A survey of California's income inequality and migration

According to the 2021gini coefficient listing, California is the fourth most unequal state in the country and while literature exists on recent California-specific income dispersion and historic changes in variation in income across the country, limited studies explore California's income inequality across decades. In terms of population mobility, California has also experienced sustained net loss of domestic migrants starting 1989. While many studies have analyzed this exodus, this current study attempts to examine if California's income inequality and migration are related to each other. Heterogeneity amongst emigration groups is also explored to compare demographics of emigrants and their correlation to the changing income inequality.

Introduction

The study begins with observing and analyzing income inequality within California as a whole and then goes on to delve deeper into income inequalities across smaller geographies within California. It then looks at trends in immigration into and emigration out of California over the years. The next subsection uses a different dataset to classify emigrants into various demographic groups and track changes through the years. Finally, the last section employs a simple VAR model to estimate correlations between income inequality, immigration, and the various emigration groups to explore how they impact one another. Analyzing data from 2005 to 2019, worsening income inequality and total emigration are found to be associated with reduced immigration. Apart from emigration, in some demographic groups that show positive correlation amongst themselves, most other types of emigration are found to worsen income inequality, with lags of a year or two.

Section 1: Income Inequality

This section discusses income inequality in California at the state and MSA level applying an analysis that employs coefficient of income variation pre-tax, with and without transfers, to understand how income inequality has changed over the decades. Subsections briefly explore income inequalities within California, across MSAs and within the Los Angeles metro area using multiple inequality metrics.

Section 1.a.: State-level Income Inequality in California:

Coefficient of variation (CoV) of income in all of California is calculated using methodologies similar to the ones employed by Gaubert et al. (2021). CoV is a measure of dispersion implying that higher CoV reflects worsening income inequality.

Per capita personal income drawn from the Bureau of Economic Analysis (BEA). Pre-tax income equals wages, employer-provided benefits, proprietors' income, dividends, interest, and rent and excludes capital gains and thereby corporate retained earnings. Social Security includes Social Security Disability Insurance. Transfers include all major government transfers including Social Security and Medicaid. We use a time-series of this data from 1976 to 2019.

The county level per capita pre-tax income, A is calculated as the ratio of pre-tax income and population for the whole country. A is log-transformed to find log of per capita pre-tax income, B. Similarly, the county level per capita pre-tax income + transfers, C is calculated as the ratio of pre-tax income + transfer and population. C is log-transformed to arrive at the log of per capita pre-tax income + transfer, D.
Next, mean and standard deviations of all counties within California (B_{sd}', B_{mean}', D_{sd}', D_{mean}') are calculated to obtain singular standard deviation and mean values for each year. The coefficients of variation are then calculated as

$$\text{Coefficient of Variation, CoV of pre-tax income} = B_{sd}'/\ B_{mean}'$$
$$\text{Coefficient of Variation, CoV of pre-tax income + transfers} = D_{sd}'/\ D_{mean}'$$

