Separating Rate from Composition: A Zero-Loss Decomposition of Electoral Change

Donald Moratz

February 16, 2026

Abstract

How do we separate the effects of changing voter preferences from changing demographic composition? Existing decomposition methods produce residual errors exceeding one million votes because they conflate these distinct dynamics. I introduce the Rate, Composition, and Volume Decomposition (RCVD), a zero-loss framework that precisely partitions total vote change into three components: changes in group loyalty (Rate), group size (composition), and overall participation (volume). Through sequential rather than simultaneous calculation, RCVD eliminates residual error while preserving substantive interpretability. Applied to U.S. presidential elections from 2016 to 2024, RCVD reveals a critical finding: the 2024 result was driven by decoupling of rate and composition effects among Hispanic voters. Substantial rightward shifts in Hispanic partisan preferences neutralized the demographic advantage from their growing share of the electorate. The analysis demonstrates that cleanly separating these components is essential for understanding when and how compositional advantages erode.

1 Introduction

When election outcomes shift, scholars seek to understand why: Did voters change their preferences, or did the electorate's demographic composition change? Decomposing these effects has proven challenging. Recent work by Marble et al. (2024) represents a major advance, providing the first comprehensive framework for translating survey data into group-level vote estimates and decomposing changes across elections. Building on their Equation 5 specification, they employ what I term a Linear Difference Approach (LDA) that attempts to isolate preference changes from demographic shifts. However, when components are summed, their decomposition yields a residual of 1.2 million votes—revealing that the approach cannot fully account for total vote change. This gap stems from the omission of interaction terms that arise when both voter preferences and demographic composition change simultaneously.

This limitation reflects a broader methodological challenge. Current approaches to decomposition—whether using linear differences (Marble et al., 2024; Fraga, Y. R. Velez, et al., 2024) or regression-based methods (Hill et al., 2021)—face a key tension: cleanly separating preference shifts (which I term rate effects) from demographic changes (composition effects) requires either accepting uninterpretable interaction terms or imposing restrictive functional form assumptions. When scholars omit these interaction terms to maintain interpretability, as is standard practice, the resulting estimates systematically misattribute vote changes. In close elections, residuals of this magnitude can reverse substantive conclusions about which factors drove the outcome.

I introduce the Rate, Composition, and Volume Decomposition (RCVD), a method that decomposes total vote change into three components without requiring interaction terms. RCVD accomplishes this through sequential calculation: first isolating how vote switching within groups affects outcomes (Rate), then measuring how shifts in group sizes matter (composition), and finally capturing how overall turnout scales the result (volume). This approach yields exact decompositions—zero residual error—while maintaining clear substantive interpretations for each component. Critically, RCVD makes transparent the assumptions embedded in any decomposition: the order of calculation determines what counterfactual world we are comparing against.

The method's value extends beyond technical precision. Applied to recent U.S. presidential elections, RCVD reveals that the 2024 result was driven by a decisive decoupling of rate and composition effects among Hispanic voters. While demographic growth continued to favor Democrats compositionally, substantial rightward shifts in Hispanic voting preferences—Rate effects—neutralized this advantage. Existing methods, by conflating these distinct dynamics through interaction terms, obscure this critical pattern. The analysis demonstrates a key insight: compositional advantages erode when group voting preferences shift, and only methods that cleanly separate these components can reveal when such decoupling occurs.

The article proceeds as follows. Section 2 reviews existing decomposition methods and introduces the RCVD framework, demonstrating why standard approaches require uninterpretable interaction terms. Section 3 presents a worked example with simulated data, showing step-by-step how RCVD differs from existing methods and validating the approach through controlled scenarios. Section 4 applies RCVD to U.S. presidential elections from 2016 to 2024, revealing how the method exposes data quality issues and captures Hispanic voting dynamics that existing approaches miss. Section 5 concludes.

2 Historical Approaches and the RCVD Framework

Understanding how group-level dynamics shape election outcomes has been a central concern in political science since Campbell (1960) demonstrated that voters' attachments to social groups fundamentally structure electoral behavior. While much subsequent work has focused on individual-level determinants of vote choice, understanding how the composition of the electorate across demographic groups shifts over time remains essential for explaining aggregate election outcomes (Axelrod, 1972). This is particularly relevant in an era where, despite increasing capacity for micro-targeted appeals (Hersh, 2015), electoral success still requires building broad coalitions across identity groups (Sides et al., 2019).

Contemporary scholarship has increasingly emphasized the role of identity-based coalitions in shaping electoral outcomes. Achen and Bartels (2016) argue that political science needs to rethink how we approach studying democracy, moving away from an exclusive focus on individuals and toward greater attention to identity groups. Sides et al. (2019) demonstrate that the Trump campaign of 2016 succeeded in part through reshaping the coalitions of both parties around identity politics. In this context of increasing polarization, building electoral coalitions based on group membership has become a central campaign strategy (Lemi, 2021), making accurate decomposition of compositional and preference effects increasingly important for understanding electoral dynamics.

Recent scholarship has developed various approaches to decompose vote changes into compositional and preference components. Hill et al. (2021) uses a regression-based approach to attribute vote share changes to either composition or conversion, while Zingher (2019) focuses on estimating how changes in group size affect party support based on underlying group dynamics. However, regression-based approaches impose strict linear specifications on the functional form of the relationship. Misspecification of the functional form can introduce significant bias in the estimated coefficients, even when the relationship is otherwise well-specified (Wooldridge, 2010). Additionally, as Engelhardt (2019) points out, distributional shifts across groups within the electorate and attitudinal shifts within groups can mask each other, making interpretations of coefficients from regressions difficult and obscuring the true drivers of electoral change.

Other analyses have employed simpler extrapolation methods to estimate compositional effects. Fraga, McElwee, et al. (2021) examines how different compositions and turnout rates across racial groups affected Clinton's performance in the 2016 election by applying 2012 turnout rates to estimate counterfactual vote shares. Carmines et al. (2016) analyze how shifts in political coalitions and changes in turnout affect overall vote share. While these studies contribute to our understanding of compositional effects, they do not provide a systematic framework for fully decomposing the interactions between group shifts and electoral outcomes.

The importance of accurately decomposing these effects extends to election forecasting. Calvo et al. (2024) demonstrate that even perfect knowledge of demographic shifts provides insufficient information to guarantee accurate forecasts, highlighting the need for methods that can cleanly separate compositional from behavioral changes. Similarly, Grimmer, Knox, et al. (2024) argue that properly evaluating election forecasts requires understanding how demographic and preference shifts interact across multiple election cycles. As the discipline develops more sophisticated approaches to election forecasting, accurately decomposing the impacts of demographic shifts on election results becomes increasingly critical.

Building on this work, the most comprehensive methodological framework to date for decomposing compositional and preference effects is provided by Marble et al. (2024). Composition here can be defined as the relative size of groups within the electorate, which contrasts with rate, which I define as the two-party vote margin within a group. For example, if 60% of a group votes Republican and 40% votes Democrat, the Republican margin rate is +20% (or +0.20). They reject the linearity assumption imposed by earlier regression-based approaches and adopt a non-linear specification. They argue that the difference in votes a group contributes to the election outcome can be captured by:

$$
\begin{aligned}
\text{Diff Net}_{t,t-1}(x) = {} & [\text{Vote Share}_t(x, \text{Republican}) - \text{Vote Share}_t(x, \text{Democrat})] \\
& \text{x Turnout}_t(x) \text{ x Group Size}_t(x) \\
& - [\text{Vote Share}_{t-1}(x, \text{Republican}) - \text{Vote Share}_{t-1}(x, \text{Democrat})] \\
& \text{x Turnout}_{t-1}(x) \text{ x Group Size}_{t-1}(x)
\end{aligned} \tag{1}
$$

Where $\text{Turnout}_t(x)$ x $\text{Group Size}_t(x)$ can be thought of as the compositional component. This equation refers to the raw votes in the system, where x is the group of interest (such as racial groups) and t is time, but defined by elections rather than years[1] Using this equation, they calculate for each group a different turnout, group size and vote choice component to describe how each individual voting group contributed to the change in total vote share in the two elections for a party. To calculate the compositional effect and the vote share effect separately, they then first hold composition

[1] i.e. the 2016 US presidential election might be defined as $t = 0$ and the 2020 US presidential election might be defined as $t = 1$.

fixed from the 2012 election and calculate the impact of the shifted vote share from 2012 to 2016 and then hold vote share fixed at the 2012 election and shift composition to the 2016 election. The joint sum of these totals is the implied effect of a group on the election outcome. Building on this work, Fraga, Y. Velez, et al. (2023) examine the shift within Latino voters, conducting an analysis of this subgroup, but applying a similar methodology. They find that shifts amongst subgroups within the broader Latino category is a driver for their increased support of Trump in 2020 compared to 2016.

However, the Linear Difference Approach faces two key challenges. First, it conflates composition and volume—that is to say, it does not distinguish between changes in the relative size of groups and changes in overall turnout. Composition can be thought of as the proportion of the total electorate comprised by each group (for example, 60 percent White, 40 percent non-White), whereas volume is the total size of the electorate. The calculation in Marble et al. (2024) allows for changes in group size and turnout rate, but does not capture whether all groups are growing equally or whether some groups are growing faster than others. This means we cannot identify what component of change stems from a simple uniform increase in turnout—which affects total votes but not vote shares—versus differential growth across groups that shifts the compositional balance. By way of example, consider turnout in the 2008 presidential election. Turnout reached a 40-year high (Woolley and Peters, 2021), but aggregate turnout was only part of the story. Turnout among Black voters rose by nearly 5 percent, while turnout among White voters declined by just over 1 percent (Lopez and Taylor, 2009). The formulation from Marble et al. (2024) would show positive turnout effects for Black voters and negative effects for White voters, but would understate the importance of these differential changes because it does not explicitly model how Black voters became a larger proportion of the overall electorate.

The second—and more fundamental—challenge is that calculating rate and composition effects independently and then summing them fails to account for their interaction. When both voter preferences and group composition change simultaneously between elections, the total effect is not simply the sum of the two independent effects. To see why, consider a simplified case with a single group. Let $r(t)$ represent the margin rate at which the group votes for a candidate and $c(t)$ represent the composition (the group's share of the electorate). The total margin votes contributed by this group at time t is simply $z(t) = r(t)c(t)$. Now consider how votes change between two elections at times t_1 and t_2. The LDA approach calculates the rate effect by holding composition fixed at t_1 and allowing rate to shift: $[r(t_2) - r(t_1)]c(t_1)$. It then calculates the composition effect by holding rate fixed at t_1 and allowing composition to shift: $r(t_1)[c(t_2) - c(t_1)]$. Summing these two components yields the LDA estimate of total change.

However, the actual total change is $\Delta z(t) = z(t_2) - z(t_1) = r(t_2)c(t_2) - r(t_1)c(t_1)$. When we expand this expression algebraically, we find:

$$\begin{aligned}\Delta z(t) &= z(t_2) - z(t_1) \\ &= r(t_2)c(t_2) - r(t_1)c(t_1) \\ &= [r(t_2) - r(t_1)]c(t_1) + r(t_1)[c(t_2) - c(t_1)] + [r(t_2) - r(t_1)][c(t_2) - c(t_1)] \qquad (2)\end{aligned}$$

The first two terms match the LDA calculation, but the third term—$[r(t_2) - r(t_1)][c(t_2) - c(t_1)]$—represents an interaction effect. This interaction term captures the additional votes generated (or lost) when both rate and composition change together. Substantively, this interaction represents a critical dynamic: when a group both grows in size AND shifts its voting preferences, the impact is multiplicative rather than simply additive. For instance, if Hispanic voters both increase as a share of the electorate (composition shift) AND become more Republican (rate shift), the net effect for Republicans is larger than simply adding the two effects independently would suggest.

