California Leavin'
An economic and political recomposition?

Ransi Clark

September 2026

Abstract

California is amid a wave of outmigration that began in 2014, with Texas and adjoining states of Arizona and Nevada being the highest intensity receiving states. Large movements of Californians, who are generally higher income and Democratic voting than the receiving states, have led to speculations that they are raising prices for locals and skewing the political makeup more Democratic. There are also concerns that this migration is worsening economic and political sorting within the receiving states. In this paper, I use zipcode-level income bracket breakdowns from the Internal Revenue Service and precinct-level voting information to describe the extent to which these speculations can be established empirically. For the intensity of Californian migration to a zipcode, the county-to-county IRS migration matrix is upsampled. For Texas, I find that a higher cumulative net flow of Californian migrants was associated with a zipcode getting more high income and more Democratic voting. But the reverse was true for Nevada, where a higher intensity of migrants was associated with a lowering of the high income share and a very small Republican tilt. Only swing precincts in Arizona demonstrated any change with a slight Republican tilt associated with higher migration. Though there are recomposition effects, they are heterogeneous between states and within states, pointing towards a sorting than a displacement.

1 Introduction

For much of American history, California has imported more domestic migrants than it has exported. Though several of these migratory waves, such as gold speculation in the mid-1800s and the Dust Bowl migration in the 1930s, have been for speculative ventures most that came stayed. Since the 1990s following end of the Cold War, the migration trend has reversed. California has seen a net outmigration to other states. Outmigrations also happen in waves (Uhler & Garosi, 2018). The period from 1990 to 1996 saw a spike in out-migration peaking at 300,000 net outmigrants. Another wave from 2002 to 2007, peaked in 250,000 net outmigrants. The most recent wave beginning about 2014, saw its highest exodus in 2020 at about 150,000 net migrants. Though the outmigration has slowed, this wave is still ongoing.

The outmigration is driven by a variety of factors including high housing costs and high income taxes, particularly, in comparison to states such as Nevada, Arizona, and Texas. The outmigration peak in 2020 is mostly explained by the transition to remote working and a spike in high income Californians moving out, which meant that Californians previously restricted to high tax and high cost of living areas in California were now free to move to low tax states and low cost of living areas (Johnson & McGhee, 2026). The recent wave of outmigration has renewed several fiscal and political concerns in California. Since California relies heavily on income taxes, particularly from those earning above $500,000 to fund the state's expenses, loss of high income taxpayers can put fiscal pressure.

The movements also concern state Republican political operatives because the outmigration partisanship skews Republican, making it harder for California Republicans to win at the state level (McGhee & Johnson, 2025). At a national level, Democratic operatives are concerned that the outmigration to largely Republican states and the loss of Electoral College votes will make it harder for Democrats to win federally (Walters, 2024).

Californians moving out do not always find warm welcomes, either. Locked in an unexpectedly tight race for Senate reelection, Senator Ted Cruz made his 2024 campaign about protecting Texas from Californian migrants voting for Californian policies saying, 'Keep Texas, Texas'. Anti-transplant sentiment in Arizona and Nevada is more economic than political. Californians with higher home equity are blamed for driving up rents, house prices, and other costs (Madriaga, 2020; Murashima, 2024). Incoming Californians can also be selectively moving into specific areas that reflect their political leanings and preferences for public amenities which increases economic and partisan sorting within their host states, raising concerns that Congressional districts could become less competitive leading to incumbency advantages, poorer representation, and furthering of partisan extremism.

Underlying these concerns of economic and political recomposition is a fear that California's higher income, more Democratic-voting migrants will displace current residents by raising prices and voting differently. But what can we empirically say about the extent of this recompostion?

I investigate two types of compositional changes: income compositional change and partisan compositional change, which are measurable proxies for economic and political change respectively. I track income composition using the share of taxpayers in an area (usually zipcode) whose income exceeds that of $100,000 as reported by the Internal Revenue Service

(IRS). And I track political sorting using the Republican vote share for President and House for years 2016-2024.

If California's migration wave is disproportionately high-income and Democratic and has recomposition effects, we would see that zipcodes that have a higher intensity of net migration would demonstrate an increase in the high income share of zipcode and a fall in the Republican vote share in the zipcode's precincts. Null effects on recomposition would require further investigation. One explanation for no recomposition could be sorting in the settling of Californian migrants, where migrants settle in areas that economic and political characteristics that re similar to them. But heterogeneity in recomposition in different areas can also average to a null effect if there are underlying parallel trend violations.

The intensity of net migration from California is measured as both a level and rate. The first is the cumulative number of migrants coming in to a zipcode from California net of the number leaving to California. The second is this same number as a percentage of the nonmigrants in the zipcode. While the two measurements are related, they are not perfectly correlated. The raw inflows and outflows and their incomes are only available as a county-to-county matrix. Since many counties are large and highly politically and economically heterogeneous, I use an upsampling algorithm to create distributions of each of these measurements at the zipcode level. The upsampler uses the zipcode's distribution of income and the income of the inflow and outflow population to generate draws for the distribution. The distribution is then summarized to quantiles.

California is a net loser of high income taxpayers which peaks at a loss of about 50,000 in 2020, and Texas, Arizona, and Nevada are gainers of high income taxpayers. Those moving into Texas, Arizona, and Nevada were more likely to be high income taxpayers than those moving out of those states, while those moving out of California were marginally more likely to be high income taxpayers than those moving in to California, though this trend reversed in 2022.

However, this wave of higher income migrants did not necessarily make zipcodes more high income, in all states. In Texas, a higher inflow of Californians migrants made the zipcode's high income share slightly higher, but in Nevada this inflow made the high income share lower. There was no effect in Arizona. The evidence on political sorting is similarly heterogeneous across states. In Texas, for both Presidential and House elections, a high cumulative flow of migrants is associated with a vote swing towards Democrats. In Nevada, more cumulative migrants increased Republican vote share very slightly. In Arizona, no effect was observed.

These overall recomposition effects (or lack thereof) may be confounded, because of differences in trends of income and partisanship within the states. When decomposing based on baseline compositions of each of the outcomes, however, both Texas and Nevada are robust although the sizes of effects vary within this decomposed samples.

Recent studies tracking voter record files show that Californians moving tend to be more Republican-leaning than the population of voters they leave behind (Gimpel & Shaw, 2024; McGhee & Johnson, 2025). But actual voting behavior can differ from registration data for a variety of reasons. For one, a sizable number of Californians moving in do not record a partisanship. In the sample of Gimpel and Shaw (2024), about 26 percent of the movers from California were unaffiliated to one of the major parties. For another, voters may distinguish between state parties. A Californian Republican might align closer to a Texas Democratic party candidate than a Texas Republican party candidate. Precinct-level partisan measures indicate that despite a more balanced composition of movers, a high concentration of such movers are associated with a Democratic swing. However, this same observation could also be caused by a concomitant change in native migration or generational turnover, rather than Californian migration. Teasing these forces out remains a difficult endeavor, particularly in Texas which does not record party membership in their voter files.

The implications of 'voting with feet' have long been a subject of academic interest (Tiebout, 1956). Tiebout sorting predicts that residents will migrate to jurisdictions that provide their preferred set of services. Some find that this type of amenity and residential sorting can inadvertently create partisan sorting (Bishop & Cushing, 2009; Gimpel et al., 2026; Martin & Webster, 2020). But others find limits to how much domestic migration results in partisan sorting, citing that amenity and affordability concerns are stronger than partisan concerns and there are too few residential choices to create strong partisan sorting effects (Abrams & Fiorina, 2012; Mummolo & Nall, 2017). This study considers a related but different implication of this voting with feet phenomenon, that is, whether those moving cause a economic and political recomposition.

The rest of the paper will expand on the empirical details starting with data used and then describe the upsampling algorithm used to create zipcode-level distributions. The results section will then describe the high-income bias of migrations out of California, and report quantile-level recomposition effects using the upsampled zipcode level data.

2 Data

I draw on three sources of data for the analysis. The best source of migration and geographic income data in the United States comes from the Internal Revenue Service (IRS). The precinct level data for Texas, Arizona, and Nevada were initially downloaded from The Redistricting Hub. The correspondence files between zipcode and county are from the US

Department of Housing and Urban Development (HUD).

IRS income and migration information comes with about a four year lag, so the latest information I observe is for the tax year 2022[1]. The data series begins in 2012 [2]. The HUD correspondence files are issued every quarter. The precinct files are available for every House and Presidential election from 2016 to 2024.

Because I analyze information at various geographic levels, I will use the following subscript convention to make these easily traceable in the text. A subscript of s refers to the state, c refers to the county, z to the zipcode, and p to the election precinct. A subscript of t refers to the year, an a subscript r refers to a congressional district race. Additionally, a subscript of b refers to income brackets.

2.1 Zipcode and county income data

At both the zipcode and county level, the IRS reports that year's number of taxpayers and their adjusted gross income (AGI) in 6 brackets. These brackets are observed consistently from 2012 and the IRS does not correct them for inflation[3]. The brackets are as follows[4]:

- 1: from 0 to under 25,000
- 2: 25,000 to 50,000
- 3: 50,000 to 75,000
- 4: 75,000 to 100,000
- 5: 100,000 to 200,000
- 6: 200,000 and over.

For the years 2012 to 2022, I observe the following information,

- X_{zbt}: number of taxpayers in zipcode z in bracket b in year t[5].

[1] The latest IRS migration data is for taxes processed during 2023 and generally corresponding to activity in 2022, and was released on March 16 of 2026.

[2] Although IRS reports this income before the 2012 tax year, in 2011 the reporting period was changed from a October to September to a December to January, so any information at the transition of 2011 to 2012 is not comparable.

[3] A \$100,000 in 2013 has much different purchasing power in 2022, but the reporting uses the nominal value.

[4] County-level data is reported in the same brackets, but bracket 1 at the county-level is more granular and has three categories. These three were coalesced to category 1 here. All other categories match the county data.

[5] In all cases, when I refer to the number of taxpayers, I refer to the number of joint returns.

- G_{zbt} : aggregate income of zipcode z in bracket b in year t.
- X_{zt} : number of taxpayers in zipcode z in year t across all income brackets.
- G_{zt} : aggregate income in zipcode z in year t across all income brackets.
- X_{cbt}: number of taxpayers in county c in bracket b in year t.
- G_{cbt}: aggregate income in county c in bracket b in year t.
- X_{ct}: number of taxpayers in county c in year t across all income brackets.
- G_{ct}: aggregate income in county c in year t across all income brackets.

The bracket I refer to as high income is the combination of the brackets 5 and 6. When I specifically refer to this combined bracket, I will use the superscript H. For example, X^H_{zt} refers to the number of taxpayers in the zipcode, in that year that are high income. I compute the high income share of the zipcode, which is the main variable used for income composition tracking, as 100 times the ratio of X^H_{zt} on X_{zt}.

Counties are generally stable geographic units, but zipcodes are subject to change yearly. Most residential zipcodes undergo no change, but because this empirical exercise requires comparison over a decade, I need to ensure that the geographical units that I am comparing information over is very similar. Useful for this purpose is information on how the zipcodes map to counties. The US Department of Housing and Urban Development (HUD) provides quarterly crosswalks for zipcode-county ratio of residential units and county-zipcode ratio of residential units. These are not necessarily what the IRS would have used to allocate to zipcodes, but is a good enough approximation once some quality checks are done. Using this information I do three data quality and consistency checks to ensure stable units. First check is to ensure that zipcode boundary changes are minimal, the second is to remove zipcodes that belong to more than one county, and the third is to ensure counties are not missing too much zip information. The details of the checks are in Appendix A. This can eliminate a number of zipcodes with the highest being in Texas whose counties are small than the other four states, so opportunity for zipcode crossover through counties is high. But this only affected about 10 percent of mostly rural zipcodes.

2.2 Taxpayer migration data

The most granular migration data is reported at the county level. The IRS also reports a state-to-state migration matrix. Each of these matrices contain both inflow and outflow information (number of taxpayers and income). However, the migration matrices are not decomposed to income brackets. Moreover, IRS migration data is reported for matched returns only, that is, returns that were observed in two consecutive years. The IRS also includes the number and income of the taxpayers that did not move between the two years. I observe the following information from the IRS matrices,

- I_{st} : number of taxpayers moving into the state s from a different state in year t.
- G^I_{st} : aggregate current income of taxpayers in year t moving into state s from a different state in year t.
- O_{st} : number of taxpayers moving out of the state s in year t.
- G^O_{st} : aggregate current income of taxpayers in year t moving out of state s in year t.
- I_{ct} : number of taxpayers moving into the county c from a different county in year t.
- G^I_{ct} : aggregate current income of taxpayers in year t moving into county c from a different county in year t.
- O_{ct} : number of taxpayers moving out of the county c in year t.
- G^O_{ct} : aggregate current income of taxpayers in year t moving out of county c in year t.
- S_{ct} : number of taxpayers in county c in years $t-1$ and t that did not migrate (non-migrants).
- G^S_{ct} : aggregate current (year t) income of non-migrant taxpayers in county c at year $t-1$ and $t-1$.

Using this data, I can also calculate the number of migrants from a specific state that moved into a specific county in a different state, which I denote I_{sct}. Also I can denote the number that moved from a county back to a different state as O_{sct}. From these, I can calculate the movement from one state into a county as a share of all movement into that county. I denote this r^I_{sct}, which is computed as the ratio I_{sct} on I_{ct}. Similarly r^O_{sct} is the share of movement back to state s from county c and is computed as the ratio O_{sct} on O_{ct}[6].

State-to-state inflow and outflow files can be used to derive a state's inflow-outflow ratio and net inflow, which encodes information of the attractiveness of the state to taxpayers. These are notated as,

[6]The county-to-county migration matrix masks information whenever the county-to-county movement is below 20 migrations. Because of this the ratios r^I_{sct} and r^O_{sct} are somewhat underestimated. Since Californian counties are generally larger than Texas counties, the underestimation on r^I_{sct} is smaller than the underestimation on r^O_{sct}, where s in California

- E_{st} : ratio of number of taxpayers moving into state s on number of taxpayers moving out of state s at time t. This is derived from I_{st}/O_{st}.
- E_{ct} : ratio of number of taxpayers moving into county c on number of taxpayers moving out of county c at time t. This is derived from I_{ct}/O_{ct}.
- M_{st} : number of taxpayers moving into state s net of number of taxpayers moving out of state s at time t. This is derived from $I_{st} - O_{st}$.
- M_{ct} : number of taxpayers moving into county c net of number of taxpayers moving out of county c at time t. This is derived from $I_{ct} - O_{ct}$.

A inflow-outflow ratio above 1 shows that more taxpayers enter the state than leave and higher the ratio more attractive the state is for inbound movement. The top 10 states with this ratio consistently above 1 are South Carolina, North Carolina, Tennessee, Delaware, Texas, Idaho, Florida, Montana, Arizona, and Nevada. The bottom 10 states with the ratio consistently low are California, New York, Louisiana, Illinois, Massachusetts, New Jersey, Maryland, Michigan, Connecticut, and Mississippi.

Appendix Figure A.1 demonstrates the yearly ratio from 2012 to 2022. Most of the high net outbound states see their largest drops in 2020 which then recover somewhat to their 2019 ratios $E_{s,2019}$. A converse pattern holds for most high net inbound states. There is notable yearly heterogeneity in the high net outbound states too. Illinois and New York were both losing population since 2012 with ratios as low as 0.7 in 2013. California's ratios began at about 0.95 in 2012, but soon declined rapidly to below 0.7 in 2019 and as low as 0.6 in one of the largest declines in the country.

Figure 1 shows this fall in raw number terms with the thick grey line. The net outflow exacerbates from about 10,000 taxpayers in 2012 to 50,000 in 2019. Outflows have recovered from 2020's dip of 150,000 but only to the 2019 levels. The same figure shows which states are receiving this outgoing population of taxpayers. Texas is the largest receiver, followed by Nevada, and Arizona. All three of these states have received over 10,000 net migrants from California since 2019. Other western states like Oregon, Washington, and Idaho also record an influx of Californians, but in smaller numbers.

Within California, Los Angeles county has seen the largest outflow of net taxpayers, losing over 60,000 in net taxpayers in 2020[7]. Los Angeles is followed by Orange and Santa Clara counties both losing about 10,000 net taxpayers since 2018. Other county losses are comparatively small.