1 sayantani.sayantani@anderson.ucla.edu

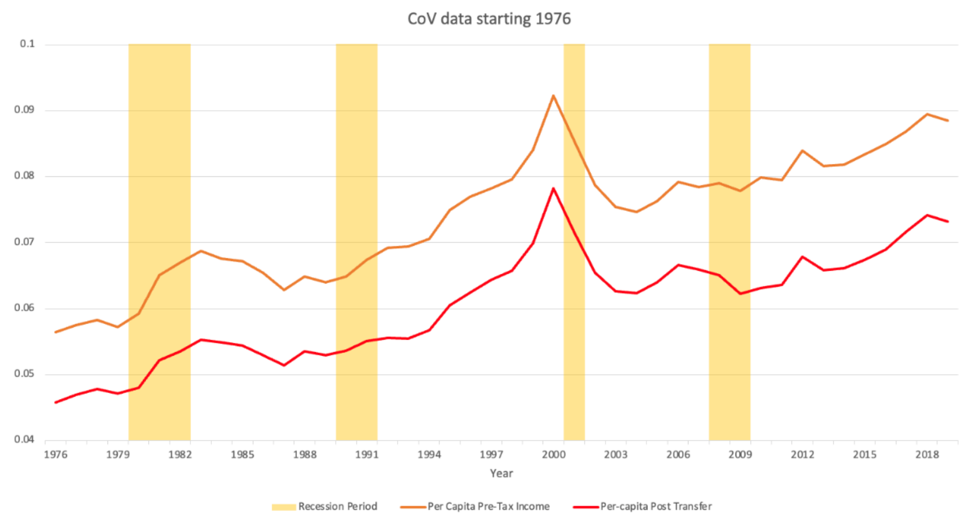

Fig 1: Coefficient of Variation of log (per-capita income)

The CoV graph shows inequality in pre-tax income (with transfers) increasing from 0.06 in 1976 to 0.09 in 2019. This is a substantial 56.6% increase. In other words, while the mean income increased over the years, the variation of income across different counties within California increased more leading to a substantial increase in the ratio of income variation to mean income.

Section 1.b.: Income Inequality within CA

MSA-level variation:

Estimating the mean and standard deviations of across all counties within each California metropolitan statistical area (MSA) allows us to track MSA-level variations in coefficient of income variation across California counties. Figures 2 and 3 show that the San Francisco-Oakland-Hayward MSA had the highest magnitude of CoV all along. However, San Jose-Sunnyvale-Santa Clara and Los Angeles-Long Beach- Anaheim show the greater worsening of income inequality San Francisco-Oakland-Hayward MSA. Riverside-San Bernardino-Ontario MSA shows gradual improvement in income inequality ending with the lowest CoV magnitude by 2019.

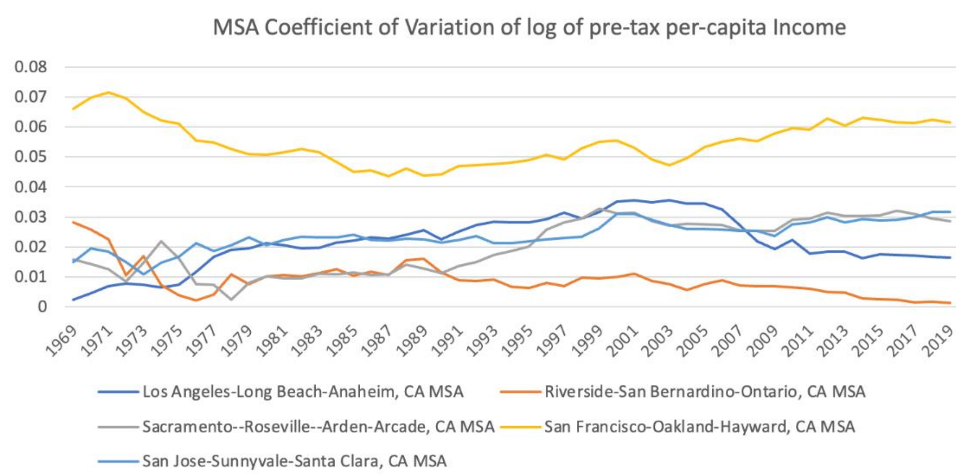

Fig 2: CoV of log (per capita pre-tax income)