When scholars using the LDA omit this interaction term, the residual must go somewhere. By omitting these terms, the LDA implicitly assumes they are zero or trivially small. However, as I demonstrate in Section 4, these interaction terms can represent millions of votes in real elections. In practice, as noted in the introduction, this creates unexplained vote totals—the 1.2 million vote residual in Marble et al. (2024)'s analysis of the 2012 to 2016 election. The problem becomes more severe when we add a third component, volume. With three variables—rate $r(t)$, composition $c(t)$, and volume $v(t)$—the complete decomposition requires not just three main effects but also three two-way interactions and one three-way interaction:

$$\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
&= r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) \\
&= [r(t_2) - r(t_1)]c(t_1)v(t_1) + r(t_1)[c(t_2) - c(t_1)]v(t_1) + r(t_1)c(t_1)[v(t_2) - v(t_1)] \\
&+ [r(t_2) - r(t_1)][c(t_2) - c(t_1)]v(t_1) + [r(t_2) - r(t_1)]c(t_1)[v(t_2) - v(t_1)] + r(t_1)[c(t_2) - c(t_1)][v(t_2) - v(t_1)] \\
&+ [r(t_2) - r(t_1)][c(t_2) - c(t_1)][v(t_2) - v(t_1)]
\end{aligned} \tag{3}$$

As can be seen, accurately capturing the total change requires including multiple interaction terms. The first line contains the three main effects—these are what the LDA attempts to calculate. The second and third lines contain the interaction terms—the products of changes across multiple components. In practice, other work has not directly calculated the impact of volume, generating an implicit assumption that any remaining votes that are left unaccounted for must be driven by changes in volume.

As a result, methods that rely only on calculating the individual component effects do not account for how these components interact when they change simultaneously. This generates error because the LDA specification is, in essence, treating the problem as if changes in each component can be evaluated independently, when in fact the underlying relationship is multiplicative. If rate and composition changes are small, or if they move in offsetting directions, the interaction terms may be relatively minor and the LDA provides a reasonable approximation. However, as the magnitude of changes grows—or when rate and composition shifts reinforce each other—the interaction terms become substantial, and the LDA approximation deteriorates. The challenge is illustrated in Figure 1, which shows how approximations that ignore interaction effects accumulate error as the changes being measured become larger.

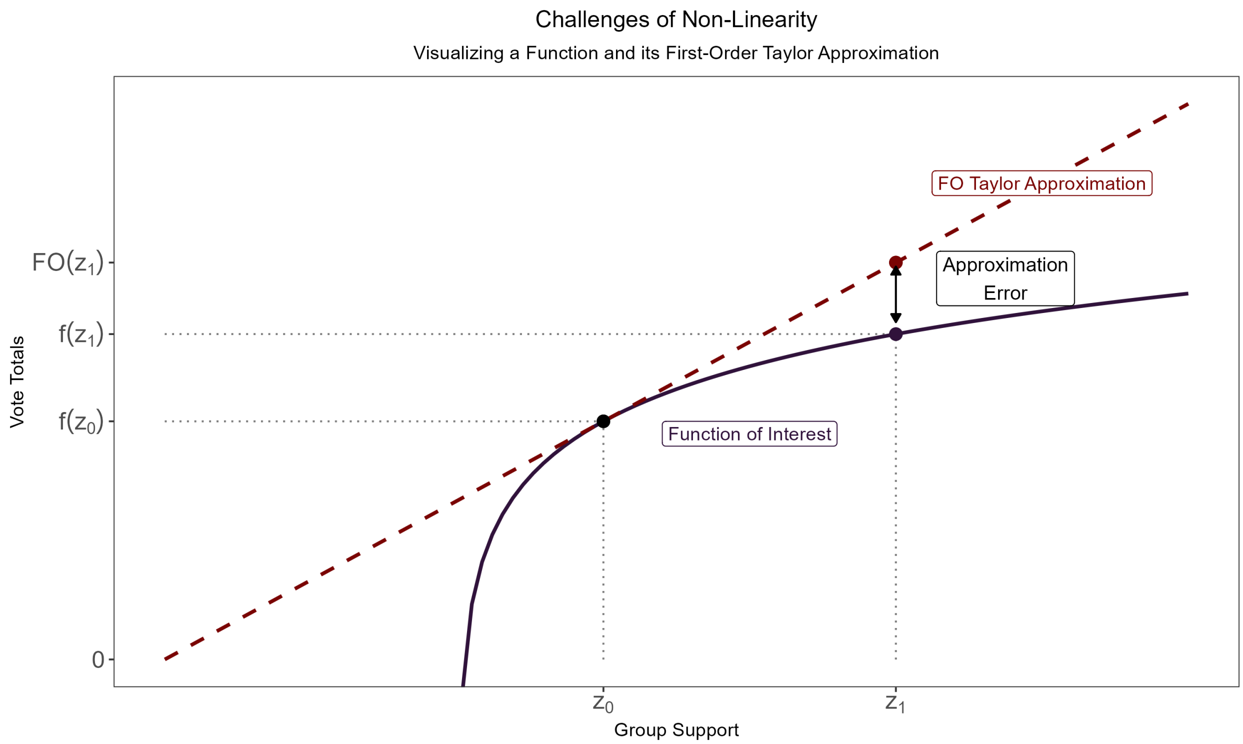

Figure 1: This figure illustrates how methods that treat changes as independent (ignoring interaction effects) accumulate error as the magnitude of change increases. The figure is based on Figure 4.3.1 in Reed (1998), showing that linear approximations become less accurate when the underlying function involves multiplicative relationships between components.

I propose an alternative specification of the method for calculating the difference between vote shares that instead treats the problem as sequential. The underlying intuition of why the problem can be described as sequential in nature is developed below, but intuitively, by calculating the effects of composition using prior rates, the LDA misses that the rate a group votes for a candidate can be changed fundamentally by who is turning out to vote. The impact of composition can only be measured by who comprises the new voter pool, not the individuals who composed the prior voting pool. This is particularly important if Citrin et al. (2003) were correct that the difference between non-voters and voters in a given election can be very different than it is in other elections. As the most important takeaway, the bias of the LDA estimate can be both substantial and even yield the wrong sign of the impact of composition when a candidate goes from winning (losing) a group to losing (winning) that group.

I argue that the specification proposed in Equation 5 in Marble et al. (2024) serves as the best starting point for understanding shifts in the electorate. To clearly identify the components of shifts in the electorate though, I will redefine the terms. I begin by defining $z(t)$ as the total number of votes a group gives to a particular candidate, where t is still defined as an election. I am interested in describing how that group's contribution to the candidate's vote total compares to their contribution in the prior period and I want to show how that contribution is broken into rate, composition and volume. Given that those are the three bins I would like to explain, then $z(t)$ must be defined as a function of those three bins:

$$z(t) = r(t)c(t)v(t) \tag{4}$$

Where $r(t)$ is the rate at which a group votes for the candidate, $c(t)$ is the proportion of the electorate that the group comprises and $v(t)$ is the total number of voters in the election. This specification differs from the one used in equation 3 in Marble et al. (2024) and Fraga, Y. R. Velez, et al. (2024) in that emphasis is placed on total volume. This is done for three reasons. First, total votes can, in and of itself, prove meaningful in analysis of election results. Second, total votes provides the only meaningful metric by which we can gauge the plausibility of any analysis. That is to say, since we rely on estimates of group compositions and vote choice, the only firm vote total against which we can compare our estimates of the group dynamics in an election is the final reported vote tally. I will show the importance of this in section 4. Third, as discussed in Marble et al. (2024), using this specification allows researchers to quickly apply the results of exit polls against vote tallies to understand the shifts in elections almost immediately after their conclusion.

I propose an alternative approach that resolves the interaction term problem through sequential calculation rather than attempting to isolate each component independently. The underlying intuition is straightforward: when both voter preferences and demographic composition change between elections, we cannot meaningfully measure the impact of composition changes while holding preferences fixed at their old values. The rate a group votes for a candidate can be fundamentally changed by who is turning out to vote. The impact of composition can only be measured by who comprises the new voter pool, not the individuals who composed the prior voting pool. This is particularly important if Citrin et al. (2003) were correct that the difference between non-voters and voters in a given election can be very different than it is in other elections. By calculating effects sequentially—first allowing preferences to shift, then composition, then overall turnout—we can decompose total change without requiring the uninterpretable interaction terms that plague the LDA.

I begin by defining $z_i(t)$ as the net margin votes that group i contributes to a party or candidate, where t is defined as an election and i indexes demographic groups (e.g., White voters, Hispanic voters). Following Marble et al. (2024), I work with vote margins—the difference between votes for and against a candidate within each group—rather than raw vote totals for a single candidate. I am interested in describing how that group's contribution to the candidate's vote margin compares to their contribution in the prior period, and I want to show how that contribution is broken into rate, composition and volume. Given that those are the three components I would like to explain, then $z_i(t)$ must be defined as a function of those three components:

$$z_i(t) = r_i(t)c_i(t)v(t) \tag{5}$$

Where $r_i(t)$ is the margin rate at which group i votes for the candidate (the difference in vote shares between the two parties), $c_i(t)$ is the proportion of the total electorate that group i comprises, and $v(t)$ is the total number of voters in the election. Note that while $r_i(t)$ and $c_i(t)$ vary across groups (indexed by i), $v(t)$ is constant—the total electorate size is the same for all groups.

To calculate the impact of rate, composition, and volume separately, I propose the following sequential decomposition:

$$
\begin{aligned}
\Delta z_i(t) =& [r_i(t_2) - r_i(t_1)]c_i(t_1)v(t_1) \\
&+ r_i(t_2)[c_i(t_2) - c_i(t_1)]v(t_1) \\
&+ r_i(t_2)c_i(t_2)[v(t_2) - v(t_1)]
\end{aligned}
\tag{6}
$$

The first term calculates the rate effect for group i: how many margin votes change when that group's preferences shift from $r_i(t_1)$ to $r_i(t_2)$, holding composition and volume at their initial values. The second term calculates the composition effect: how many margin votes change when group i's share of the electorate shifts from $c_i(t_1)$ to $c_i(t_2)$, using the new margin rate $r_i(t_2)$ but holding volume at its initial value. The third term calculates the volume effect: how many margin votes change when overall turnout shifts from $v(t_1)$ to $v(t_2)$, using the new margin rate and new composition for group i. While initially counterintuitive, the proof that this sum equals $z_i(t_2) - z_i(t_1)$ exactly—with zero residual—is presented in the Appendix. I call this the Rate, Composition, and Volume Decomposition (RCVD).

There are two noteworthy features of this approach. First, turnout, a fundamentally important feature of election results, affects both $c(t)$ and $v(t)$, as does population growth. To see this for $c(t)$, consider that the percentage of the total voters that a group comprises is a function of both the proportion of the population comprised of the group and the turnout rate of that group. For $v(t)$, a similar challenge exists—an increase in the total number of potential voters in a group (population size) and the total number of realized voters (turnout) both affect the final total size of the voting population. While a decomposition of the formula into the component parts of turnout and population change is achievable, I argue it is not necessary for the purposes of this paper for two main reasons. First, across the simple case of two time periods only four years apart, underlying changes in the racial composition of the potential electorate are unlikely to be meaningful, at least compared to the impact of turnout. Second, turnout changes are only interesting insofar as they are differential. If all groups increase turnout at the same rate (thus preserving the relative group sizes), then there are no compositional effects. With this specification, any changes that are universal (overall population growth, increased turnout across all groups) are captured in volume, while any changes that are differential are captured in composition.

