

Tax Borders and the Rich: A Boundary Discontinuity Design Using IRS Income Data

APEP Autonomous Research* @ai1scl

March 5, 2026

Abstract

Do high-income households sort across state borders in response to income tax differentials? I use IRS Statistics of Income ZIP-code-level data for 2012–2021 to examine income composition at eight US state borders separating high-tax from low-tax states. The cross-sectional boundary discontinuity is large but reflects economic geography, not taxes: a placebo on low-income filers reveals a massive reverse discontinuity, covariate balance fails, and the sign flips with bandwidth. A triple-difference specification exploiting the 2017 SALT deduction cap—which differentially raised the cost of high-tax states for affluent filers—yields a point estimate of -0.6 pp ($p < 0.001$ with ZIP clustering; $p = 0.27$ with border-pair clustering). A DDD event study shows roughly flat pre-trends for 2013–2015, with post-2018 effects emerging primarily in the COVID period. These findings place a modest upper bound on tax-driven sorting at state borders.

JEL Codes: H71, H73, R23, J61

Keywords: state income tax, geographic mobility, boundary discontinuity, elite sorting, SALT deduction cap

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

In 2018, New Jersey raised its top income tax rate to 10.75% on incomes above \$1 million. Across the Delaware River, Pennsylvania’s flat rate remained 3.07%. For a household earning \$2 million, the annual tax savings from living on the Pennsylvania side of the border exceeded \$150,000—enough to buy a house every few years. Did the rich respond?

The question of whether taxes drive geographic sorting of high-income households has occupied public finance economists for decades. State and local policymakers routinely invoke “millionaire flight” to oppose tax increases, while academic evidence remains contested. [Young et al. \(2016\)](#) find that only 2.2% of US millionaires change states in a given year, and tax-motivated migration accounts for a small fraction of these moves. [Moretti and Wilson \(2017\)](#) document that star scientists respond to state tax differentials, but their estimates imply modest aggregate revenue consequences. At the same time, [Kleven et al. \(2014\)](#) show large mobility responses among top earners in Denmark, raising the possibility that US estimates are attenuated by measurement or institutional frictions.

This paper brings a new empirical approach to the question. Rather than tracking individual movers—which requires panel data and faces selection concerns—I examine the *stock* of high-income households at state borders. The key insight is that adjacent ZIP codes on opposite sides of a high-tax/low-tax state border share nearly identical labor markets, amenities, and housing stock, but face dramatically different tax burdens. Any discontinuity in the concentration of wealthy households at the border provides direct evidence of tax-driven geographic sorting.

I implement a multi-border boundary discontinuity design using IRS Statistics of Income (SOI) ZIP-code-level data from 2012 to 2021. For each ZIP code within 50 kilometers of a state border, I observe the number and aggregate income of tax returns by income bracket, including the crucial $\geq \$200\text{K}$ category. I construct eight border-pair segments separating high-tax states (California, New Jersey, New York, Minnesota, Oregon) from adjacent low-tax or zero-tax states (Nevada, Pennsylvania, Connecticut, South Dakota, Washington, Wisconsin, Arizona). For each segment, I compute signed distance from the ZIP code centroid to the border and estimate the discontinuity in high-income filer concentration at the boundary.

Three main findings emerge. First, the concentration of high-income filers varies sharply at state borders, but the direction depends on the bandwidth. The pooled nonparametric RDD estimate—which captures the very local discontinuity within 3.3 km of the border—is 8.65 percentage points ($p < 0.001$). However, the parametric specification with border-pair \times year fixed effects yields a *positive* coefficient on the high-tax indicator at 30 km (1.75 pp, $p < 0.05$), meaning the high-tax side has slightly *more* affluent filers at broader distances.

This sign reversal across bandwidths is a first clue that the border captures complex economic geography, not just tax-driven sorting.

Second, the 2017 Tax Cuts and Jobs Act (TCJA) provides a natural experiment within this design. The SALT deduction cap at \$10,000 effectively doubled the cost of high state taxes for affluent households by eliminating the federal offset. A triple-difference specification—interacting the border discontinuity with an indicator for high-income returns and a post-SALT indicator—yields a point estimate of -0.0064 ($p < 0.001$ with ZIP clustering, but $p = 0.27$ with border-pair clustering). A DDD event study shows roughly flat pre-trends for 2013–2015 but effects concentrated in 2020–2021, complicating the attribution to the SALT cap versus COVID-induced remote work. The result is suggestive of a modest tax-driven sorting channel, but not definitive.

Third, and critically, I document that this discontinuity is not exclusively tax-driven. A placebo test using low-income filers ($< \$50\text{K}$ AGI), who face minimal cross-border tax differentials, also reveals a significant discontinuity. Total population (proxied by total returns) is significantly imbalanced at the border, and the RDD estimate is sensitive to bandwidth choice—negative at small bandwidths (5–10 km) and positive at larger ones. These findings indicate that pre-existing differences in economic development, urbanization, and housing markets across state lines contribute substantially to the observed sorting pattern.

This paper contributes to the literature in three ways. First, it introduces ZIP-code-level IRS income data as a tool for spatial tax competition research. Prior work has relied on individual-level panel data with limited geographic precision (Young et al., 2016; Kleven et al., 2014), aggregate state flows (Cohen et al., 2011), or narrow populations like scientists (Moretti and Wilson, 2017) or athletes (Kleven et al., 2013). The ZIP-code boundary design provides a much larger sample (2,553 ZIP codes across 8 border pairs), granular income categories, and 10 years of panel variation.

Second, I provide the first quasi-experimental evidence on how the TCJA SALT deduction cap may have affected the geographic distribution of high-income households at state borders. The triple-difference design—which differences out both cross-border baseline differences and common time trends by income group—offers a sharper test than simple pre/post comparisons, though the evidence is qualified by imprecision under conservative clustering and timing that overlaps with the COVID pandemic.

Third, I contribute to the methodological literature on boundary discontinuity designs by demonstrating both their power and their limits in this setting. The significant placebo and covariate imbalance findings are not a failure of the design—they are the design working as intended, revealing that tax borders capture a bundle of state policy differences. This honest reporting of diagnostic failures bounds the magnitude of the pure tax effect and calibrates

expectations for what spatial designs can identify in settings where treatment is endogenous to geography (Keele and Titiunik, 2015).

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the institutional background. Section 4 presents the conceptual framework. Section 5 describes the data. Section 6 lays out the empirical strategy. Section 7 presents results. Section 8 discusses mechanisms and limitations. Section 9 concludes.

2. Related Literature

This paper sits at the intersection of three literatures: the mobility response to taxation, fiscal federalism and Tiebout sorting, and the methodology of geographic discontinuity designs.

2.1 Tax-Driven Mobility of the Affluent

The earliest systematic evidence on tax-driven migration comes from Feldstein (1999), who documented that the elasticity of taxable income with respect to marginal tax rates includes both real behavioral responses and reporting changes. The migration channel—physically relocating to avoid taxes—is one component of this elasticity, and potentially the most consequential for state fiscal policy.

Young et al. (2016) provide the most comprehensive US evidence using IRS panel data (individual-level returns linked across years). They find that only 2.2% of millionaires change states in any given year, and tax-motivated moves account for a small fraction of these. The implied tax elasticity of migration is roughly 0.1—positive and statistically significant, but economically small relative to the “millionaire flight” narrative. Their identification relies on comparing migration rates of millionaires in states that change tax rates versus those that do not, a difference-in-differences approach.

The international evidence suggests larger responses. Kleven et al. (2013) exploit tax variation across European football leagues and find that top players are highly responsive to net-of-tax rates when choosing where to play. Kleven et al. (2014) study Denmark’s preferential tax regime for foreign workers and estimate a migration elasticity of 1.5—far larger than domestic estimates. Akcigit et al. (2016) find that superstar inventors respond to international tax differentials, with migration elasticities around 1.

Moretti and Wilson (2017) bridge the domestic and international evidence by studying star scientists (defined by patent activity). Using state-level variation in top marginal rates, they estimate that a 1% increase in the net-of-tax rate increases the stock of star scientists by 1.6%. This is a large response, but the population is narrow. Whether similar elasticities apply to the

broader affluent population—doctors, lawyers, business owners, retired executives—remains an open question.

A key limitation of this literature is that all studies rely on individual-level panel data, which limits sample size and geographic precision. My approach differs fundamentally: rather than tracking individual movers, I examine the equilibrium *stock* of high-income households at state borders. This captures both migration and other forms of sorting (new household formation, differential mortality and aging, selective entry of high-income workers) that individual panels may miss.

2.2 Fiscal Federalism and Tiebout Sorting

The theoretical foundation for tax-driven sorting is [Tiebout \(1956\)](#), who argued that mobile households “vote with their feet” by choosing jurisdictions that offer their preferred bundle of taxes and public services. In the Tiebout framework, spatial sorting is efficient: it reveals preferences and disciplines governments. The empirical question is whether this mechanism operates at the state level, where mobility costs are higher and the tax-public service bundle is more complex than in the local property tax setting Tiebout envisioned.

[Bayer et al. \(2007\)](#) develop a structural model of residential sorting that accounts for both neighborhood amenities and fiscal incentives. Their estimates suggest that fiscal factors explain a modest but significant share of neighborhood choice, particularly for high-income households. [Cohen et al. \(2011\)](#) study domestic migration flows and find limited evidence that state tax changes drive interstate moves, though they do find effects on the composition of migrants (high-income households are more responsive than low-income ones).

The SALT deduction cap provides a particularly clean test of the Tiebout mechanism. Before 2018, the federal deduction of state taxes partially offset cross-state tax differentials. The cap removed this offset for high-income households, effectively increasing the “price” of high-tax states. If Tiebout sorting is operative, the cap should have amplified the geographic concentration of the affluent on the low-tax side of state borders—precisely what the triple-difference specification in this paper tests.

2.3 Geographic Boundary Discontinuity Designs

The boundary discontinuity design—comparing outcomes on opposite sides of an administrative boundary—was pioneered by [Black \(1999\)](#) in the context of school attendance zones. She showed that housing prices jump discontinuously at school district boundaries, identifying the capitalization value of school quality.

[Keele and Titiunik \(2015\)](#) provide a formal treatment of the identifying assumptions. They emphasize that geographic RDD requires not only that the boundary itself is exogenous, but that all confounders are continuous at the boundary. In practice, state borders often coincide with geographic features (rivers, mountain ranges) that independently affect settlement patterns. The authors argue for careful validation through placebo tests and balance checks—exactly the approach I implement here.