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

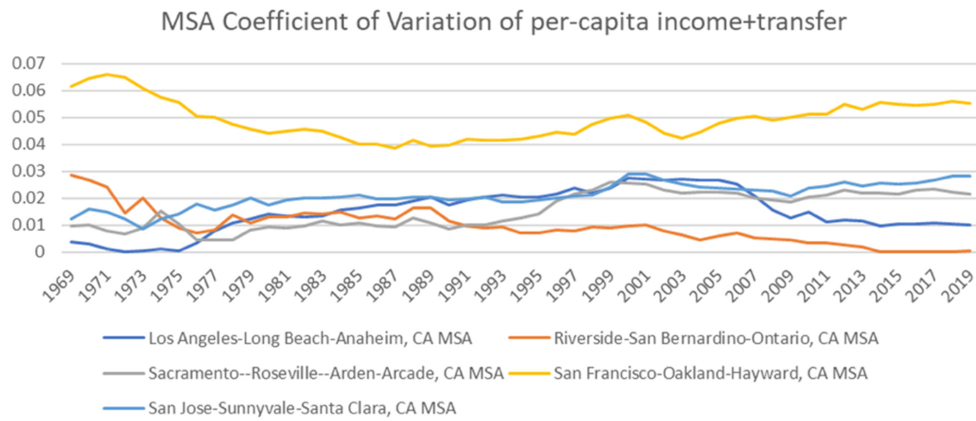

Fig 3: CoV of log (per capita pre-tax income + transfers)

Variation with LA Metro Area:

Using Census Bureau ACS 1-year panel data for metro and micro areas within California starting 2010 to 2021, we find trends in mean values of quintiles, upper limits of quintiles and household shares of quintiles. Mean value of lowest income quintile is mean income of the income quintile containing the lowest 20% of the income held by households in the area. Mean value of fourth income quintile is mean income of the income quintile containing 60% - 80% of the income held by households in the area.

Within the LA metro area, the mean value of the fourth income quartile seems to be increasing at a much higher rate than the mean of the lowest quintile. However, the year over year % change calculated as (($QuintileMeanIncome_{Year\_t}$ – $QuintileMeanIncome_{Year\_t-1}$)/ $QuintileMeanIncome_{Year\_t-1}$) show that the lowest quintile had a lower rate of increase in mean income the fourth quintile up until 2017, after which the mean income of the lowest quintile increased at a higher rate (till 2020). This suggests that the inequality rose till 2017 and then got subdued. The % change from base year 2010 is calculated as ($QuintileMeanIncome_{Year\_t}$ – $QuintileMeanIncome_{2010}$) / $QuintileMeanIncome_{2010}$, shows the fourth income quintile mean rising higher than the lowest income quintile mean. Thus, starting from a base of 2010, the income inequality worsened.

Upper limit of the lowest income quintile is the upper limit of the lowest 20% of the income held by households in the area. Upper limit of the fourth income quintile is the upper limit of the income quintile containing 60% - 80% of the income held by households in the area.

Within the LA metro area, the upper limits of income quintiles seem to be increasing at a steeper rate for the fourth quintile relative to the lowest quintile, suggesting increasing inequality. However, the year over year % change calculated as (($QuintileUpperLimit_{Year\_t}$– Quintile $UpperLimit_{Year\_t-1}$)/ Quintile $UpperLimit_{Year\_t-1}$) show that the lowest quintile kept up with the rate of increase (sometimes even exceeding) the fourth quintile rates of increase, up until 2019. Post 2019, both the rates increase dropped, with the lowest quintile drop being significantly more pronounced than the fourth quintile drop, suggesting more negative trends in income inequality. The % change from base year 2010 is calculated as ($QuintileUpperLimit_{Year\_t}$ – Quintile $UpperLimit_{2010}$) / Quintile $UpperLimit_{2010.}$ Similar to the income quintile means, the upper limits too show that starting from a base of 2010, the income inequality worsened.

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

Section 2: The story of California Migration

Section 2.a.: Migration to and from California

This section uses yearly state-to-state migration flows from 2005 to 2019 sourced from the American Community Survey (ACS) 1-Year Estimates. The data universe is population 1 year and over. Note that immigration and emigration all refer to movements within the USA. In other words, immigration henceforth would refer to domestic immigration.