A second crucial feature is that this sequence is not unique. The calculation can be performed using any order-combination of rate, composition, and volume. However, different orderings yield different interpretations of the estimated coefficients, as each ordering implicitly defines a different counterfactual comparison. To address this, in the next section I highlight how changing the sequencing can affect the interpretation of a simulated election result. There are two key arguments for why the rate → composition → volume ordering is the most interpretable. First, by moving rate first, the rate calculation exactly matches all previous work that has been done on the impact of shifts in rates—which is to say, calculating the impact of shifts in rates on the first period's composition and volume. This ensures continuity with existing literature while correcting its flaws. Second, by moving volume third, the impact of composition is also calculated on the first period's volumes. This ensures that the final calculation, volume, reflects only a perfectly proportional shift from the previous period. In essence, this order rearranges all of the components of the calculation according to the levels observed in the first period, and only after rearranging those components does volume move last, leaving it as a simple proportional scaling of the newly specified outcome. This approach causes the shift in volume to function as a scaling factor that preserves the underlying relationships between rate and composition effects.

3 Proof of Concept: Validation through Controlled Simulation

This section moves from theoretical framework to empirical application. The previous section introduced the mathematical structure of RCVD and explained why sequential calculation resolves the interaction term problem that affects the Linear Difference Approach. Here, I demonstrate how these approaches perform in practice. I begin with a simple worked example using hypothetical data to illustrate exactly how the LDA generates a residual while RCVD achieves zero error. I then apply both methods to increasingly complex scenarios: first, a two-group analysis of the 2016-2020 election showing the magnitude of the LDA's residual in real data; second, a controlled scenario isolating the rate-composition interaction by holding volume constant; and finally, a justification for the rate → composition → volume calculation sequence based on theoretical interpretability.

A Simple Worked Example

To see precisely how the LDA generates a residual while RCVD does not, consider a straightforward hypothetical scenario with two voter groups across two elections. Following Marble et al. (2024), we work with vote margins—the difference between votes for Candidate X and votes for Candidate Y within each group. Table 1 presents the data, using round numbers throughout to allow readers to verify each calculation step.

Table 1: Hypothetical Two-Group Election Data

	Election 1				Election 2			
	Voters	% of Total	Margin	Margin Rate	Voters	% of Total	Margin	Margin Rate
Group A	30	30%	+6	+20%	40	33.3%	-4	-10%
Group B	70	70%	-14	-20%	80	66.7%	-8	-10%
Total	**100**	**100%**	**-8**	–	**120**	**100%**	**-12**	–

Note: Margin = (Votes for X - Votes for Y). Margin Rate = Margin/Total Voters.
Election 1 Group A: 18 votes for X, 12 for Y → Margin = +6 (20% margin rate)

In Election 1, Group A favors Candidate X (18 votes for X versus 12 for Y, yielding a +6 vote margin and a +20% margin rate), while Group B favors Candidate Y (28 votes for X versus 42 for Y, yielding a -14 vote margin and a -20% margin rate). The total margin across both groups is -8 votes, meaning Candidate Y leads by 8 votes overall. By Election 2, both groups shift toward Candidate Y. Group A now opposes X (18 votes for X versus 22 for Y, yielding a -4 vote margin and a -10% margin rate), and Group B continues opposing X but less strongly (36 votes for X versus 44 for Y, yielding an -8 vote margin and a -10% margin rate). The total margin is now -12 votes, meaning Candidate X's margin declined by 4 votes between the two elections. This change resulted from three factors: shifts in margin rates within groups (Rate), changes in group sizes (composition), and an increase in overall turnout (volume). Table 2 shows how the LDA and RCVD decompose this 4-vote margin change.

Table 2: LDA vs RCVD Decomposition of 4-Vote Margin Change

Component	**LDA Calculation**	**RCVD Calculation**
Rate Effect:		
Group A	$(-0.30) \times 0.30 \times 100 = -9.0$	$(-0.30) \times 0.30 \times 100 = -9.0$
Group B	$(+0.10) \times 0.70 \times 100 = +7.0$	$(+0.10) \times 0.70 \times 100 = +7.0$
Total	**-2.0 votes**	**-2.0 votes**
Composition Effect:		
Group A	$(+0.20) \times (+0.033) \times 100 = +0.667$	$(-0.10) \times (+0.033) \times 100 = -0.333$
Group B	$(-0.20) \times (-0.033) \times 100 = +0.667$	$(-0.10) \times (-0.033) \times 100 = +0.333$
Total	**+1.333 votes**	**0.0 votes**
Volume Effect:		
Group A	$(+0.20) \times 0.30 \times 20 = +1.2$	$(-0.10) \times 0.333 \times 20 = -0.667$
Group B	$(-0.20) \times 0.70 \times 20 = -2.8$	$(-0.10) \times 0.667 \times 20 = -1.333$
Total	**-1.6 votes**	**-2.0 votes**
Sum of Components	**-2.267 votes**	**-4.0 votes**
Actual Change	**-4.0 votes**	**-4.0 votes**
Residual Error	**-1.733 votes**	**0.0 votes**

RCVD running totals: -10 votes after Rate, -10 after Composition, -12 after Volume

Both methods calculate the same rate effect (-2.0 votes), but they diverge in the composition and volume calculations. The LDA holds all components at Election 1 values when calculating each effect independently. For the composition effect, it applies Election 1's margin rates (+20% for Group A, -20% for Group B) to the compositional shift (+3.3 percentage points for Group A, -3.3 for Group B), producing a +1.333 vote effect. For the volume effect, it applies Election 1's margin rates and composition to the 20-voter increase, producing a -1.6 vote effect. When summed, these components total -2.267 votes, leaving a -1.733 vote residual.

In contrast, RCVD calculates each component sequentially, using updated values at each step. After the rate effect reduces the margin from -8 to -10 votes, the composition calculation uses the *new* margin rates (-10% for both groups) rather than the old rates. Because both groups have identical margin rates in Election 2, the compositional shift from 30%/70% to 33.3%/66.7% produces zero net effect—moving voters between two groups that vote identically has no impact on the overall margin. The volume effect then scales from 100 to 120 voters using both the new margin rates and new composition, reducing the margin by an additional 2.0 votes (from -10 to -12). The RCVD components sum exactly to -4.0 votes with zero residual.

This simple example demonstrates why RCVD is necessary: even with straightforward data and small changes, the LDA generates residuals because it cannot properly account for how changes in rate, composition, and volume interact. In real elections with larger shifts and more groups, these residuals grow substantially, as the following sections demonstrate.

Two-Group Simulation: Quantifying the Residual

The simple example above used hypothetical data to illustrate the mechanics of the LDA's residual problem. Applying the same analysis to actual election data reveals that these residuals grow substantially larger in real-world applications. Table 3 presents a simplified analysis of the 2016 to 2020 presidential election, collapsing racial groups into White and non-White voters. The data is derived from the framework introduced by Marble et al. (2024), with vote totals rounded for clarity of presentation.[23]

Table 3 shows the magnitude of the change we must decompose: a total shift of 9 million votes for the Democratic candidate (from Clinton's +8 million margin to Biden's +17 million margin). This shift is characterized by changes in all three fundamental parameters. With respect to rate, there was a 7.5% decline in margin

[2]The Democratic party is consistently set as the reference group across all analyses. Positive numbers indicate increased support/votes for Democrats; negative numbers indicate decreased support. No political endorsement should be inferred from this mathematical designation.

[3]I have replicated the process used by Marble et al. (2024) in creating this data, which is based upon the data from ANES (2021). The difference in their reported votes from reality seems to be related to their estimation of voter support for candidates based on the ANES. Vote totals are rounded to the nearest half-million for ease of presentation; Section 4 provides more precise figures.

within the non-White bloc, countered by a 6.6% improvement in margin within the White bloc. In terms of composition, there was a 3.6% shift of the electorate towards non-White voters. Finally, there was a large increase in volume of 26 million additional total votes cast.

Table 3: Comparing Two Elections — Basic Data

Election	Group	Candidate 1	Candidate 2	Total	Margin	Margin Rate (%)	Composition (%)
1st Election (2016)	Non-White	29.5	7.5	37.0	22.0	59.5	28.7
	White	39.0	53.0	92.0	−14.0	−15.2	71.3
	Total	68.5	60.5	129.0	8.0	6.2	100.0
2nd Election (2020)	Non-White	38.0	12.0	50.0	26.0	52.0	32.3
	White	48.0	57.0	105.0	−9.0	−8.6	67.7
	Total	86.0	69.0	155.0	17.0	11.0	100.0
Comparing the Two Elections	Non-White	8.5	4.5	13.0	4.0	−7.5	3.6
	White	9.0	4.0	13.0	5.0	6.6	−3.6
	Total	17.5	8.5	26.0	9.0	4.8	0.0

Note: Votes are reported in millions. Candidate 1 refers to Clinton (2016) and Biden (2020); Candidate 2 refers to Trump.

Figure 2 demonstrates how the LDA fails to fully account for the total 9 million vote change. When applied to this two-group system, the LDA generates a Mean Absolute Error (MAE) of just over 1.4 million votes. This unallocated portion is the residual term—real votes that contributed to the 2020 outcome but are mathematically unassigned to any specific component (Rate, composition, or volume).

Figure 2: This figure shows that even in a simple case with only two voter groups, the LDA calculation can yield misattributed votes on an order of magnitude of over one million votes. The underlying numbers reflected in this figure are in Appendix A.

This non-zero residual creates interpretive problems. Since previous decomposition studies have left the volume term unspecified, they have, in essence, folded the residual into an error term that also contained the correctly calculated volume effect. As demonstrated in Figure 3, this leads to systematic misattribution. The benefit Biden received from the change in the racial composition of the electorate is understated by 650,000 votes using the LDA, while a large portion of the 1.4 million residual is incorrectly assigned, masking the true contribution of the volume effect.

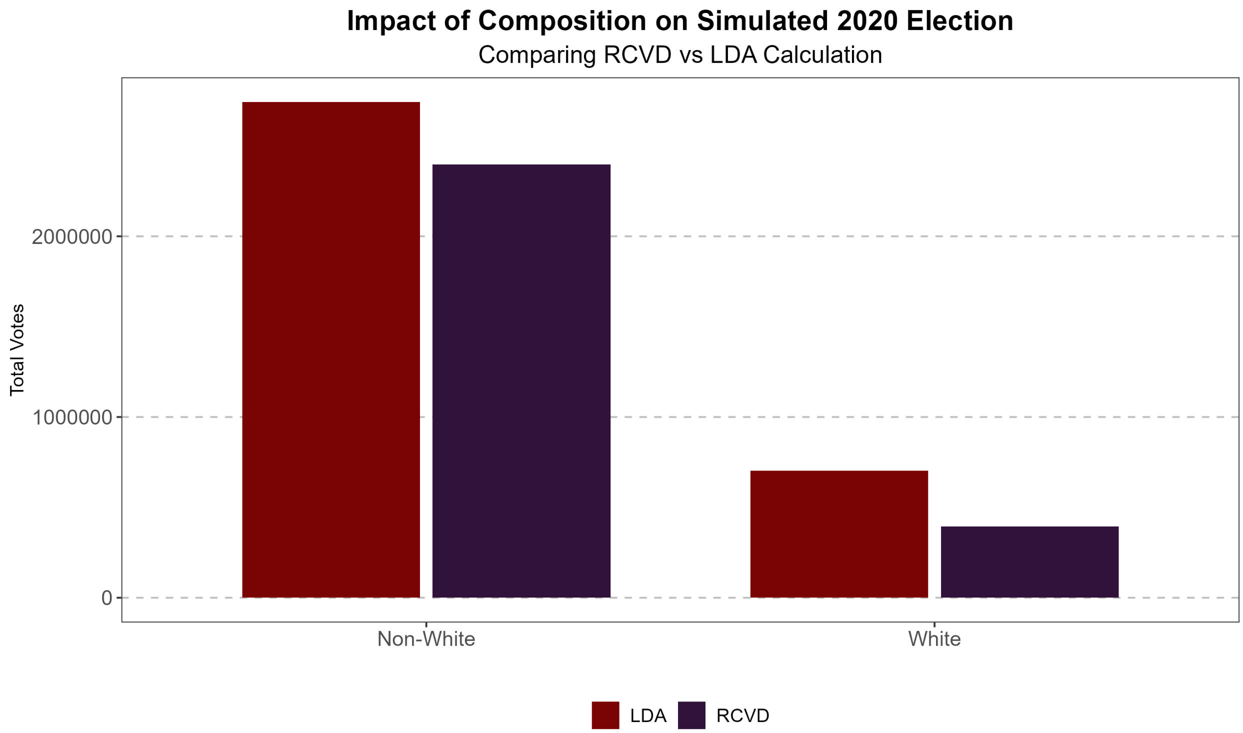

Figure 3: This figure shows that even in a simple case with only two voter groups, the LDA calculation can misattribute the impact of shifts in composition on an order of magnitude of hundreds of thousands of votes. The underlying numbers reflected in this figure are in Appendix A.