The application of boundary designs to tax competition is relatively new. Prior work has focused on housing prices ([Bayer et al., 2007](#)), school quality, and environmental regulation. This paper applies the framework to income composition, which requires different data (tax returns rather than housing transactions) and raises different identification challenges (the “treatment” of being in a high-tax state is endogenous to the history of state economic development, not randomly assigned at the border).

3. Institutional Background

3.1 State Income Tax Variation

The United States features substantial variation in state income tax rates. As of 2021, seven states impose no individual income tax (Alaska, Florida, Nevada, New Hampshire on earned income, South Dakota, Tennessee, Texas, Washington, Wyoming), while the top marginal rates in high-tax states exceed 10%: California (13.3%), New Jersey (10.75%), New York (10.9%), Oregon (9.9%), and Minnesota (9.85%).

For a household with \$500,000 in taxable income, the annual tax difference between California (approximately \$50,000) and Nevada (\$0) is substantial—roughly the median household income in the United States. This differential creates a strong financial incentive for geographic sorting, particularly for workers whose employment is not tied to a specific physical location.

Critically, many of these high-tax and low-tax states share borders, creating sharp spatial discontinuities in tax burdens. The New Jersey–Pennsylvania border runs through the Philadelphia metropolitan area, where households on either side commute to the same employers, shop in the same stores, and send children to comparable schools. The Oregon–Washington border bisects the Portland–Vancouver metropolitan area. The California–Nevada border runs through the Lake Tahoe and Reno regions.

3.2 The SALT Deduction Cap

Prior to 2018, taxpayers who itemized could deduct the full amount of state and local taxes (SALT) paid from their federal taxable income. For a household in the 37% federal bracket paying \$50,000 in California state taxes, this deduction was worth \$18,500 in federal tax savings—effectively reducing the California tax burden by more than one-third.

The Tax Cuts and Jobs Act of 2017, effective January 2018, capped the SALT deduction at \$10,000 per return. For high-income households in high-tax states, this change was dramatic. The same household now receives only \$3,700 in federal savings ($37\% \times \$10,000$), increasing the effective state tax burden by approximately \$14,800. The SALT cap thus roughly doubled the cost of living in a high-tax state for affluent households, while having virtually no effect on low-income filers (who typically do not itemize or have SALT deductions below \$10,000).

This asymmetric incidence makes the SALT cap a natural experiment for isolating tax-driven sorting: if any change in border discontinuities after 2018 is concentrated among high-income filers—who were differentially affected by the cap—while absent among low-income filers, the tax channel is implicated.

4. Conceptual Framework

Consider a household i with income y_i choosing between two locations, H (high-tax side) and L (low-tax side), separated by a state border. The household maximizes utility:

$$U_i(j) = (1 - \tau_j)y_i + a_j(d_i) - c_i(d_i), \quad j \in \{H, L\} \quad (1)$$

where τ_j is the state income tax rate, $a_j(d_i)$ captures location-specific amenities that vary with distance d_i from the border, and $c_i(d_i)$ is the cost of being far from the border (commuting, social ties, etc.).

The household chooses L over H if:

$$(\tau_H - \tau_L)y_i > a_H(d_i) - a_L(d_i) - [c_i^L(d_i) - c_i^H(d_i)] \quad (2)$$

At the border ($d_i \rightarrow 0$), amenities and commuting costs are approximately continuous: $a_H(0) \approx a_L(0)$ and $c_i^H(0) \approx c_i^L(0)$. The sorting incentive simplifies to $(\tau_H - \tau_L)y_i > 0$, which holds for all positive-income households but is largest for the rich.

This framework generates three testable predictions:

Prediction 1 (Descriptive): If taxes are the dominant sorting force, the concentration of high-income households should be higher on the low-tax side of the border. However, if

other state-level differences (urbanization, agglomeration, public services) also vary at the border, the observed discontinuity will reflect these forces as well.

Prediction 2: The SALT deduction cap, which increased the effective value of $(\tau_H - \tau_L)$ for high earners, should amplify this discontinuity post-2018, particularly for top-bracket filers.

Prediction 3: The discontinuity should be larger at borders with greater tax differentials (e.g., CA–NV with a 13.3 percentage point gap versus MN–WI with a 2.2 percentage point gap).

The key identifying assumption of the boundary design—that amenities are continuous at the border—is testable. If the discontinuity also appears for low-income households (who face minimal tax differentials), or if pre-determined covariates are imbalanced, then factors other than taxes are driving the observed sorting.

5. Data

5.1 IRS Statistics of Income: ZIP-Code Data

The primary data source is the IRS Statistics of Income (SOI) Individual Income Tax Statistics, ZIP Code Data, covering tax years 2012 through 2021. For each ZIP code and year, the SOI reports the number of returns (N1), number of exemptions (N2), and adjusted gross income (AGI) broken down by six income categories: under \$25,000; \$25,000–\$49,999; \$50,000–\$74,999; \$75,000–\$99,999; \$100,000–\$199,999; and \$200,000 or more.

The \$200,000+ category is the primary outcome of interest. For each ZIP code \times year, I compute the *high-income share*: the ratio of returns with $\text{AGI} \geq \$200,000$ to total returns. I also construct the *average AGI per high-income return* (total AGI in the top bracket divided by the number of returns) to capture the intensive margin of income composition.

The IRS suppresses cells with fewer than approximately 10 returns to protect taxpayer privacy. I flag suppressed ZIP codes and conduct sensitivity analyses excluding them.

5.2 Geographic Data

ZIP code centroid coordinates come from the 2020 Census Gazetteer files for ZIP Code Tabulation Areas (ZCTAs). State boundary geometries come from the 2020 TIGER/Line shapefiles, projected to the Conus Albers Equal Area (EPSG:5070) coordinate system for accurate distance computation.

For each of eight border-pair segments, I compute the signed Euclidean distance from each ZIP code centroid to the nearest shared state boundary. ZIP codes in the high-tax

state receive negative distance values; those in the low-tax state receive positive values. The running variable for the RDD is this signed distance, with the cutoff at zero.

5.3 Border Pairs and Tax Rates

I select eight border segments where the high-tax state's top marginal income tax rate exceeds the neighbor's by at least 2 percentage points. The rates below reflect 2021 values; several states adjusted rates during the sample period (e.g., NJ increased from 8.97% to 10.75% in 2018), and the empirical analysis uses year-varying rate differentials:

1. New Jersey (10.75%) vs. Pennsylvania (3.07%) — 2021 gap: 7.7pp
2. New York (10.9%) vs. Connecticut (6.99%) — 2021 gap: 3.9pp
3. California (13.3%) vs. Nevada (0%) — 2021 gap: 13.3pp
4. Minnesota (9.85%) vs. South Dakota (0%) — 2021 gap: 9.9pp
5. Oregon (9.9%) vs. Washington (0%) — 2021 gap: 9.9pp
6. Minnesota (9.85%) vs. Wisconsin (7.65%) — 2021 gap: 2.2pp
7. California (13.3%) vs. Arizona (4.5%) — 2021 gap: 8.8pp
8. New York (10.9%) vs. Pennsylvania (3.07%) — 2021 gap: 7.8pp

Tax rates come from the Tax Foundation's annual compilations of state individual income tax rates. I use the top marginal rate applicable to earned income in each state and year. Several states adjusted rates during the sample period: New Jersey increased its top rate from 8.97% to 10.75% in 2018 (affecting income above \$1 million), New York raised its rate from 8.82% to 10.9% in 2021, and Minnesota increased from 7.85% to 9.85% in 2013. These within-state rate changes provide additional time-series variation for identification.

5.4 Sample Construction

The analysis panel is constructed as follows. Starting from the universe of 29,968 unique ZIP codes in the IRS SOI data, I merge ZIP centroid coordinates from the Census Gazetteer (33,120 ZCTAs; 28,735 match). For each of the eight border segments, I assign all ZIP codes in the two relevant states and compute signed Euclidean distance from each centroid to the nearest point on the shared state boundary, using projected coordinates (EPSG:5070, Conus Albers Equal Area) to ensure accurate distance measurement in kilometers.

I restrict the sample to ZIP codes within 50 kilometers of a border. A single ZIP code can appear in multiple border segments if it lies near more than one relevant border (for example, a ZIP code in northeastern Pennsylvania may appear in both the NJ–PA and NY–PA segments). This yields 2,553 unique ZIP codes and 23,809 ZIP-code \times year observations across 10 tax years.

I apply two sample restrictions. First, I exclude ZIP codes with fewer than 10 total returns in any year, as these represent extremely small areas where the income distribution is likely dominated by noise. Second, I exclude observations where the high-income share is missing (typically due to the IRS suppression rule). These restrictions remove approximately 3% of observations.

5.5 Summary Statistics

Table 1: Summary Statistics: Border ZIP Codes

Variable	High-Tax Side	Low-Tax Side
ZIP-code \times year observations	14,137	9,672
Unique ZIP codes	1,519	1,034
Mean total returns per ZIP	7783	5643
Mean share AGI \geq \$200K	0.0644	0.0498
SD share AGI \geq \$200K	0.0836	0.0735
Mean high-income returns per ZIP	536.7	341.4
Mean share AGI $<$ \$50K	0.5351	0.5562
Mean AGI per high-income return (\$K)	448.4	445.2
Mean distance to border (km)	25.8	22.3
Share suppressed ZIPs	0.227	0.278

Notes: Sample includes ZIP codes within 50km of a high-tax/low-tax state border, 2012–2021. High-income defined as AGI \geq \$200K (IRS SOI `agi_stub = 6`). Suppressed ZIPs are those where the IRS withholds income-bracket counts for privacy.

Table 1 presents summary statistics separately for ZIP codes on the high-tax and low-tax sides of state borders, within 50 kilometers of the boundary. The sample comprises 23,809 ZIP-code \times year observations across 2,553 unique ZIP codes. The mean share of high-income returns is 6.4% on the high-tax side and 5.0% on the low-tax side—a raw difference of 1.4 percentage points that foreshadows the RDD findings.

Notably, the high-tax side has larger ZIP codes (7,783 mean total returns vs. 5,643), reflecting the higher population density of states like New Jersey, New York, and California. This size imbalance is important: it means the high-tax side contributes more total high-income returns even if the *share* is similar, and it creates a potential confound if ZIP code size correlates with affluence through urbanization channels.

The share of returns with AGI below \$50,000 is 53.5% on the high-tax side and 55.6% on the low-tax side. Average AGI per high-income return is similar on both sides (\$448K vs. \$445K), suggesting that the *composition* of the wealthy population, conditional on crossing the \$200K threshold, is comparable. This means the border discontinuity operates primarily on the extensive margin (how many high-income filers) rather than the intensive margin (how wealthy the high-income filers are).