[7] This loss of taxpayers is not necessarily out of California

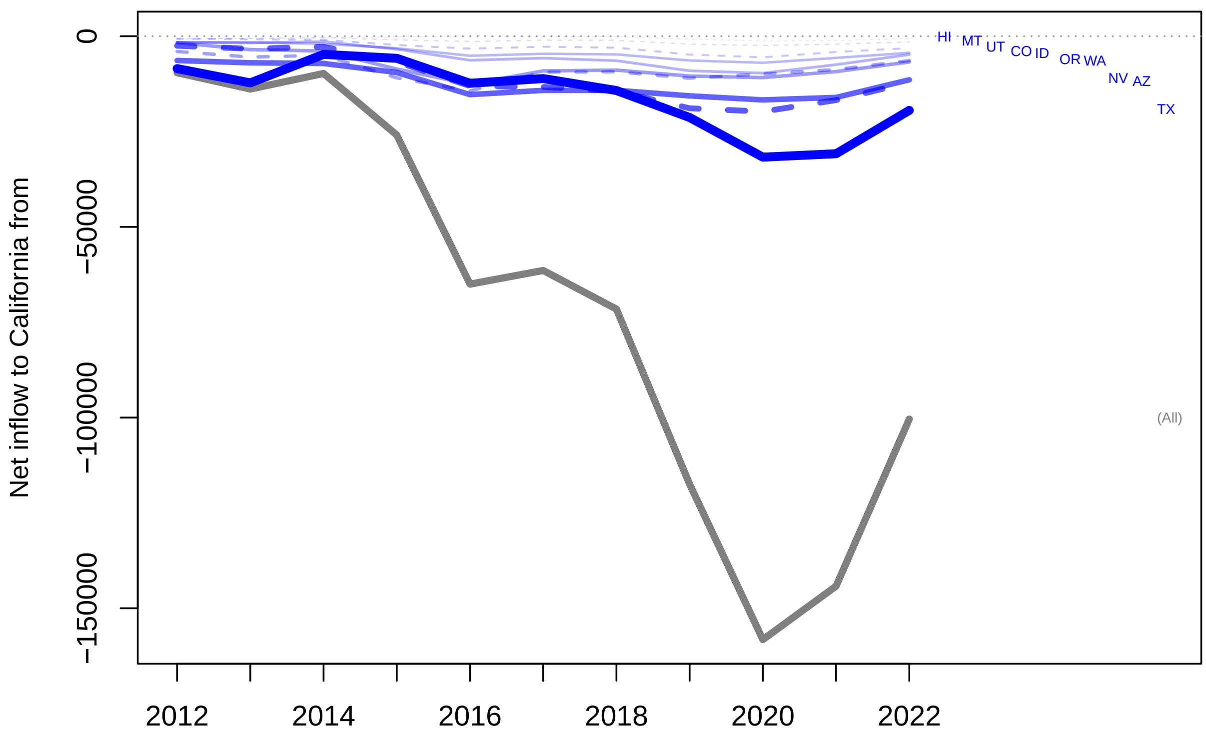


Figure 1: Top 10 outbound states for Californians

The Appendix Figure A.2 plots the dynamic of the inflow-outflow ratio for the top 15 inbound and outbound Californian counties. Of the 10 counties generally considered as part of Southern California, 6 are in this top 15 outbound counties. Only one Southern Californian county, Riverside county, is in the top 15 inbound counties. Three of the Bay area counties — San Mateo, Monterey, and Santa Clara — are also in the top 15 outbound counties. And only one Bay area county, San Benito, is in the top 15 inbound. The cluster of ten counties including and surrounding Sacramento county are all in the top 15 inbound counties. The net inflows to these counties are comparatively smaller than the net outflows from the Southern Californian counties, so any gains in the Sacramento cluster are offset by the much larger losses out of Southern California.

Figure 3 shows the net inflow of cumulative migrants for each county from the period 2013 to 2022 as a percentage of the non-migrants remaining in the counties in 2022. Notationally this is calculated as $100 \times \sum_{t=2013}^{2022} M_{ct}/S_{c,2022}$. The counties are ordered by their non-migrant population. Most counties were net gainers of population. However, several counties (Lander, White Pine, Elko, and Humboldt) also lost population. The residents of these counties did

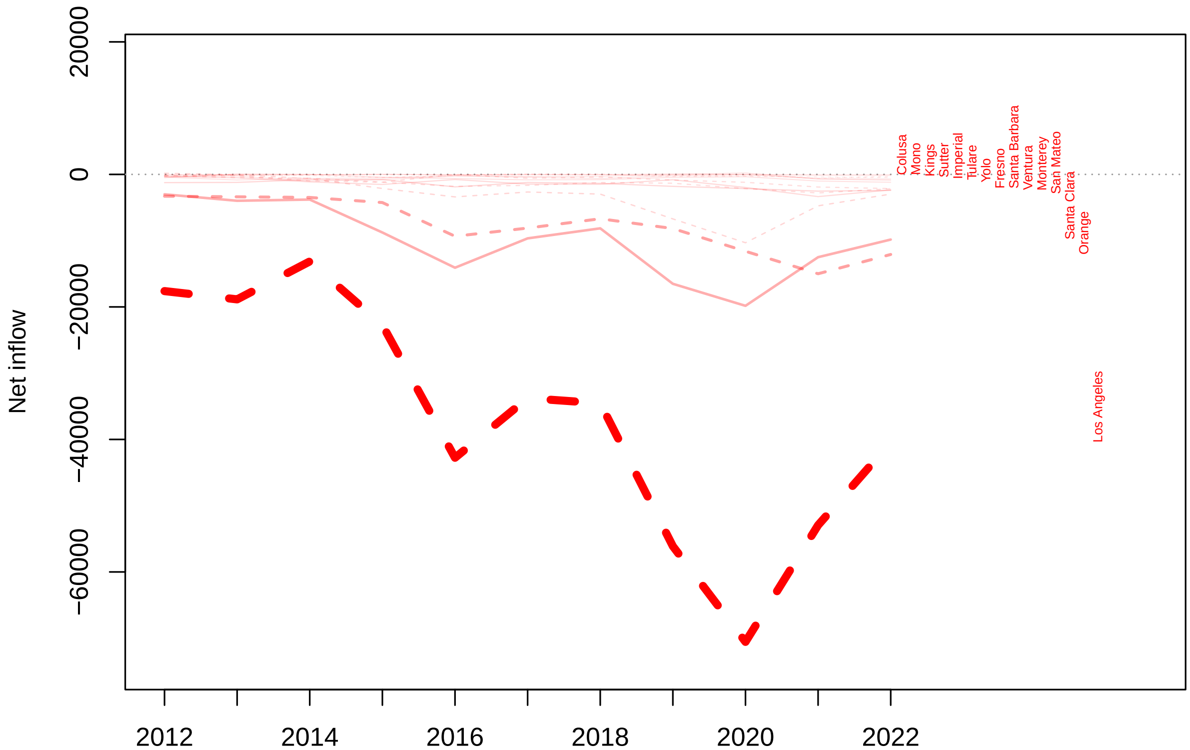


Figure 2: Top 15 California counties losing population

not necessarily leave to other states. They may have instead migrated to other counties within Nevada.

This figure also demonstrates the distinction between the level of internal migration and rate of internal migration (in relation to the counties population). While Clark county (county where Las Vegas sits) has the largest net inflow of migrants, its rate of migration was only just over a third of the neighboring Nye county. These different indicators of migration indicate different effects on the host county. A low volume but a high rate can have greater impact on the housing markets because smaller markets are more likely to be supply constrained.

Figure 4 distinguishes between the net inflow of all migrants to and from counties with the net inflow of California migrants. This value is calculated as 100 times the sum of of $I_{CA,c,t} - O_{CA,c,t}$ from 2013 to 2022 divided by the counties non-migrants in $S_{c,2022}$. The denominator of this percentage and that of the previous Figure 3 is the same. But they need not be the same sign. Elko county, for example, had a net outflow of migrants but has a positive inflow of Californian migrants because while more Californians came into Elko than left in that period, there were more leaving Elko in general than coming in. This is also true of all other counties that had negative cumulative inflows from the county. Even if the sign

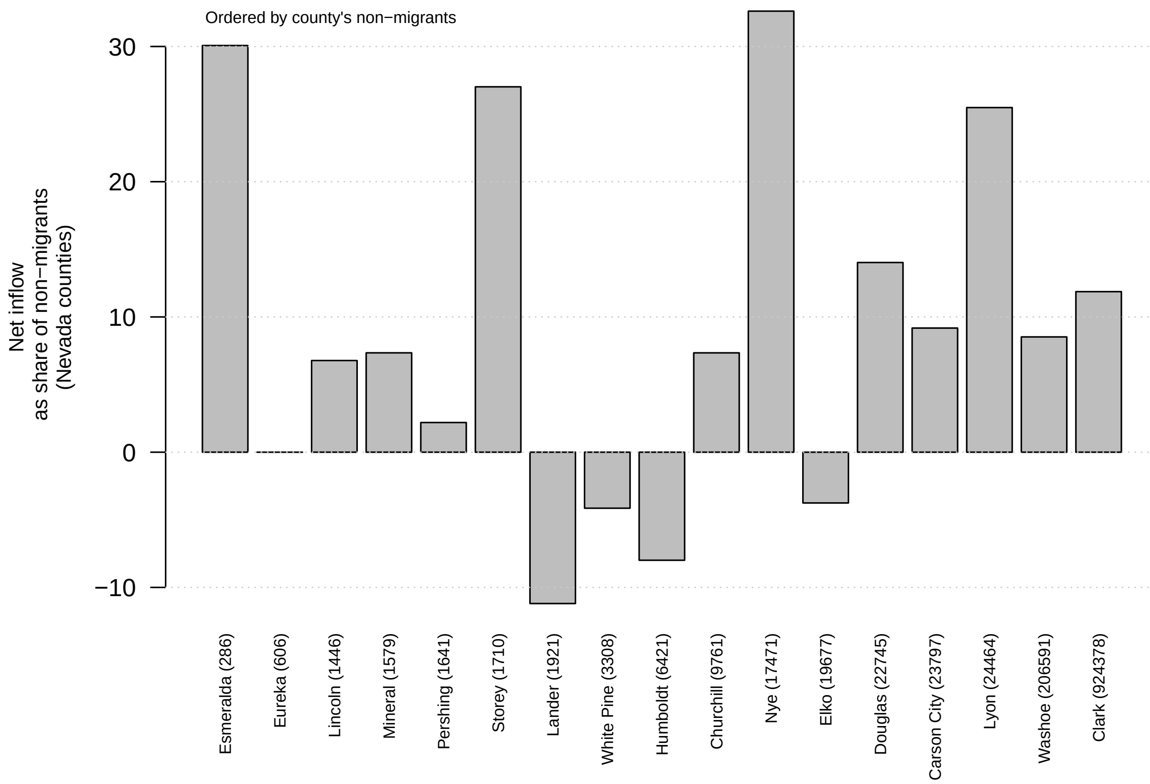


Figure 3: Nevada's county-level cumulative net inflow as a percentage of non-migrants in 2022

doesn't change counties can have a higher rate of Californian net migration than the overall rate of migration. These county differences provide us the variation necessary to distinguish between migrations and other concomitant changes in income and partisanship happening within the state unrelated to migration.

The corresponding figures for Arizona and Texas are in the Appendix. Appendix figure A.3 and figure A.4 show the net inflow and the net inflow related to California, respectively. As with Nevada, several Arizona counties had a net outflow (negative percentage), but all counties had a positive net inflow percentage of Californians. Maricopa county had the highest inflow of migrants, but both the overall rate and California-specific rate were below 10 percent. The much smaller Mohave county, adjoining California, has a much higher rate of about 25 percent of both percentages.

The cumulative net inflow percentage for Texas is in Figure A.5. And the cumulative net California inflow numbers for Texas is in Figure A.6. Because Texas has a very large number of counties, many of which are sparsely populated, I limit the figures to show the

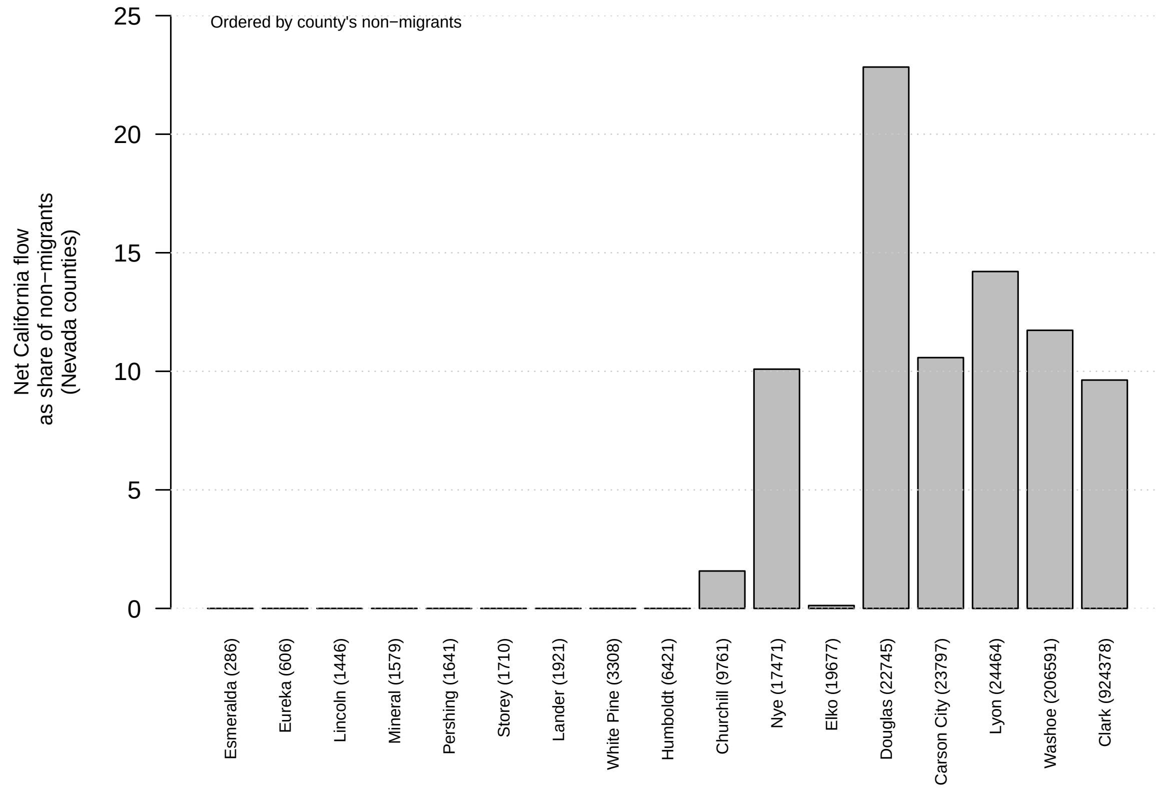


Figure 4: Nevada's county-level cumulative net inflow from California as a percentage of non-migrants in 2022

counties with the highest non-migrant population. The smallest county of the reported is Brown county with 13,400 non-migrants. Almost all smaller counties see very little migratory flows, so there rates are zero or very close to zero. Because of the larger number of counties, there is also greater heterogeneity in migration percentages in Texas. Comal, Williamson, Montgomery counties has a very high net inflow rate of domestic migrants, while the largest counties of Harris and Dallas have lost migration population. But Harris and Dallas counties were still net receivers of Californians.

Texas was the highest receiver of Californians in raw number terms, but as a percentage of its non-migrant population Californians formed a smaller percentage. For example, Harris county received 17000 Californians, but this was less than 1 percent of the counties total population.

2.3 Precinct voting data

As with zipcodes, precincts can change slightly over time. Some changes can be quite drastic, particularly after redistricting. When analyzing precinct data as a panel, precinct comparisons are best made between precincts of similar geography. To enforce a panel structure, I first construct a stable geography of precincts, starting with precincts that were the same between 2016 and 2024. Then for the remaining geography, I create a correspondence between precincts that had a geographic overlap of 85 percent. Much else of the geography remaining even after that comprises of larger precincts that were split in 2024. In a few occasions, smaller precincts in 2016 were combined to larger precincts in 2024. Correspondences are built between these. To this stable geography, precinct level Presidential and House results are added. Some precincts go through drastic changes and these needed to be dropped. This was most common in Nevada, particularly areas in the northern reaches of Las Vegas, where new master-planned communities require additional geography for vote-tallying. But some rural Nevada precincts also were removed due to large geographic changes.

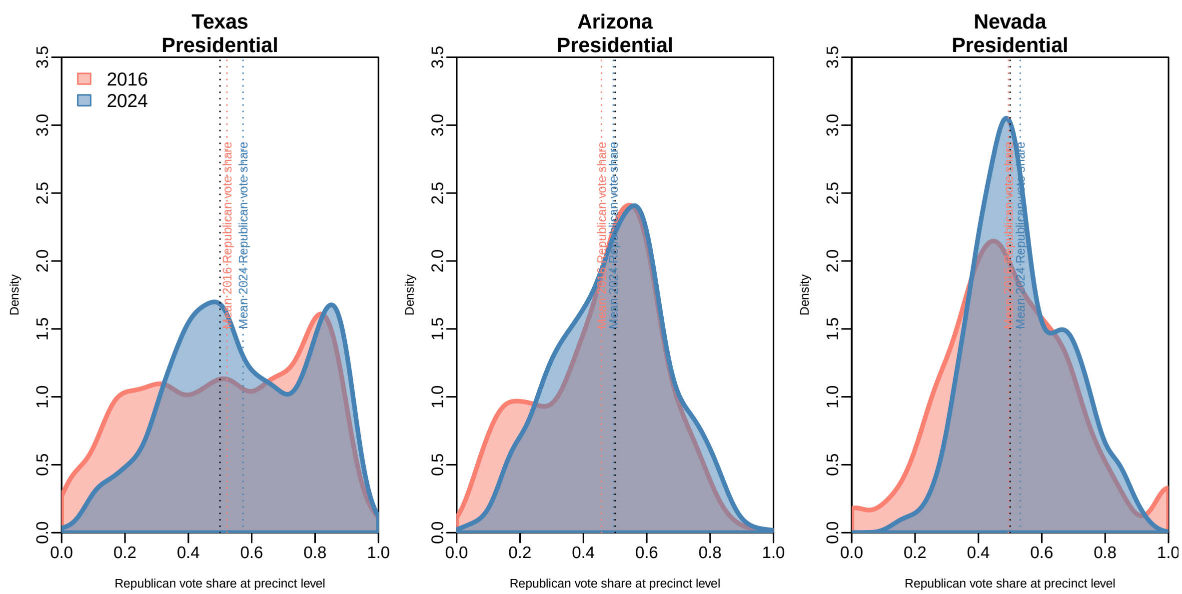


Figure 5: A density plot of precinct-level Presidential Republican vote share in 2016 and 2024

Figure 5 shows the precinct level distribution of Republican vote share in 2016 and 2024. The unweighted mean of the precincts of the years are marked in dotted horizontal lines[8]. In all states, there is a Republican gain between 2016 and 2024 Presidential elections[9]. In Texas,

[8]These are means of precinct's Republican vote shares and not the state's total vote share received by the Republican candidate.