By contrast, the RCVD approach leaves a MAE of 0, precisely accounting for every vote in the system. The RCVD approach retains the highly intuitive property of explaining rate in a manner identical to the LDA, meaning that the large body of existing work focused solely on rate changes remains valid, while the problematic composition and volume terms are corrected.[4][5]

Non-Uniqueness and the Fixed RCVD Path

As noted in the conclusion of Section 2, the Rate, Composition, and Volume Decomposition (RCVD) can be calculated in any order. Mathematically, there are six possible permutations of the rate (ΔR), composition (ΔC), and volume (ΔV) terms that sat-

[4]See Appendix for the robustness of the method to partitions.

[5]A natural question is whether the LDA's residual might simply result from omitting the volume component rather than from mismodeling interaction terms. The appendix addresses this directly by analyzing a controlled scenario where volume is held constant at zero. Even with no volume change, the LDA still produces a residual of 1.7 million votes, definitively demonstrating that the residual stems from the interaction term problem, not from volume omission.

isfy the zero-loss property of the RCVD framework. This path-dependence is inherent to correctly modeling the interaction effects that the LDA omits. However, different orderings yield different interpretations of each component's contribution, as each ordering implicitly defines a different counterfactual comparison. Therefore, the chosen order ($\Delta R \rightarrow \Delta C \rightarrow \Delta V$) requires substantive justification.

Figure 12 illustrates an alternative zero-loss order (composition $\rightarrow$ rate $\rightarrow$ volume) when applied to the simple 2016-2020 simulation. While the total MAE remains 0, the calculated contribution of rate and composition shifts: the alternative order assigns a slightly larger role to composition and a slightly smaller role to rate.

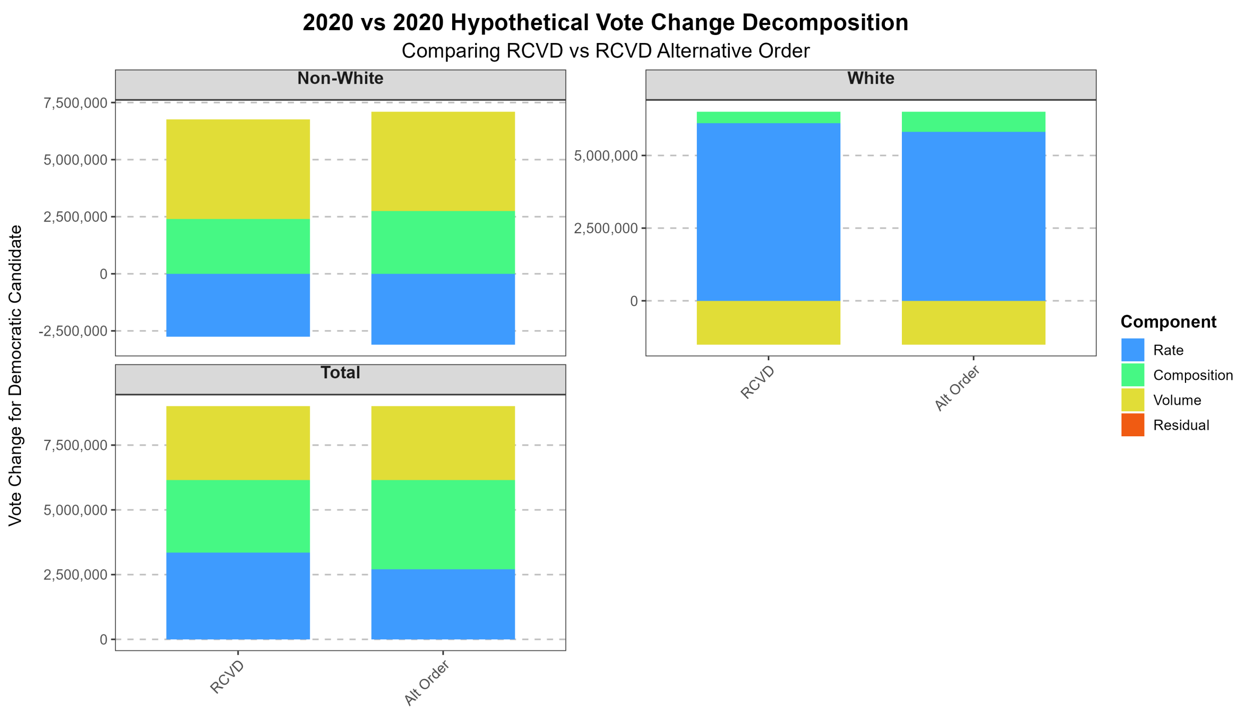

Figure 4: This figure shows that shifting the order in which RCVD is calculated yields no error term, but can shift the interpretation of the impacts of both rate and composition.

The $\Delta R \rightarrow \Delta C \rightarrow \Delta V$ ordering provides the most interpretable decomposition for two reasons. First, this order ensures continuity with existing literature. The convention in electoral studies has always been to calculate the change in group loyalty (Rate) based on the prior period's electorate composition. By sequencing rate first, the ΔR term captures preference changes holding composition and volume at their initial values. This means the rate calculation in RCVD produces exactly the same value as in the LDA, ensuring that the large body of existing work focused on rate changes remains valid while the composition and volume terms are corrected.

Second, this order ensures that volume functions as a proportional scaling factor. By sequencing volume last, we calculate it after rate and composition have already been updated. This means volume captures only the impact of expanding or contracting the electorate while preserving the relationships established by the new rate and new composition. Intuitively, if we shifted composition first (say, increasing the non-White share of the electorate) before calculating volume, the volume effect would be confounded with compositional changes. By placing volume last, it represents a clean proportional increase or decrease in the system's size, distributed across the electoral configuration determined by rate and composition.

In sum, the $\Delta R \to \Delta C \to \Delta V$ ordering ensures that each component has a clear, theoretically meaningful interpretation that aligns with how electoral scholars conceptualize these dynamics.

4 Real-World Application: ANES, Pew, and the 2024 Hispanic Vote

The preceding section established the rate, composition, and volume Decomposition (RCVD) framework and demonstrated that the Linear Difference Approach obscures the mechanisms driving electoral change. This section applies RCVD to analyze recent U.S. presidential elections, with particular attention to the Hispanic electorate.

The analysis proceeds in three parts. First, I apply RCVD to the 2020 election using ANES data (ANES, 2021), showing how it reveals different interpretations of compositional effects compared to LDA methods. Second, I compare ANES and Pew data for the 2020 election, demonstrating that Pew provides greater descriptive accuracy for analyzing Hispanic voter behavior. This comparison supports recent findings by Fraga, Y. R. Velez, et al. (2024) on Hispanic electoral volatility. Finally, I analyze three electoral cycles (2016–2020, 2020–2024, and 2016–2024) using Pew data, isolating how shifts in voter preferences (Rate), demographic composition (composition), and overall turnout (volume) shaped the 2024 election outcome.

The 2020 Baseline: Compositional Effects in the ANES

Political analysis of the 2020 presidential election typically relies on changes in two-party vote margins within demographic groups.[6] However, a change in a group's aggregate vote margin may result from a shift in partisan preference (Rate), a change in that group's share of the electorate (composition), or overall turnout changes (volume). Standard decomposition methods fail to separate these effects cleanly, hindering our understanding of what drove the election outcome.

I begin by applying the RCVD framework to data from the 2020 American National Election Studies (ANES) (ANES, 2021), which represents the discipline's historical "gold standard." While ANES has limitations in describing recent election outcomes (discussed in the next subsection), it provides a useful starting point for demonstrating RCVD's value.

Conventional analysis suggests that the Democratic victory was driven by a modest rate shift among non-White voters combined with significant compositional advantages. The RCVD decomposition reveals a more nuanced picture. By correctly partitioning the total vote shift, I find that compositional effects are significantly less beneficial to Democrats than LDA analyses suggest. Figure 5 shows how the LDA overstates the positive impact of compositional shifts for Democrats by roughly 365,000 votes.

[6] Following Marble et al. (2024), I work with two-party vote margins throughout this analysis. A margin represents the difference in vote shares between parties: for example, if Democrats receive 72% and Republicans 28% of a group's votes, the Democratic margin is +44 percentage points. While I report some results in terms of vote shares for interpretability, the RCVD framework decomposes margin changes.

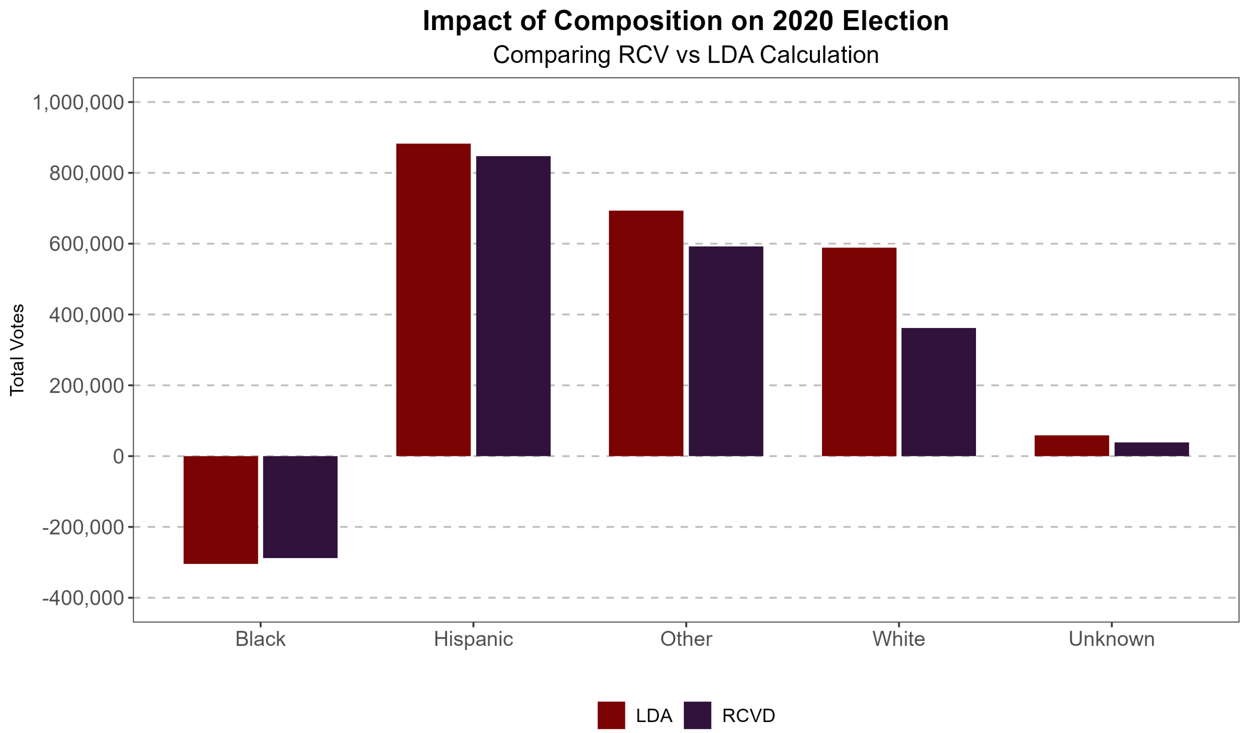

Figure 5: This figure shows how shifting from the LDA calculation to the RCVD calculation changes the evaluation of the impact of composition on the 2020 election. In particular, the LDA calculation overemphasizes the benefits of composition towards Democrats by several hundred thousand votes.

