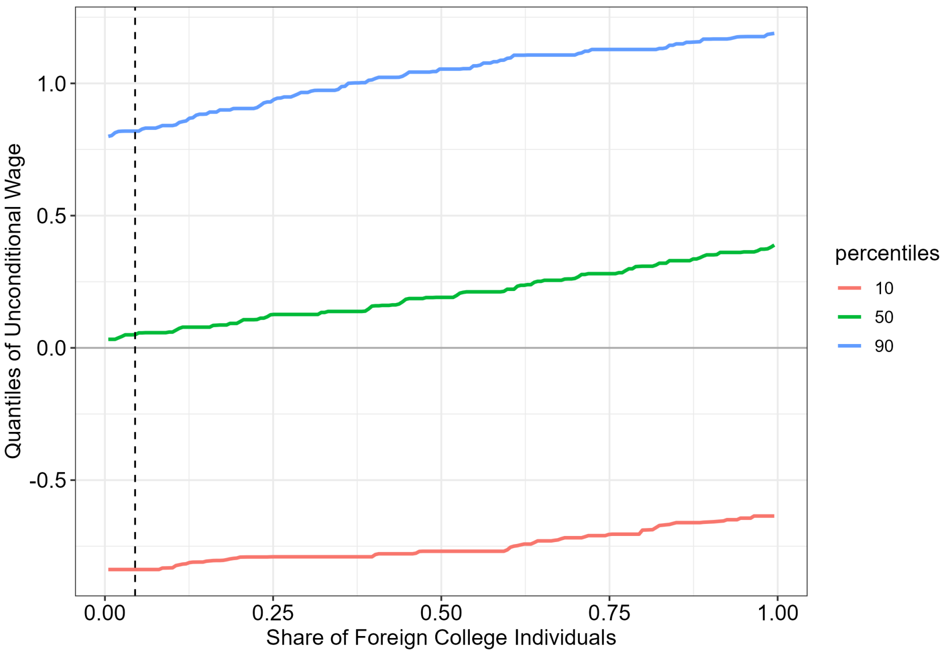


Figure 10: Unconditional Quantiles of Age Residualized Log Wage by Proportion of College-Educated Foreigners in 2001

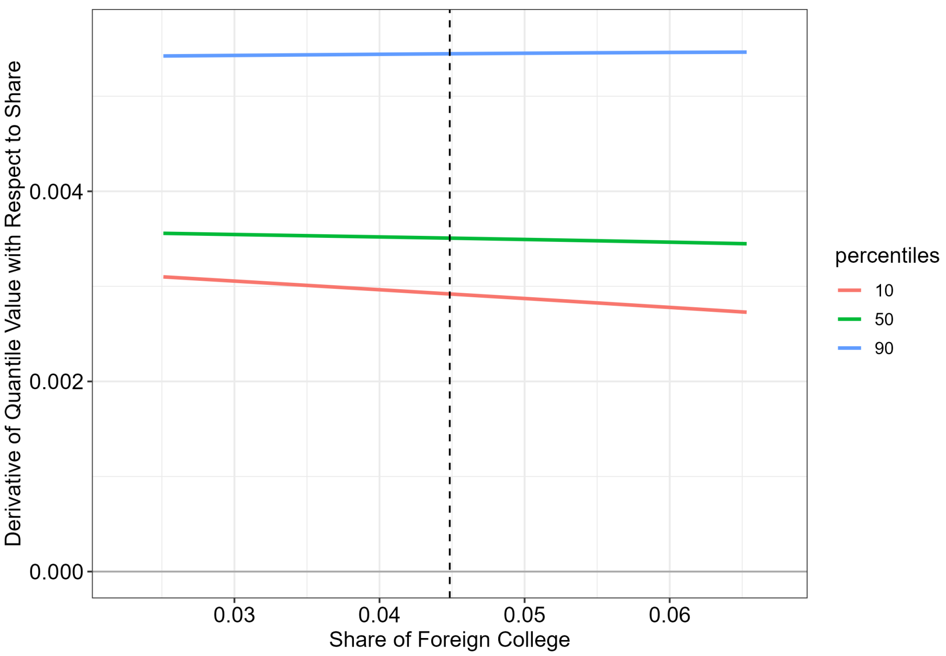


Figure 11: Derivative of Unconditional Quantiles of Log Wage by Proportion of College-Educated Foreigners in 2001

4.2.1 Comparison to Conditional Quantile

(a) Unconditional Quantile Regression Results by Quantile (2001)

(b) Conditional Quantile Regression Results by Quantile (2001)