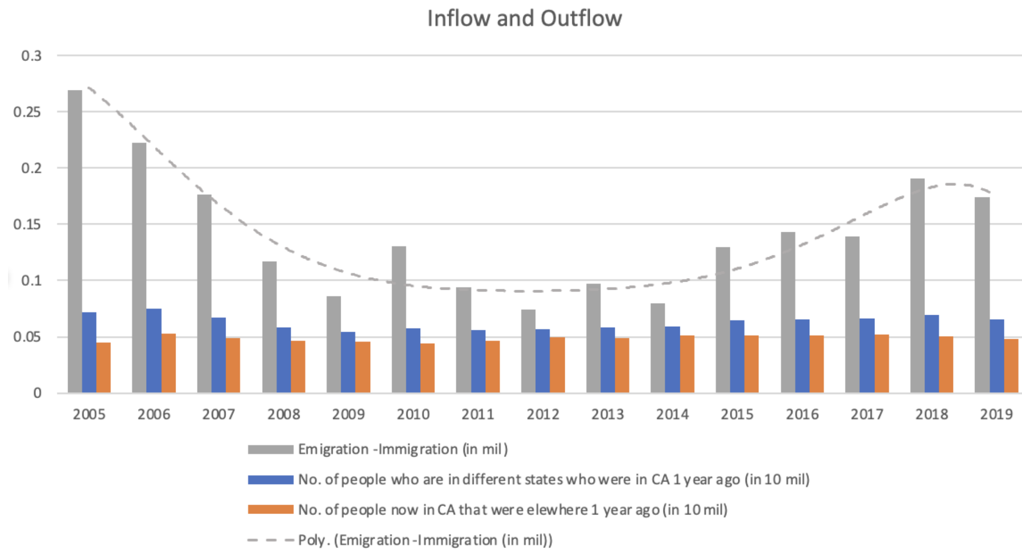

Fig 4: Immigration from and Emigration to CA w/ trendline

The emigration out of California is consistently higher than the immigration into CA. However, between 2006 and 2018, there's an increase in immigration. Also starting 2006 but ending by 2013, is a sustained decrease in emigration. Together, the difference between emigration and immigration observes a fall from 2005 to 2010, holds steady at around 90,000 and then increases till 2018, dipping slightly in 2019.

While there are many factors associated with the increase in California emigration (minus domestic immigration), the most influential may be discernible increases in California's price levels, particularly in housing rents, that are not necessarily matched by improved housing conditions. Housing is an important determinant of relocation-related decisions and the average housing conditions in California are arguably inferior relative to other US states as evidenced by County Health Rankings and Roadmaps published by the University of Wisconsin Population Health Institute in 2023. According to the release, using data between 2015-2019, California counties have an average of 26% households with at least 1 out of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities; relative to a national average of 13% (excluding California). Despite that, in 2023, according to the Bureau of Economic Analysis (BEA), California had the highest rates of Regional Price Parity in housing rents. The BEA releases a Regional Price Parity (RPP) measure that considers 170 personal consumption expenditure categories (including housing) at the county level, across states, to estimate the differences in price levels across states and metropolitan areas for a given year. These are expressed as a percentage of the overall national price level, and California is among the toppers in the list of high RPPs, with a trend that's mostly been increasing since 2010.

Particularly after 2014, rising housing prices are an important piece of the emigration puzzle, as evidenced by California's Housing Price Index. The Federal Housing Finance Agency's House Price Index (HPI) incorporates tens of millions of home sales and offers insights about house price fluctuations at the national, census division, state, metro area, county and ZIP code, levels. The HPI shows significant correlation (0.78) to the (emigration-immigration) measure between 2005 and 2019. Other research connects rising housing prices and income inequality. For instance, Albouy et. al. (2016) discuss how real income inequality has been exacerbated by rising relative rents, and how this affects the poor disproportionately. While there may be multiple factors endogenous to housing prices that simultaneously correlate to both income inequality and emigration, studying them at the same time requires lengthy

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

and intricate economic models. The current study abstracts from the complexity and attempts to isolate inequality and emigration in a simple and concise model.

Section 2. b.: Migration between states originating from California

For the following section, IPUMS USA person level data on demographics of California individuals who moved between states for the years 1970, 1980, 1990, 2000 and 2001-2019 has been used. Observations that had missing demographic information for one or more variables were dropped. For all the values pertaining to years 1970, 1980, 1990 and 2000, the data counts reflect everyone who migrated in the last 5 years. Starting 2001, the data counts reflect everyone who migrated in the last 1 year. The cumulative numbers for 1970, 1980, 1990 and 2000 are therefore higher. The following emigration graphs all show a drop in 1980 (emigrated in last 5 years), a low level of emigration from 2001-2004 and then a relatively higher level starting 2005.