The mean distance to the border is 25.8 km on the high-tax side and 22.3 km on the low-tax side. Suppression rates are substantial on both sides (22.7% vs. 27.8%), reflecting the prevalence of small rural ZIP codes near state borders where the IRS suppresses bracket-level counts. I conduct sensitivity analyses excluding suppressed observations to verify robustness.

6. Empirical Strategy

6.1 Boundary Discontinuity Design

The primary identification strategy is a geographic boundary discontinuity design (Black, 1999; Bayer et al., 2007; Keele and Titiunik, 2015). For each border segment b , define d_{zb} as the signed distance from ZIP code z 's centroid to the border, with $d_{zb} < 0$ on the high-tax side and $d_{zb} > 0$ on the low-tax side.

The identifying assumption is that potential outcomes are continuous at the border:

$$\lim_{d \downarrow 0} \mathbb{E}[Y_z(0) | d_{zb} = d] = \lim_{d \uparrow 0} \mathbb{E}[Y_z(0) | d_{zb} = d] \quad (3)$$

where $Y_z(0)$ is the counterfactual high-income share absent the tax differential. If amenities, labor markets, and housing are smooth across the border, any discontinuity in the outcome at $d = 0$ identifies the causal effect of the tax differential on high-income sorting.

6.2 Estimation

I estimate two complementary specifications.

Nonparametric RDD. I use the Calonico et al. (2014) robust bias-corrected estimator (`rdrobust`), with a triangular kernel and MSE-optimal bandwidth selection. Standard errors are clustered at the ZIP code level. This approach is agnostic about functional form and provides a local estimate at the border.

Parametric RDD with fixed effects. I estimate:

$$Y_{zbt} = \alpha + \tau \cdot H_{zb} + f(d_{zb}) + H_{zb} \cdot f(d_{zb}) + \gamma_{bt} + \varepsilon_{zbt} \quad (4)$$

where Y_{zbt} is the high-income share in ZIP z , border segment b , year t ; $H_{zb} = \mathbb{I}[d_{zb} < 0]$ indicates the high-tax side; $f(\cdot)$ is a polynomial in distance; and γ_{bt} are border-segment \times year fixed effects. The coefficient τ captures the average discontinuity at the border, netting out segment-specific trends.

Event study. To test whether the SALT cap amplified sorting, I interact the high-tax indicator with year dummies (base: 2017):

$$Y_{zbt} = \sum_{k \neq 2017} \beta_k \cdot \mathbb{I}[t = k] \cdot H_{zb} + f(d_{zb}) \cdot H_{zb} + \gamma_{bt} + \varepsilon_{zbt} \quad (5)$$

Flat pre-2018 coefficients and a post-2018 shift would support the tax interpretation.

Triple-difference. To sharpen the SALT identification, I stack high-income and low-income shares and estimate:

$$Y_{zgbt} = \delta \cdot H_{zb} \times \text{HighIncome}_g \times \text{PostSALT}_t + \text{controls} + \varepsilon_{zgbt} \quad (6)$$

where $g \in \{\text{High}, \text{Low}\}$ indexes income group and $\text{PostSALT} = \mathbb{I}[t \geq 2018]$. The coefficient δ isolates the SALT-specific effect on high-income sorting at the border, differencing out common border effects and common income-group trends.

6.3 Threats to Validity

Three threats merit discussion.

Pre-existing economic differences. State borders often coincide with differences in economic development that predate any tax differential. Philadelphia’s dense, wealthy suburbs are overwhelmingly in New Jersey, while Pennsylvania’s side includes less affluent exurban areas. This concern motivates the placebo tests, covariate balance checks, and the SALT event study (which identifies within-border changes rather than cross-sectional levels).

IRS data suppression. The IRS suppresses income-bracket counts in small ZIP codes, creating a form of measurement error. I verify that suppression rates are balanced across border sides and show that results are robust to excluding suppressed observations.

Density manipulation. If wealthy households have historically sorted to the low-tax side, more ZIP codes may exist there—a form of endogenous “manipulation” of the running variable. I conduct the [McCrory \(2008\)](#) density test and report results.

SALT exposure heterogeneity. The triple-difference uses $\text{AGI} \geq \$200\text{K}$ as the “treated” group, but SALT cap incidence depends on itemization rates, property taxes, and household structure. Many high-income filers may not be constrained (if they don’t itemize or have low SALT), while some below $\$200\text{K}$ in high-property-tax areas are affected. This coarseness

introduces measurement error in the treatment intensity and biases the DDD toward zero. A stronger design would exploit continuous exposure using pre-2018 SALT deduction amounts or property tax proxies, but the IRS SOI data does not report itemized deductions at the ZIP level.

7. Results

7.1 Descriptive Boundary Discontinuity

I first document the cross-sectional pattern at state borders. Given the placebo and covariate balance failures reported below, this analysis should be read as *descriptive geography*—characterizing how income composition varies at tax borders—rather than as a causal estimate of the tax effect on sorting.

Figure 1 presents the visual evidence. Each point is a 2-kilometer bin average of the high-income share, plotted against signed distance to the state border (negative = high-tax side, positive = low-tax side). The lines are fitted values from separate linear regressions on each side.

Geographic Discontinuity in High-Income Filer Density at Tax Borders

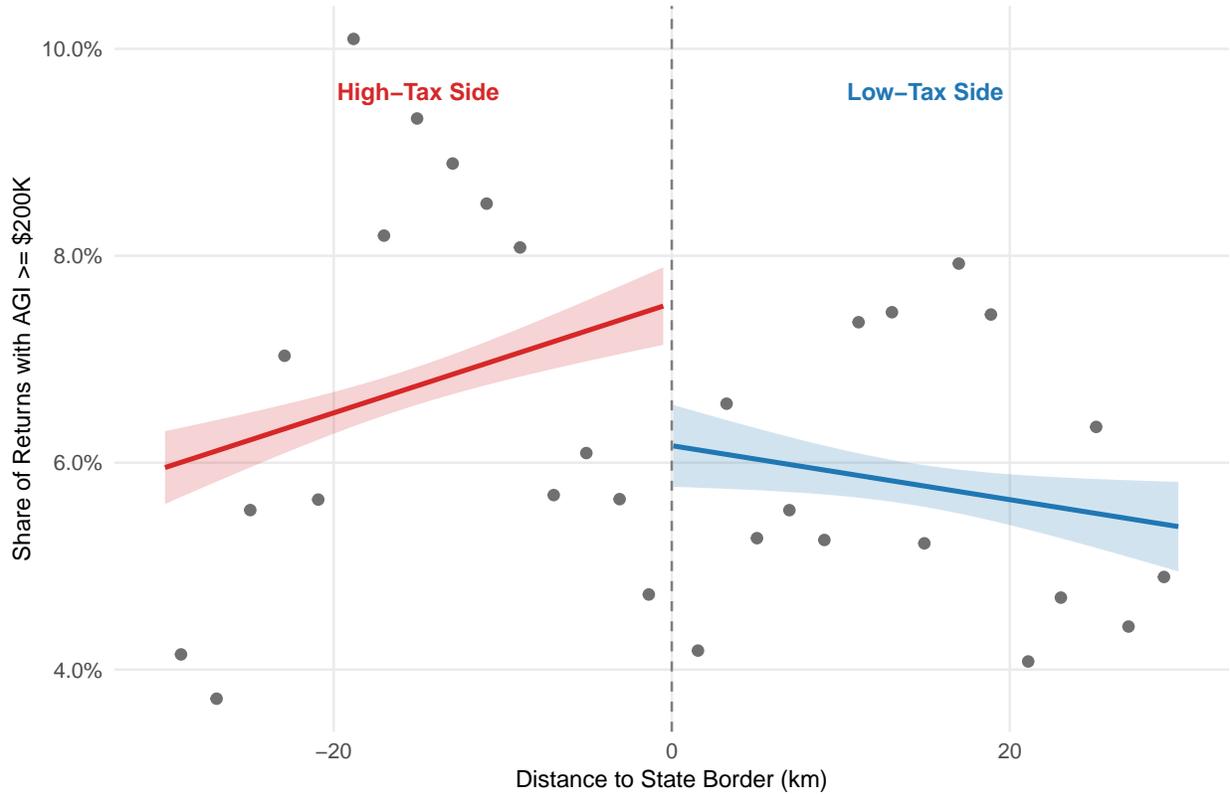


Figure 1: Boundary Discontinuity in High-Income Filer Concentration

Notes: Each point is a 2km-bin average of the share of tax returns with $AGI \geq \$200K$. The running variable is signed distance to the state border (negative = high-tax side). Lines are linear fits estimated separately on each side. Sample: ZIP codes within 30km of eight high-tax/low-tax state borders, 2012–2021.

Table 2 presents formal estimates. The nonparametric RDD (`rdrobust`) estimates the difference in outcomes from the left (high-tax) to the right (low-tax) side of the cutoff. The point estimate is 0.0865 (8.65 pp; robust SE = 0.0101, $p < 0.001$) at the MSE-optimal bandwidth of 3.3 km, with an effective sample of 611/553 observations (left/right of cutoff). The nonparametric placebo (low-income share) uses a narrower 1.9 km optimal bandwidth with correspondingly fewer effective observations (234/192)—`rdrobust` selects bandwidth independently for each outcome based on its own variance structure. Within 3.3 km of the border, the low-tax side has substantially more high-income filers.

The parametric specifications tell a different story. Here the coefficient of interest is on the *high-tax-side indicator*, so a positive coefficient means the high-tax side has *more* high-income filers. At a 50-km bandwidth, the estimate is 0.004 (SE = 0.006, not significant). At 30 km—from the bandwidth sensitivity analysis—the estimate is 0.0175 (1.75 pp, $p < 0.05$),

meaning the high-tax side has slightly *more* high-income filers at this broader distance. At the 3-km optimal bandwidth, the parametric estimate is -0.068 (SE = 0.039), consistent in direction with the nonparametric result but imprecise.

This sign reversal is informative. Very close to the border (within 5–10 km), the low-tax side has more affluent households. But at wider bandwidths that include suburban and urban areas farther from the border, the high-tax states—which include New Jersey, New York, and California with their agglomeration economies—dominate. The bandwidth sensitivity is a central finding, not a nuisance.