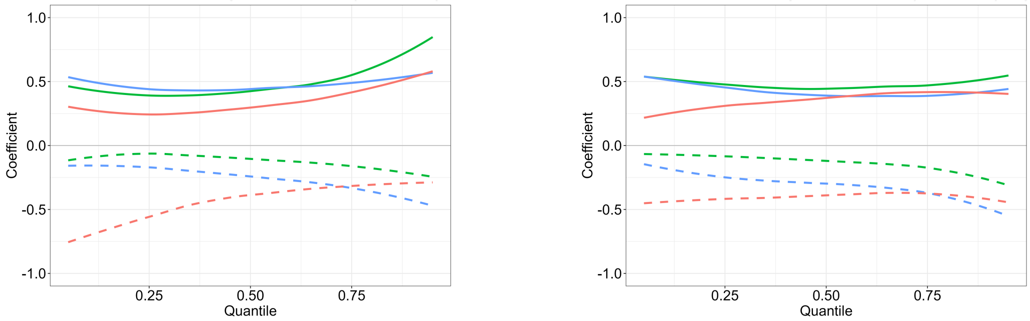


Figure 12: Quantile Regression Results Comparison

4.3 Variance Decomposition

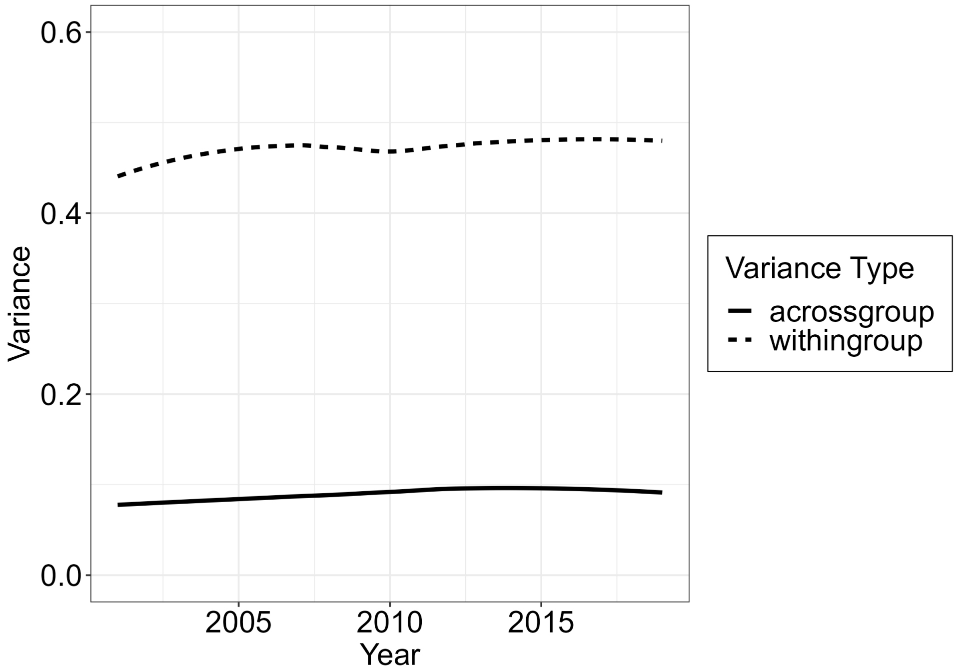


Figure 13: Within and Across Group Variance in Age Residualized Log Wage for Six Education-Migration Groups Over Time