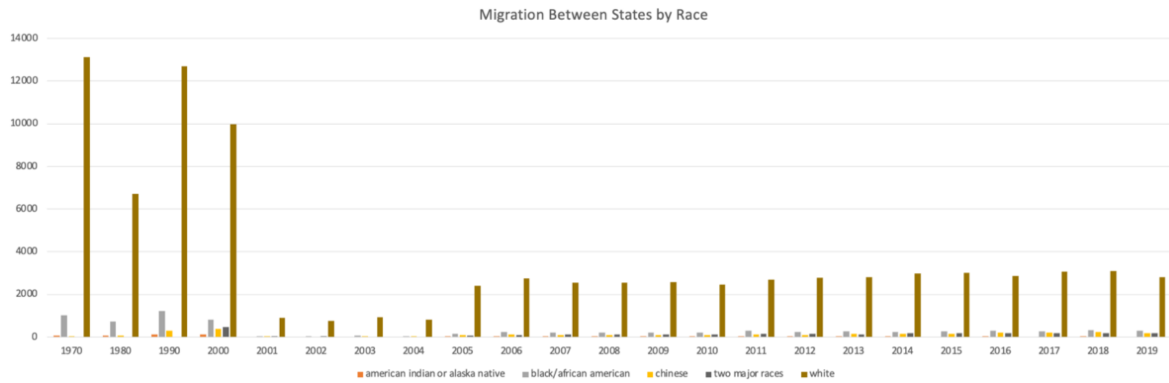

Fig 5: Emigration from CA by race

By race, the migration of white people away from California seems to be consistently higher than the migration by any other race, followed by the migration of Black people.

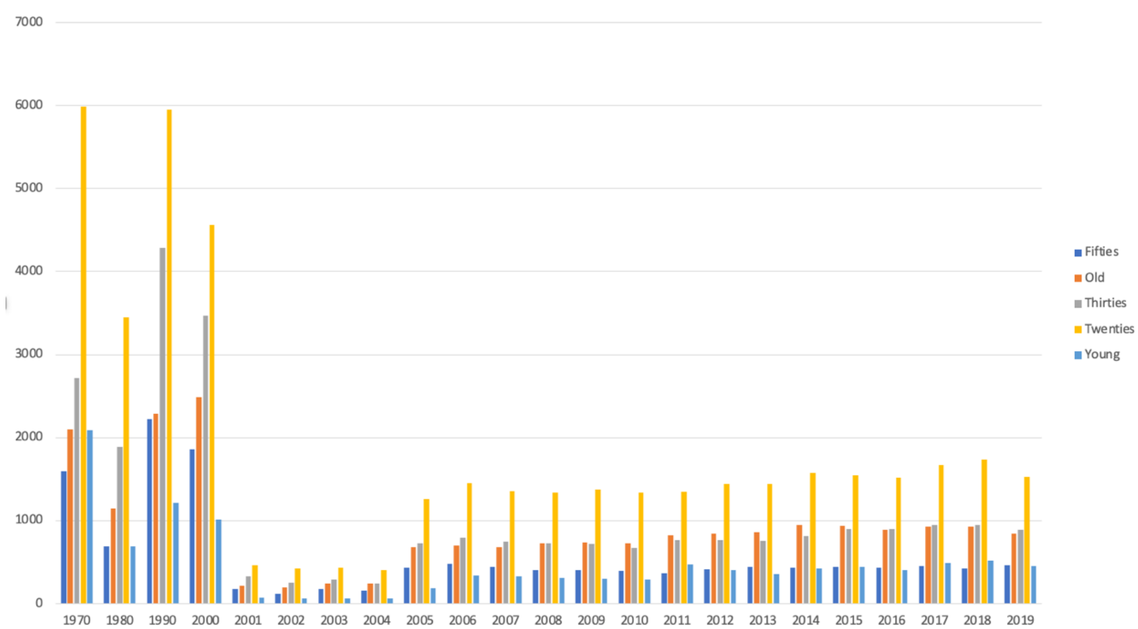

Fig 6: Emigration from CA by age

By age, the migration of people who are in their twenties seems to be consistently relative to other age groups, followed in decreasing order by migration of people in their thirties, people over 60 years of age and people in their fifties.