The RCVD framework provides a zero-loss allocation, ensuring that the sum of the rate, composition, and volume components precisely equals the observed overall change in the national vote margin. This contrasts with LDA methods that leave unexplained residual terms. Figure 6 shows that for White voters, the LDA's error actually exceeds the total impact of composition on the final outcome. The total error across all groups is roughly 700,000 votes versus a total compositional impact (from RCVD) of roughly 1.5 million votes.

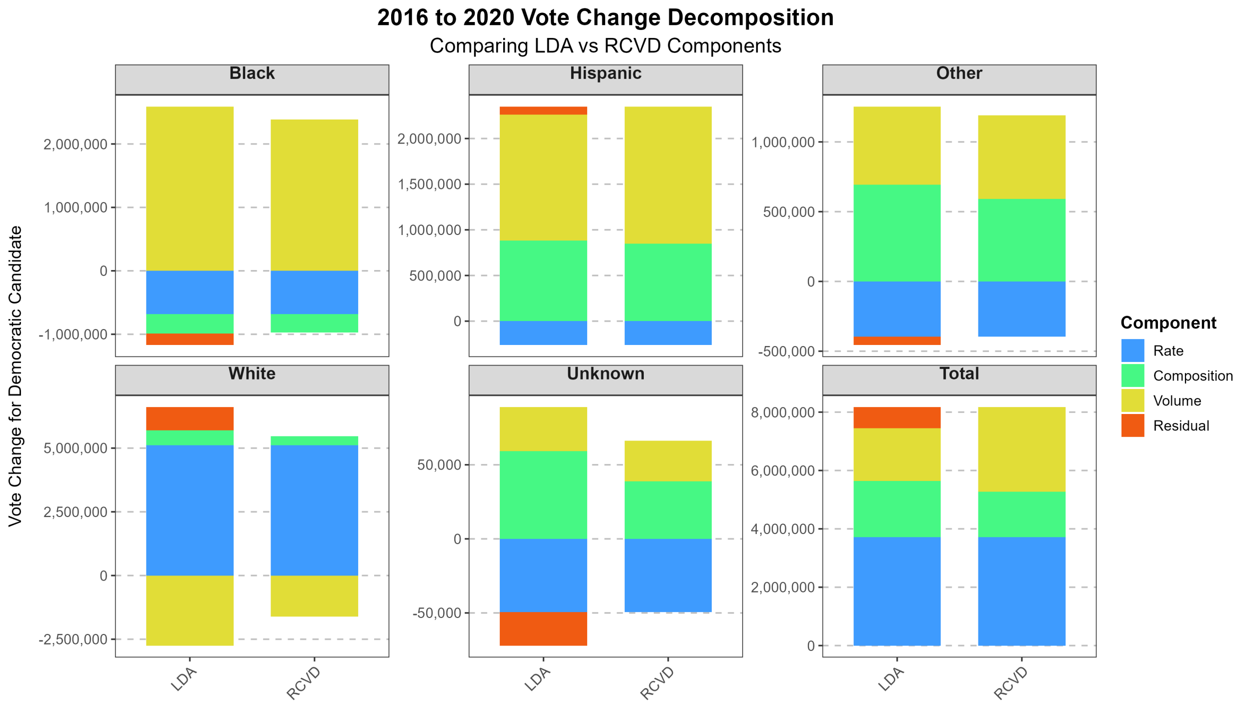

Figure 6: This figure shows how shifting to RCVD shifts interpretations of the 2020 election results when compared to 2016. It highlights that the magnitude of the error associated with the LDA is oftentimes several hundred thousand votes, and in the case of White voters, actually exceeds the total impact of composition.

The application of RCVD to the 2020 ANES data demonstrates that standard methods risk substantial bias in understanding how composition affects election outcomes. This finding motivates the subsequent analysis, where I show that Pew data provides greater accuracy in describing recent elections and use it to analyze the increasingly salient Hispanic electorate.

Survey Divergence: Comparing ANES and Pew

The robustness of any electoral decomposition depends on the descriptive accuracy of the underlying survey data. While ANES remains the discipline's preferred benchmark, its ability to capture rapidly evolving preferences among the Hispanic electorate has been questioned (Fraga, 2016). I therefore compare ANES to Pew Research Center's validated voter data (Hartig et al., 2023), which employs a different sampling strategy and may achieve better resolution for non-White subpopulations.

Figure 7 compares the two datasets' accuracy and their RCVD decompositions for

Hispanic voters. The lower panel shows that ANES overstates Democratic support by nearly 4 million votes, while Pew understates it by under 600,000 votes (Wooley and Peters, 2025)—suggesting Pew provides a more accurate picture of the 2020 electorate. The upper panel reveals a fundamental difference in interpretation. ANES shows a modest Republican gain in rate among Hispanic voters but attributes the overall Democratic increase from this group to composition—— a potentially persistent, demographic-driven effect. In contrast, Pew isolates a significantly larger rate effect driving the net impact of Hispanic voters, with composition playing a minimal role. This suggests ANES may underestimate preference volatility among Hispanic voters, supporting the argument in Fraga, Y. R. Velez, et al. (2024) about the magnitude of Hispanic electoral shifts.

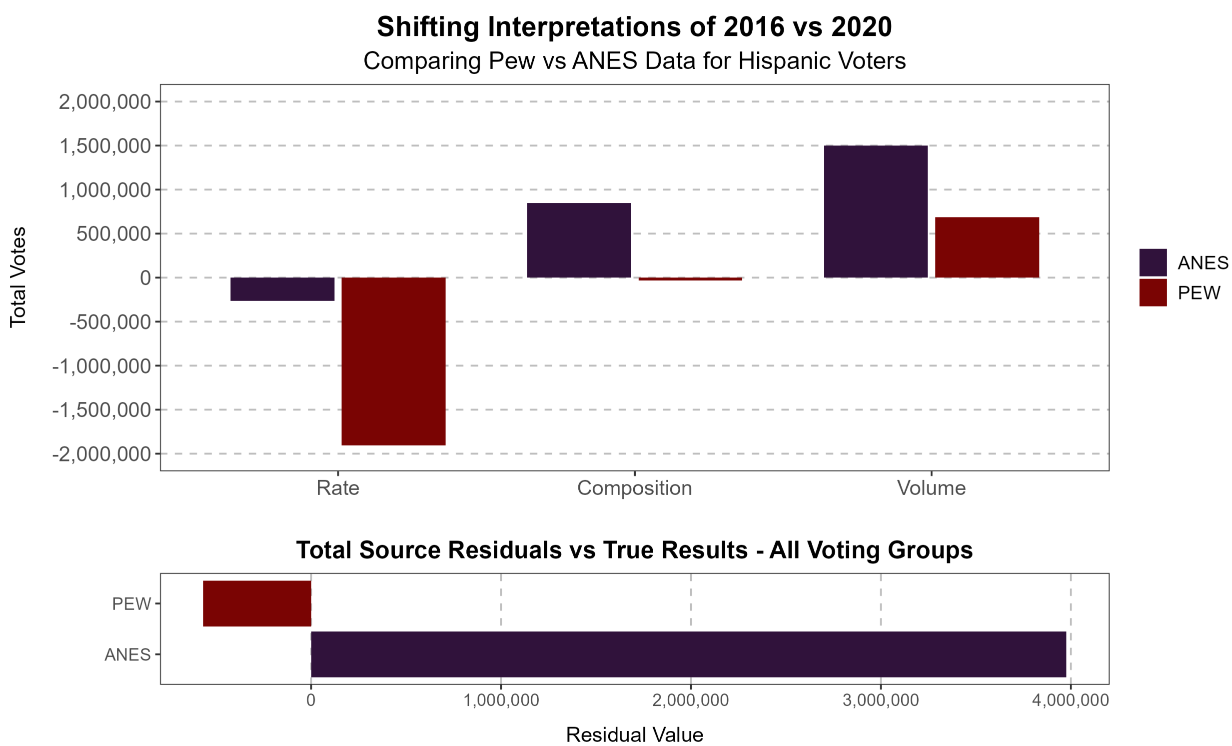

Figure 7: This figure compares the implied election results from ANES data versus Pew data. The upper portion of the figure shows that utilizing Pew data fundamentally alters the interpretation of the impact of Hispanic voters on the 2020 election. In particular, it shows that there is no evidence that Democrats benefited from a shift into Hispanic voters in 2020 and instead suffered a significant rate decline. The lower portion of the figure shows that the implied error in votes of using ANES data to evaluate the 2020 election is nearly 4 million votes, whereas using the Pew data has an error of less than 600,000.

This difference in interpretation carries strategic implications. If the change is mostly compositional (as ANES suggests), the Democratic challenge is primarily organizational. If the change is driven by rate effects (as Pew suggests), the challenge is fundamentally one of persuasion. The Pew data thus refutes the notion that Democrats can rely on demographic changes alone (Klein, 2024), revealing deeper vulnerability to Republican outreach than previously recognized.

This comparison serves as a methodological caution: researchers should carefully assess whether their data reflects reality, as argued in Grimmer, Marble, et al. (2022). By embracing the superior descriptive validity of Pew for the Hispanic electorate, the RCVD framework establishes that Hispanic volatility is driven less by passive demographic turnover and more by active shifts in partisan loyalty—a crucial baseline for understanding the dramatic shifts in the 2024 election.

The Full Picture: Decomposition Across Three Electoral Cycles

I now apply the RCVD framework to analyze vote change across three electoral cycles: 2016–2020, 2020–2024, and the full 2016–2024 interval. This analysis demonstrates how the compounding effects of rate and composition determined the 2024 electoral outcome.