[9]In 2016, the Democratic candidate Clinton won Nevada but lost Arizona.

precinct level vote shares have started to look increasingly bimodal, possibly indicating a spatial sorting pattern among its precincts.

In the Appendix, Figure A.7 demonstrates the corresponding information for House elections. For House elections, the Republican shift is more muted, but the bimodality in Texas is still observable. The number of House districts that were not contested by Democrats have also increased. In Arizona, the number of districts not contested by Democrats reduced.

2.4 Upsampled zipcode level taxpayer migration matrix

Counties can drastically differ in their sizes, particularly if some of these counties have urban cores. Harris, Dallas, Travis, Clark, Maricopa, and Los Angeles counties all contain dense urban cores that make them much larger and economically heterogeneous. So migration to and from such counties will be also be heterogeneous. Despite this restriction, since I have access to income of migrants and zipcode income brackets, I can take sampling draws for the zipcode's contribution to the county-level movement[10].

I produce three output distributions from this upsampling procedure for each zip-bracket and year. The draws from these distributions can them be summarized as quantiles. The outputs are the zip-bracket draws of inflow of taxpayers of a bracket into the zipcode from a different county I^E_{zbt}, and the the outflow of taxpayers from the zipcode to a county outside the zipcode, O^E_{zbt}[11]. Additionally, it also samples the zip-bracket distribution of non-migrants in each year, notated S^E_{zbt}.

While I observe the zipcode level zip-bracket taxpayer numbers X_{zbt}, the non-migrant distribution S^E_{zbt} is further sampled, because X_{zbt} is sensitive to changes in the underlying taxpayer population, and not the movers itself. This was particularly observable in 2020 where the number of filers in Los Angeles county increased in 2020 from 2019 because those that usually would not have filed for taxes, filed in the pandemic year to claim Covid payments through the IRS. This increase in X_{zt} occurred despite the large net outflow of taxpayers from Los Angeles. In contrast to X_{ct} numbers, non-migrant numbers S_{ct} fell. Distinguishing between these populations is useful to reduce the effect of short-term fluctuations in the taxpayer population.

From these three outputs, I can obtain the following information to analyze changes in income caused by moving and the intensity of migration from outside the county to the zipcode.

[10] The most granular data to track migration would be matched voter registration data. But I do not have access to individual voter registration data to match across states.

[11] Note that this value is the inflow into the zipcode from outside the county net of the outflow from the zipcode to outside the county. Movement within the zipcode cancels out in the county level totals.

1. M^E_{zt}: the number taxpayers inbound to zipcode z from outside the zipcode's county c across all brackets ($\sum_b I^E_{zbt}$) minus the number of taxpayers outbound from the county originating in zipcode z ($\sum_b O^E_{zbt}$).

2. $M^{E,H}zt = \sum_{b=5}^{6}(I^E_{zbt} - O^E_{zbt})$: the net migration as before but for high income taxpayers.

3. $\pi^{S,H}_{zt}$: the probability a non-migrant of the zipcode and county is a high income taxpayer.

4. $\pi^{I,H}_{zt}$: the probability a mover to the zipcode from outside the zipcode's county is a high income taxpayer.

5. $\pi^{O,H}_{zt}$: the probability a mover out of the zipcode's county is a high income taxpayer.

6. $M^{E,CA}_{zt}$: the number of taxpayers inbound to zipcode z from California net of the number of taxpayers outbound from the zipcode z to California. The number is calculated as a difference of $r^I_{CA,zt} * I^E_{zbt}$ and $r^O_{CA,zt} * O^E_{zbt}$.

7. $M^{cum,E,CA}_{zt}$: the cumulative of $M^{E,CA}_{zt}$ to time t starting from 2013.

8. $rM^{cum,E,CA}_{zt}$: the percentage of cumulative migrants of the non-migrant population in year t of the zipcode z.

Several other data consistency rules are imposed when the draws are made for the upsampler.

A.1 $X_{ct} = \sum_b X_{cbt}$: the number of taxpayers across brackets in county must sum to the county total[12].

A.2 $X_{zt} = \sum_b X_{zbt}$: the number of taxpayers across brackets in zipcode must sum to the zipcode total.

A.3 $G_{ct} = \sum_b G_{cbt}$: aggregate income across brackets in county must sum to aggregate income of county.

A.4 $G_{zt} = \sum_b G_{zbt}$: aggregate income across brackets in zipcode must sum to aggregate income of zipcode.

[12] The totals reported at the county level also includes negative incomes due to large business losses, but the totals reported at the zipcode level does not include this negative income bracket. I drop this bracket for counties and aggregate only positive incomes

A.5 $\sum_{z\in c} X_{zbt} \leq X_{cbt}$: across the zipcodes in a county, the number of taxpayers in each zipcode bracket must be less than or equal to that number of taxpayers in the county bracket. The bracket totals between zipcodes and the county of the zipcodes do not reconcile exactly for several reasons. When this is the case, a residual county group is added to force these to reconcile. See the Appendix section A.2 for these details.

The most primitive type of upsampler that can be used is to take sample draws for the number of migrations into the zipcode-bracket based on the observed zip-bracket compositions X_{zbt}/X_{ct} (for stayers or non-migrants) and X_{zbt-1}/X_{ct-1} (for leavers and incoming) and reject any zip-bracket draw combinations whose sum $\sum_b \sum_{z\in c} I^E_{zbt}$ do not add up to the relevant county's I_{ct}. Then these draws can be used to take quantiles of the distribution. But because of additional information on the income of the county migrants G^I_{ct}, I am able to make these draws sharper by including additional constraints on the income of the draws, so that the draws retained also conform to the constraint that $|\hat{G}^{I(s)}_{ct} - G^I_{ct}| \leq \epsilon^I_{ct}$ where $\hat{G}^{I(s)}_{ct}$ is the implied aggregate income of the draw $I^{E(s)}_{zbt}$, and similarly for draws $O^{E(s)}$ and $S^{E(s)}$. The draws are also made to satisfy a further constraint that the zipcode's current population X_{zt} is higher than the number of those that stayed $S^{E(s)}$, plus those who arrived, $I^{E(s)}$. The draws that survive these constraints are then used to obtain the various measurements such as $M^{E,H}_{zt}$, from which quantiles for those measurements are derived. For exact details of the algorithm that makes these draws, consult the Appendix A.3.

3 Results and Discussion

3.1 Income sorting

As a first step, I demonstrate the high income taxpayer movements to and from the states. The IRS data does not breakdown the movement of taxpayers by income bracket, though they provide the overall income of the movers. The upsampling algorithm breaks these movements down by making draws from zipcodes and income brackets within those zipcodes. The zipcode-bracket draws, $\{M^E_{zbt}\}_{b=5,6}$, can be aggregated across the two brackets and all zipcodes and counties to get a statewide net movement of high income taxpayers. The left column of Figure 6 contains these draws $M^{E,H}_t$ summarized to state-level quantiles. The quantiles range from 10th to 90th[13]. The grey line in each of these left column figures denote the number of net taxpayers in the state from the zipcodes that were usable[14]. A positive

[13] The scale of the y-axis makes it difficult to see the difference between the quantiles.

[14] Some zipcodes are masked by the IRS due to too few returns

number means that more taxpayers moved into the state than moved out, while a negative number means more taxpayers moved out than in during the year.

As has been observed before, California is a net loser of taxpayers since 2013 while all other states are net gainers. Unsurprisingly, this pattern is also true of high income taxpayers. California's largest loss of high income taxpayers coincided with its largest loss of all taxpayers. This was true of Texas, but in the opposite direction, where it experienced a spike in the net inflow of taxpayers and high income taxpayers. In Nevada and Arizona, even while the net migration numbers were dropping, the high income taxpayer numbers were increasing or constant.

Given these observations, the immediate follow-up would be were high income taxpayers more likely to leave California, and were they more likely to move into Nevada or Texas than any other taxpayer. The plots in Figure 6's right panel provides some clues. For each z, t, the upsampler makes draws of the probablity that a mover into the zipcode is a high income taxpayer, $\pi_{z,t}^{I,H}$, and the probability that a mover out of the zipcode is a high income taxpayer, $\pi_{z,t}^{O,H}$. Another measurement is the probability that a stayer is a high income taxpayer, $\pi_{z,t}^{S,H}$. These draws are summarized to state-level quantiles across zipcodes and reported in the right panel for each state. The line marked 'Stay' denotes the quantile summaries for $\pi_t^{S,H}$. The line marked 'In' denotes the quantile summaries for $\pi_t^{I,H}$, and the line marked 'Out' denotes the quantile summaries for $\pi_t^{O,H}$.

In the baseline year of 2013, Californians were more likely to be high income than all other states at a rate of about one in five taxpayers being high income (0.2). Nevada had the smallest ratio of high income taxpayers at 0.12. Arizona and Texas were between 0.16 and 0.18. The upward trend in all these quantities is an artifact of wage growth as more and more taxpayers earn over $100,000. Since it is difficult to separate out this wage growth trend from our desired trend of whether high income taxpayers are more likely to move in or out, the discussion must make comparisons within the year, rather than across the year. In California, a outgoing taxpayer was more likely to be high income than an incoming taxpayer across the period of observation. But this difference was very slight and reversed in 2022. The difference was more drastic in 2020, where the outgoing taxpayers were much more likely to be high income than incoming or staying. The exact converse pattern occurred in Nevada, where an incoming taxpayer was more likely to be high income than an outgoing taxpayer, and incoming taxpayers were more likely to be high income than even those staying in 2020. In Texas and Arizona, incoming taxpayers were more likely to be high income than those outgoing. But these two states did not experience a pandemic spike in high income taxpayers relative to other taxpayers.

These previous measurements show that incoming taxpayers to Arizona, Nevada, and

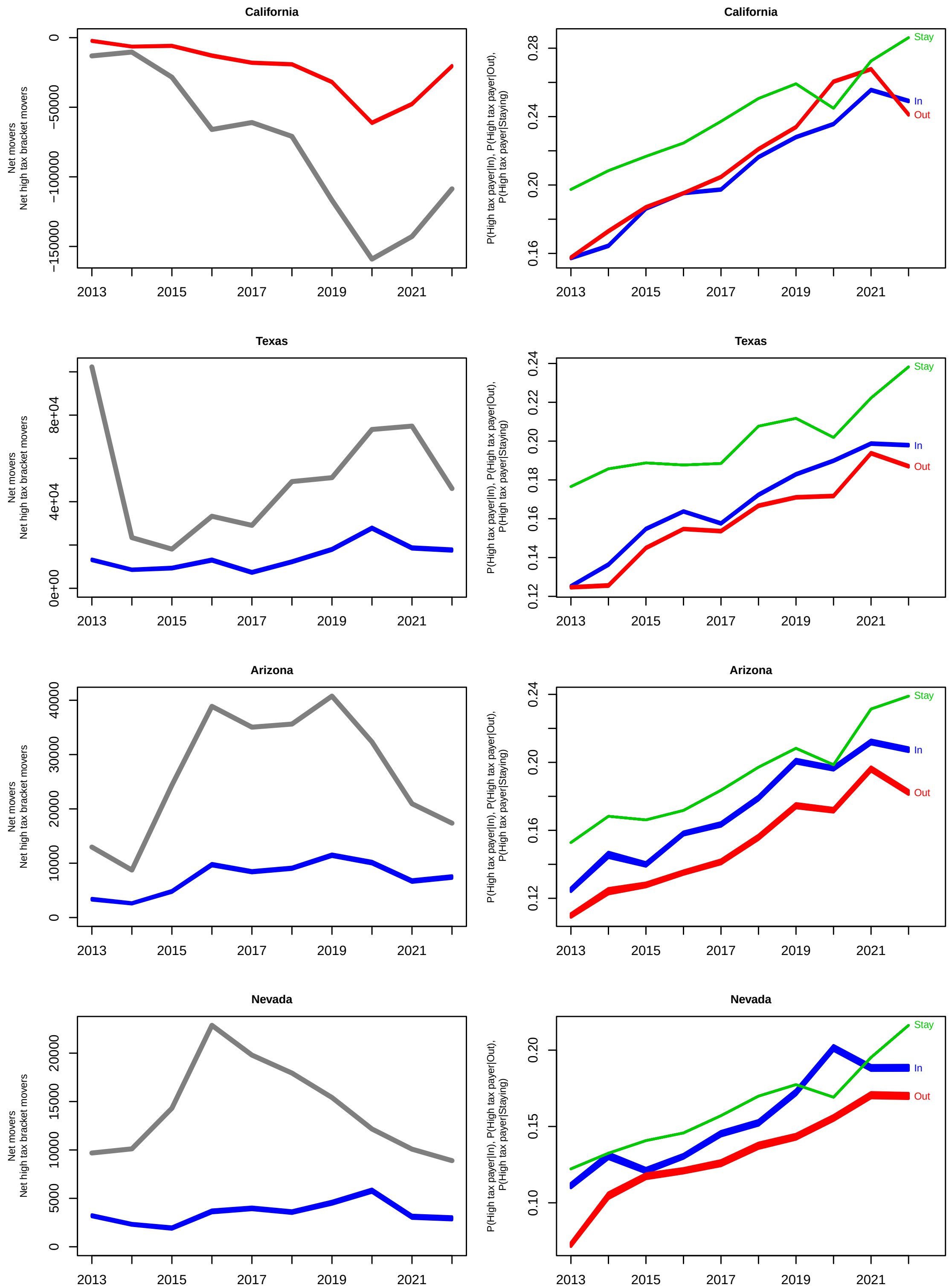


Figure 6: High income taxpayer movements to and from the four states Arizona, California, Nevada, and Texas

Texas are more likely to be richer. But there is limited evidence that a convergence is happening between the states. As the left panel of Appendix Figure A.9 shows almost no reduction in the high income share gap between the California and other states.

We still cannot necessarily rule out that the Californian migration has no impact on income compositions in the states. For one, the state-level aggregates might be obscuring local clustering by incoming migrants into high income areas in such a way that, on average, incomes look the same. For another, the trends could be driven by outmigration from the state rather than the migration in by Californians. To distinguish between these various possibilities, I analyze the variation of Californian migration rates into the zipcodes as drawn by the upsampler against the high income share of the zipcode. The dependent variable in this analysis is calculated as the percentage of each zipcode's taxpayers that are high income, i.e. $100 * X_{zt}^{H}/X_{zt}$. This quantity is derived purely from the IRS reported data without any input from the upsampler.

The quantities $M_{zt}^{cum,E,CA}$ and $rM_{zt}^{cum,E,CA}$ encode the intensity of migration to each zipcode from 2013 to year t. The draw $M_{zt}^{cum,E,CA}$ is the cumulative net migration *level* upto year t in a zipcode and the draw $rM_{zt}^{cum,E,CA}$ is the cumulative net migration *rate* upto year t, where rate is based on the current resident population at year t of the receiving zipcode. As observed in the previous discussion surrounding Figure 4, the ordering of counties according to these level and rate measurements do not coincide. This is also true of zipcodes. In other words, a zipcode receiving a high level of Californians ($M_{zt}^{E,CA}$) might not necessarily have a higher rate of Californian migration compared to its resident population. Another zipcode might have a very high rate of Californian migration, but have a low level of migration.

These measures give different types of information. To see this, consider Appendix Figures A.10 and A.11. Respectively, these figure plot the cumulative intensity measurements $M_{zt}^{cum,E,CA}$ and $rM_{zt}^{cum,E,CA}$ against the dependent variable $100 * X_{zt}^{H}/X_{zt}$ in 2013 for each state. In each state's plot, a line of best fit is drawn to summarize the association. In Texas, the association is slightly positive under both the cumulative level and rate. That is, high income zipcodes saw a higher level and a higher rate of migration from California. However, in both Arizona and Nevada, this associations between levels and rates had different signs. In Arizona, higher income zipcodes were more likely to receive a high inflow of cumulative migration, but at the same time higher income zipcodes were less likely to have high rate of cumulative migration. In Nevada too, the associations had opposite signs, but the inverse of Arizona. Poorer zipcodes in Nevada received higher flows from California, but poorer zipcodes also saw a lower rate of migration from California compared to their own resident population.