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

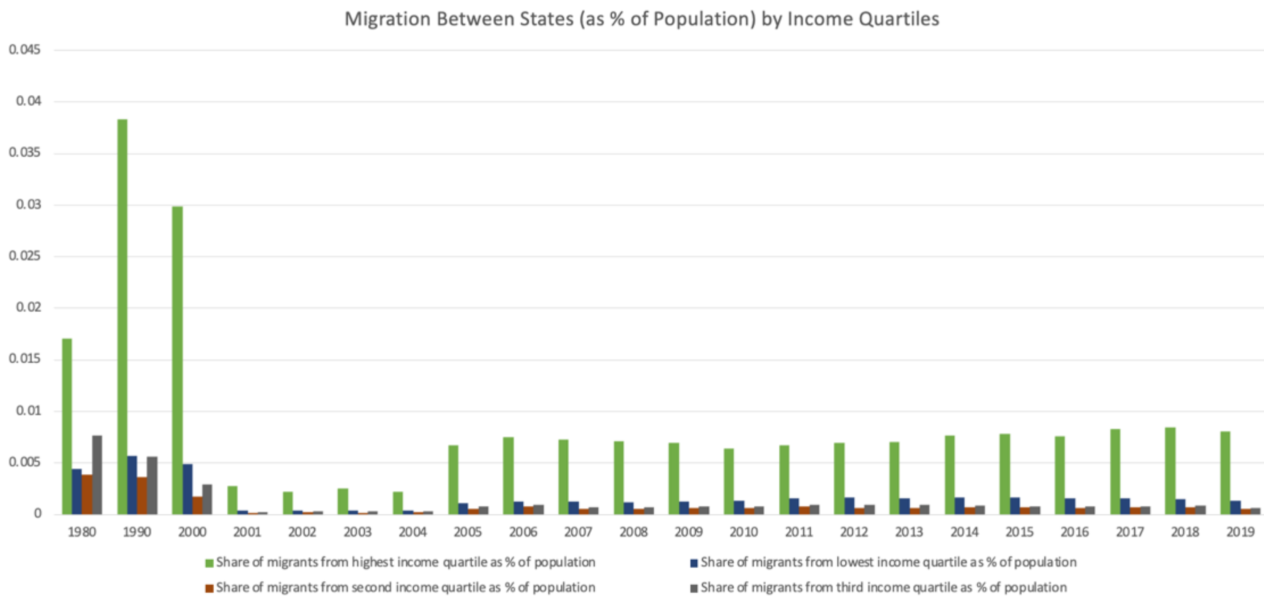

Fig 7: Emigration from CA by income quartile

By income quartile, the migration of people who earn the highest quartile of income outnumber the migration of people in other income quartiles.

Section 3: Income Inequality and Migration

This section attempts to identify if and how the changes in California's income inequality correlate with changes in migration of people to and from California. It is expected that income inequality may impact and may be impacted migration of people to and from California. It is also expected that it can take time for these variables to impact themselves and each other. Consequently, this section employs a vector auto-regression model to explore associations between income inequality and migration.

Vector autoregression (VAR) is a statistical model used to capture the relationship between multiple quantities as they change over time. VAR is a type of stochastic process model that generalizes the single-variable (univariate) autoregressive models by allowing for multivariate time series.