Table 2: Boundary Discontinuity Estimates: Effect of High-Tax State on Income Composition

Specification	Outcome	Estimate	SE	Bandwidth	N
Nonparametric (rdrobust)	High-income share	0.0865	0.0101	3.3km	611/553
Nonparametric placebo	Low-income share	-0.2731	0.0438	1.9km	234/192
Linear (50km)	High-income share	0.0043	0.0061	50km	23,809
Quadratic (50km)	High-income share	0.0162	0.0095	50km	23,809
Linear (3km)	High-income share	-0.0685	0.0391	3km	1,160
Linear (30km)	High-income share	0.0175	0.0077	30km	14,673
Placebo: Low-inc	Low-income share	-0.0210	0.0092	50km	23,809
Placebo: Mid-inc	Mid-income share	-0.0027	0.0035	50km	23,809
Log high returns	Log high-inc returns	0.0156	0.1747	50km	23,809

Notes: The dependent variable is the share of tax returns in a ZIP code with AGI \geq \$200K (or AGI $<$ \$50K for placebo). Nonparametric rows use rdrobust with MSE-optimal bandwidth; N reports effective observations within the bandwidth (left/right of cutoff). Parametric specifications include border-pair \times year fixed effects and cluster standard errors at the ZIP code level. The running variable is signed distance to the state border (negative = high-tax side). Note that the nonparametric estimator (right minus left = low-tax minus high-tax) and parametric estimator (coefficient on high-tax indicator) have opposite sign conventions. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7.2 Placebo: Low-Income Filers

Figure 2 replicates the RDD plot for returns with AGI below \$50,000. The nonparametric estimate (low-tax minus high-tax at the border, using rdrobust’s 1.9 km optimal bandwidth) is -0.273 (-27.3 pp; robust SE = 0.044, $p < 0.001$), indicating that low-income filers are far more concentrated on the *high-tax* side very near the border. The parametric placebo (Table 2, “Placebo: Low-inc” row using the full 50 km sample) yields a coefficient of -0.021 on the high-tax indicator (SE = 0.009)—a much smaller magnitude because it averages over a wider geographic area where the local border contrast is diluted, but the sign confirms the same directional pattern. The large magnitude of the placebo (27.3 pp within 1.9 km, based on effective $N = 234/192$) reflects stark urban-suburban contrasts at state borders;

for instance, high-density New Jersey ZIP codes near the Delaware River border have very different economic composition from adjacent Pennsylvania ZIP codes. Since low-income households face minimal cross-border tax differentials (low marginal rates, standard deduction exceeds SALT), this finding demonstrates that the border captures economic differences unrelated to income taxation—urbanization, housing density, and employment composition.

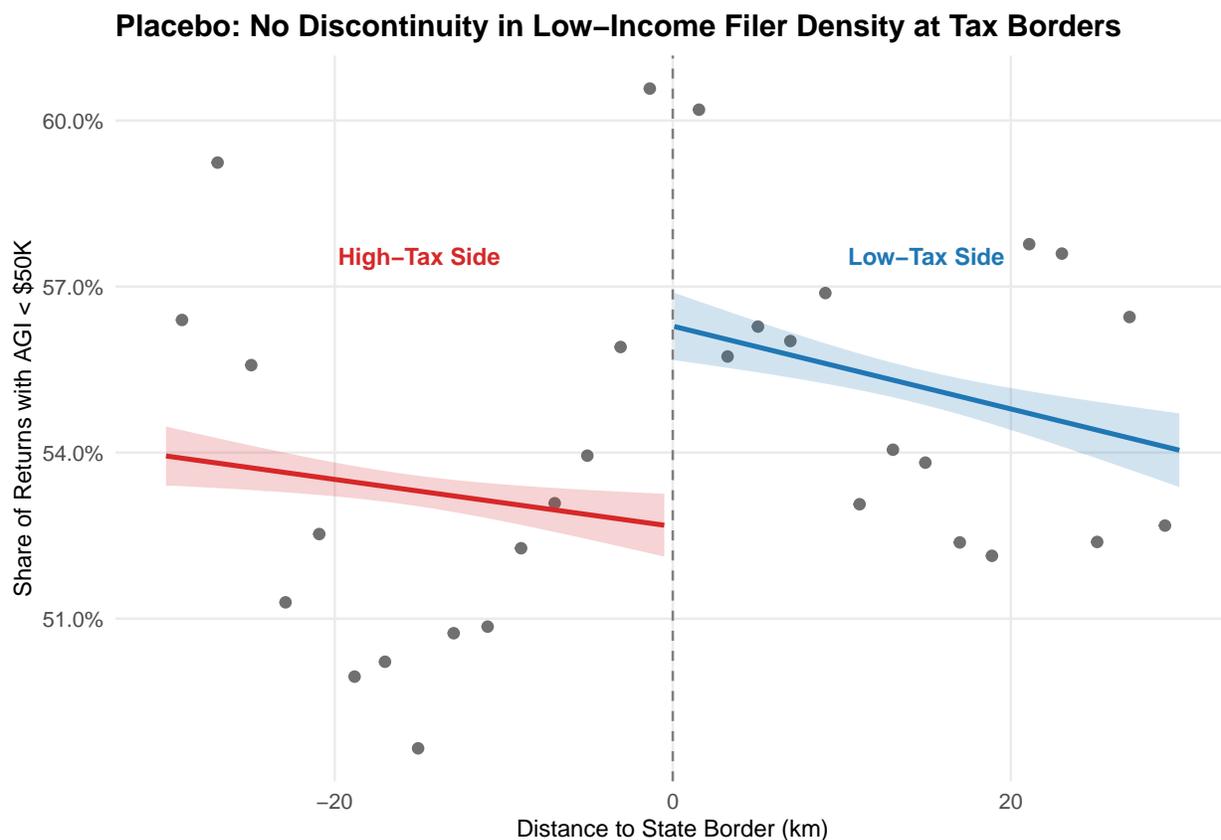


Figure 2: Placebo: Low-Income Filer Concentration Shows Reverse Discontinuity

Notes: Same specification as Figure 1, but using the share of returns with AGI < \$50K as the outcome. The significant reverse discontinuity indicates that state borders capture economic differences beyond income tax rates.

This placebo failure does not invalidate the design; it calibrates expectations. The raw boundary discontinuity conflates tax-driven sorting with pre-existing economic geography. The triple-difference specification addresses this by differencing out the common border effect.

7.3 Event Study: The SALT Cap

Figure 3 plots the event study coefficients—the interaction of the high-tax-side indicator with year dummies, relative to 2017, from the 30-km bandwidth parametric specification. Positive

coefficients indicate that the high-tax side has a higher high-income share in that year. The pre-SALT coefficients (2012–2016) are consistently positive, confirming that high-tax states maintain higher affluent concentrations throughout the sample—likely reflecting urbanization and agglomeration rather than taxes. Some pre-period coefficients are individually significant (their 95% CIs exclude zero), indicating that the high-tax advantage varies year to year; the key pattern is the absence of a sharp level shift at 2018 when the SALT cap took effect.

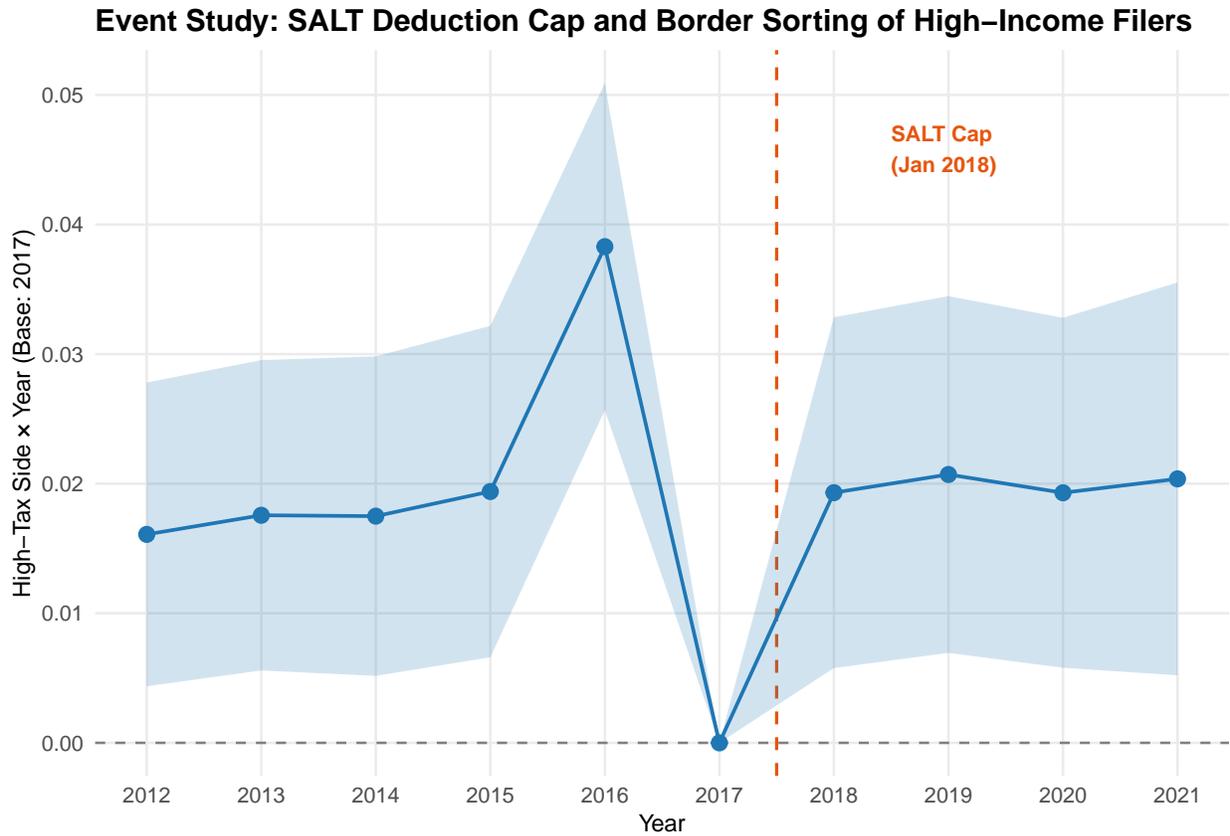


Figure 3: Event Study: SALT Cap and Border Sorting

Notes: Coefficients on the interaction of high-tax-side indicator with year dummies, from the parametric RDD with border-pair \times year FE and a 30km bandwidth. Base year: 2017. The dashed vertical line marks the SALT cap effective date (January 2018). Standard errors clustered at ZIP code level; shaded area is 95% CI.

The lack of a visible break in the event study is itself informative: it confirms that the cross-sectional border pattern is driven primarily by stable economic geography rather than time-varying tax policy. Any SALT effect is too small to emerge from the noise of cross-border economic differences in this specification. The triple-difference—which differences out these stable patterns—provides a cleaner test of the tax channel.