Since these two measures of migration intensity are different, I report results for both.

For each quantile of these measures, I run a regression on $100 * X^H_{zt}/X_{zt}$ with fixed effects for the zipcode z and year t. In addition, I also control for the zipcode's taxpayer population S^E_{zt} of the respective quantile. So nine regressions are run for each quantile from 10th quantile to 90th quantile. For each quantile q, such a regression would recover the marginal effect β of the relevant indicator of migration intensity from

$$100 * X^H_{zt}/X_{zt} = \alpha_z + \alpha_t + \beta^q \cdot [\text{Mig Intensity}_{zt}]^q + \delta^q[\cdot S^E_{zt}]^q \tag{1}$$

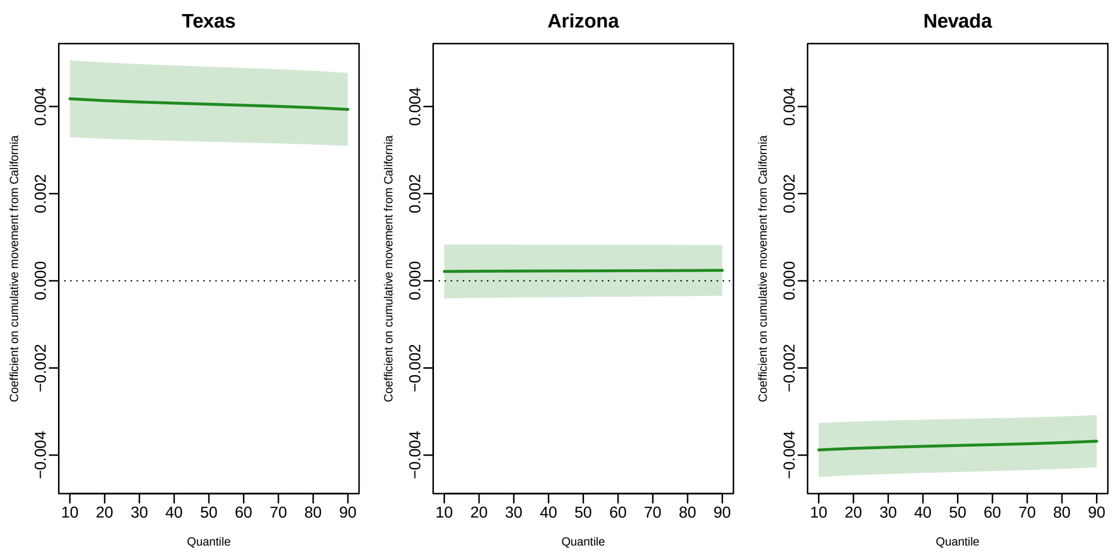


Figure 7: The effect of cumulative Californian migration on zipcode's high income share for each quantile of $M^{cum,E,CA}_{zt}$

Figure 7 reports the coefficient β when migration intensity is the number of cumulative net migrants from California to the zipcode, $M^{cum,E,CA}_{zt}$. Since $M^{cum,E,CA}_{zt}$ is upsampled from county-level data, the coefficient β is reported for every 10th quantile from 10 to 90. The results show that states responded differently to Californian migtation. In Texas, a higher cumulative flow of migrants increased the high income share. In Arizona, no effect was observed, while in Nevada a higher cumulative flow was associated with a decline in high income share. All in all, however, the effects were marginal. In Texas zipcodes, a cumulative flow of 1000 Californian migrants on average increased the high income share by roughly 4 percentage points. And in Nevada zipcodes, a cumulative flow of 1000 Californian migrants, decreased the high income share by roughly 4 percentage points.

The S^E_{zt} controls for any baseline sorting of migrants into high population zipcodes. In general, the direction of this coefficient δ was positive, indicating that higher population zipcodes brought in more Californian migrants, but was mostly insignificant because any of these zipcode differences were absorbed by the zipcode fixed effect.

Though the coefficient β is recovered with controls for baseline factors such as zipcode's high income share in 2013 and year-level factors that can affect this such as the 2020 pandemic, zipcodes can differ in their dynamic trends. Trend differences between high income and low income zipcodes can affect the recovered effect β which aggregates effects across both high income and low income zipcodes. Appendix Figures A.12 and A.13 split the zipcode sample by each zipcode's 2013 high income share and re-implements the regression 1 in these disaggregated samples. The results in these disaggregated samples match the pattern in Figure 7. The Appendix Figure A.12 reports the marginal coefficient for zipcodes whose high income share in 2013 was less than 20 percent. The effect in Texas was again roughly a 3 to 4 percentage point increase in zipcode high income share, and the effect in Nevada was a decline of zipcode high income share of 4 percentage points. Figure A.13 reports the marginal coefficient for the sample of zipcodes whose 2013 high income share was between 20 and 40 percent, and the effect's direction matches the low income zipcodes, but with a weaker effect.

This disaggregated analysis also provides information on the local sorting effect of the migratory wave. While there seems to be some evidence of income sorting into zipcodes, the direction of sorting differed between states with high income zipcodes more likely to see higher inflows in Texas, and low income zipcodes more likely to see higher inflows in Nevada (see Appendix Figures A.10). Domestic migration, however, did not necessarily worsen this sorting. Zipcodes in Texas, on average, regardless of income share, saw a similar increase in their high income share. And all zipcodes in Nevada, on average, regardless of income share, saw a decline in high income share. No effects were noticeable in Arizona, even when disaggregated by zipcode's baseline income share. However, important to note, is that these effects are very small: even a 1000 person increase on average only shifts high income share by 4 percentage points.

When migration intensity is measured by the rate of cumulative migration rather than the flow, the effects on the zipcode high income share become even weaker. As Figure 8 shows, a one percent increase in cumulative migration rate within the zipcode, shifts the high income share of the zipcode by 0.8 percentage points. In both Nevada and Arizona, the effect of such a percentage point increase is very close to zero.

The empirical evidence of income sorting between the specific states is weak, even though Californians leaving are more likely to be higher income than the resident populations (See right panel of Appendix Figure A.9). The above analysis shows that even if there is sorting it is heterogeneous between states (Texas vs Nevada). There are several possible explanations at to why the higher incomes of Californians are not changing income compositions. The

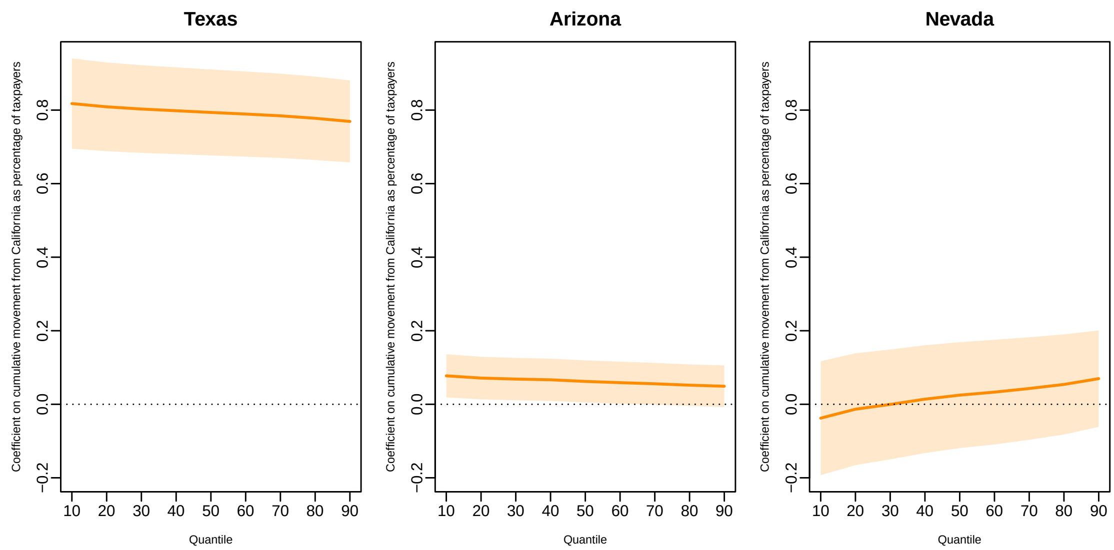


Figure 8: The effect of rate of cumulative Californian migration on zipcode's high income share for each quantile of $rM_{zt}^{cum,E,CA}$

first is the difference in demographic groups moving which also explains the state level heterogeneity. Another explanation might be that moving Californians are drawing a lower income once they settle into their new state. This is particularly the case with retirees who will move to lower cost states like Nevada. But, even families with young children may be willing to trade off cheaper housing costs and lower taxes for lower gross salaries.

Empirical evidence for within state sorting is even weaker. Californians do not necessarily make richer zipcodes richer, and poor zipcodes poorer. In Texas, both high income and low income zipcodes that saw high migration saw a slight increase in high income share, that is, they got slightly richer. In Nevada, both high income and low income zipcodes that saw high migration saw a decline in high income share, although this was not true when migration intensity was measured by rates rather than levels.

3.2 Partisan sorting

Like income sorting, partisan sorting can occur between states and within states. Given California's large Democratic voting population, Republicans in host states worry they will face greater electoral competition in statewide offices. Even if the migratory flow is too small to effect state-level changes, local Republican incumbents worry that Democratic voters from California flocking to marginal areas are making their races unwinnable. Democratic incumbents in marginal areas may have similar worries about local sorting. Those tracking voter registration files, however, find that those leaving California are predominantly Republican registered voters (McGhee & Johnson, 2025). But the composition of those entering Texas from California, seems to be more evenly divided between Democrats, Republicans, and the unaffiliated (Gimpel & Shaw, 2024).

Voter registration is observable at an individual migrant level whereas voting behavior can only be observed at a precinct level. However, voter registration in one state is not perfectly predictive of the partisan voting behavior in a different state, either. Platforms and candidate quality can sufficiently differ between states in both major parties. So a registered Republican Californian arriving in Texas could vote for a Democrat. The change in Republican vote share from the 2016 Presidential election to 2024 Presidential election as show in Figure 5 indicate a Republican swing in all three receiving states. However, this cannot be necessarily attributed to Californian migration.

Registration information also shows patterns spatial sorting that reinforces partisanship at their destinations. Gimpel and Shaw (2024) find that Californian Democrats to the Houston area moved closer to the city, while Californian Republicans chose destinations further out in the greater Houston area which leaned more Republican. Figure 5 also offers some clues of this type of partisan sorting. Texas precincts in 2024 have a bimodal distribution, with one peak left of the 50 percent Republican vote share and another peak on the right of 50 percent R vote share. This could be indicative of a type of local partisan sorting. The zipcode level migration intensity measurements do not allow me to distinguish the partisanship of the arriving Californians. But while this is useful information to have, registration information can also be misleading, particularly for Californian voters who do not necessarily vote in Presidential primaries. I analyze partisan sorting as I did for income sorting, exploiting the variation in migration intensity rather than the partisanship of the migratory flow.

Using the 2016 Presidential Republican vote share as a baseline, I investigate whether Californian migrants were more likely to sort into Democratic-leaning zipcodes. Appendix Figure A.14 shows a weak association between the cumulative migration flow in 2022 for the zipcode and the Republican vote share in 2016 of the precincts in that zipcode for both Texas and Arizona. In Nevada, Democratic leaning zipcodes in 2016 were more likely to see cumulative migration from 2013 to 2022. Appendix Figure A.15 demonstrates the association between cumulative migrant flow rate in 2022 and R vote share in 2016. The associations have the same sign, but are weaker.

As with income sorting, I calculate the marginal effect of the cumulative migratory flow. I only observe this migration flow up to 2022. But instead of eliminating the 2024 observation, I implement a regression on the lagged migratory flow with the assumption that the cumulative flows will not change drastically had I had access to the newer information. For each quantile

q, the marginal effect β of the relevant indicator of migration intensity is recovered from the regression,

$$100 * \text{RVoteShare}_{pt+2} = \alpha_p + \alpha_r + \beta^q \cdot [\text{Mig Intensity}_{zt}]^q + \delta^q[\cdot S_{zt}^E]^q \tag{2}$$

where p denotes the precinct and r denotes the race, which is just the year fixed effect (t+2) for Presidential elections but the district and year (t+2) fixed effect for House elections.

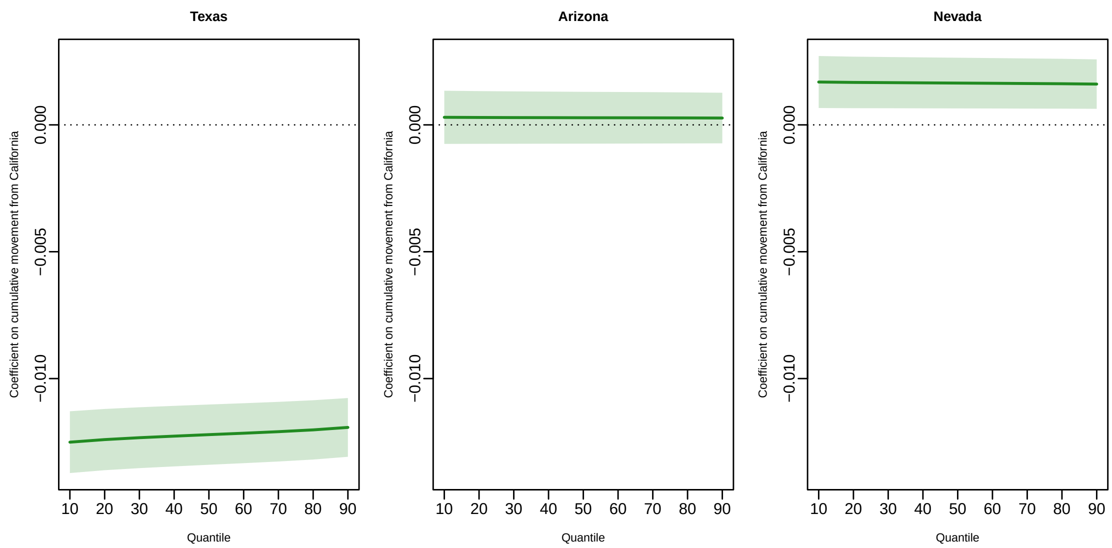


Figure 9: The effect of precinct zipcode's cumulative Californian migration on precinct's Republican vote share for each quantile of $M_{zt}^{cum,E,CA}$

Figure 9 shows that a higher cumulative migration is associated with a Democratic voting effect in Texas with no effect in Arizona, and a very small Republican voting effect in Nevada. The recovered Democratic voting effect in Texas is about 1.2 percentage point marginal effect per additional 100 cumulative migrants, on average. The corresponding results for the cumulative rate is in Appendix Figure A.19 and corroborates the same pattern. Texas precincts saw a 2.5 percentage point increase in precinct level Democratic vote percentage for every 1 percentage point increase in cumulative migration rate.

One concern for the interpretation of the coefficient on migration intensity in the regression implemented in 2 is the differences in dynamic trends of partisanship between Democratic, Republican, and marginal precincts. If the native population is sorting into co-partisan areas, even without migration Democratic precincts will be getting more Democratic, while Republican precincts are getting more Republican. To account for possible violations of parallel trends, as with income, I split each state sample to three groups of precincts based on the 2016 Presidential vote share. The three groups are predominantly Republican voting precincts (over 60 percent Republican vote percentage), predominantly Democratic voting precincts (below 40 percent Republican vote share), and marginal precincts (those above 40 percent Republican vote share but below 60 percent).

In Texas's predominantly Democratic voting precincts, a high migration flow increased Democratic vote share by 2 percentage points for every 100 additional migrant inflow (Appendix figure A.18), and by 4 percentage points for every 1 percentage point increase in the rate of cumulative migration (Appendix figure A.20) . In both marginal (40 to 60 Republican vote share) and Republican leaning precincts (more than 60 Republican vote share) in Texas, the marginal effect was a 1 percentage point increase in Democratic vote percentage for every 100 additional cumulative migrants, and a 2 percentage point increase in the Democratic vote percentage for every 100 additional cumulative migrants.

For both Arizona and Nevada, the overall marginal effect and the partisan effect are both very close to zero. Nevada had a very small Republican voting swing in zipcodes that received a high cumulative flow, but not when the indicator of migration intensity was the cumulative rate. When disaggregated by 2016 Presidential voting, all three marginal effects on the cumulative rate of migration had a Democratic swing, but the same marginal effects on cumulative level was null. Arizona had mostly null results, except its marginal precincts had a Republican swing with greater migration. The effect was however small. In these marginal precincts, a cumulative inflow of 1000 migrants increased Republican vote percentage by 2 points and a cumulative rate increase of 1 percentage point increased Republican vote swing by 0.5 percentage points.

3.3 Further improvements

I have analyzed California's migratory wave as a homogeneous flow, focusing on its intensity rather than the differences across time and origin. But this simplification is due to a lack of more granular data on the migrants. In reality, the migratory flow is heterogeneous in relation to both income and partisanship. The heterogeneity of its impact on the different states is another clue that these differences exist.

Individual matched voter records can sharpen the conclusions we make about migratory flows, particularly if we were to answer more specific questions about income sorting such as 'Do high income Californians go to high income areas and make them more high income'. Individual voter income is difficult to observe, but voter age and exact origin address can help deduce the trends.