Assuming that migration demographics and income inequality are endogenous variables that affect one another up until lags of 2 years[1], we are estimating the following model:

$$Y_{k,t} = c + A_{1k,t-1}Y_{k,t-1} + A_{2k,t-2}\, Y_{k,t-2} + \Sigma_j(A_{1j,t-1}Y_{j,t-1} + A_{2j,t-2}\, Y_{j,t-2}) + \varepsilon_t. \qquad \forall\, j \neq k$$

Where j and k are the following variables:

- Income inequality (Coefficient of Variation of log per-capita pre-tax income)
- Emigration (Total count of people who left CA in the last 1 year)
- Immigration (Total count of people who came into CA in the last 1 year)
- Black/White/Asian Emigration: Total count of people who left CA of the mentioned race.
- Count of emigrants aged 20-30: Total count of people aged 20-30 who left CA.
- % emigrants earning the lowest quartile of income: emigrants earning the lowest quartile of income as a percentage of total California population
- % emigrants earning the highest quartile of income: emigrants earning the highest quartile of income as a percentage of total California population

The results are shown for each major (dependent) variable and those correlated variables which showed significant VAR results. The lag numbers show how long (in years) it took the correlated variable to affect the major variable

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

and the sign shows which direction the correlated variable impacted the major variable. Hence if variable Y_j positively affects variable Y_k with a lag of 2 years, it implies that variable Y_k in year *t* changes positively with positive changes in variable Y_j that happened 2 years before, in year *t-1*.

Note: Correlated variables show up only corresponding to the most significant out of the 2 lags considered. If directions vary across lags, for ease of understanding, only variables that have (relatively more) significant and higher degree (coefficient magnitude) of correlation with the major variable are shown. *, **, *** implies 10%, 5% and 1% significance level of the correlated variable.

Emigration, Immigration, and pre-tax income inequality VAR Results from ACS 1-year estimates:

Dependent Variable	**Independent Variable**		
	Income Inequality	***Emigration***	***Immigration***
Income Inequality	Lag 2: Positive**	-	Lag 2: Negative*
Emigration	-	Lag 2: Positive**	Lag 1: Positive***
Immigration	Lag 1: Negative*	Lag 2: Negative**	Lag 1: Positive*

From the VAR results it seems that immigration negatively impacts income inequality while it positively affects emigration, both with a lag of 2 years. Also, income inequality and emigration both negatively impact immigration with a lag of 1 year and 2 years, respectively. In other words, after a year of reduction of immigration, income inequality worsens, and emigration reduces too. However, worsening income inequality itself leads to lower immigration after a year. Among own variable correlations, we notice that income inequality and emigration move in the same direction as it did 2 years ago; while immigration moves in the same direction as it did 1 year ago.

Emigration demographics and pre-tax income inequality VAR Results from IPUMS USA estimates:

Dependent Variable	Independent Variable						
	Income Inequality	*Black Emigration*	*White Emigration*	*Asian Emigration*	*Emigrants aged 20-30*	*% Emigrants earning the lowest quartile of income*	*% Emigrants earning the highest quartile of income*
Income Inequality	Lag 1: Positive*	Lag 1: Positive*	Lag 2: Positive**	Lag 2: Positive***	Lag 2: Negative***	Lag 2: Negative*	-
Black Emigration	Lags 1 & 2: Negative***	Lag 2: Negative***	Lag 1: Negative***	Lag 2: Positive***	Lag 1: Positive***	Lags 1 & 2: Positive***	Lag 1: Positive***
White Emigration	Lags 1 & 2: Negative***	Lag 2: Negative***	Lag 1: Negative***	Lag 2: Positive***	Lag 1: Positive***	Lag 2: Positive***	Lag 2: Positive***
Asian Emigration	Lags 1 & 2: Negative***	Lag 2: Negative***	Lag 1: Negative***	Lag 2: Positive***	Lag 2: Positive***	Lag 2: Positive***	Lag 2: Negative***
Count of emigrants aged 20-30***	Lags 1 & 2: Negative***	Lag 1: Positive***	Lag 1: Negative***	Lag 2: Positive***	Lag 1: Positive***	Lag 2: Positive***	Lag 1: Negative***
% Emigrants earning the lowest quartile of income	Lags 1 & 2: Negative***	Lag 1: Positive***	Lag 1: Negative***	Lag 2: Positive***	Lag 1: Positive***	Lag 2: Positive***	Lag 2: Negative***
% Emigrants earning the highest quartile of income	Lags 1 & 2: Negative***	Lag 1: Positive***	Lag 1: Negative***	Lag 2: Positive***	Lag 1: Positive***	Lag 2: Positive**	Lag 2: Negative***