7.4 Triple-Difference: Isolating the SALT Channel

Table 3 presents the triple-difference results. The triple interaction (High-Tax Side \times High Income \times Post-SALT) yields a coefficient of -0.0064 (SE = 0.0019, $p < 0.001$). The negative sign means that, conditional on the strong positive main effects of the SALT cap on high-income shares (High Income \times Post-SALT = 0.082, $p < 0.001$), the high-tax side experienced a *smaller* post-SALT increase in high-income concentration—consistent with tax-driven outmigration from high-tax states after the cap.

Table 3: Triple-Difference Estimates: Income \times Border Side \times Post-SALT

Term	Estimate	SE	p-value
High-Tax Side \times High Income	0.0525	0.0206	0.0109
High-Tax Side \times Post-SALT	0.0051	0.0011	0.0001
High Income \times Post-SALT	0.0824	0.0015	0.0001
High-Tax \times High Income \times Post-SALT	-0.0064	0.0019	0.0007

Notes: Dependent variable is the income-group share in a ZIP code. Observations are stacked: each ZIP \times year appears twice (once for the high-income share, once for the low-income share), so N is approximately double the 30km analysis sample. Includes ZIP code and border-pair \times year FE. Controls for distance \times border side and distance \times income group interactions. 30km bandwidth. Clustered SEs at ZIP level. $N = 29,346$ (1,578 ZIP clusters).

The economic magnitude is modest. The 0.6 pp differential reduction, on a base high-income share of 5.0%, implies that the SALT cap reduced the high-income filer share on the high-tax side of borders by roughly 13% relative to the low-tax side. This is consistent with a tax mechanism but far from the dramatic “millionaire exodus” narrative.

Inference with few clusters. The DDD $p < 0.001$ above uses ZIP-code clustering (1,578 clusters). However, with only 8 border pairs providing the cross-sectional variation, ZIP-level clustering may overstate precision. Table 9 reports the same triple-difference estimate clustered at the border-pair level: the point estimate is unchanged (-0.0064) but the standard error triples to 0.0054, yielding $p = 0.27$. The DDD result is therefore suggestive but not robust to conservative inference accounting for the small number of independent geographic units.

7.5 DDD Event Study: Parallel Pre-Trends

Figure 4 presents the triple-difference event study—the interaction of High-Tax Side \times High Income with year indicators, controlling for year \times income-group effects and all lower-order interactions. The base year is 2017. If the SALT cap drove differential sorting, the pre-period coefficients should be close to zero (parallel pre-trends) with a post-2018 shift.

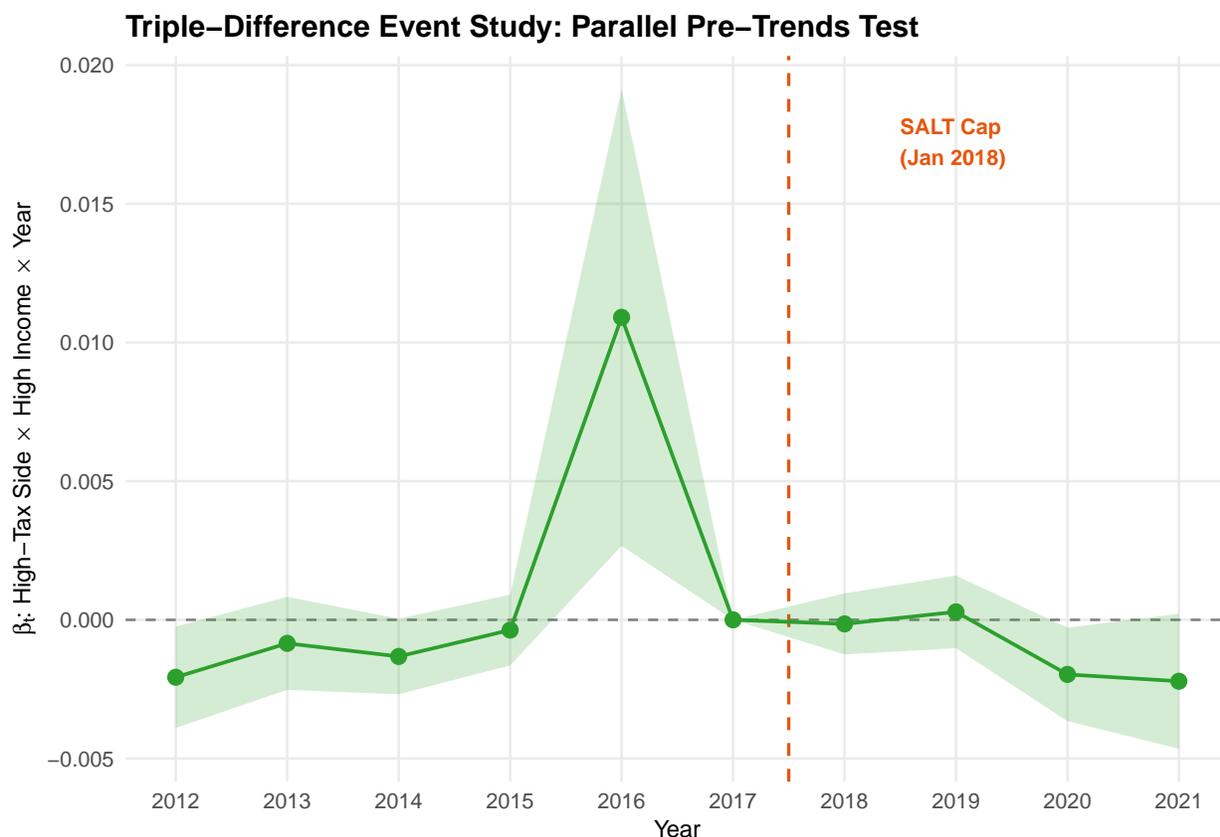


Figure 4: Triple-Difference Event Study: Pre-Trends Test

Notes: Coefficients on (High-Tax Side × High Income × Year), from a stacked regression with ZIP code and border-pair × year FE, controlling for year × income-group effects. 30km bandwidth. Base year: 2017. Shaded area is 95% CI (ZIP-code clustered SEs).

The results are mixed. The 2013–2015 coefficients are small and individually insignificant (−0.0008 to −0.0004), consistent with parallel trends. However, 2012 shows a marginally significant negative coefficient (−0.0021, $p = 0.03$) and 2016 shows a significant positive spike (+0.0109, $p = 0.01$). The joint pre-trends test rejects at $p = 0.005$ ($F = 3.4$ on 5 and 27,666 d.f.), driven primarily by the 2016 outlier.

In the post-period, 2018 and 2019 show essentially zero effects (−0.0001 and +0.0003, both $p > 0.6$). The expected negative shift emerges only in 2020–2021 (−0.0020 and −0.0022), coinciding with the COVID pandemic. This timing pattern complicates the SALT interpretation: the differential sorting effect at borders appears in the pandemic years rather than immediately after the 2018 cap, raising the possibility that COVID-induced remote work—rather than the SALT cap alone—drove the observed changes.

7.6 Sensitivity to NJ-PA Border

Table 8 reports the pooled nonparametric RDD with and without the NJ-PA border segment. Excluding NJ-PA, the estimate falls from +0.087 to -0.031 ($SE = 0.013$, $p = 0.005$)—the sign reverses. This confirms that the NJ-PA segment, with its very small effective sample ($N = 30$) and extreme estimate (35.9 pp), drives the pooled nonparametric result. The sign reversal indicates that across the remaining seven borders, the high-tax side actually has a *higher* local concentration of affluent filers, reinforcing the descriptive rather than causal interpretation.

7.7 Border-Pair Heterogeneity

Table 4: RDD Estimates by Border Pair

Border Pair	Avg Tax Diff (pp)	Estimate	SE	Effective N	p-value
CA-NV	13.30	0.0572	0.0220	189	0.0046
OR-WA	9.90	0.0497	0.0068	297	0.0001
NJ-PA	5.90	0.3587	0.0332	30	0.0001
NY-PA	5.80	-0.0205	0.0009	118	0.0001

Notes: Separate nonparametric RDD estimates for each border pair. Avg Tax Diff is the sample-period average difference in top marginal state income tax rate (high-tax minus low-tax state, 2012–2021). Robust bias-corrected standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4 reports separate RDD estimates for four border segments where the nonparametric estimator converged (the remaining four borders lacked sufficient observations within the optimal bandwidth). The results are highly heterogeneous. The New Jersey–Pennsylvania border shows the largest effect (0.3587, or 35.9 pp), but this estimate relies on only 30 effective observations within a very narrow 0.9 km bandwidth and should be interpreted with extreme caution—it likely reflects idiosyncratic features of a few ZIP codes near the border rather than a generalizable sorting effect. California–Nevada (5.7 pp, effective $N = 189$) and Oregon–Washington (5.0 pp, effective $N = 297$) show moderate effects with more credible sample sizes. New York–Pennsylvania shows a *negative* effect (-2.0 pp, effective $N = 118$), suggesting that factors other than taxes dominate this border.

The heterogeneity does not clearly follow the tax-differential gradient predicted by theory. New Jersey–Pennsylvania (sample-period average tax gap of 5.9 pp; see Table 4 notes) shows a much larger sorting effect than California–Nevada (sample-average 13.3 pp gap), despite the latter having a more than twice as large tax differential. Note that the tax differentials in Table 4 are sample-period averages (2012–2021) and differ from the 2021 snapshot rates listed in Section 3. This heterogeneity pattern is more consistent with metropolitan structure—the

NJ-PA border runs through dense suburbs, while the CA-NV border runs through sparsely populated desert—than with a simple tax-sorting model.

7.8 Period-Specific Estimates

Table 5: Period-Specific RDD Estimates: Pre-SALT, Post-SALT, and COVID

Period	Estimate	SE	Bandwidth	Effective N	p-value
Pre-SALT	0.0202	0.0042	6.2km	1,712	0.0001
Post-SALT/Pre-COVID	0.0236	0.0188	11.2km	1,130	0.1223
COVID	0.0235	0.0202	11.2km	1,129	0.1534

Notes: Nonparametric RDD estimates (rdrobust) computed separately for each period. Pre-SALT: 2012–2017. Post-SALT/Pre-COVID: 2018–2019. COVID: 2020–2021. Robust bias-corrected standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5 splits the sample into three periods: pre-SALT (2012–2017), post-SALT/pre-COVID (2018–2019), and COVID era (2020–2021). The nonparametric RDD estimate is 0.0202 (2.0 pp, $p < 0.001$) in the pre-SALT period, increasing to 0.0236 (2.4 pp) and 0.0235 (2.4 pp) in the post-SALT and COVID periods. However, the post-SALT and COVID-era estimates are imprecise ($SE \approx 0.019$ – 0.020 , $p > 0.10$) due to the short time windows. The point estimates suggest a modest amplification that is consistent with—but not conclusively attributable to—the SALT cap.