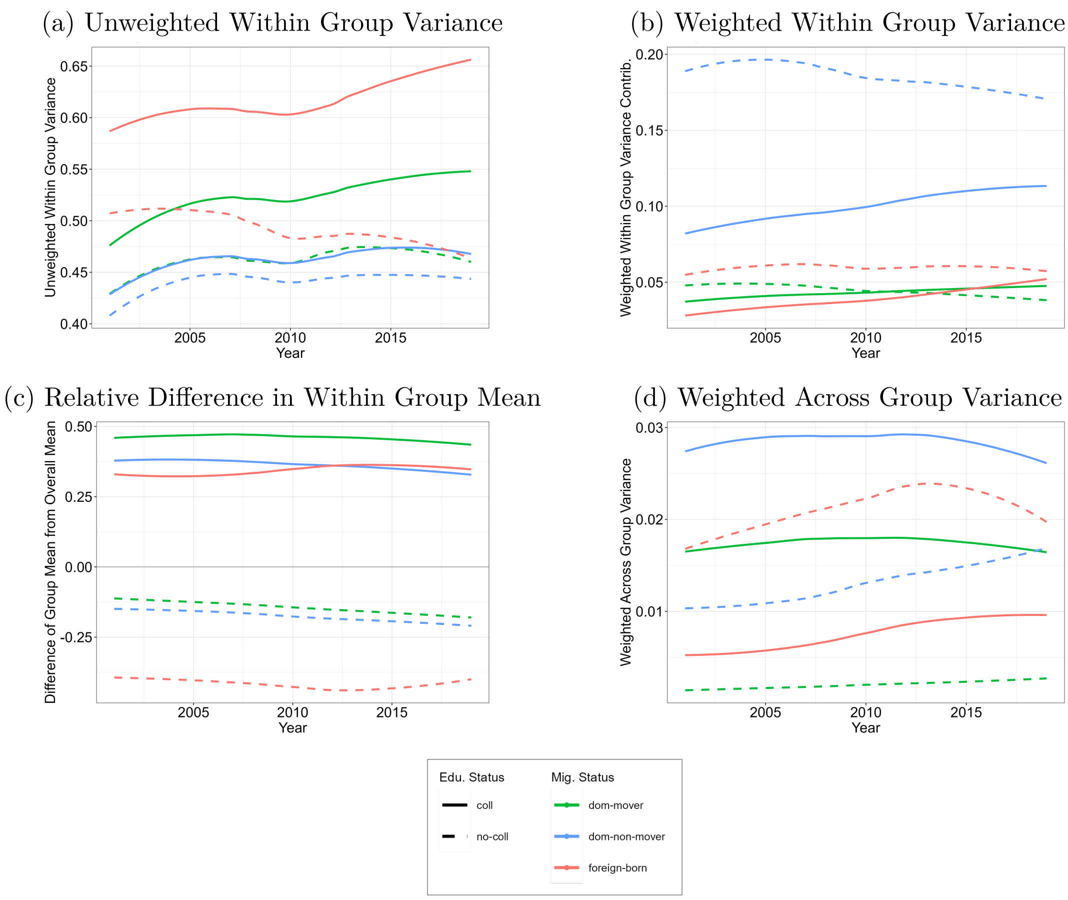


Figure 14: Weighted and Unweighted Within and Across Group Variance in Log Income (Age Controlled)

4.4 Total and Mean Effects

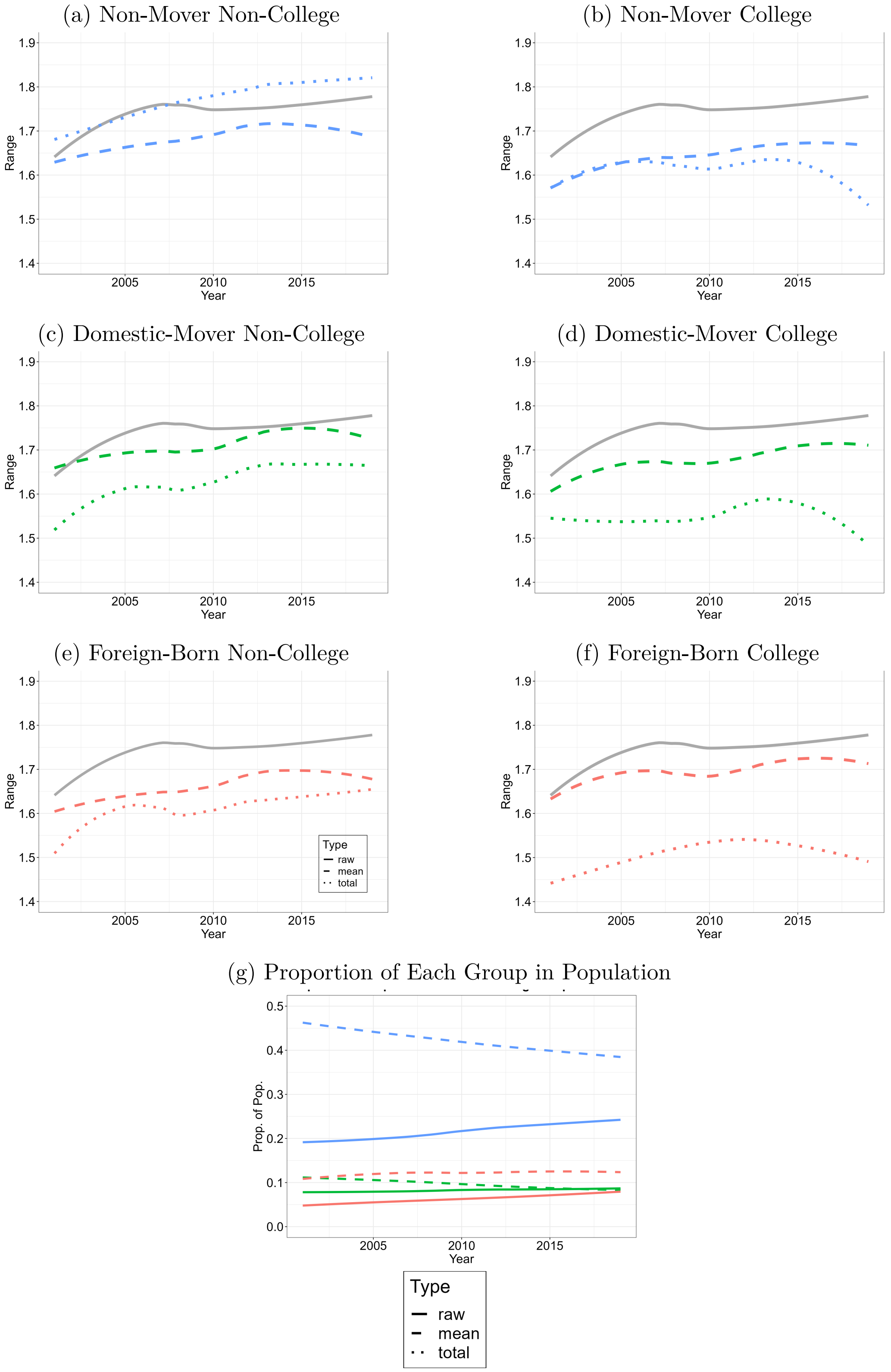
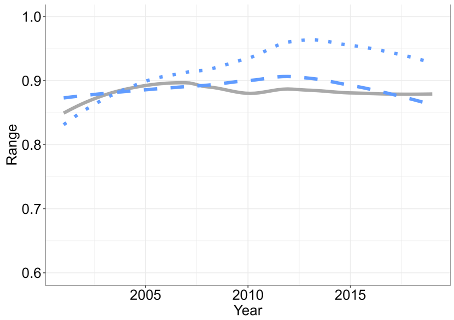
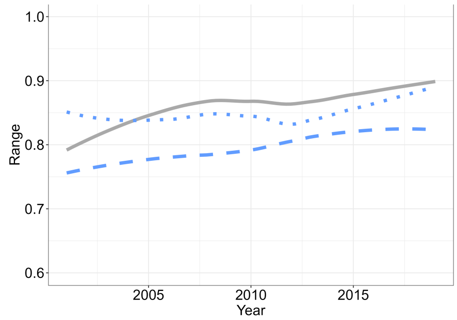