Note that in the above model, data points are only used from 200-2019, since the difference in emigration measurement in 1970, 1980 and 1990 could lead to biases in the estimation.

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

To summarize the results:

1. Determinants of income inequality:
   Income inequality seems to worsen after 2 years of increase in white and Asian emigration and after 2 years of decrease in emigrants aged 20-30 and emigrants earning the lowest quartile of income. Black emigration worsens income inequality with a lag of a year. Each year, income inequality moves in the same direction as it did a year back.
2. Determinants of emigration by race:
   a. Black, white and emigration are impacted similarly by most factors other than % emigrants earning the highest quartile of income.
   b. Increase in income inequality reduces Black, white and Asian emigration after both 1 and 2 years.
   c. While white emigration and Black emigration impacts themselves negatively after 1 and 2 years respectively; asian emigration impacts itself positively.
   d. While higher Black and white emigration is associated with higher emigration of people earning the highest quartile of income, higher Asian immigration is associated with lower emigration of people earning the highest quartile of income.
3. Determinants of emigration of people aged 20-30:
   a. Increase in income inequality reduces emigration of people aged 20-30 after both 1 and 2 years.
   b. Emigration of people aged 20-30 positively impacts itself after 1 year.
   c. Increased Black emigration, Asian emigration and emigration of people earning the lowest quartile of income seems to be associated with increased emigration of people aged 20-30.
   d. With a lag of a year, increased white emigration and emigration of people earning the lowest quartile of income seems to be associated with decreased emigration of people aged 20-30.
4. Determinants of emigration by income quartile:
   a. Increase in income inequality reduces emigration of people earning both the lowest and highest quartiles of income after 1 and 2 years.
   b. While the % value of emigrants earning the lowest quartile of income positively impacts itself after 2 years; with the same time lag, the % value of emigrants earning the highest quartile of income negatively impacts itself.
   c. % Emigrants earning the lowest quartile of income drops after a year of increase in white emigration and decrease in emigration of people aged 20 – 30 years. It also drops after 2 years of increase in Asian emigration and increase in % emigrants earning the highest quartile of income.
   d. % Emigrants earning the highest quartile of income drops after a year of increase in white emigration and decrease in emigration of people aged 20 – 30 years. It also drops after 2 years of increase in Asian emigration and decrease in % emigrants earning the lowest quartile of income.

Conclusion

Using a measure of dispersion, over the decades, income inequality is found to be increasing. While the worsening income inequality is found to be associated with decreased domestic immigration, its impact on total emigration seems to be ambiguous and insignificant. Emigration also seems to impact domestic immigration negatively. Apart from increased emigration of emigrants earning the lowest quartile of income and emigrants aged 20-30, all other types of emigration are found to worsen income inequality. Surprisingly, worsening income inequality reduces emigration in all emigrant demographic groups. Also, interestingly, the percentage of emigrants earning the lowest quartile of income leads to an increase in and is itself increased by the number emigrants aged 20-30.

Reference

Gaubert, Cecile, Patrick Kline, Damián Vergara, and Danny Yagan. 2021. "Trends in US Spatial Inequality: Concentrating Affluence and a Democratization of Poverty." AEA Papers and Proceedings, 111: 520-25.

Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. IPUMS USA: Version 14.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D010.V14.0

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.

Albouy, David and Ehrlich, Gabriel and Liu, Yingyi, Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis (November 2016). NBER Working Paper No. w22816, Available at SSRN: https://ssrn.com/abstract=2868914

[1] FHFA uses a fully transparent methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.