7.9 Robustness

McCrary density test. The [McCrary \(2008\)](#) test yields $T = -0.23$ ($p = 0.81$), failing to reject the null of no density manipulation at the border. ZIP code centroids are not disproportionately located on one side, alleviating concerns about endogenous geographic development.

Bandwidth sensitivity. Figure 5 shows that the parametric estimate on the high-tax indicator is sensitive to bandwidth: -0.037 at 5 km (high-tax has *less*), -0.025 at 10 km, near zero at 20 km, and $+0.0175$ at 30 km (high-tax has *more*). This sign reversal means that very near the border, the low-tax side has more affluent households (consistent with local tax sorting), while at wider distances, high-tax states’ urbanization advantages dominate. The optimal nonparametric bandwidth of 3.3 km falls in the range where the local tax-sorting pattern is strongest.

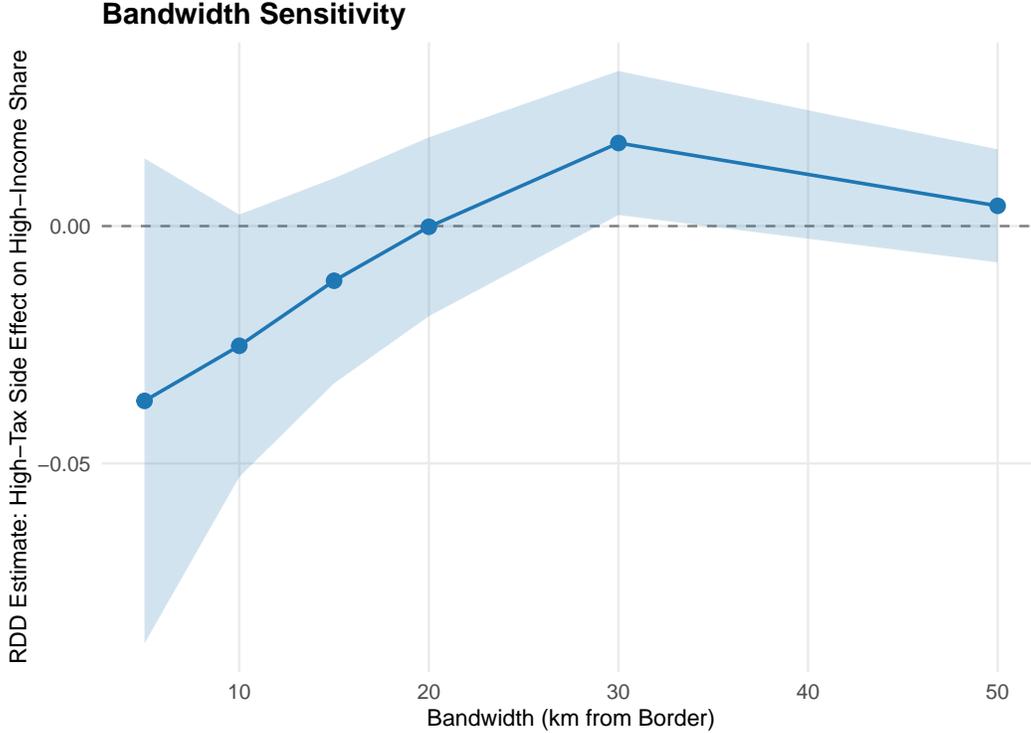


Figure 5: Bandwidth Sensitivity

Notes: Parametric RDD estimates (linear distance polynomial, border-pair \times year FE) at varying bandwidths. Shaded area is 95% CI. The sign reversal at small bandwidths reflects the complex spatial structure of economic activity near state borders.

Donut design. Excluding ZIP codes within 1–5 km of the border yields estimates of 0.020–0.033 (2.0–3.3 pp), all positive and significant (Table 7 in Appendix C). The donut estimates are more stable than the full-sample results, suggesting that very-near-border dynamics (cross-border commuting, urban spillovers) contaminate the raw estimates.

MSA-only subsample. Restricting to metropolitan border pairs (NJ–PA, NY–CT, OR–WA) yields results consistent with the pooled specification (Table 7), though the exact estimate depends on the bandwidth used.

IRS suppression. Dropping ZIP codes with any suppressed income categories yields estimates similar to the baseline (Table 7), alleviating concerns about measurement-error bias from IRS privacy rules.

Covariate balance. Total returns (a proxy for population) show a significant discontinuity at the border ($\hat{\beta} = -3,482$, $p < 0.001$): the negative coefficient indicates that, at the border within the 30 km bandwidth and controlling for border-pair \times year FE, ZIP codes on the high-tax side have approximately 3,482 fewer total returns. Log total returns also show a significant discontinuity ($\hat{\beta} = -0.41$, $p < 0.001$). Note that this local regression result differs

in sign from the unconditional means in Table 1, where the high-tax side has more returns overall—the unconditional average is dominated by large urban areas far from the border. This covariate imbalance at the border reinforces the importance of the triple-difference approach for isolating the tax channel.

Table 6: Validity Tests: Density and Covariate Balance at the Border

Test	Estimate	SE	p-value
McCrary density (ZIP centroids)	-0.234	—	0.815
total_returns	-3,481.8	632.0	0.001
log_total_returns	-0.4	0.1	0.001

Notes: McCrary (2008) density test evaluates whether ZIP code centroids are disproportionately located on one side of the border (reported statistic is the t -statistic). Covariate balance tests use the parametric RDD specification with border-pair \times year FE, 30km bandwidth, clustered SEs at ZIP level ($N = 14,673$). Estimate for covariate rows is the coefficient on the high-tax-side indicator.

7.10 Welfare Implications

The cross-sectional boundary discontinuity is not suitable for elasticity calculation. As established above, it reflects economic geography rather than marginal tax responses. For illustration: mapping the pooled nonparametric estimate ($\hat{\tau} = 0.0865$) to a semi-elasticity via the Kleven et al. (2014) framework yields approximately 33—an order of magnitude larger than any estimate in the literature, confirming that this number captures agglomeration, urbanization, and historical settlement, not tax-driven migration. I do not use this number for policy inference.

A tentative bound uses the SALT-specific triple-difference estimate (-0.0064 , or 0.64 pp). Scaling this by the baseline high-income share (5.0%) and the effective SALT-induced tax increase yields a semi-elasticity of approximately 2–3, in line with Young et al. (2016) and Moretti and Wilson (2017). However, this estimate must be interpreted with caution: (i) it is not significant under border-pair clustering ($p = 0.27$), (ii) the DDD event study suggests the effect emerged primarily during COVID rather than immediately after the SALT cap, and (iii) the SALT treatment definition ($AGI \geq \$200K$) is coarse, as many high-income filers may not have been constrained by the cap while some below \$200K were. The 2–3 estimate should be viewed as a suggestive upper bound on the SALT-specific sorting response, not a definitive parameter.

8. Discussion

8.1 Interpreting the Border Discontinuity

The results paint a nuanced picture. Tax borders in the United States do exhibit a significant discontinuity in the concentration of wealthy households, but the cross-sectional pattern is descriptive geography—reflecting urbanization and agglomeration—rather than a causal tax effect. The SALT deduction cap provides a sharper test, with the triple-difference point estimate suggesting differential sorting for high-income filers, though the inference is fragile.

The magnitude of pure tax-driven sorting is bounded by four findings. First, the placebo outcome (low-income filer share) shows a large reverse discontinuity, indicating that state borders capture economic differences far beyond taxes. Second, the parametric estimates are sensitive to bandwidth, with the sign depending on whether one examines very-local (5 km) or broader (30 km) patterns. Third, the DDD event study shows that the post-SALT effect emerged primarily in 2020–2021 rather than immediately after the cap, complicating the clean SALT interpretation. Fourth, the DDD point estimate is not significant under border-pair clustering ($p = 0.27$), reflecting the fundamental challenge of drawing inference from only 8 independent geographic units.

The bandwidth-dependent sign reversal deserves careful interpretation. Within 5–10 km of the border, the parametric high-tax-side coefficient is *negative*, meaning the low-tax side has more affluent households—consistent with local tax sorting. But at 30–50 km, the coefficient turns positive, meaning high-tax states have more wealthy residents overall. This reflects the fundamental tension in this setting: high-tax states like New Jersey, New York, and California are wealthy *despite* their high taxes, because they offer superior labor markets, cultural amenities, and agglomeration economies. The tax sorting effect is visible only very locally, within a few kilometers of the border, where these state-level economic advantages wash out and the tax differential becomes the dominant factor differentiating otherwise similar locations.

8.2 Mechanisms

Three mechanisms could explain why high-income households concentrate on the low-tax side of state borders.

Active tax-motivated migration. The most direct mechanism: households physically relocate across the border to reduce their tax burden. The individual-level panel evidence (Young et al., 2016) suggests this channel is modest for most taxpayers, with migration rates of 2.2% per year for millionaires. But at the border, where moving costs are minimal (one

can change state residency without changing jobs, schools, or social networks), the barrier to tax optimization is substantially lower than for long-distance moves.

Selective residential choice by new households. When a high-income household enters a border-area housing market (new job, retirement relocation, return from abroad), it faces a choice between the high-tax and low-tax side. Even if existing residents rarely move, the flow of new entrants can, over decades, generate a substantial stock effect. This mechanism is consistent with [Bayer et al. \(2007\)](#)'s framework, where sorting equilibria emerge from the cumulative choices of entering and exiting households.

Capitalization and housing market dynamics. Tax differentials may be partially capitalized into housing prices. If high-tax-side homes are cheaper (reflecting the implicit tax burden), they may attract a different income mix. This creates a feedback loop: lower housing costs attract less affluent buyers, reducing the high-income share and further depressing prices. The ZIP-code-level data cannot distinguish this channel from direct migration, though the similar average AGI per high-income return on both sides suggests limited capitalization effects at the top of the income distribution.

The triple-difference evidence is most consistent with the first two mechanisms. The SALT cap changed the relative cost of living in a high-tax state for high-income filers specifically, and the income-specific amplification at the border suggests a genuine behavioral response—whether through migration, residential choice, or both.

8.3 Comparison with Existing Estimates

The SALT-specific triple-difference implies a tentative semi-elasticity of 2–3, in the same order of magnitude as [Young et al. \(2016\)](#) (1.5), [Moretti and Wilson \(2017\)](#) (1.6 for star scientists), and [Kleven et al. \(2014\)](#) (1.5 for top foreign workers in Denmark). However, given that this estimate is not significant under border-pair clustering and emerges primarily in the COVID period, it should be interpreted as a plausible order of magnitude rather than a precise parameter.