Figure 15: Top Panel: Effect on the 90/10 Quantile Range of Age Residualized Log Wage of Removing Each Education-Migration Group (Total) and Regressing on Each Education-Migration Group (Mean) Relative to Raw Log Income Range (Raw); Bottom Panel: Proportion Relative to Total Population of Each Education-Migration Group

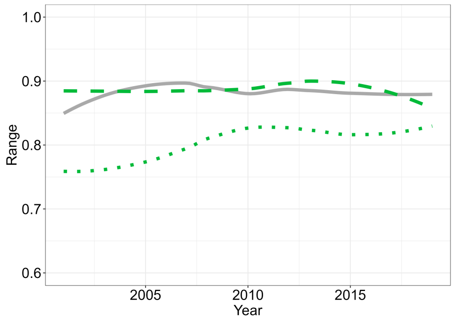
(a) Non-Mover No-College, 50/10



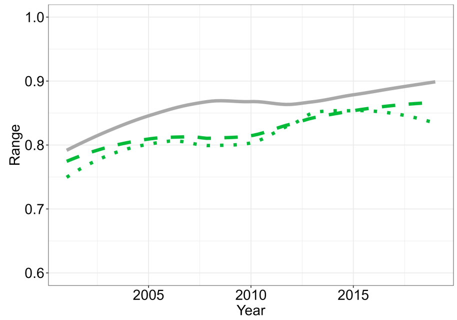
(b) Non-Mover No-College, 90/50



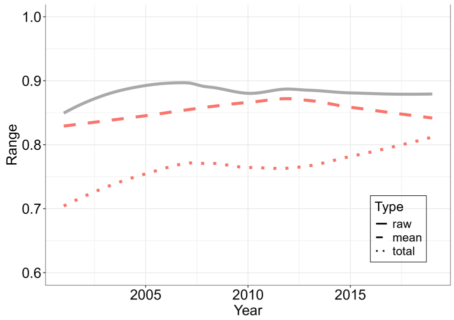
(c) Domestic Mover No-College, 50/10



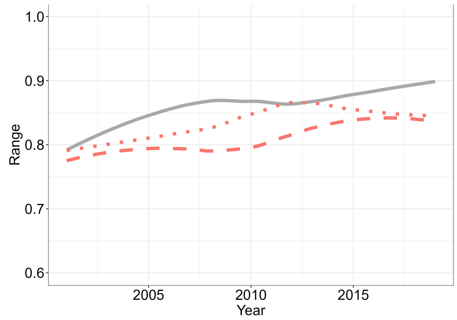
(d) Domestic Mover No-College, 90/50



(e) Foreign-Born No-College, 50/10



(f) Foreign-Born No-College, 90/50



(g) Proportion of Each Group in Population

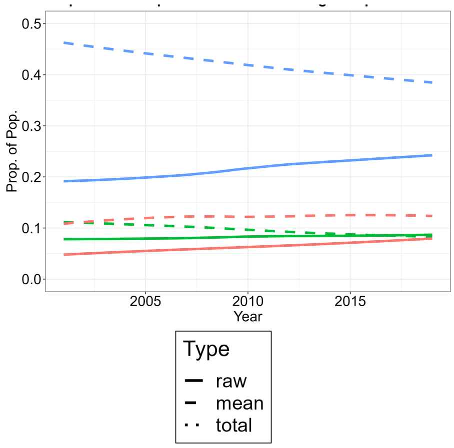


Figure 16: Top Panel: Effect on the 50/10 and 90/50 Quantile Range of Age Residualized Log Income of Removing Each Non-College Educated Migration Group (Total) and Regressing on Each Non-College Educated Migration Group (Mean) Relative to Raw Log Income Range (Raw); Bottom Panel: Proportion Relative to Total Population of Each Education-Migration Group