Another challenge to identification of a migration's impact are concomitant changes in the underlying native population. Voter file movements can help distinguish some of these trends too, particularly if the partisanship change is a function of age and generational turnover in an area. Migrants might be attracted to areas that are undergoing change, and in order to disentangle these two forces more granular data is needed.

Still another challenge is change in the definition of partisanship itself. Californian Democrats and Texas Democrats differ in their policy positions. So any analysis tracking partisanship change caused by a migratory flow must also have a model of how partisanship is perceived across different states. This is much harder to do, but split-ticket voting patterns between Presidential and House members can help illuminate these even if they are observed at only the precinct level.

4 Conclusion

Fear and loathing is a common reaction to Californian migrants to Las Vegas, Phoenix, Houston, and Dallas. Incoming Californians frequently encounter 'Don't California my [State]' bumper stickers. Underlying such sloganeering is a fear that higher income and Democratic-leaning Californians are changing the economic and partisan makeup of the local areas, raising prices and rents for current residents and changing the politics. This paper empirically tested whether such a recomposition is occurring.

Californians moving out are more likely to be high income (earning over $100,000) than the existing taxpayers in the receiving states, but their effect on the high income share on zipcodes differs across states. California's migratory flows had little impact on the zipcode income composition in both Arizona, reduced the high income share in Nevada, and raised the high income share in Texas. However, the recomposition effect size was small even when split by baseline income composition. There could be several reasons for this heterogeneity. One reason might be that the incoming migrants find states and zipcodes with similar income composition to them, so that even with higher volumes of migration income compositions do not change. Another reason could be that incoming Californians earn similarly to host populations in those labor markets, which means they accept a lower gross income but still make the move because their costs are lower (Fischer & White, 2026).

Partisan compositional changes were of larger size than income compositional change, but displayed the same state level heterogeneity. Areas with high inflow of migrants saw a Democratic tilt in Texas, a slight Republican tilt in Nevada, and no effect in Arizona. Among Democratic leaning precincts in Texas, those that received a higher migration flow became even more Democratic. These compositional changes agreed directionally when migration intensity was measured in rates.

Individual voter files can improve on the identification challenges inherent to the exercise.

One of this is concomitant changes in native population through generational turnover which can be traced through new voter registrations. Voter files can also help identify when the specific origins and age of the incoming voters so that the heterogeneity of the receiving population can be better understood.

A final observation concerns the remarkable absorption of even large migratory flows in modern United States. While any large movement of people puts upward pressure on rents and house prices, this recent pandemic-related movement seems to have been absorbed quite well by the housing and rental markets in receiving metropolitan areas. In fact, rents in many of these receiving metros were falling by 2025, rather than increasing (Winters, 2025). Though I observe several significant compositional changes, those changes are much less disruptive than what I would have observed in the last century. The reasons could be various. I speculate that domestic movers are more informed movers now and choose their destinations to reflect their economic status more closely.

References

Abrams, S. J., & Fiorina, M. P. (2012). "The Big Sort" that wasn't: A skeptical reexamination. *PS: Political Science & Politics*, *45*(2), 203–210.

Bishop, B., & Cushing, R. G. (2009). *The Big Sort: Why the clustering of like-minded america is tearing us apart*. Houghton Mifflin Harcourt.

Fischer, B., & White, E. (2026). *Priced out: Relocation amidst California's affordability crisis*. California Policy Lab, University of California. https://capolicylab.org/priced-out-relocation-amidst-californias-affordability-crisis/

Gimpel, J. G., Newton, J., & Reeves, A. (2026). Short-haul moves and the political geography of partisanship: Intrametropolitan migration as a force for change in US politics. *Urban Affairs Review*, *62*(4), 975–1018.

Gimpel, J. G., & Shaw, D. R. (2024). Long distance migration as a two-step sorting process: The resettlement of Californians in Texas. *Political Behavior*, *46*(3), 1837–1864.

Johnson, H., & McGhee, E. (2026). Who's leaving California—and who's moving in? *PPIC Blog*. https://www.ppic.org/blog/whos-leaving-california-and-whos-moving-in/

Madriaga, M. (2020). San Diegans head to Phoenix. *San Diego Reader*. https://www.sandiegoreader.com/news/2020/may/27/cover-san-diegans-who-enjoy-phoenix/

Martin, G. J., & Webster, S. W. (2020). Does residential sorting explain geographic polarization? *Political Science Research and Methods*, *8*(2), 215–231.

McGhee, E., & Johnson, H. (2025). California's Republican exodus. *PPIC Blog*. https://www.ppic.org/blog/californias-republican-exodus/

Mummolo, J., & Nall, C. (2017). Why partisans do not sort: The constraints on political segregation. *The Journal of Politics*, *79*(1), 45–59.

Murashima, C. (2024). Why Nevada's housing crisis is about more than Californians driving up home prices. *OPB*. https://www.opb.org/article/2024/09/25/why-nevada-s-housing-crisis-is-about-more-than-californians-driving-up-home-prices/

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, *64*(5), 416–424.

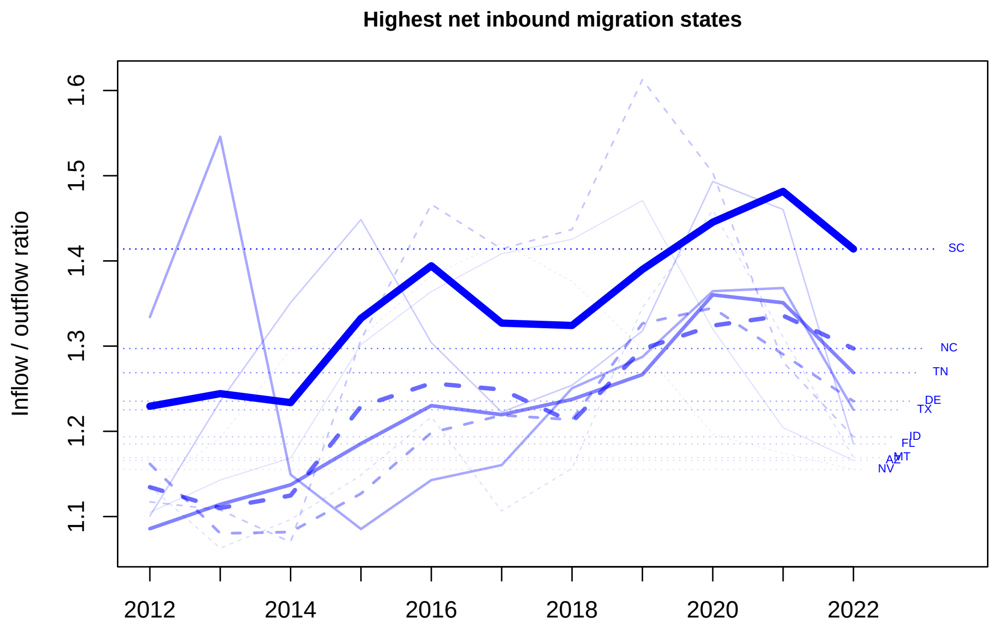
Uhler, B., & Garosi, J. (2018). California losing residents via domestic migration. *California Economy and Taxes*. https://lao.ca.gov/laoecontax/article/detail/265

Walters, D. (2024). California's political clout will fade as long as population growth remains slow. *CalMatters*. https://calmatters.org/commentary/2024/11/california-political-clout-population-growth/

Winters, M. (2025, December). *Rents are falling in these major U.S. cities heading into 2026—one of the more "renter-friendly periods" in a decade, says expert*. Retrieved September 10, 2026, from https://www.cnbc.com/2025/12/26/rents-are-falling-in-these-major-us-cities-heading-into-2026.html

Appendix A

A.1 State and county net-inflow information



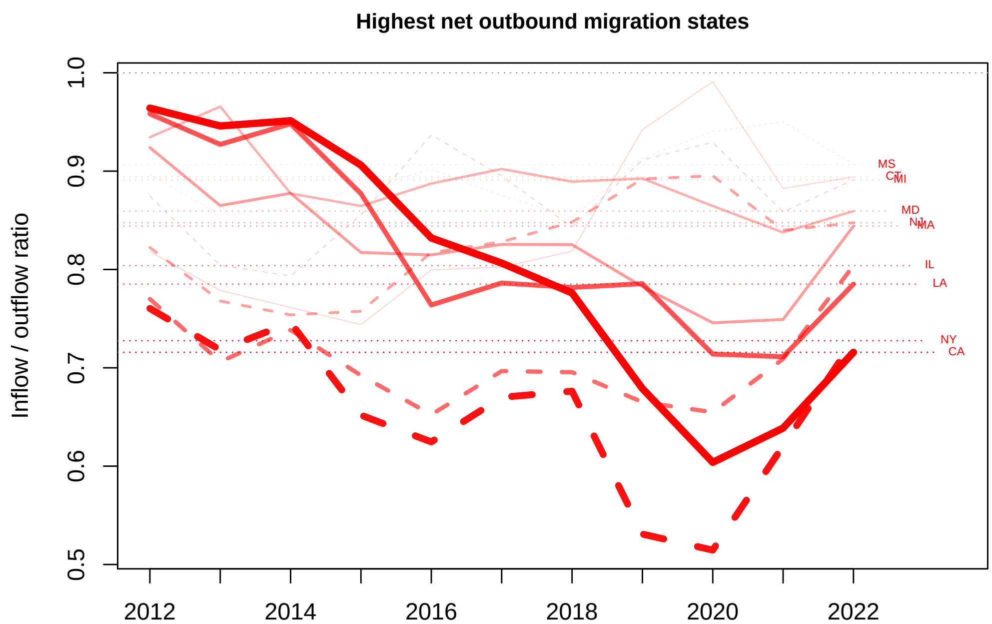
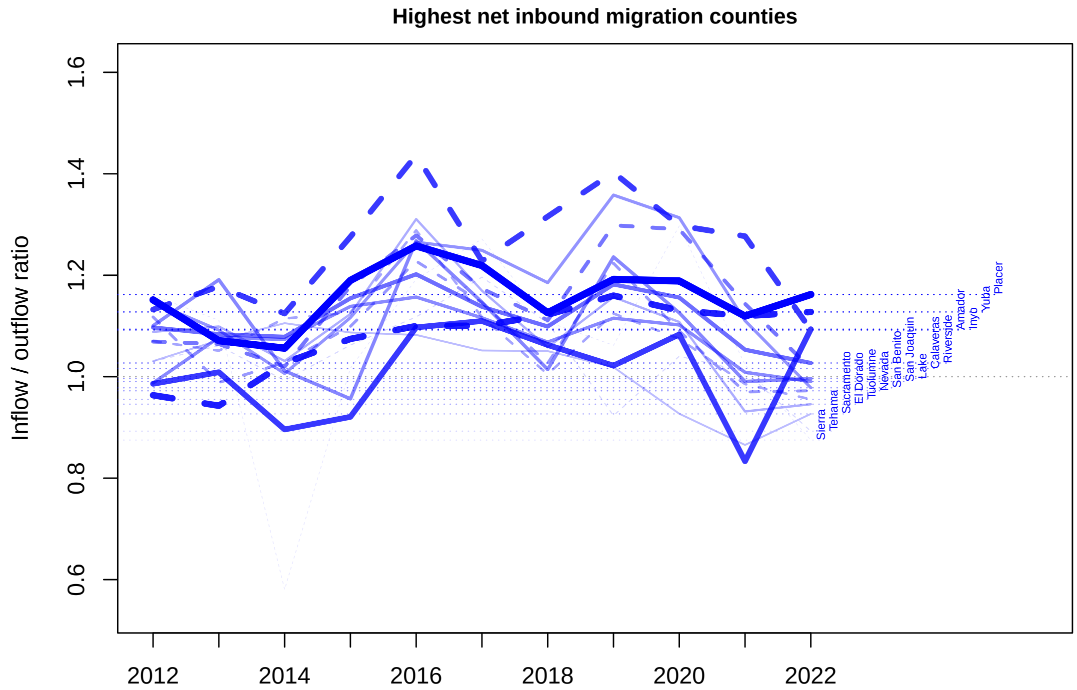


Figure A.1: Top 10 inbound states and Top 10 outbound states according to their inflow over outflow ratio



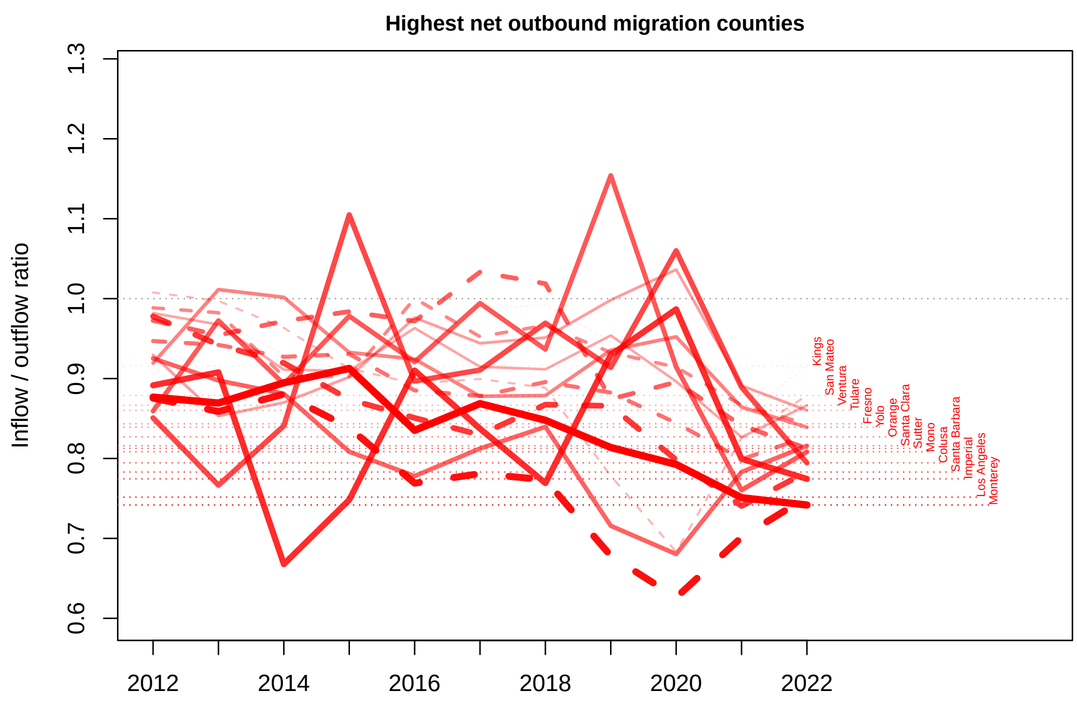


Figure A.2: Top 15 inbound California counties and Top 15 outbound California counties according to their inflow over outflow ratio

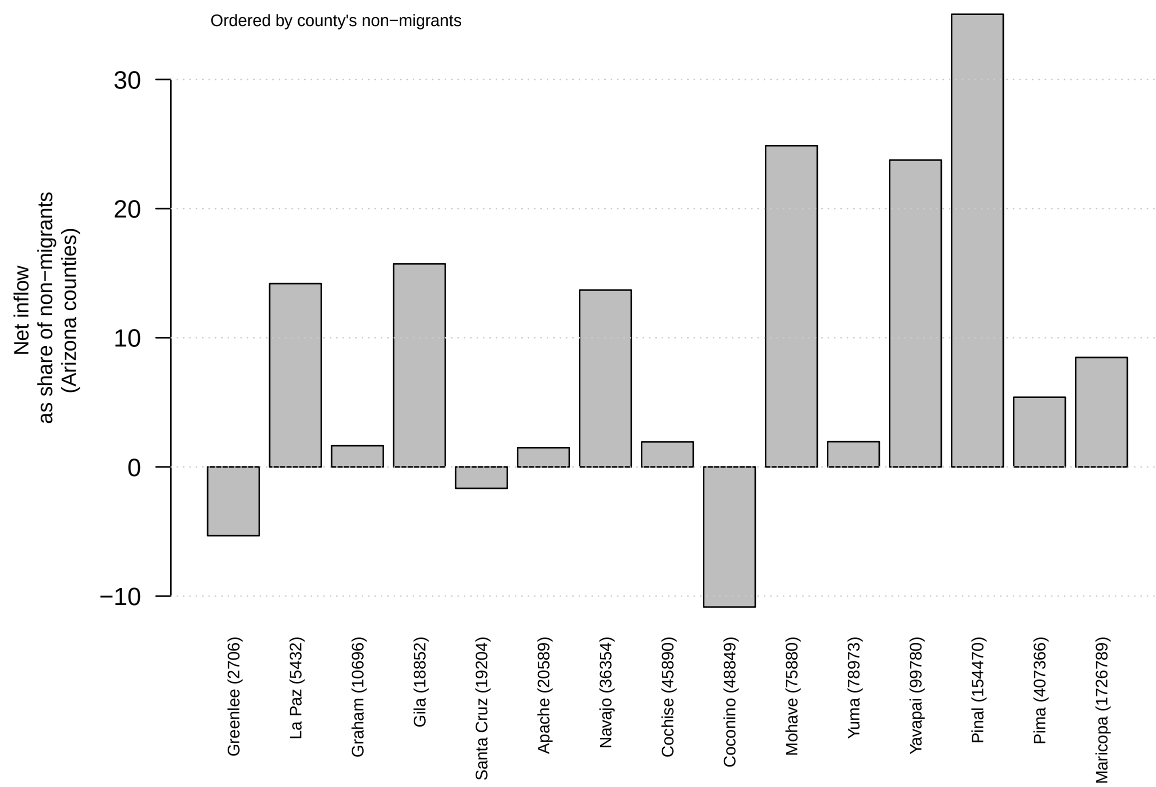


Figure A.3: Arizona's county-level cumulative net inflow as a percentage of non-migrants in 2022