The 2016 to 2020 Cycle— Figure 8

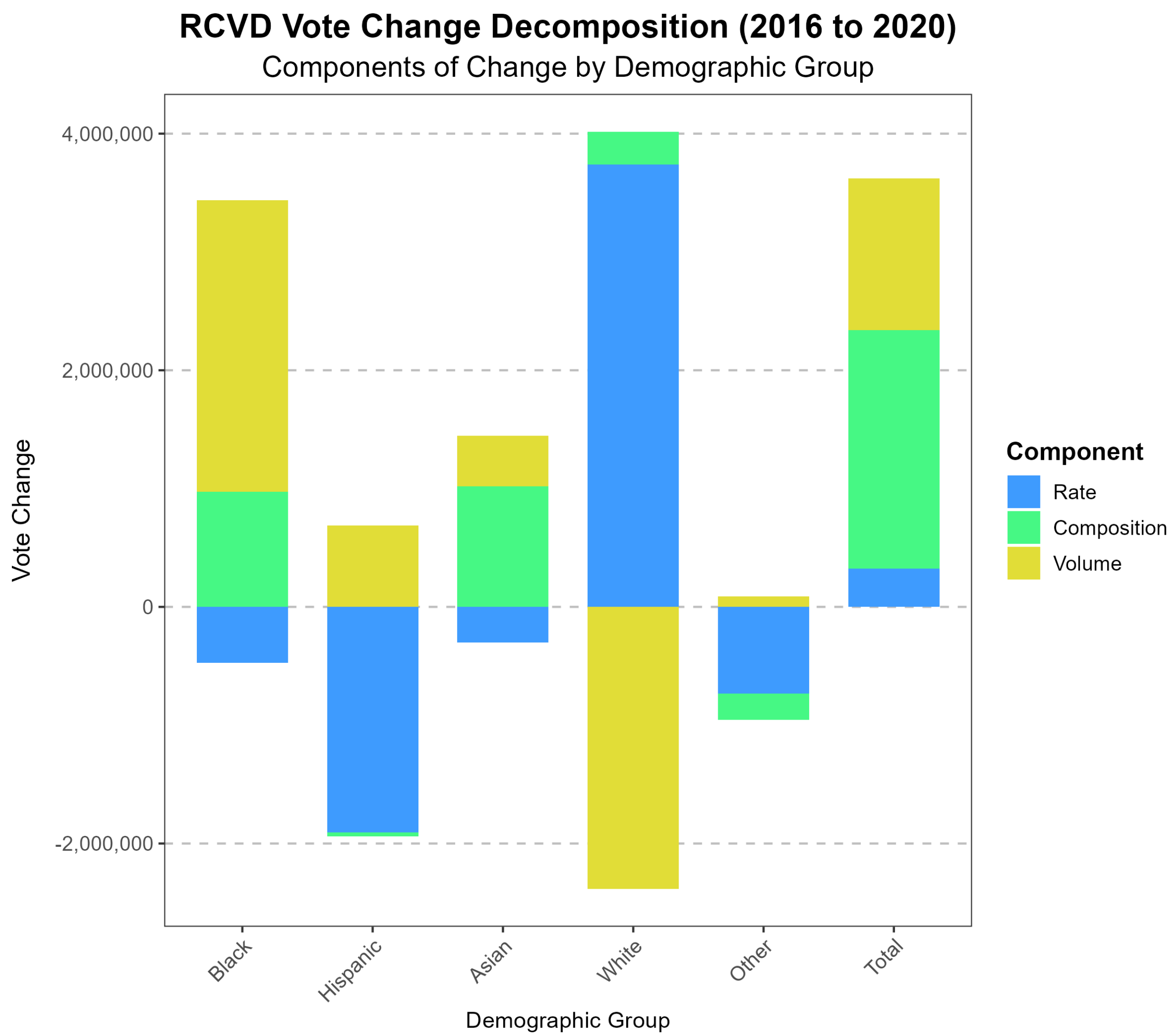

Figure 8: This figure provides an evaluation of the 2020 election compared to the 2016 election using Pew data and the RCVD approach. It shows that composition had a large positive impact for Democrats as a result of shifting out of White voters and into Black and Asian voters. There was also a large positive impact for Democrats driven by improved margins among White voters. However, this impact was almost entirely offset by declining Democratic margins among Hispanic voters, and to a lesser extent among Black, Asian and other voters.

The 2016 to 2020 period represents the Democratic bounce-back from their 2016 loss. While accounting for the translation of the popular vote to the Electoral College involves challenges (Grimmer, Knox, et al., 2024), Democrats clearly performed better in 2020. The victory was characterized by favorable compositional forces (increasing share of non-White voters, particularly Black and Asian voters) combined with positive rate effects among White voters. The RCVD decomposition confirms that mobilization efforts and demographic trends worked synergistically with preference shifts to yield a substantial Democratic gain. This period represents the high-water mark of the conventional wisdom that demographic change inherently favors Democrats. However, even in 2020, warning signs appeared: all non-White groups voted for Democrats at lower rates, with Democratic vote share among Hispanic voters declining noticeably (from 72% to 61%, a margin shift from +44% to +22%).

The Critical 2020 to 2024 Swing— Figure 9

The 2020 to 2024 period represents a critical realignment that challenges established assumptions about the stability of the non-White Democratic coalition. The RCVD analysis reveals a dramatic reversal: Democrats suffered a severe net loss attributable overwhelmingly to the rate component across all non-White voting groups.

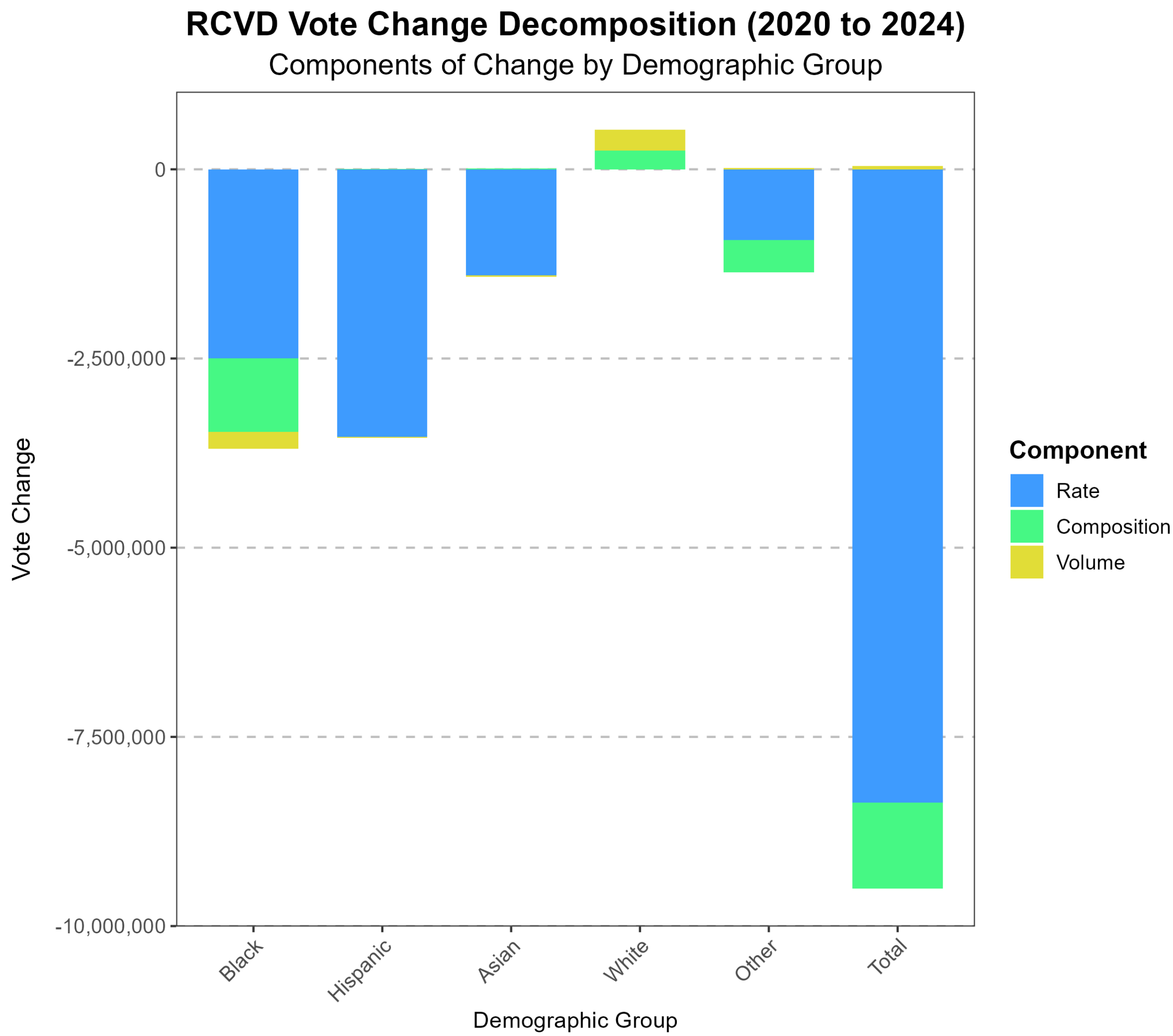

Figure 9: This figure provides an analysis of the 2024 election compared to the 2020 election. It shows that Democrats suffered a dramatic decrease in vote margins among all non-White voting groups, most prominently among Hispanic voters. The decline in support was so significant that even with continued shifts away from White voters amongst the composition of the electorate, the total impact of composition in 2024 was a net negative for Democrats.

The rate effect was negative and substantial among Hispanic and Black voters, signaling erosion of Democratic loyalty and successful Republican persuasion. This negative rate swing was so significant that, even with continued shifts away from White voters in the electorate's composition, the total composition effect in 2024 was net negative for Democrats. This finding underscores the core mechanism of the 2024 result: declining base loyalty (Rate) entirely neutralized and then overcame the structural advantages of demographic change (composition).

The Cumulative 2016 to 2024 Arc—Figure 10

The eight-year view from 2016 to 2024 provides the final synthesis. While the analysis demonstrates some favorable impacts from composition over the long run, this benefit was dwarfed by the cumulative negative trend in the rate component.

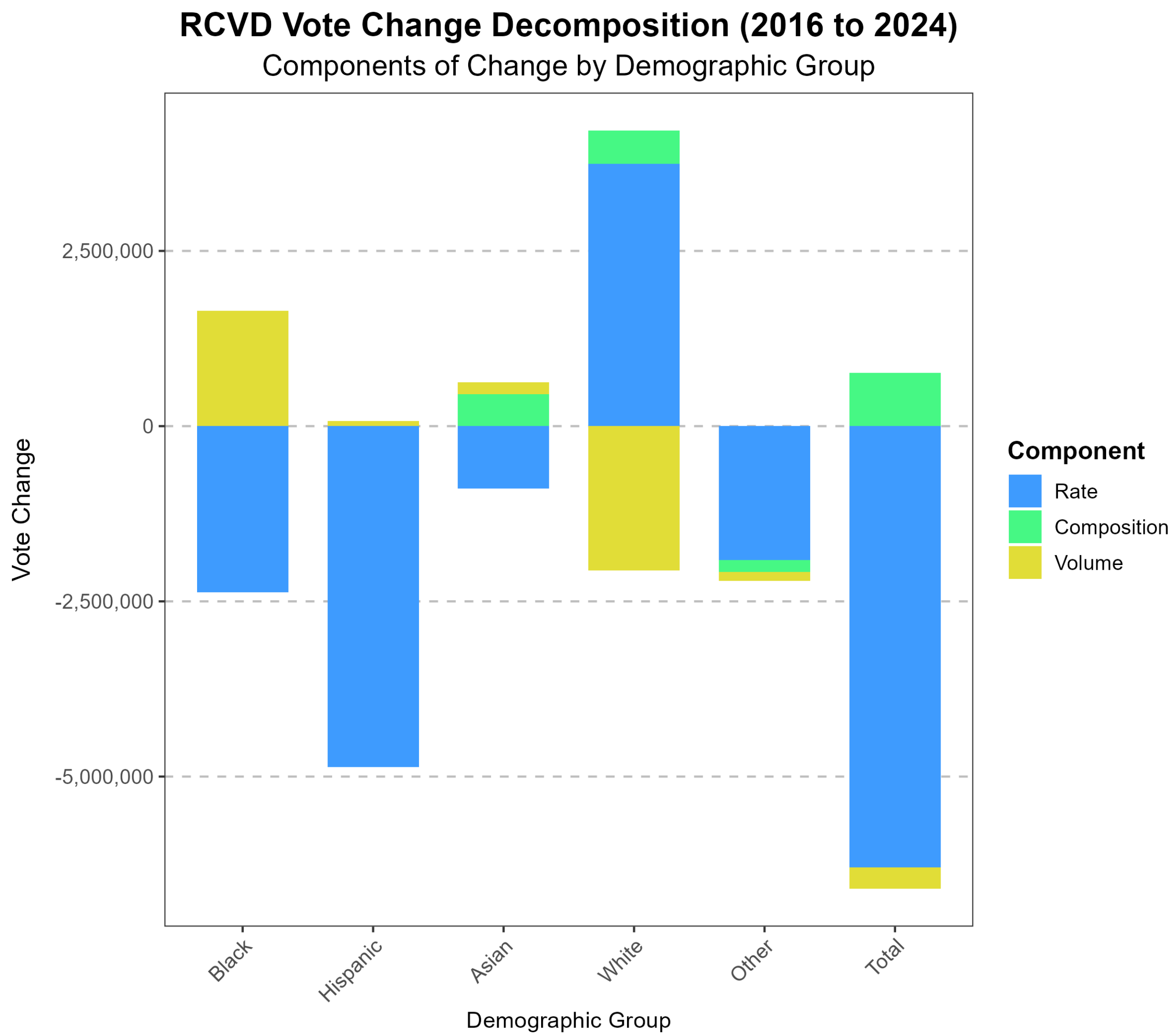

Figure 10: This figure compares the 2024 election to the 2016 election. It highlights the strong negative trend in Democratic vote margins among Hispanic voters, and to a lesser extent among other non-White voters. While this analysis demonstrates some favorable impacts from composition, it highlights the shortcomings of a reliance on these demographic shifts to promote future success for the Democratic party.