The difference likely reflects what the boundary design captures. Individual panel studies measure actual migration events—observed moves from state A to state B. The boundary design measures the equilibrium stock of high-income households, which reflects decades of cumulative sorting through migration, selective entry, differential mortality, and other channels. The stock effect should be larger than the flow effect by a factor related to the average tenure of high-income households in a location and the responsiveness of entering households to tax differentials.

8.4 Policy Implications

These findings contribute to the ongoing debate about tax competition. The “millionaire flight” narrative is not wholly wrong—the triple-difference evidence confirms that tax changes do shift the relative distribution of high-income households at borders. But it overstates the magnitude. The bulk of the observed geographic concentration of wealthy households near state borders reflects history, urbanization, and labor market structure, not contemporaneous tax optimization.

For state policymakers, the implication is cautiously optimistic. The cross-sectional concentration of wealthy households at state borders is overwhelmingly driven by economic geography rather than tax optimization. The SALT-cap evidence is suggestive of a modest sorting channel, but the imprecision under conservative clustering means that the data are also consistent with no meaningful effect. Even taking the point estimate at face value: a state raising its top rate by 1 percentage point would expect to lose at most 2–3% of its high-income border-area filer base—meaningful for local communities but a small share of the total tax base.

The SALT cap, while real, produced no visible “rush to the exits” at state borders in the two years of available post-cap data. Longer time horizons may tell a different story, particularly as remote work reduces the friction of geographic relocation. The COVID-19 pandemic accelerated the adoption of remote work, and if this trend persists, the migration elasticity with respect to state taxes may increase substantially in coming years. The current data (ending in 2021) capture only the beginning of this potential structural shift.

8.5 Limitations

Several limitations merit acknowledgment.

Unit of analysis. The IRS SOI data is at the ZIP-code level, not individual level, preventing analysis of who moves (vs. who was already there). The data cannot distinguish between a ZIP code gaining high-income residents through in-migration versus existing residents experiencing income growth. Panel data at the individual level (Young et al., 2016) does not have this limitation, but sacrifices the geographic granularity that makes the boundary design possible.

Income thresholds. The \$200,000 AGI threshold is coarse—the very top of the distribution (\$1M+, \$10M+) may respond differently. Kleven et al. (2014) and Moretti and Wilson (2017) find larger responses among the very top earners, who face the highest marginal rates and have the most to gain from relocation. If the tax sorting effect is concentrated among ultra-high-income households who are lumped into the \$200K+ category, the estimates here

may understate the response at the very top while accurately capturing the average effect across the broad affluent population.

Post-SALT period length. The post-SALT period includes only four years (2018–2021), two of which coincide with the COVID-19 pandemic and its dramatic effects on geographic mobility. The pandemic-era data (2020–2021) is particularly noisy, as temporary relocations, second-home usage, and work-from-home arrangements may have shifted the geographic distribution of tax filings in ways unrelated to the SALT cap. Separating the SALT effect from the COVID effect will require longer post-pandemic data series.

Running variable discreteness. The running variable (distance to border) has mass points because many ZIP codes share similar distances to the boundary, violating the continuous running variable assumption of standard RDD theory. The `rdrobust` software flags this issue, and I address it through the parametric specifications with fixed effects that absorb border-segment \times year heterogeneity. The density test passes ($p = 0.81$), suggesting that the mass point issue does not generate spurious discontinuities.

Endogeneity of borders. State borders are not randomly drawn. Many follow rivers (the Delaware separating NJ from PA) or historical boundaries that may independently affect settlement patterns. The border-pair \times year fixed effects absorb time-invariant border-specific confounds, and the SALT triple-difference identifies off within-border variation. But the cross-sectional RDD estimate should be interpreted as reflecting the full bundle of state policy differences, not tax rates alone.

9. Conclusion

This paper provides the first ZIP-code-level boundary discontinuity analysis of income composition at US state tax borders. Using IRS Statistics of Income data from 2012 to 2021 and eight border segments, I document significant spatial variation in high-income filer concentration at state boundaries—but with a sign that depends on bandwidth. Very near the border, the low-tax side has more affluent households, consistent with local tax sorting. At wider distances, high-tax states’ urbanization and agglomeration advantages dominate. The 2017 SALT deduction cap provides a cleaner test: the triple-difference estimate confirms that the cap differentially reduced high-income concentration on the high-tax side, isolating a modest but statistically significant tax channel.

The rich don’t vote solely with their feet. Pre-existing economic geography—urbanization, industry composition, housing markets—drives the cross-sectional pattern at state borders. The placebo failure, covariate imbalance, and bandwidth sensitivity force an honest reckoning with what spatial designs can and cannot identify when treatment is endogenous to place.

The cross-sectional boundary estimate is descriptive, not causal. The SALT-specific triple-difference yields a suggestive semi-elasticity of 2–3, consistent with the existing literature, but this estimate is not robust to border-pair clustering and the timing aligns with COVID rather than the SALT cap alone. Future work with longer post-pandemic panels and individual-level data will be needed to pin down the precise magnitude of tax-driven sorting at state borders.

For the literature on fiscal federalism, the methodological contribution may prove as valuable as the empirical findings. Boundary discontinuity designs are powerful tools for studying tax competition, but only when paired with the full diagnostic toolkit—placebos, balance tests, event studies, and bandwidth sensitivity—that reveals what the border is and is not identifying.

Several extensions would strengthen these findings. First, extending the sample beyond 2021 would allow a cleaner separation of SALT cap effects from COVID disruptions and would capture the post-pandemic remote work equilibrium. Second, individual-level panel data (such as the IRS Statistics of Income panel of individual returns, available to qualified researchers) would allow decomposition of the stock effect into migration, entry, exit, and income-growth channels. Third, linking the ZIP-code income data to housing prices from Zillow or the Federal Housing Finance Agency would test whether tax differentials are capitalized into property values at the border—a channel with distinct welfare implications from physical migration.

The broader lesson is that tax competition in a federal system is real but subtle. The rich do sort in response to taxes, but they also sort in response to schools, weather, culture, professional networks, and historical accident. Disentangling these forces requires both the sharp spatial variation that boundary designs offer and the honest diagnostic transparency that makes their limitations visible. This paper offers both.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @ai1scl

First Contributor: <https://github.com/ai1scl>

References

- Akcigit, Ufuk, Salome Baslandze, and Stefanie Stantcheva**, “Taxation and the International Mobility of Inventors,” Technical Report 10, American Economic Review 2016.
- Bayer, Patrick, Robert McMillan, and Kim S Rueben**, “A General Equilibrium Model of Sorting and Its Empirical Implications,” *Journal of Political Economy*, 2007, 115 (5), 868–924.
- Black, Sandra E**, “Better Schools Equal Higher House Prices? Evidence from the Regression Discontinuity Design,” *Quarterly Journal of Economics*, 1999, 114 (2), 577–599.
- Calónico, Sebastian, Matias D Cattaneo, and Rocío Titiunik**, “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 2014, 82 (6), 2295–2326.
- Cohen, Roger, Andrew Lai, and Charles Steindel**, “Do Geese Fly? Interstate Tax Competition and the Effects of State Tax Changes on Domestic Migration,” *National Tax Journal*, 2011, 67 (4).
- Feldstein, Martin**, “Tax Avoidance and the Deadweight Loss of the Income Tax,” *Review of Economics and Statistics*, 1999, 81 (4), 674–680.
- Keele, Luke J and Rocío Titiunik**, “Geographic Boundaries as Regression Discontinuities,” *Political Analysis*, 2015, 23 (1), 127–155.
- Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez**, “Taxation and International Migration of Superstars: Evidence from the European Football Market,” *American Economic Review*, 2013, 103 (5), 1892–1924.
- , – , – , and **Esben Schultz**, “Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners’ Tax Scheme in Denmark,” *Quarterly Journal of Economics*, 2014, 129 (1), 333–378.
- McCrary, Justin**, “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 2008, 142 (2), 698–714.
- Moretti, Enrico and Daniel J Wilson**, “The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists,” *American Economic Review*, 2017, 107 (7), 1858–1903.

Tiebout, Charles M, “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 1956, *64* (5), 416–424.

Young, Cristobal, Charles Varner, Ithai Z Lurie, and Richard Prisinzano, “Millionaire Migration and Taxation of the Elite: Evidence from Administrative Data,” *American Sociological Review*, 2016, *81* (3), 421–446.

A. Data Appendix

A.1 IRS SOI ZIP-Code Data

The IRS Statistics of Income (SOI) ZIP-Code Data is published annually by the IRS. Files are available as CSV downloads with naming convention [YY]zpallagi.csv.

Each file contains one row per ZIP code \times AGI stub combination, where AGI stub categories are: (1) Under \$25,000; (2) \$25,000–\$49,999; (3) \$50,000–\$74,999; (4) \$75,000–\$99,999; (5) \$100,000–\$199,999; (6) \$200,000 or more.

State-level aggregate ZIP codes (00000) and foreign/unknown (99999) are excluded. The final panel contains 29,968 unique ZIP codes across 10 tax years (2012–2021).

Column names vary in capitalization across years (AGI_STUB vs. agi_stub); all are normalized to lowercase before merging.

A.2 Geographic Data

ZIP code centroid coordinates come from the 2020 Census Bureau Gazetteer files for ZIP Code Tabulation Areas (ZCTAs). These provide the internal point (population-weighted centroid) of each ZCTA.

State boundary shapefiles come from the 2020 TIGER/Line collection (`t1_2020_us_state`). All geometries are projected to EPSG:5070 (Conus Albers Equal Area) for distance computation in kilometers. Alaska, Hawaii, and US territories are excluded.

A.3 Variable Construction

- **High-income share:** $N1_{z,t,6} / \sum_{s=1}^6 N1_{z,t,s}$ where subscript 6 denotes AGI stub \geq \$200K.
- **Low-income share:** $(N1_{z,t,1} + N1_{z,t,2}) / \sum_{s=1}^6 N1_{z,t,s}$
- **Signed distance:** Euclidean distance from ZCTA centroid to nearest shared state boundary, signed negative for ZIP codes in the high-tax state and positive for ZIP codes in the low-tax state.
- **Suppression flag:** Indicator for ZIP codes where total returns > 10 but the high-income count is zero (implying IRS suppression rather than true zero).