Note: Loess method used for smoothing. Regression equation uses dummy for the education-migration group and four dummy variables for age

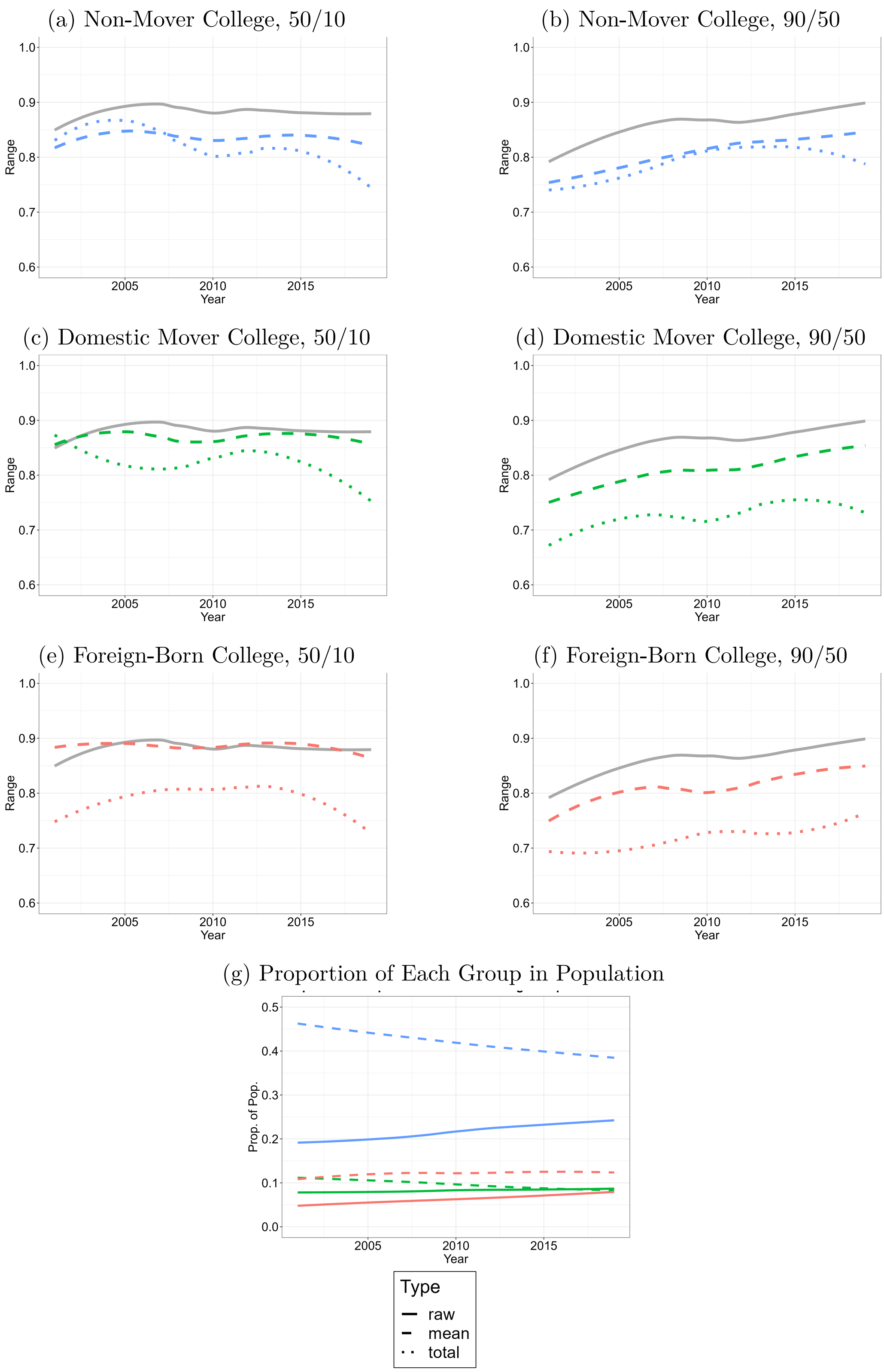


Figure 17: Top Panel: Effect on the 50/10 and 90/50 Quantile Range of Age Residualized Log Income of Removing Each College Educated Migration Group (Total) and Regressing on Each College Educated Migration Group (Mean) Relative to Raw Log Income Range (Raw); Bottom Panel: Proportion Relative to Total Population of Each Education-Migration Group

Note: Loess method used for smoothing. Regression equation uses dummy for the education-migration group and four dummy variables for age