The negative rate trend among Hispanic voters is the single most important factor identified by the RCVD framework over this period. While Democrats showed positive growth among White voters, the framework cleanly partitions the positive effect of an increasing non-White share from the negative effect of these groups becoming less reliably Democratic. The final outcome is explained not by failure of the electorate to grow (volume) or shift demographically (composition), but by failure of the Democratic Party to retain the loyalty of the electorate that did grow (Rate).

The application of RCVD across these three cycles demonstrates that the 2024 result was driven by a decisive decoupling of composition and rate effects. The long-predicted demographic dividend (composition) was neutralized by a collapse of partisan preference (Rate) within the fastest-growing segments of the electorate, a phenomenon masked by aggregate analysis and only revealed by the rate, composition, and Volume Decomposition.

5 Conclusion

As demographic shifts reshape the U.S. voting landscape, understanding how changes between groups affect election outcomes is crucial for both scholars and policymakers. Previous methods for explaining these shifts have introduced significant interpretive errors, frequently amounting to millions of misallocated votes, due to the conflation of preference shifts with compositional change. The Rate, Composition, and Volume Decomposition (RCVD) approach, as defined and applied in this paper, resolves this fundamental ambiguity. It provides a zero-loss methodology that precisely specifies how changes in partisan preference (Rate), the relative size of groups (composition), and overall participation (volume) influence electoral outcomes. The RCVD framework is broadly applicable, extending beyond U.S. elections to any context where subdividing populations into distinct groups helps explain dynamics in time-series cross-sectional data.

My empirical application of RCVD to the 2016, 2020, and 2024 presidential elections yields a core finding that fundamentally reconfigures the prevailing ”demographics are destiny” narrative. The 2024 election was driven not by the expected long-term compositional benefit for the Democratic Party, but by a decisive decoupling of composition and rate effects. The framework isolates a substantial negative rate effect—the loss of partisan loyalty—among the fastest-growing non-White electorates, particularly Hispanic voters. This preference volatility neutralized the structural compositional advantage, confirming that electoral success depends less on passive demographic shifts and more on active defense of coalition loyalty. The RCVD framework is thus crucially diagnostic, directing attention from organizational success (volume and composition) toward persuasion challenges (Rate).

Future work will explore the robustness and granularity of the RCVD approach in two key areas. First, I will investigate its ability to account for nested subgroups, demonstrating how shifts within smaller partitions—such as regional or age-based splits within the Hispanic electorate contribute to broader electoral changes, building on the work of Fraga, Y. R. Velez, et al. (2024). Second, I will extend the analysis across multiple non-presidential time periods to capture long-term trends, showing how RCVD can reveal the evolving durability and fragility of group shifts over time. These extensions will further validate RCVD as a tool for analyzing historical shifts in political support and for understanding electorate dynamics in future elections.

References

Achen, Christopher H. and Larry M. Bartels (2016). *Democracy for realists: why elections do not produce responsive government*. Princeton Studies in Political Behavior. Princeton: Princeton University Press. ISBN: 9780691169446.

ANES (2021). *American National Election Studies*. ANES 2020 Time Series Study Full Release [dataset and documentation]. February 10, 2022 version.

Axelrod, Robert (Mar. 1972). "Where the Votes Come From: An Analysis of Electoral Coalitions, 1952–1968". en. In: *American Political Science Review* 66.1, pp. 11–20. ISSN: 0003-0554, 1537-5943. DOI: 10.2307/1959275. URL: https://www.cambridge.org/core/product/identifier/S0003055400138882/type/journal_article (visited on 03/12/2024).

Calvo, Richard, Vincent Pons, and Jesse Shapiro (Oct. 2024). *Pitfalls of Demographic Forecasts of US Elections*. en. Tech. rep. w33016. Cambridge, MA: National Bureau of Economic Research, w33016. DOI: 10.3386/w33016. URL: http://www.nber.org/papers/w33016.pdf (visited on 10/16/2025).

Campbell, Angus, ed. (1960). *The American voter*. eng. 2. Aufl. Chicago, Ill.: Univ. of Chicago Press. ISBN: 978-0-226-09253-9.

Carmines, Edward G., Michael J. Ensley, and Michael W. Wagner (Dec. 2016). "Ideological Heterogeneity and the Rise of Donald Trump". en. In: *The Forum* 14.4, pp. 385–397. ISSN: 1540-8884. DOI: 10.1515/for-2016-0036. URL: https://www.degruyter.com/document/doi/10.1515/for-2016-0036/html?lang=en (visited on 03/12/2024).

Citrin, Jack, Eric Schickler, and John Sides (Jan. 2003). "What if Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections". en. In: *American Journal of Political Science* 47.1, pp. 75–90. ISSN: 0092-5853, 1540-5907. DOI: 10.1111/1540-5907.00006. URL: https://onlinelibrary.wiley.com/doi/10.1111/1540-5907.00006 (visited on 10/13/2024).

Engelhardt, Andrew M. (July 2019). "Trumped by Race: Explanations for Race's Influence on Whites' Votes in 2016". In: *Quarterly Journal of Political Science* 14.3, pp. 313–328. ISSN: 15540634. DOI: 10.1561/100.00018068. URL: http://www.nowpublishers.com/article/Details/QJPS-18068 (visited on 03/12/2024).

Fraga, Bernard (Jan. 2016). "Candidates or Districts? Reevaluating the Role of Race in Voter Turnout". en. In: *American Journal of Political Science* 60.1, pp. 97–122. ISSN: 0092-5853, 1540-5907. DOI: 10.1111/ajps.12172. URL: https://onlinelibrary.wiley.com/doi/10.1111/ajps.12172 (visited on 10/31/2025).

Fraga, Bernard, Sean McElwee, et al. (Dec. 2021). "Analysis | Why did Trump win? More whites — and fewer blacks — actually voted." en-US. In: *Washington Post*. ISSN: 0190-8286. URL: https://www.washingtonpost.com/news/monkey-cage/wp/2017/05/08/why-did-trump-win-more-whites-and-fewer-blacks-than-normal-actually-voted/ (visited on 03/12/2024).

Fraga, Bernard, Yamil Velez, and Emily West (Aug. 2023). *Reversion to the Mean, or their Version of the Dream? An Analysis of Latino Voting in 2020*. en. DOI: 10.33774/apsa-2023-764r1. URL: https://preprints.apsanet.org/engage/apsa/article-details/64c9451d9ed5166e93de2210 (visited on 04/16/2024).

Fraga, Bernard, Yamil R. Velez, and Emily A. West (May 2024). "Reversion to the Mean, or Their Version of the Dream? Latino Voting in an Age of Populism". en. In: *American Political Science Review*, pp. 1–9. ISSN: 0003-0554, 1537-5943. DOI: 10.1017/S0003055424000406. URL: https://www.cambridge.org/core/product/identifier/S0003055424000406/type/journal_article (visited on 12/08/2024).

Grimmer, Justin, Dean Knox, and Sean Westwood (Aug. 2024). *Assessing the Reliability of Probabilistic US Presidential Election Forecasts May Take Decades*. DOI: 10.31219/osf.io/6g5zq. URL: https://osf.io/6g5zq (visited on 12/17/2024).

Grimmer, Justin, William Marble, and Cole Tanigawa-Lau (June 2022). *Measuring the Contribution of Voting Blocs to Election Outcomes*. preprint. SocArXiv. DOI: 10.31235/osf.io/c9fkg. URL: https://osf.io/c9fkg (visited on 02/18/2024).

Hartig, Hannah et al. (July 2023). *Republican Gains in 2022 Midterms Driven Mostly by Turnout Advantage*. Tech. rep. Pew Research Center. URL: https://www.pewresearch.org/politics/2023/07/12/republican-gains-in-2022-midterms-driven-mostly-by-turnout-advantage/ (visited on 07/03/2025).

Hersh, Eitan (2015). *Hacking the electorate: how campaigns perceive voters*. New York, NY: Cambridge University Press. ISBN: 9781107102897 9781107501164.

Hill, Seth J., Daniel J. Hopkins, and Gregory A. Huber (Apr. 2021). "Not by turnout alone: Measuring the sources of electoral change, 2012 to 2016". en. In: *Science Advances* 7.17, eabe3272. ISSN: 2375-2548. DOI: 10.1126/sciadv.abe3272. URL: https://www.science.org/doi/10.1126/sciadv.abe3272 (visited on 02/18/2024).

Klein, Ezra (Jan. 2024). *The Ezra Klein Show: 'The Strongest Democratic Party That Any of Us Have Ever Seen' on Apple Podcasts*. en-US. URL: https://podcasts.apple.com/us/podcast/the-strongest-democratic-party-that-any-of-us-have-ever-seen/id1548604447?i=1000642890587 (visited on 05/08/2024).

Lemi, Danielle Casarez (Dec. 2021). "Do Voters Prefer Just Any Descriptive Representative? The Case of Multiracial Candidates". en. In: *Perspectives* *on Politics* 19.4, pp. 1061–1081. ISSN: 1537-5927, 1541-0986. DOI: 10.1017/S1537592720001280. URL: https://www.cambridge.org/core/product/identifier/S1537592720001280/type/journal_article (visited on 09/30/2024).

Lopez, Mark Hugo and Paul Taylor (2009). *Dissecting the 2008 Electorate: Most Diverse in U.S. History*. Tech. rep. Pew Research Center. URL: https://www.pewresearch.org/race-and-ethnicity/2009/04/30/dissecting-the-2008-electorate-most-diverse-in-us-history/ (visited on 10/12/2024).

Marble, William, Justin Grimmer, and Cole Tanigawa-Lau (Sept. 2024). "Measuring the Contribution of Voting Blocs to Election Outcomes". en. In: *The Journal of Politics*, p. 732964. ISSN: 0022-3816, 1468-2508. DOI: 10.1086/732964. URL: https://www.journals.uchicago.edu/doi/10.1086/732964 (visited on 09/30/2024).

Reed, Michael (1998). *Fundamental ideas of analysis*. eng. New York Chichester: John Wiley. ISBN: 978-0-471-15996-4.

Sides, John, Michael Tesler, and Lynn Vavreck (Dec. 2019). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton University Press. ISBN: 978-1-4008-8830-6. DOI: 10.1515/9781400888306. URL: https://www.degruyter.com/document/doi/10.1515/9781400888306/html (visited on 10/13/2024).

Wooldridge, Jefrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. ISBN: 9780262232586. URL: https://www.jstor.org/stable/j.ctt5hhcfr (visited on 03/20/2024).

Wooley, John and Gerhard Peters (2025). "The American Presidency Project." In: (). URL: https://www.presidency.ucsb.edu/ (visited on 10/31/2025).

Woolley, John and Gerhard Peters (2021). *Voter Turnout in Presidential Elections*. URL: https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections (visited on 10/12/2024).

Zingher, Joshua N. (Aug. 2019). "An analysis of the changing social bases of America's political parties: Group support in the 2012 and 2016 presidential elections". en. In: *Electoral Studies* 60, p. 102042. ISSN: 02613794. DOI: 10.1016/j.electstud.2019.04.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0261379419300253 (visited on 02/18/2024).

Appendix

On Overfitting

While zero-error terms in regression models often raise concerns about overfitting (Wooldridge, 2010), the RCVD approach is not subject to these risks. As it is not a regression or estimation technique, but rather an accounting framework, zero-error reflects the accurate decomposition of shifts rather than problematic estimation.