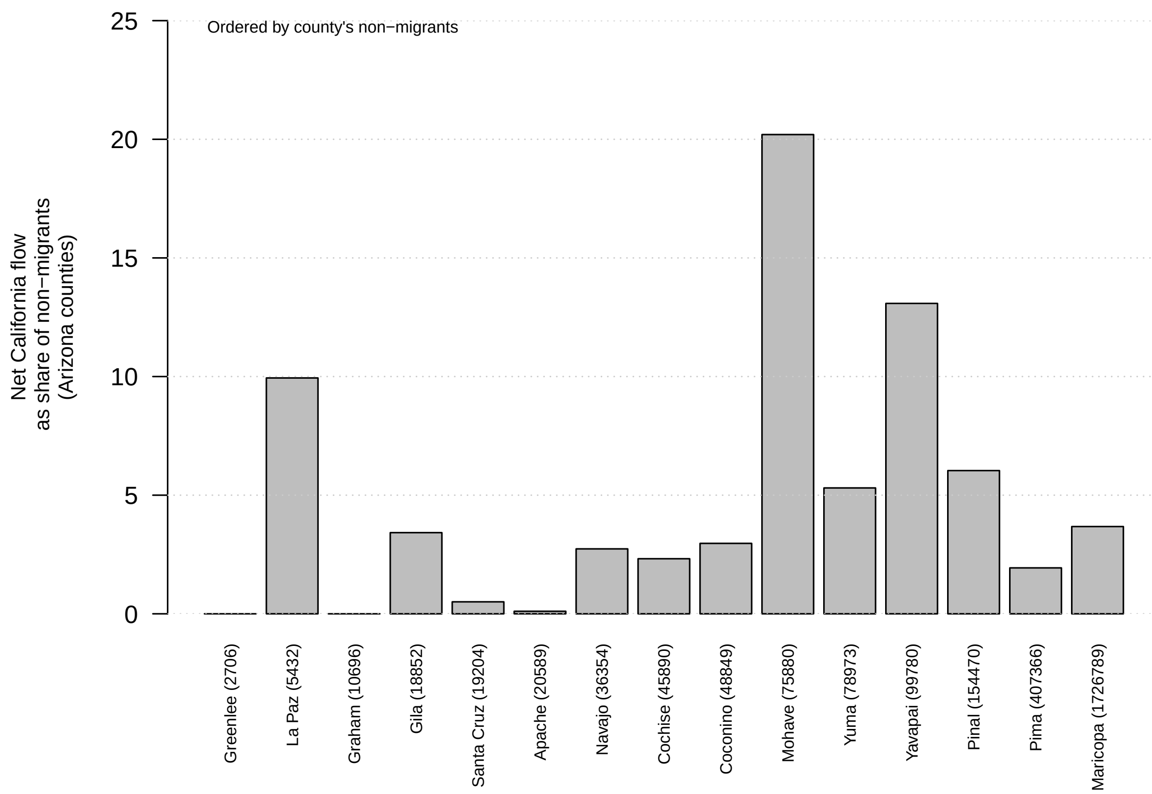


Figure A.4: Arizona's county-level cumulative net inflow from California as a percentage of non-migrants in 2022

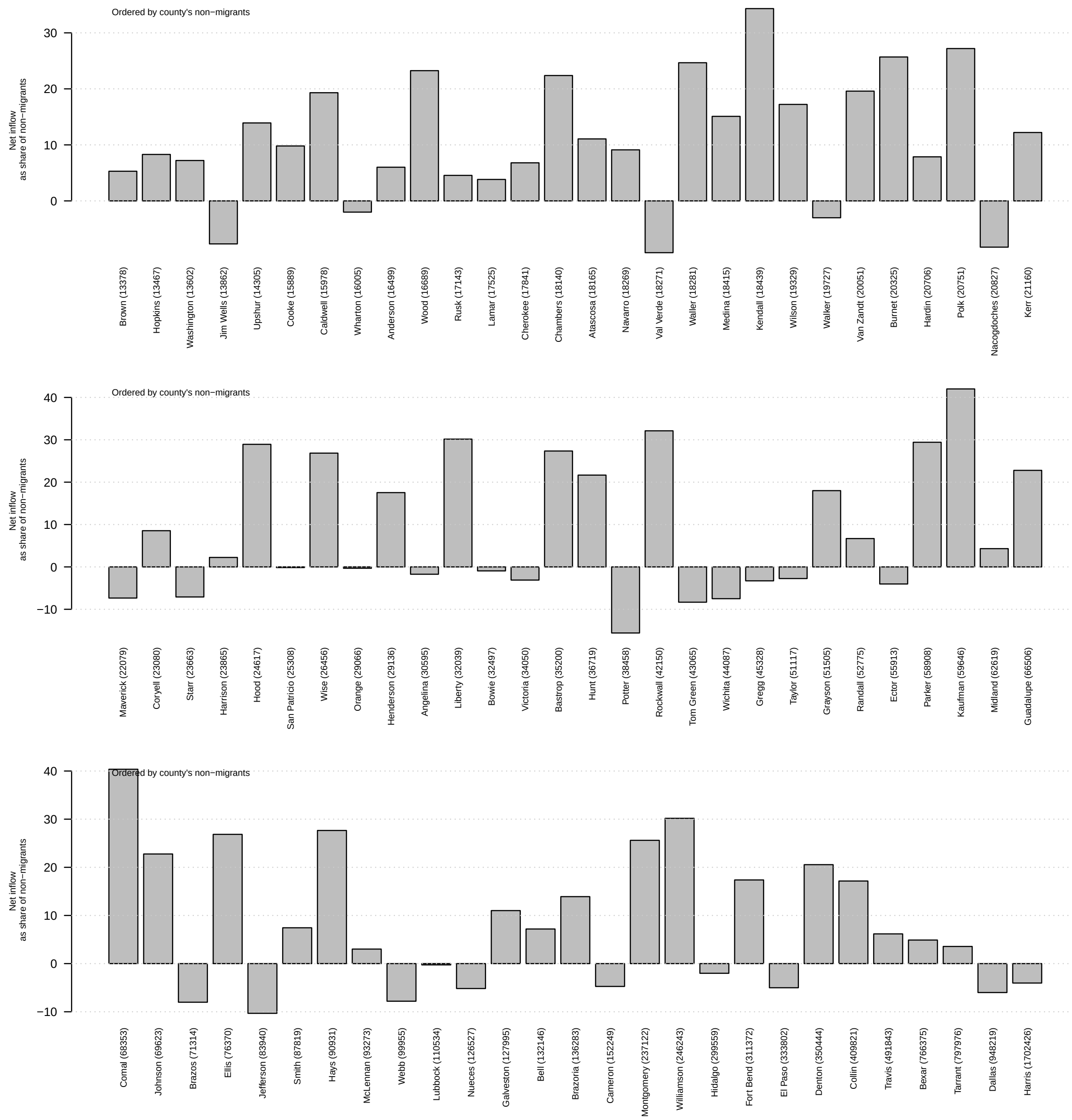


Figure A.5: Texas's county-level cumulative net inflow as a percentage of non-migrants in 2022

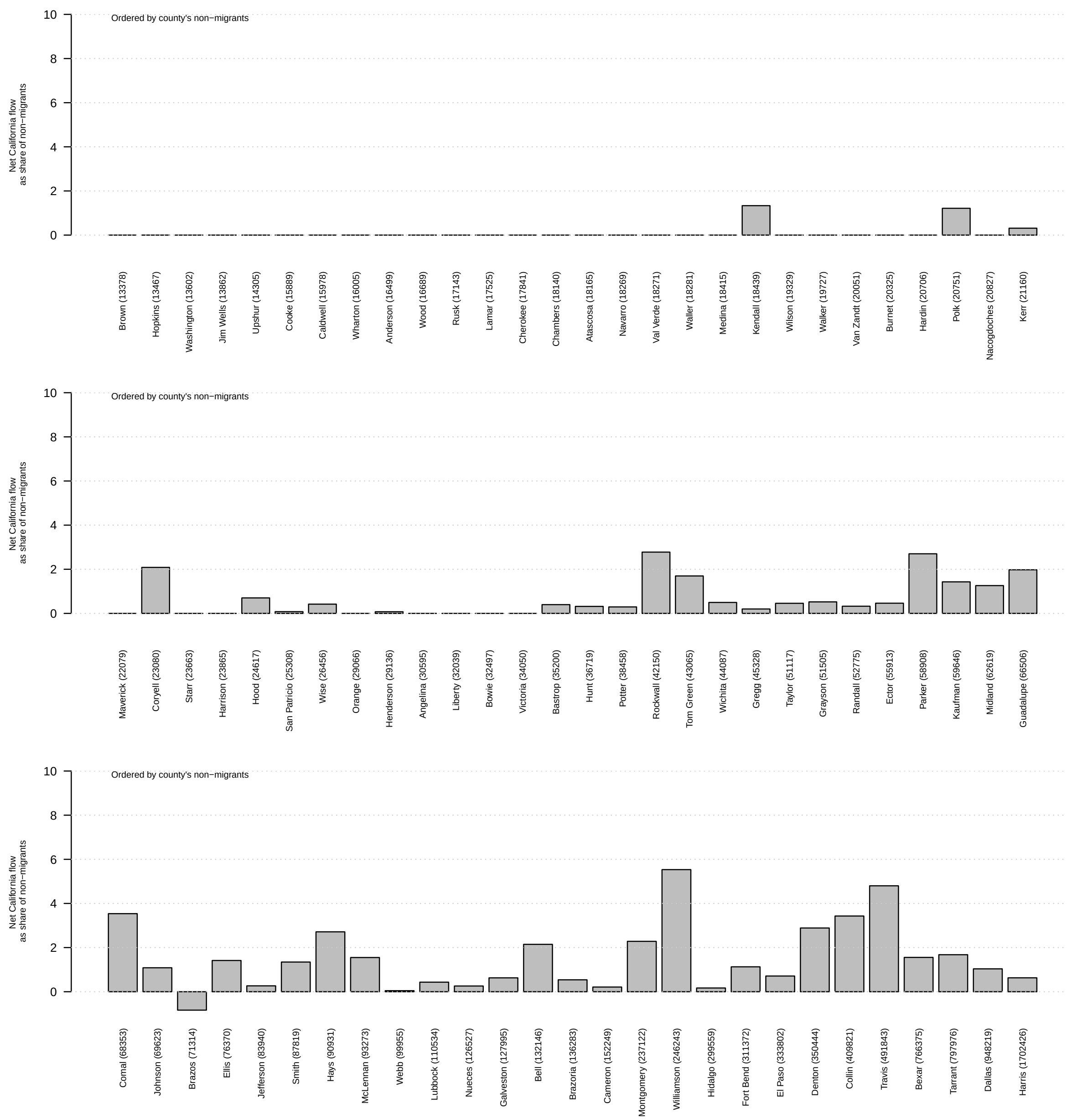


Figure A.6: Texas's county-level cumulative net inflow from California as a percentage of non-migrants in 2022

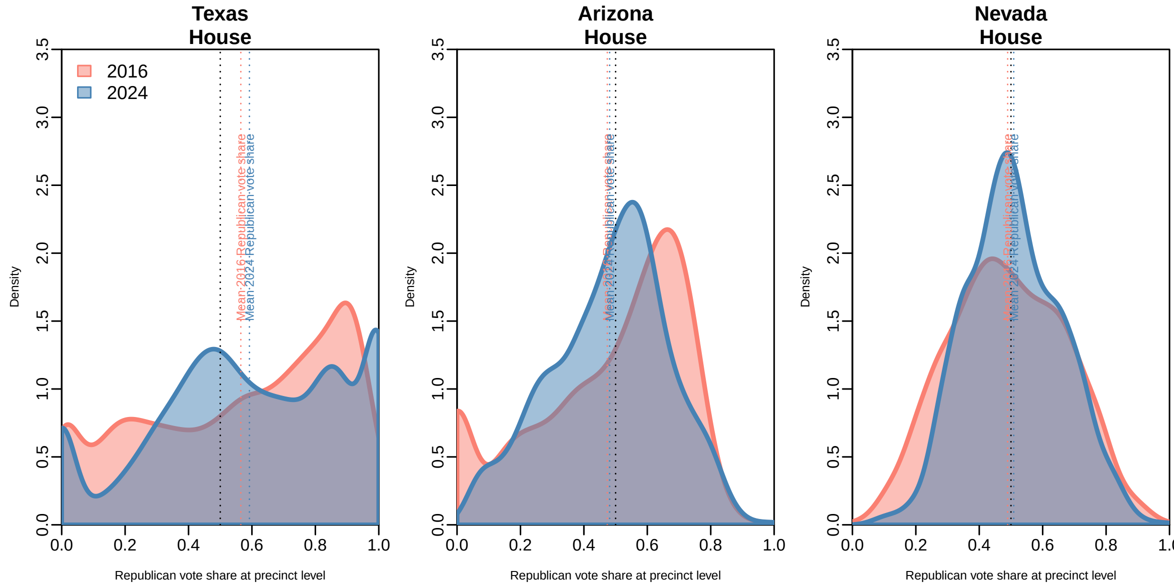


Figure A.7: A density plot of precinct-level House Republican vote share in 2016 and 2024

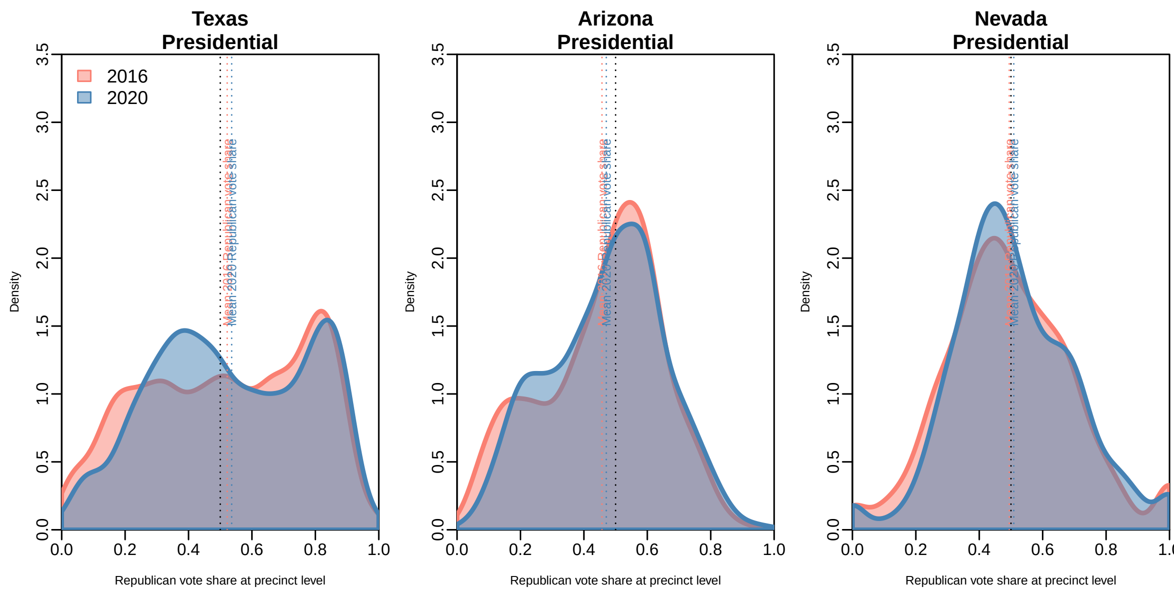


Figure A.8: A density plot of precinct-level Presidential Republican vote share in 2016 and 2020

A.2 Quality and consistency checks

The following checks were performed to ensure that the zip-county correspondence was stable over time and removed counties whose residual was too large.

Since zipcode boundaries change over time, I remove any zipcodes that seem to have substantially changed year-on-year in terms of the county-zipcode ratio of residential units. This is to ensure that when comparing zipcode information between the two years, we are not comparing very different geographic units as zipcodes. I choose a cutoff of 0.005 of absolute difference of year-on-year difference of the count-zipcode ratio of residential units, which removes 280 rows of 16,577 zipcode-county-year observations for California.

Also, a few zipcodes can straddle several two or more counties. While, these can be managed by using the zipcode-county ratio of residential units (from HUD), this type of splitting makes it difficult to incorporate other information that are aggregated to the zipcode but cannot be disaggregated to specific county portions. Also, small errors in these residential ratios makes reconciling aggregates of zipcodes and counties difficult. So these zipcodes are dropped and added to a county-specific residual. The exact filtering was if 95% of the zipcode's residential units was not within one county, that zipcode was dropped from the analysis. For California, this dropped 54 zipcodes entirely out of 1486 zipcodes.