5 Marginal Effects

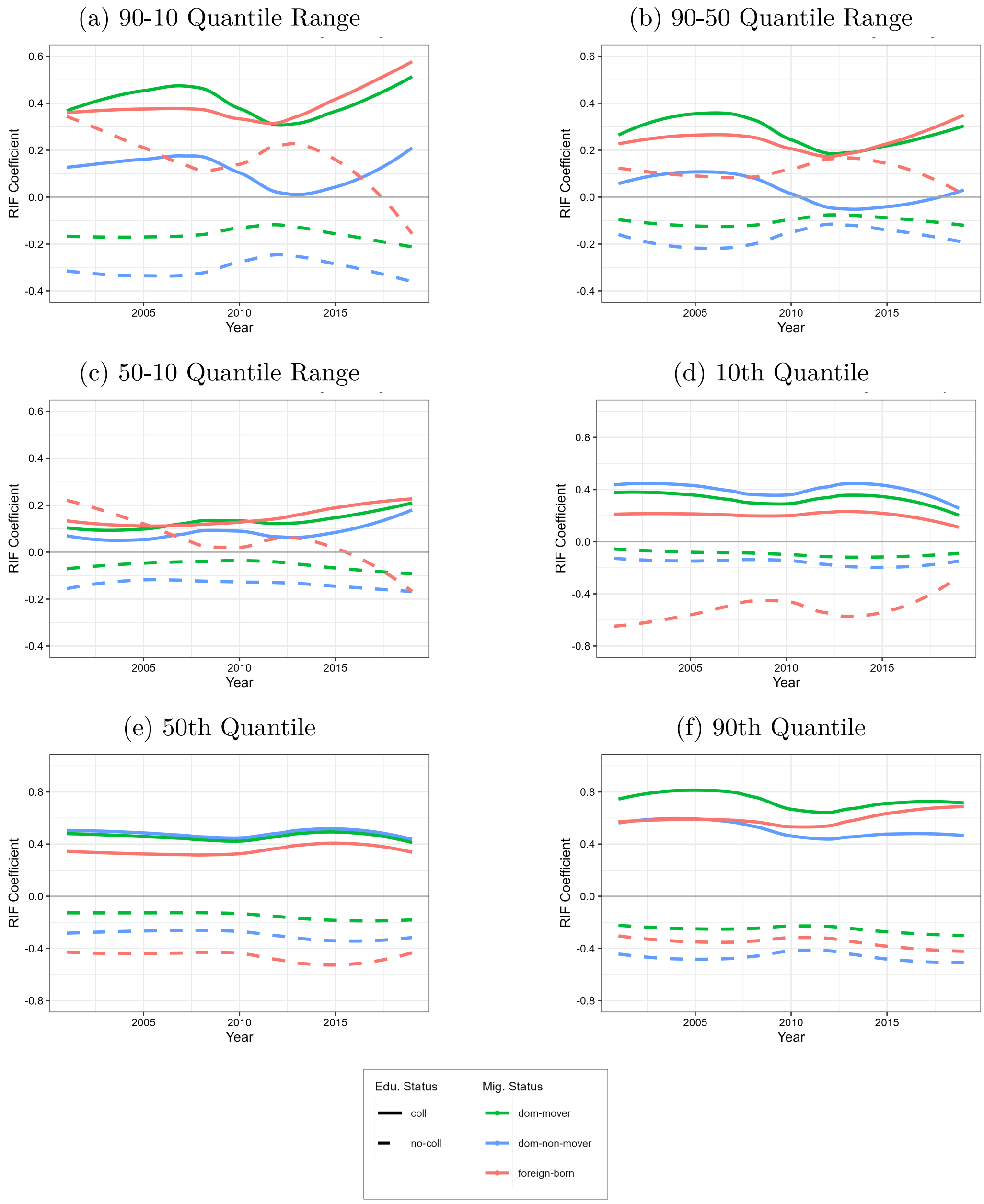
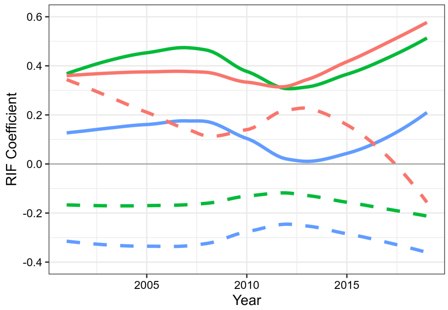


Figure 18: Marginal Effect for Each Migration-Education Group of Increasing Relative Proportion by One Percent on the Unconditional 90-10 Quantile Range and the 10th, 50th, and 90th Quantiles, as Calculated by RIF Regression

Note: Loess method used for smoothing

5.1 Compare RIF to Exact Across Years

(a) 90-10 Quantile Range, Firpo RIF



(b) 90-10 Quantile Range, Exact Formula

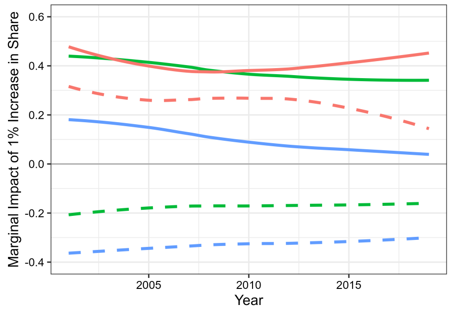


Figure 19: Marginal Effect for Each Migration-Education Group of Increasing Relative Proportion by One Percent on the Unconditional 90-10 Quantile Range and the 10th, 50th, and 90th Quantiles, as Calculated by RIF Regression

References

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics* 90, no. 2 (May 1, 2008): 300–323. https://doi.org/10.1162/rest.90.2.300.

Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *The Quarterly Journal of Economics* 118 (4): 1335–1374. https://doi.org/10.1162/003355303322552810.

Card, David. 2009. "Immigration and Inequality." *American Economic Review* 99, no. 2 (May): 1–21. https://doi.org/10.1257/aer.99.2.1.

Fernández, Raquel, and Richard Rogerson. 2004. "Sorting and Long-Run Inequality." *The Quarterly Journal of Economics* 116 (February 6, 2004): 1305–1341. https://doi.org/10.1162/003355301753265589.

Firpo, Sergio, Nicole Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica : journal of the Econometric Society* 77(3) (3): pp. 953–973. https://doi.org/10.3982/ECTA6822.

Hershbein, Brad, Melissa S. Kearney, and Luke W. Pardue. 2020. "College Attainment, Income Inequality, and Economic Security: A Simulation Exercise." *AEA Papers and Proceedings* 110 (May 10, 2020): 352–355. https://doi.org/10.1257/pandp.20201062.

Hibbs, Brian, and Gihoon Hong. 2015. "An examination of the effect of immigration on income inequality: A Gini index approach." *Economics Bulletin* 35 (1): 650–656.

Hoxby, Caroline, and Bridget Terry. 1999. *Explaining Rising Income and wage Inequality Among the College Educated.* NBER Working Paper 6873. National Bureau of Economic Research, Inc, January.

Lin, Ken-Hou, and Inbar Weiss. 2019. "Immigration and the Wage Distribution in the United States." *Demography* 56 (December): 2229–2252. https://doi.org/10.1007/s13524-019-00828-9.

Llull, Joan. 2017. "U.S. Wage-Inequality Trends and Recent Immigration." *The Review of Economic Studies* 85 (September 18, 2017): 1852–1896. https://doi.org/10.1093/restud/rdx053.

Xu, Ping, James Garand, and Ling Zhu. 2016. *Imported Inequality? Immigration and Income Inequality in the American States.*

ONLINE APPENDIX

A Framework for Decomposing Wage Dispersion with an Application to Foreign and Internal Migration in the US from 2001-2019

Bent E. Sørensen, Fan Wang, and Adelia Zytek

A Additional data details (online)

A.1 IPUMS Data

A.1.1 Sampling

The following is the approximate proportion of Americans sampled in each IPUMS ACS Survey from 2000-2020

2000: 1 in 750

2001: 1 in 232

2002: 1 in 261

2003: 1 in 236

2004: 1 in 239

2005-2020: 1 in 100

A.2 Macro Data

A.2.1 Federal Reserve Source Data

Data for CPI values (USACPICORMINMEI) in each year from 2000-2022 are downloaded using R function pdfetch_FRED, which links to the St. Louis Fed's FRED database.

A.3 Major Data Manipulations

A.3.1 Sub Sample of Interest

For our wage income analysis we limit our samples in each year to individuals who reported themselves as being employed and working at least 36 hours a week.

A.3.2 Variable Manipulations

All variable manipulations are described in detail in the the R code file and are summarised below:

Monetary Variables:

<u>Recode</u>: the IPUMS variables INCWAGE, INCTOT, INCINVST, and VALUEH are all recoded

so that values of 999999, 999998, 9999999, or 9999998 are replaced with NAs

Normalization: for each of the listed four variables, two new variables are created, one that has adjustments made for CPI (standardized to 2015 prices) and another which standardizes for both CPI and regional price parity (RPP). These variables are denoted with a "_Norm" and a "_rpp" respectively

Demographic Variables:

Bracketed Age Variables: The IPUMS variable AGE is used to create two new variables, AGEB and AGE4, which create eight and four age brackets respectively for simplification of later analysis

Region of Residence and Birth Variables: IPUMS variables BPL and STATEFIP, denoting state/country of birth and state of residence respectively, are first mapped to character variables and consolidated into six regions:

- **Northeast**: "Connecticut", "Maine", "Massachusetts", "New Hampshire", "Rhode Island", "Vermont", "New Jersey", "New York"
- **MidAtlantic**: "Pennsylvania", "Delaware", "Maryland", "Virginia", "West Virginia"
- **Midwest**: "Illinois", "Indiana", "Iowa", "Kansas", "Michigan", "Minnesota", "Missouri", "Nebraska", "North Dakota", "Ohio", "South Dakota", "Wisconsin"
- **South**: "Alabama", "Arkansas", "Florida", "Georgia", "Kentucky", "Louisiana", "Mississippi", "North Carolina", "Oklahoma", "South Carolina", "Tennessee", "Texas"
- **West**: "Alaska", "Arizona", "California", "Colorado", "Hawaii", "Idaho", "Montana", "Nevada", "New Mexico", "Oregon", "Utah", "Washington", "Wyoming"
- **Foreign**: Other, including Puerto Rico and other US territories