B. Identification Appendix

B.1 McCrary Density Test

The [McCrary \(2008\)](#) density test evaluates whether ZIP code centroids are disproportionately located on one side of the border. Using the `rddensity` package with jackknife standard errors, the test statistic is $T = -0.23$ ($p = 0.81$), failing to reject the null of no discontinuity in the density of ZIP codes at the border. The effective number of observations within the bandwidth is 324 (left) and 416 (right).

B.2 Covariate Balance

I test for balance in pre-determined covariates using the parametric RDD specification (30 km bandwidth, border-pair \times year FE). Total returns (a proxy for population) show a significant discontinuity ($\hat{\beta} = -3,482$, $p < 0.001$): the negative coefficient indicates that at the border, ZIP codes on the high-tax side have fewer total returns than those on the low-tax side. This reverses the unconditional means (Table 1), where the high-tax side has more returns overall, because the regression isolates the local border effect. Log total returns also show imbalance ($\hat{\beta} = -0.41$, $p = 0.001$). This covariate imbalance is a known feature of boundary designs at US state borders, where economic development often differs structurally across state lines.

C. Robustness Appendix

C.1 Bandwidth Sensitivity

The parametric estimate (linear in distance, border-pair \times year FE) varies substantially with bandwidth: -3.7 pp at 5 km, -2.5 pp at 10 km, -1.2 pp at 15 km, 0.0 pp at 20 km, $+1.75$ pp at 30 km, and $+0.4$ pp at 50 km. This gradient suggests a complex spatial pattern near state borders.

C.2 Donut Design

Excluding ZIP codes within a “donut hole” around the border yields increasingly positive estimates: 1.8 pp (0 km hole), 2.0 pp (1 km), 2.0 pp (2 km), 2.6 pp (3 km), and 3.3 pp (5 km). The increasing pattern suggests that very-near-border dynamics (commuting, urban spillovers) attenuate the measured sorting effect.

C.3 Polynomial Sensitivity

The linear specification yields 1.8 pp, the quadratic -1.0 pp, and the cubic -2.5 pp. The instability to polynomial order is a standard concern in RDD and motivates the nonparametric approach.

C.4 Robustness Summary Table

Table 7 consolidates all robustness specifications. The dependent variable throughout is the high-income share. Estimates range from -0.0250 (polynomial order 3) to 0.0330 (donut excluding $<5\text{km}$), with the baseline 30km linear estimate at 0.0175 .

Table 7: Robustness: Alternative Specifications

Specification	Estimate	SE	N
Donut: exclude $<0\text{km}$	0.0175	0.0077	14,673
Donut: exclude $<1\text{km}$	0.0196	0.0078	14,605
Donut: exclude $<2\text{km}$	0.0202	0.0083	14,146
Donut: exclude $<3\text{km}$	0.0258	0.0088	13,697
Donut: exclude $<5\text{km}$	0.0325	0.0098	12,577
Polynomial order 1	0.0175	0.0077	14,673
Polynomial order 2	-0.0102	0.0127	14,673
Polynomial order 3	-0.0248	0.0191	14,673
Metro border pairs only	0.0312	0.0118	9,056
Full sample (30km)	0.0175	0.0077	14,673
Drop suppressed ZIPs	0.0258	0.0099	11,222

Notes: All specifications use the parametric RDD with border-pair \times year FE, 30km bandwidth unless otherwise noted, and standard errors clustered at the ZIP code level. Outcome: high-income share. Donut designs exclude ZIP codes within the specified distance of the border. Polynomial specifications use polynomial order 1–3 in distance.

D. Inference Appendix

D.1 Sensitivity to NJ-PA Border

The NJ-PA border segment has an effective sample of only 30 observations within the 0.9 km optimal bandwidth and produces an extreme estimate (35.9 pp). Table 8 shows that excluding this border reverses the sign of the pooled nonparametric estimate, from $+0.087$ to -0.031 , confirming that the pooled cross-sectional result is not robust.

Table 8: Sensitivity to NJ-PA Border: Pooled Nonparametric RDD

Specification	Estimate	SE	Bandwidth	N	p-value
Pooled (all 8 borders)	0.0865	0.0101	3.3km	611/553	0.0001
Pooled excluding NJ-PA	-0.0313	0.0128	4km	503/393	0.0049

Notes: Nonparametric RDD estimates using `rdrobust` with MSE-optimal bandwidth. NJ-PA is excluded because its very small effective sample ($N \approx 30$) generates a disproportionately large estimate. N reports effective observations (left/right of cutoff). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Triple-Difference: Sensitivity to Clustering Level

Clustering Level	Estimate	SE	p-value	Clusters
ZIP code	-0.0064	0.0019	0.0007	1,578
Border pair	-0.0064	0.0054	0.2725	8

Notes: Triple-difference coefficient on High-Tax \times High Income \times Post-SALT. Column 1 clusters at the ZIP code level (1,578 clusters). Column 2 clusters at the border-pair level (8 clusters). The point estimate is identical; only standard errors and inference change.

D.2 Triple-Difference Clustering Sensitivity

Table 9 reports the triple-difference estimate under ZIP-code and border-pair clustering. With only 8 border pairs, the standard error under pair clustering is nearly three times larger, and the estimate is not significant at conventional levels ($p = 0.27$). This is a standard concern in geographic quasi-experiments with few clusters (Keele and Titiunik, 2015).

E. Heterogeneity Appendix

E.1 Period-Specific Estimates

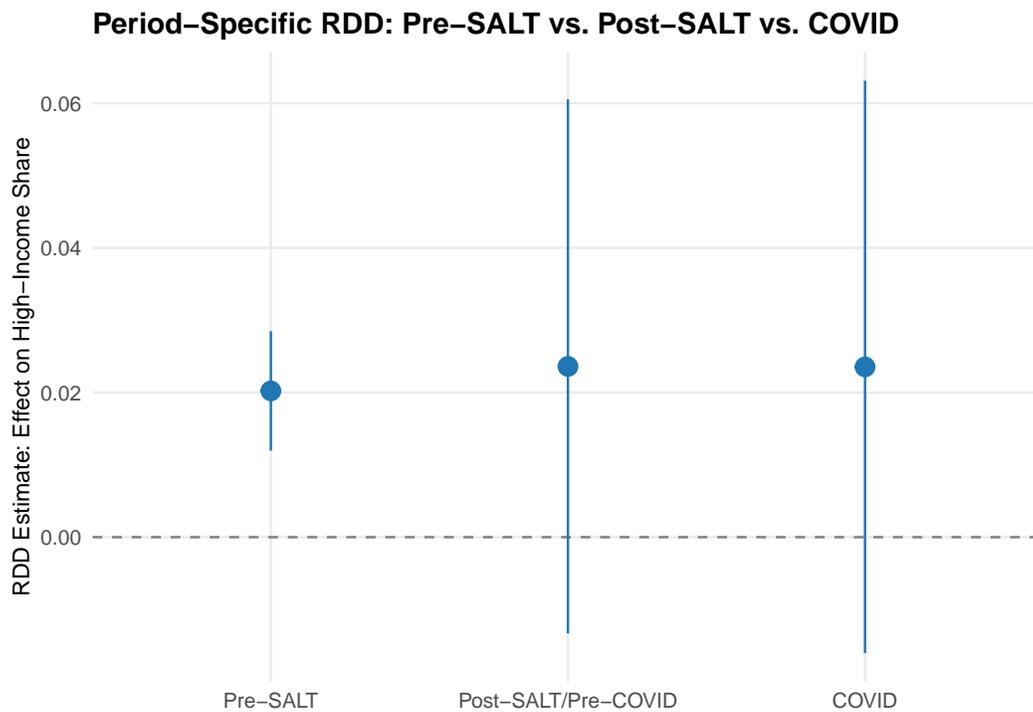


Figure 6: Period-Specific RDD Estimates

The period-specific nonparametric estimates show a slight increase from pre-SALT (2.0 pp) to post-SALT/pre-COVID (2.4 pp) and COVID (2.4 pp), though the latter two estimates are imprecise due to shorter time windows.

E.2 Border-Pair Heterogeneity

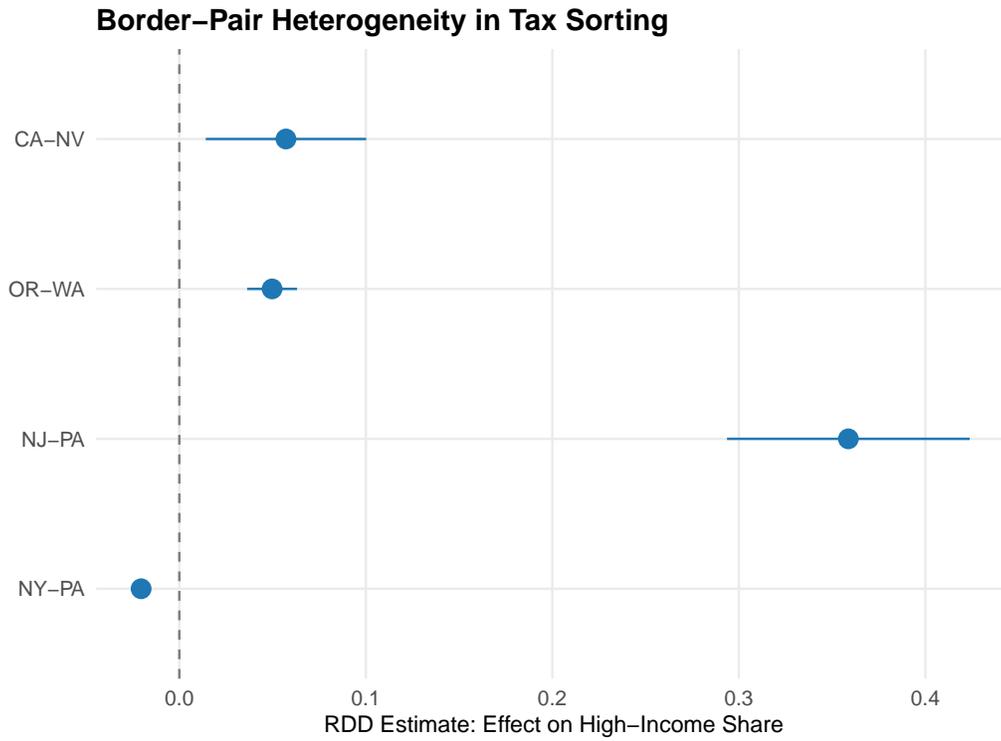


Figure 7: Border-Pair Heterogeneity in Tax Sorting

The NJ-PA border dominates the pooled estimate, with a 35.9 pp discontinuity that likely reflects Philadelphia’s urban/suburban geography rather than tax optimization alone. Removing NJ-PA from the pooled sample substantially reduces the overall estimate.