When the IRS reports information at zipcode level too small zips are suppressed. A county comprised of many such zips will have little coverage of their zip information which affects the quality of all other zips remaining in the county. The county's coverage will also be worsened by the previous two quality checks that removed rapid year-on-year zip boundary changes and removed zips that are distributed across several counties. For each county, a residual of the county-zip residential ratio was calculated after all these changes, and if this residual represented over 50% of a county-year's coverage, these county-years were removed. For California, this dropped 61 county-years with the worst being for Sierra, Mono, Lake, Glenn counties.

County totals and zip totals will sometimes not reconcile for several reasons. One is that IRS removes zipcodes that have too few taxpayers. As mentioned, I also remove zipcodes that straddle several counties which further reduces the coverage. So the difference between X_{cbt} and $\sum_{z \in c} X_{zbt}$ was positive 92% of the time and never zero. But for some county years one or more zipcode brackets sums exceeded that of the county bracket sum. This occurs because of IRS possibly undercounting returns in counties. This undercounting mostly affected the bracket number 5 of those earning between 100,000 and 200,000, but there were also few cases of such discrepancies in the totals X_{ct} and X_{zt}. Since such a negative discrepancy can affect the sampling procedure, for each county a discrepancy row z^* was added and any negatives in this row were zeroed out by rolling them into the next higher bracket so as to keep the X_{z^*t} in the row constant before the rolling in. If rolling up to a higher bracket left that bracket negative, then it was zeroed out. This would result in the county total X_{ct} somewhat higher than what is officially recorded for these counties. This only affected small counties such as Alpine, so any reconciled discrepancy is small.

A.3 Sampling algorithm

- Calculate zip-bracket mean aggregate income $\mu_{zbt} = \frac{G_{zbt}}{X_{zbt}}$.
- Calculate the directional mover means and the stayer means at the county level,

$$\bar{y}^I_{ct} = \frac{G^I_{ct}}{I_{ct}}$$
$$\bar{y}^O_{ct} = \frac{G^I_{ct}}{O_{ct}}$$
$$\bar{y}^S_{ct} = \frac{G^S_{ct}}{S_{ct}}$$

- Calculate a balancing numnber

C.1 For all $X_{zbt} > 0$, calculate the zip-bracket mean aggregate income

$$\mu_{zbt} = \frac{G_{zbt}}{X_{zbt}}$$

If $X_{zbt} = 0$, $\mu_{zbt} = 0$.

C.2 Calculate the directional mover mean aggregate income.

$$\bar{y}^I_{ct} = \frac{G^I_{ct}}{I_{ct}}$$
$$\bar{y}^O_{ct} = \frac{G^I_{ct}}{O_{ct}}$$

C.3 Calculate the p^I_{zbt} (the model's probability that an incoming taxpayer to county c is in bracket b and is destined for zipcode z) and p^O_{zbt} (the model's probability that an outgoing taxpayer is in bracket b and is originating from zipcode z) using the sampling distribution,

$$p^d_{zbt} \propto q^d_{zbt} \exp(\lambda_{dct} m^d_{zbt})$$

where $d = \{I, O\}$ and which satisfy

$$\sum_{z \in c} \sum_{b} p^d_{zbt} = 1$$

C.3.1 q^d_{zbt} is the proportion of the county's prior year taxpayers in zip z and bracket b.

$$q^I_{zbt} = q^O_{zbt} = \frac{X_{zbt-1}}{X_{ct-1}}$$

where

$$\sum_{z \in c} \sum_{b} q^d_{zbt} = 1$$

This is a baseline probability.

C.3.2 m^d_{zbt} is the model's expected mover income for zipcode z and bracket b. For incoming taxpayers this is μ_{zbt}. For outgoing taxpayers this is μ_{zbt-1}.

C.3.3 λ_{dct} is the balancing number that makes the model's expected mover income match the observed mover income at the county level for both incoming and outgoing taxpayers. A $\lambda_I > 0$ means that the incoming taxpayers are tilted towards higher income cells and conversely for $\lambda_I < 0$. See Appendix A for the optimization algorithm to calculate this multiplier.

C.4 Draw inbound and outbound zip-bracket counts. For inbound counts draw

$$\{I^{E(s)}_{zbt}\}_{z,b} \sim \text{Multinomial}(I_{ct}, \{p^I_{zbt}\}_{z,b})$$

in such a way that the draw s across brackets add to I_{ct}

$$\sum_{z,b} I^{E(s)}_{zbt} = I_{ct}$$

Calculate its implied aggregate income.

$$\hat{G}^{I(s)}_{ct} = \sum_{z,b} m^I_{zbt} I^{E(s)}_{zbt}$$

then retain draws that are

$$|\hat{G}^{I(s)}_{ct} - G^I_{ct}| \leq \epsilon^I_{ct}$$

C.5 Then for each retained draw s calculate directional totals and total net movement.

The total for incoming taxpayers is

$$I_{zt}^{E(s)} = \sum_b I_{zbt}^{E(s)}.$$

The total for outgoing taxpayers is

$$O_{zt}^{E(s)} = \sum_b O_{zbt}^{E(s)}.$$

Total external net movement is

$$M_{zt}^{E(s)} = I_{zt}^{E(s)} - O_{zt}^{E(s)}$$

C.6 Out of these draws, calculate the quantiles of net movement.

A.4 Additional results

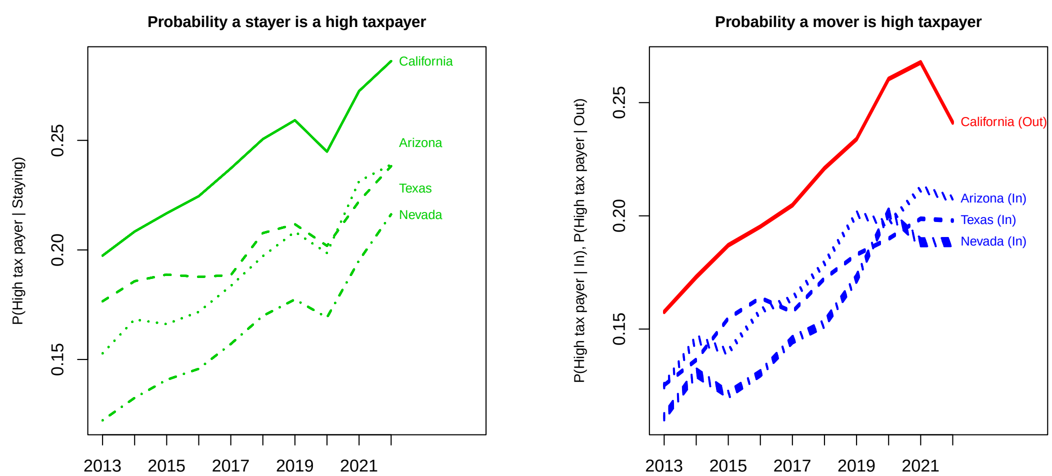


Figure A.9: High income taxpayer movements to and from the four states Arizona, California, Nevada, and Texas

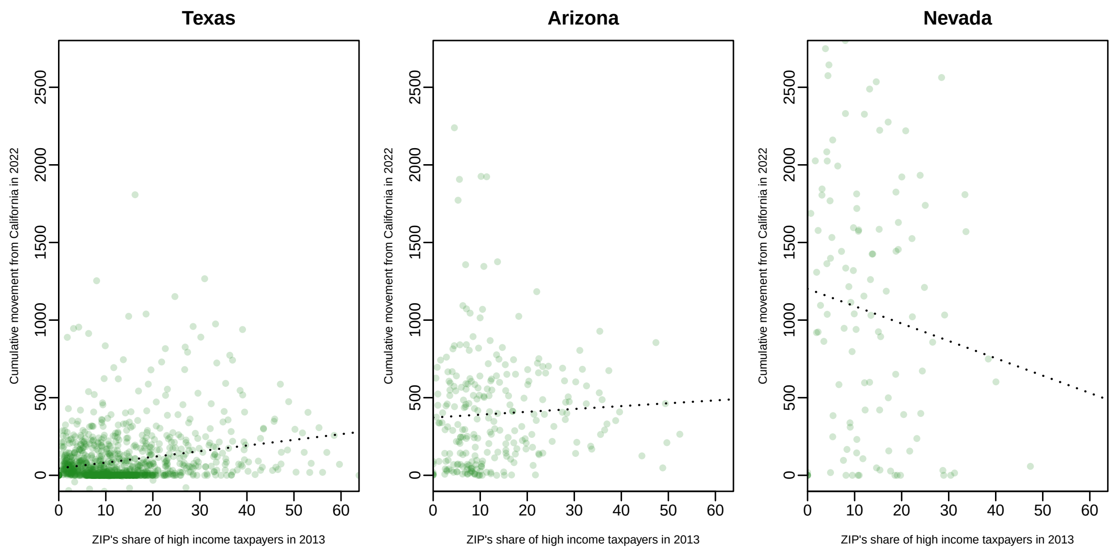


Figure A.10: The cumulative net migration volume to the zipcode from 2013 to 2022 against the high income share of the zipcode in 2013

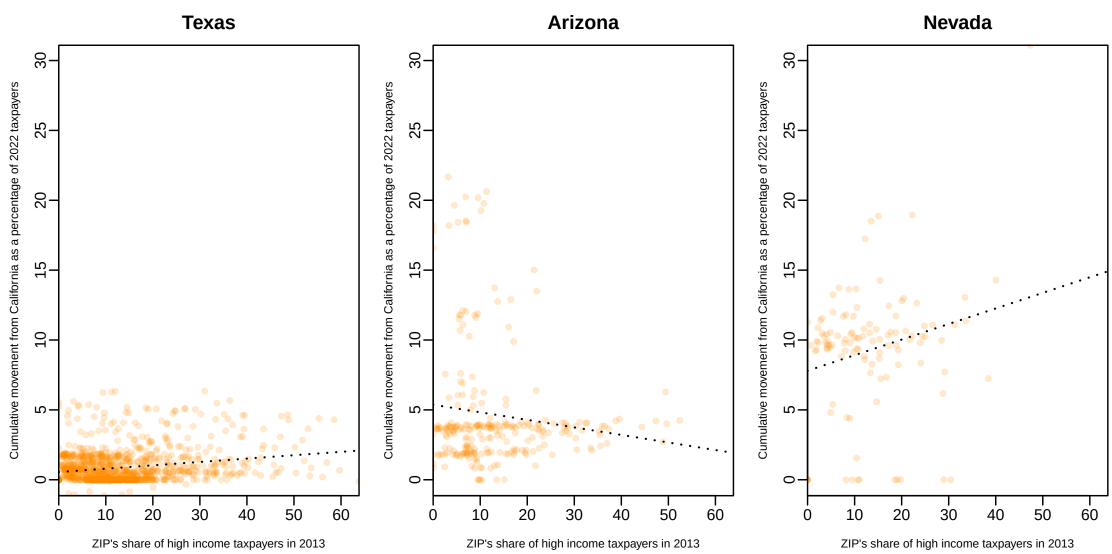


Figure A.11: The cumulative net migration as a percentage of the zipcode's nonmigrants in 2022 against the high income share of the zipcode in 2013

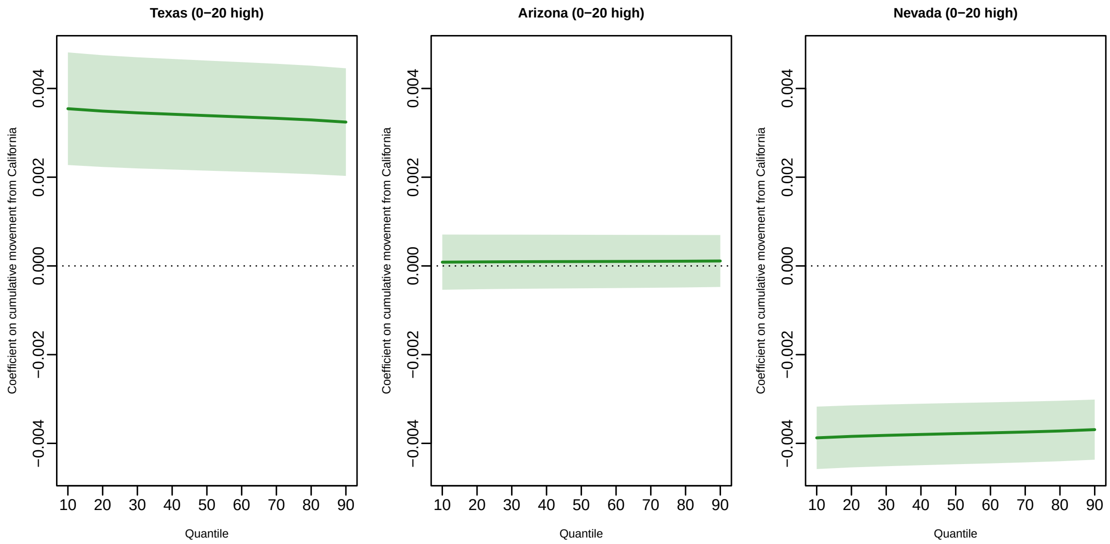


Figure A.12: The effect of cumulative Californian migration on low income zipcode's high income share for each quantile of $M_{zt}^{cum,E,CA}$ (zipcodes whose 2013 high income share is between 0 and 20 percent)

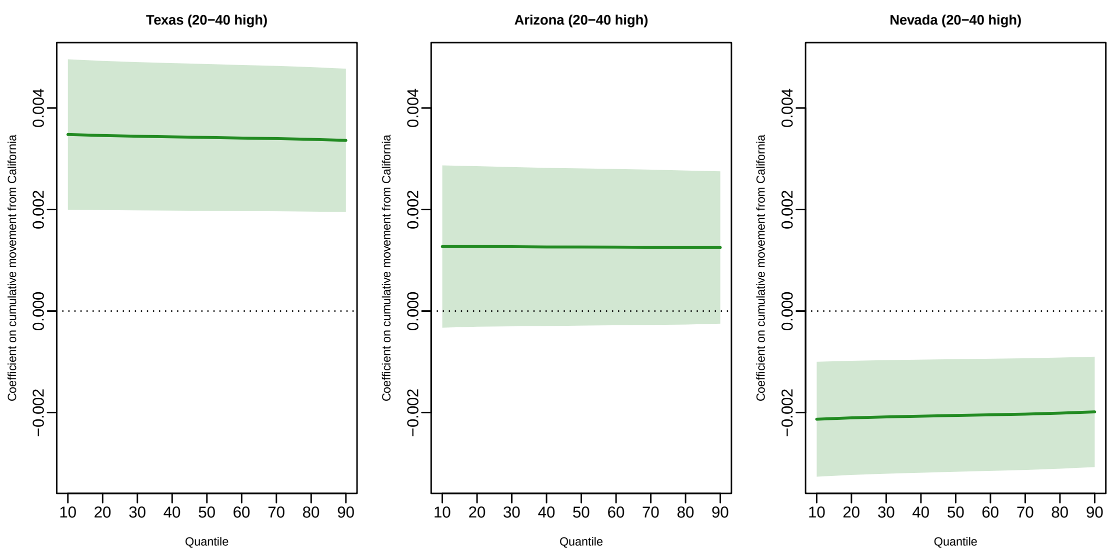


Figure A.13: The effect of cumulative Californian migration on high income zipcode's high income share for each quantile of $M_{zt}^{cum,E,CA}$ (zipcodes whose 2013 high income share is between 20 and 40 percent)

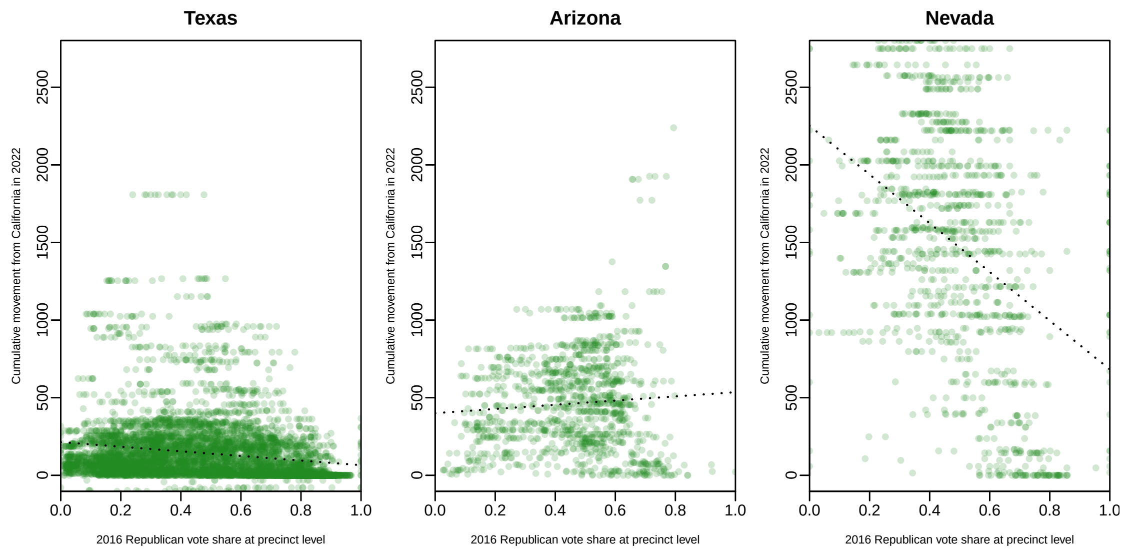


Figure A.14: The cumulative net migration volume to the zipcode from 2013 to 2022 against the Republican Presidential vote share in 2016