Recode of YRSUSA1: For those born in the United States, IPUMS variable YRSUSA1 is replaced with their AGE value (originally coded to 0)

Recode of METRO: IPUMS variable METRO originally had four categories, three of which denoted residence in different parts of a metropolitan region. Recoded to be binary between metropolitan and rural living

B Methods Appendix (online)

B.1 Variance Decomposition

We recall our overall variance decomposition equation:

$$\sigma^2 = \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot \sigma^2_{e,m} + \sum_{m \in M} \sum_{e \in E} w_{e,m} \cdot (\mu_{e,m} - \mu)^2 \ , \tag{B.1}$$

where $\omega_{e,m}$ is a group specific weight equal to the relative proportion of that group in the population, $\mu(\cdot)$ is the mean of the variable of interest, and σ^2 is the variance of the variable of interest. In figure 14a we show the unweighted within group variance which is equivalent to the values of $\sigma^2_{e,m}$ for each education-migration group. The weighted variances shown in figure 14b is equivalent to the value of $\omega_{e,m} \cdot \sigma^2_{e,m}$ for each education-migration group. The mean difference values shown in figure 14c are equivalent to the values of $\mu_{e,m} - \mu$ and the weighted across group variance values shown in figure 14d is equivalent to $\omega_{e,m} \cdot (\mu_{e,m} - \mu)^2$

B.2 Mean Effects

B.2.1 Derivation of Mean Effect Formula

In order to derive the value of the mean effect of a particular education-migration group $\{e, m\}$, first, denote, $Y^{e,m}_{i,t}$ to be the log wage of individual i in year t belonging to the education group e and migration group m. We are interested in the effect of group's mean wage on national inequality, independent of differences in the age distribution of a group. Therefore, we first run the following regression on age:

$$Y^{e,m}_{i,t} = \sum_{j=1}^{J} \beta^j_t \cdot G^j_{i,t} + \epsilon_{i,t} \ , \tag{B.2}$$

where $G^j_{i,t} = \mathbf{1}\{a_j \leq A_{i,t} < a_{j+1}\}$ is a dummy indicating if individual i observed in year t belongs to age group j—with inclusive lower age bound a_j and exclusive upper age bound a_{j+1}—among J exhaustive and equi-distance age bins for at individuals who are at least age 25 and at most age 65, and $\epsilon_{i,t}$ is the log-wage error term.

We then consider an individuals age adjusted log income in year t as:

$$Y^{age,e,m}_{i,t} = \epsilon_{i,t} + \mu(Y^{e,m}_{i,t}) \ , \tag{B.3}$$

or, in words, their age residual in addition to the national average wage in year t.

We then proceed by running a regression year by year of age adjusted log income on our six mutually exclusive migration-education dummy variables. This regression takes the form:

$$Y_{i,t}^{e,m,age} = \sum_{(e,m)\in M\times E} \gamma_t^{e,m} \cdot D_{i,t}^{e,m} + \delta_{i,t} \ , \tag{B.4}$$

where $D_{i,t}^{e,m} = \mathbf{1}\{M_{i,t} = m, E_{i,t} = e\}$ is an interaction dummy indicating if individual i observed in year t belongs to migration group m and education group e. By the nature of a linear regression without an intercept, for each migration-education groups, $\gamma_t^{e,m}$ will be the mean age adjusted log income for group $\{e, m\}$ in year t.

Next, we normalize the means of all individuals in the population in the following way:

$$Y_{i,t}^{e^*,m^*,mean} = Y_{i,t}^{e,m,age} - D_{i,t}^{e^*,m^*}(\gamma_t^{e^*,m^*} + \mu(Y_{i,t})) \ , \tag{B.5}$$

where $\mu(Y_{i,t})$ is the average age adjusted log income across all individuals in a particular year. Therefore, for individuals not belonging to group $\{e^*, m^*\}$, their mean adjusted log wage will simply be their age adjusted wage, whereas individuals in group $\{e^*, m^*\}$ will we given the national average log wage for their age group.

The mean impact of group $\{e^*, m^*\}$ is then calculated as:

$$F_{mean,90/10}^{e^*,m^*,t} = IPR_{90/10}(\{Y_{i,t}^{age,}\}) - IPR_{90/10}(\{Y_{i,t}^{e^*,m^*,mean}\}). \tag{B.6}$$

Comparing the magnitude of the *mean impact* across time and groups gives insight into the "between group inequality" mechanism in explaining shifts in inequality overtime. The magnitude of the *mean impact* is jointly determined by the magnitude of the log mean wage gap between all other population groups and joint group m and e as well as the size of joint group m and e in the overall population.