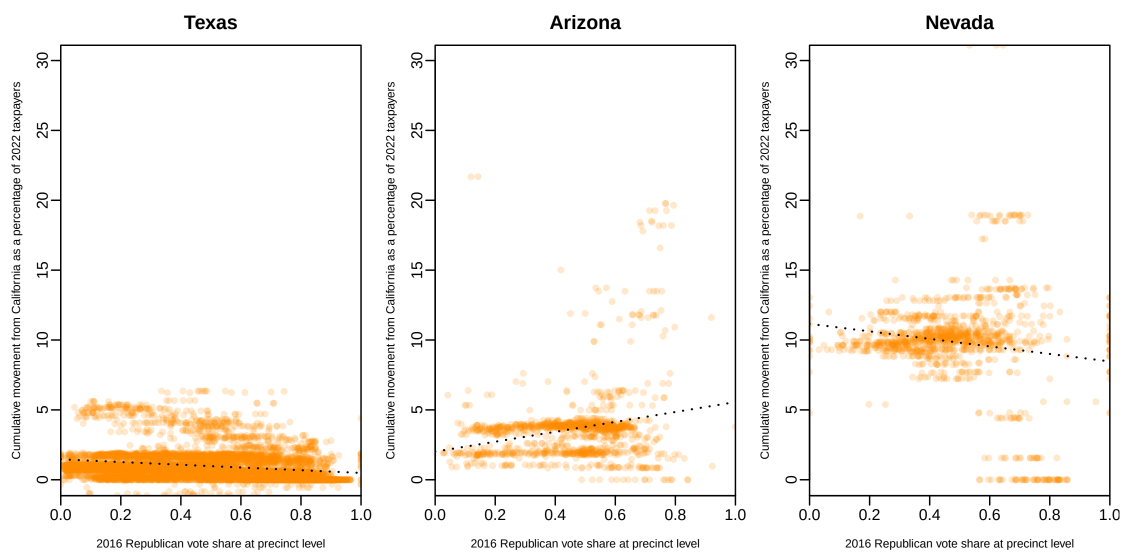


Figure A.15: The cumulative net migration as a percentage of the zipcode's nonmigrants in 2022 against the the Republican Presidential vote share in 2016

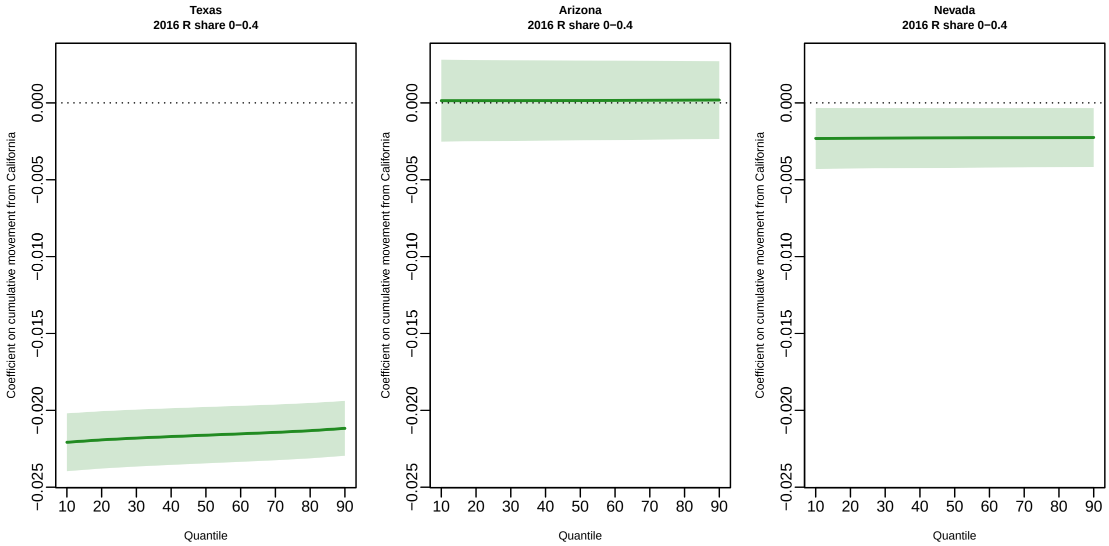


Figure A.16: The effect of precinct zipcode's cumulative Californian migration on precinct's Republican vote share for each quantile of $M_{zt}^{cum,E,CA}$ for precincts voting below 40 percent Republican in 2016

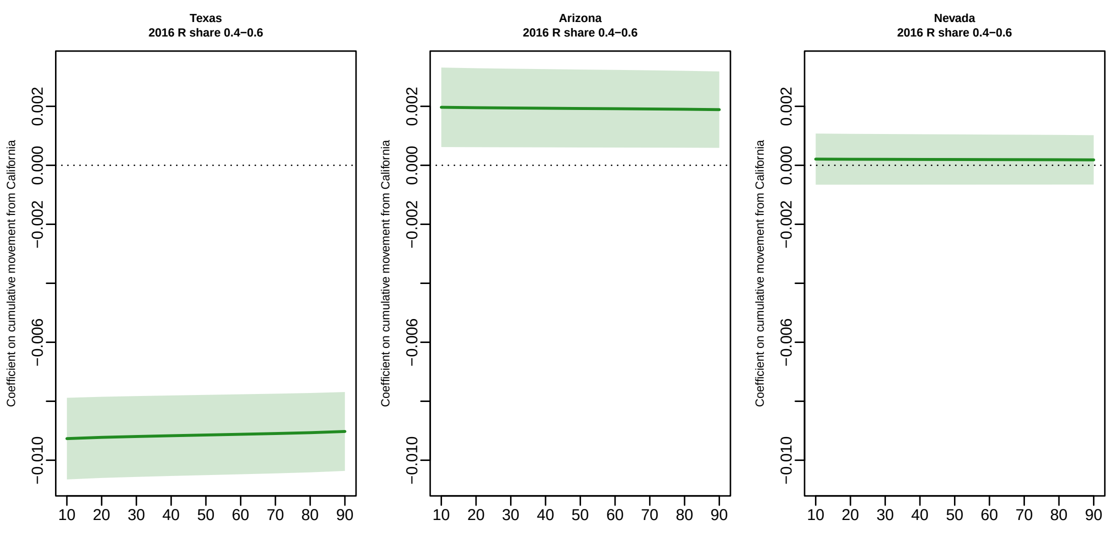


Figure A.17: The effect of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $M_{zt}^{cum,E,CA}$ for precincts voting between 40 percent and 60 percent Republican in 2016

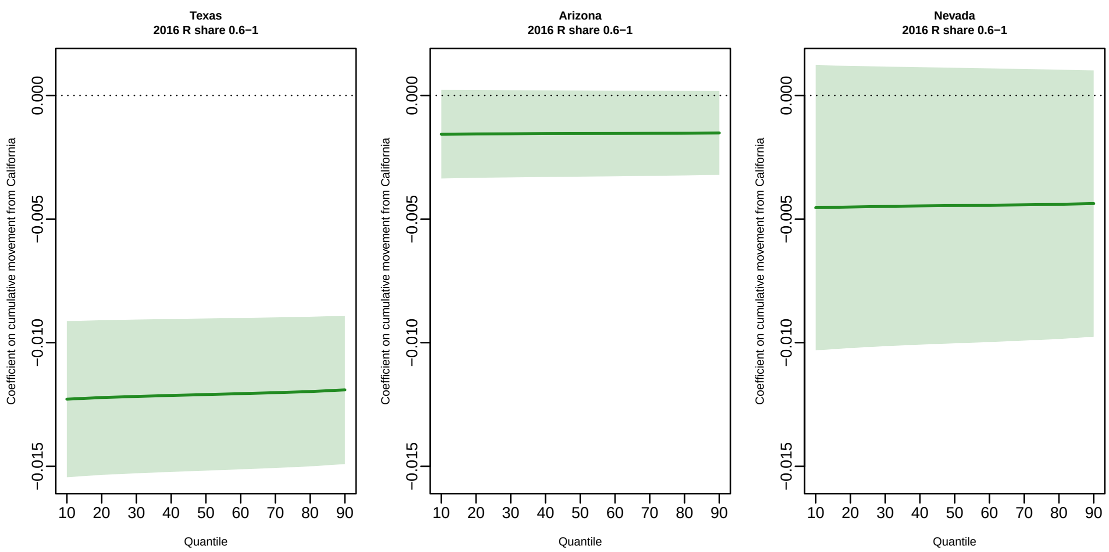


Figure A.18: The effect of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $M_{zt}^{cum,E,CA}$ for precincts voting above 60 percent Republican in 2016

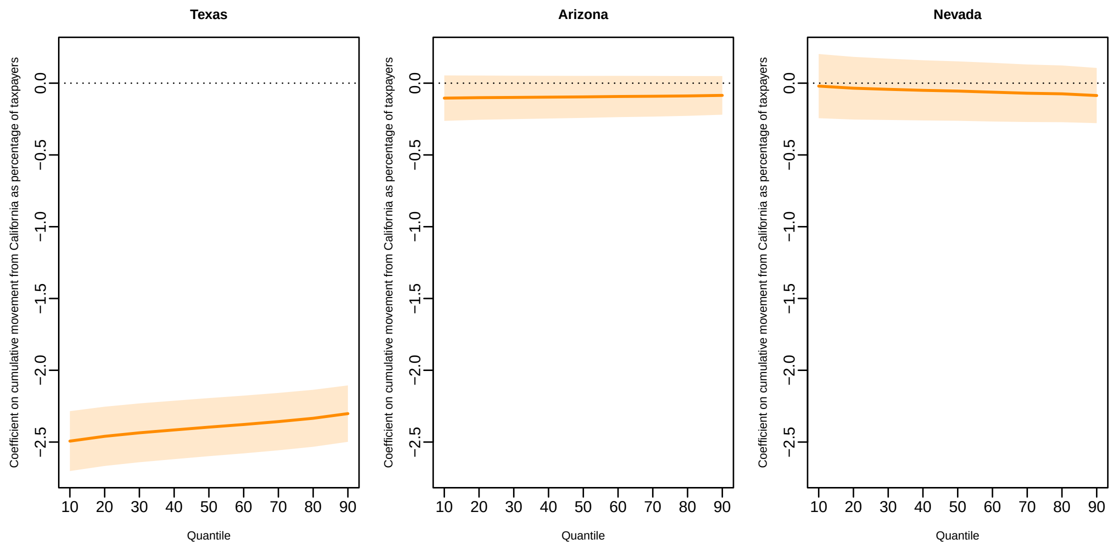


Figure A.19: The effect of rate of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $rM_{zt}^{cum,E,CA}$

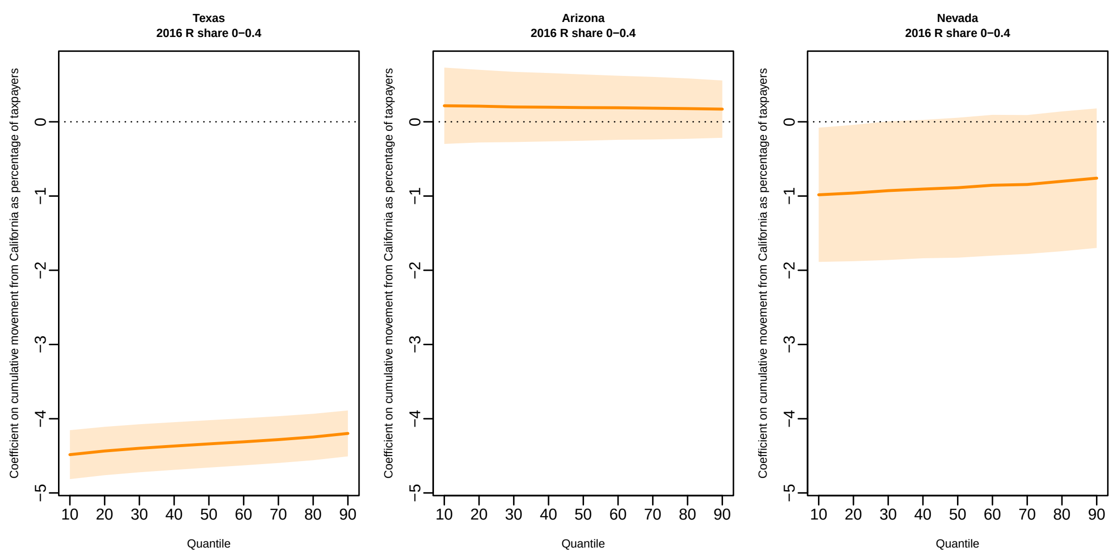


Figure A.20: The effect of rate of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $rM_{zt}^{cum,E,CA}$ for precincts voting below 40 percent Republican

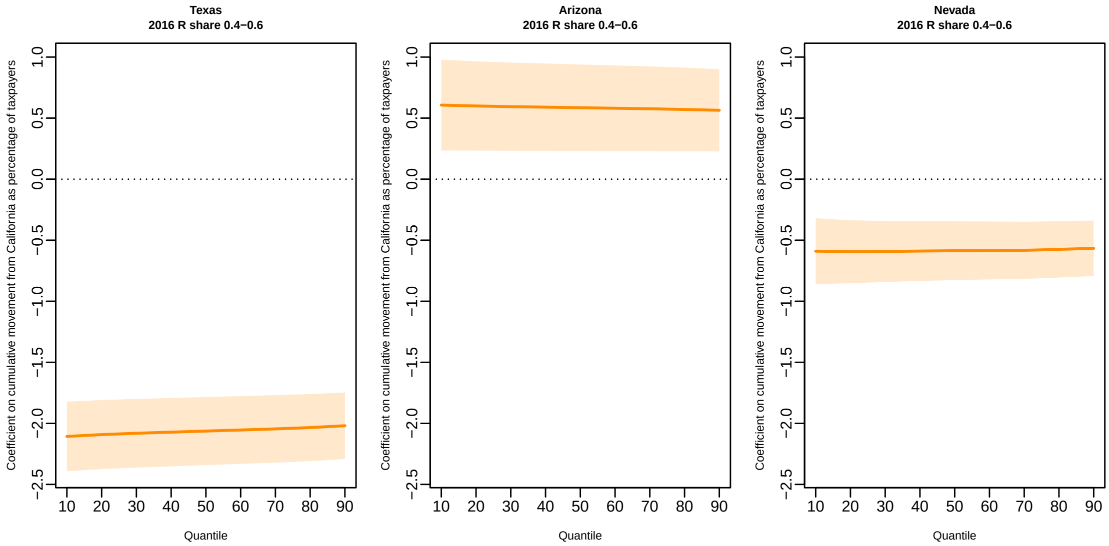


Figure A.21: The effect of rate of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $rM_{zt}^{cum,E,CA}$ for precincts voting between 40 percent Republican and 60 percent Republican

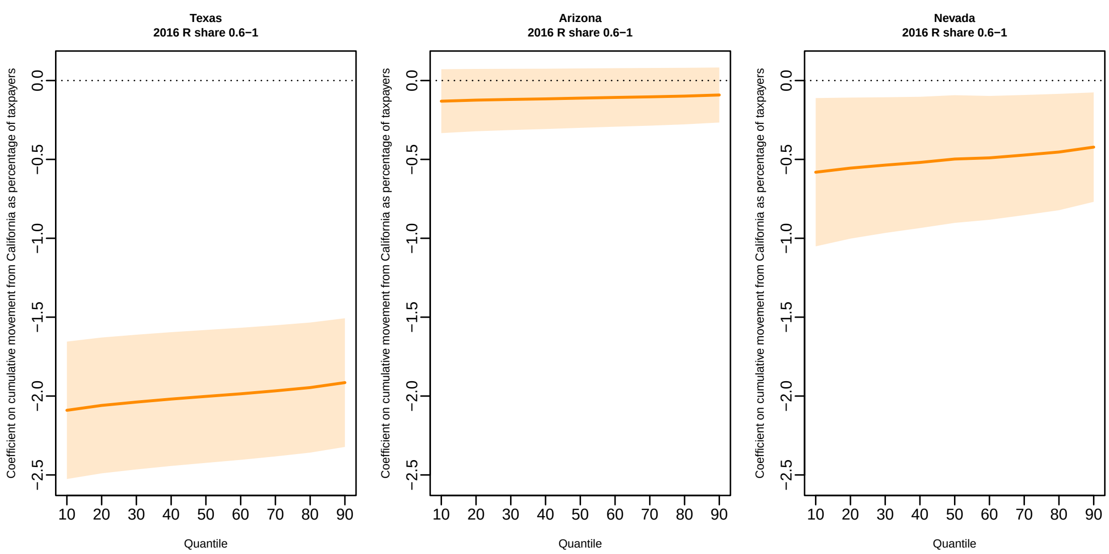


Figure A.22: The effect of rate of precinct zipcode's cumulative Californian migration on precinct's Democratic vote share for each quantile of $rM_{zt}^{cum,E,CA}$ for precincts voting above 60 percent Republican