Proof of Equation 3

$$
\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
&= r(t_2)c(t_2) - r(t_1)c(t_1) \\
r(t_2)c(t_2) - r(t_1)c(t_1) &= r(t_2)c(t_2) + r(t_1)c(t_1) - r(t_1)c(t_1) + r(t_2)c(t_1) - r(t_2)c(t_1) + r(t_1)c(t_2) - r(t_1)c(t_2) - r(t_1)c(t_1) \\
&= [r(t_2)c(t_1) - r(t_1)c(t_1)] + [r(t_1)c(t_2) - r(t_1)c(t_1)] + [r(t_2)c(t_2) - r(t_1)c(t_2) - r(t_2)c(t_1) + r(t_1)c(t_1)] \\
r(t_2)c(t_2) - r(t_1)c(t_1) &= [r(t_2) - r(t_1)]c(t_1) + r(t_1)[c(t_2) - c(t_1)] + [r(t_2) - r(t_1)][c(t_2) - c(t_1)]
\end{aligned}
$$

Proof of Equation 4

$$
\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
&= r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) \\
r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) &= r(t_2)c(t_2)v(t_2) - r(t_2)c(t_1)v(t_1) + r(t_2)c(t_1)v(t_1) \\
&- r(t_1)c(t_2)v(t_1) + r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_2) + r(t_1)c(t_1)v(t_2) \\
&- r(t_2)c(t_2)v(t_1) + r(t_2)c(t_2)v(t_1) - r(t_2)c(t_1)v(t_2) + r(t_2)c(t_1)v(t_2) \\
&- r(t_1)c(t_2)v(t_2) + r(t_1)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) \\
&= 3r(t_1)c(t_1)v(t_1) - 3r(t_1)c(t_1)v(t_1) \\
&+ r(t_2)c(t_1)v(t_1) - 2r(t_2)c(t_1)v(t_1) \\
&+ r(t_1)c(t_2)v(t_1) - 2r(t_1)c(t_2)v(t_1) \\
&+ r(t_1)c(t_1)v(t_2) - 2r(t_1)c(t_1)v(t_2) \\
&+ r(t_2)c(t_2)v(t_1) + r(t_2)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_2) \\
&+ [r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) - r(t_2)c(t_1)v(t_2) + r(t_2)c(t_1)v(t_1) \\
&- r(t_1)c(t_2)v(t_2) + r(t_1)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
&= [r(t_2)c(t_1)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
&+ [r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)]
\end{aligned}
$$

$$
\begin{aligned}
&+ [r(t_1)c(t_1)v(t_2) - r(t_1)c(t_1)v(t_1)] \\
&+ [r(t_2)c(t_2) - r(t_2)c(t_1) - r(t_1)c(t_2) + r(t_1)c(t_1)]v(t_1) \\
&+ [r(t_2)v(t_2) - r(t_2)v(t_1) - r(t_1)v(t_2) + r(t_1)v(t_1)]c(t_1) \\
&+ [c(t_2)v(t_2) - c(t_2)v(t_1) - c(t_1)v(t_2) + c(t_1)v(t_1)]r(t_1) \\
&+ [r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) - r(t_2)c(t_1)v(t_2) + r(t_2)c(t_1)v(t_1) \\
&- r(t_1)c(t_2)v(t_2) + r(t_1)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)]
\end{aligned}
$$

$$
\begin{aligned}
r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) &= [r(t_2) - r(t_1)]c(t_1)v(t_1) + r(t_1)[c(t_2) - c(t_1)]v(t_1) + r(t_1)c(t_1)[v(t_2) - v(t_1)] \\
&+ [r(t_2) - r(t_1)][c(t_2) - c(t_1)]v(t_1) + [r(t_2) - r(t_1)]c(t_1)[v(t_2) - v(t_1)] + r(t_1)[c(t_2) - c(t_1)][v(t_2) - v(t_1)] \\
&+ [r(t_2) - r(t_1)][c(t_2) - c(t_1)][v(t_2) - v(t_1)]
\end{aligned}
$$

Proof of Equation 4 with Description and Color Coding:

$$
\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
&= r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1)
\end{aligned}
$$

The core expansion begins here. I begin by color coding terms which are introduced that cancel

$$
\begin{aligned}
r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) &= r(t_2)c(t_2)v(t_2) + 2[r(t_2)c(t_1)v(t_1) - r(t_2)c(t_1)v(t_1)] \\
&+ 2[r(t_1)c(t_2)v(t_1) - r(t_1)c(t_2)v(t_1)] + 2[r(t_1)c(t_1)v(t_2) - r(t_1)c(t_1)v(t_2)] \\
&+ [r(t_2)c(t_2)v(t_1) - r(t_2)c(t_2)v(t_1)] + [r(t_2)c(t_1)v(t_2) - r(t_2)c(t_1)v(t_2)] \\
&+ [r(t_1)c(t_2)v(t_2) - r(t_1)c(t_2)v(t_2)] + 3[r(t_1)c(t_1)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
&- r(t_1)c(t_1)v(t_1)
\end{aligned}
$$

After rearranging, I drop colors for terms that will be completely reconfigured, only maintaining color groupings for terms that are essentially finalized

$$
\begin{aligned}
= &\ 3r(t_1)c(t_1)v(t_1) - 3r(t_1)c(t_1)v(t_1) \\
&+ r(t_2)c(t_1)v(t_1) - 2r(t_2)c(t_1)v(t_1) \\
&+ r(t_1)c(t_2)v(t_1) - 2r(t_1)c(t_2)v(t_1) \\
&+ r(t_1)c(t_1)v(t_2) - 2r(t_1)c(t_1)v(t_2) \\
&+ r(t_2)c(t_2)v(t_1) + r(t_2)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_2) \\
&+ [r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) - r(t_2)c(t_1)v(t_2) + r(t_2)c(t_1)v(t_1) \\
&- r(t_1)c(t_2)v(t_2) + r(t_1)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)]
\end{aligned}
$$

After additional rearranging, I apply color groupings for terms as they will appear in the final proof

$$
\begin{aligned}
&= [r(t_2)c(t_1)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
&\quad + [r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
&\quad + [r(t_1)c(t_1)v(t_2) - r(t_1)c(t_1)v(t_1)] \\
&\quad + [r(t_2)c(t_2) - r(t_2)c(t_1) - r(t_1)c(t_2) + r(t_1)c(t_1)]v(t_1) \\
&\quad + [r(t_2)v(t_2) - r(t_2)v(t_1) - r(t_1)v(t_2) + r(t_1)v(t_1)]c(t_1) \\
&\quad + [c(t_2)v(t_2) - c(t_2)v(t_1) - c(t_1)v(t_2) + c(t_1)v(t_1)]r(t_1) \\
&\quad + [r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) - r(t_2)c(t_1)v(t_2) + r(t_2)c(t_1)v(t_1) \\
&\quad -r(t_1)c(t_2)v(t_2) + r(t_1)c(t_1)v(t_2) + r(t_1)c(t_2)v(t_1) - r(t_1)c(t_1)v(t_1)] \\
r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) &= [r(t_2) - r(t_1)]c(t_1)v(t_1) + r(t_1)[c(t_2) - c(t_1)]v(t_1) + r(t_1)c(t_1)[v(t_2) - v(t_1)] \\
&\quad + [r(t_2) - r(t_1)][c(t_2) - c(t_1)]v(t_1) + [r(t_2) - r(t_1)]c(t_1)[v(t_2) - v(t_1)] \\
&\quad + r(t_1)[c(t_2) - c(t_1)][v(t_2) - v(t_1)] \\
&\quad + [r(t_2) - r(t_1)][c(t_2) - c(t_1)][v(t_2) - v(t_1)] \Rightarrow \\
\Delta z(t) &= \Delta r(t)c(t_1)v(t_1) + r(t_1)\Delta r(t)v(t_1) + r(t_1)c(t_1)\Delta v(t) \\
&\quad + \Delta r(t)\Delta c(t)v(t_1) + \Delta r(t)c(t_1)\Delta v(t) \\
&\quad + r(t_1)\Delta c(t)\Delta v(t) \\
&\quad + \Delta r(t)\Delta c(t)\Delta v(t)
\end{aligned}
$$

Proof of Key Equation 7

$$
\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
z(t_2) - z(t_1) &= r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) \\
&= r(t_2)[c(t_1)v(t_1) - c(t_1)v(t_1) + c(t_2)v(t_1) - c(t_2)v(t_1) + c(t_2)v(t_2)] - \\
&r(t_1)c(t_1)v(t_1) \\
&= r(t_2)c(t_1)v(t_1) - r(t_1)c(t_1)v(t_1) + r(t_2)c(t_2)v(t_1) - \\
&r(t_2)c(t_1)v(t_1) + r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) \\
&= [r(t_2) - r(t_1)]c(t_1)v(t_1) + \\
&r(t_2)[c(t_2) - c(t_1)]v(t_1) + \\
&r(t_2)c(t_2)[v(t_2) - v(t_1)]
\end{aligned}
\tag{7}
$$

Simple Example Data

Table A1: Concept Demonstration

Method	Group	Rate	Composition	Volume	Calc Mar Chng	Var to Act	Variance %
Derivative Calculation	Non-White	−2.76	2.74	4.43	4.42	0.42	10.4
	White	6.11	0.70	−2.82	3.99	−1.01	−20.1
	Total	3.35	3.44	1.61	8.41	−0.59	−6.5
RCV Calculation	Non-White	−2.76	2.40	4.36	4.00	0.00	0.0
	White	6.11	0.40	−1.51	5.00	0.00	0.0
	Total	3.35	2.79	2.85	9.00	0.00	0.0
Comparing the Two Approaches	Non-White		0.34	0.07	0.42		
	White		0.31	−1.31	−1.01		
	Total		0.65	−1.24	−0.59		

Note: All numbers are in millions. Variance percentages are shown with one decimal place.

The Problem is Not Volume: Isolating Rate-Composition Misspecification

In the preceding analysis, the LDA produced a significant Residual, partially driven by unassigned interaction effects and partially by the unmodeled Volume term. The traditional literature often assumes this error is synonymous with unexplainable variance or Volume. In this section, I directly challenge that assumption by demonstrating that the LDA's fundamental flaw lies in its handling of the Rate and Composition interaction, even when Volume is perfectly controlled.

To achieve this isolation, I analyze a hypothetical, zero-sum vote shift that transforms the actual 2020 Biden popular vote victory into a narrow Trump victory. I ensure this shift is achieved by manipulating only the Rate (group vote loyalty) and Composition (group size) parameters, while holding the total Volume of votes constant. Since changes in Volume can be modeled as a linear transformation (an eigenvalue applied to the data that perfectly preserves underlying relationships), controlling Volume to zero ensures that any resulting error must be driven exclusively by the misspecification of the Rate and Composition interaction.

Table A2 presents this hypothetical scenario, utilizing a five-group breakdown of the electorate based on data from Marble et al. (2024). I achieve the synthetic shift by multiplying the Rate of support for the Democratic candidate by a factor of .93 and shifting the Composition of all non-White groups by a factor of .8. The final line of the 'Total' column confirms the successful manipulation: the net change in total votes across the two scenarios is exactly 0.00 million votes. Therefore, in this controlled, synthetic example, the only two theoretical buckets for explaining the resulting ≈ 17.75 million vote margin shift are Rate within groups and Composition between them.

Table A2: Hypothetical Data, Both Rate and Composition

Election	Group	Candidate 1	Candidate 2	Total	Gross Margin	Trad Mar (%)	Trad Comp (%)
1st Election (2020)	Black	15.42	1.47	16.89	13.95	82.60	10.89
	Hispanic	13.21	4.47	17.68	8.73	49.40	11.41
	Other	8.64	5.16	13.80	3.48	25.20	8.90
	White	47.95	57.21	105.16	−9.25	−8.80	67.84
	NA	0.80	0.67	1.47	0.13	8.60	0.95
	Total	86.02	68.98	155.00	17.03	10.99	100.00
2nd Election (2020)	Black	11.47	2.04	13.51	9.43	69.82	8.72
	Hispanic	9.83	4.32	14.14	5.51	38.94	9.12
	Other	6.43	4.61	11.04	1.81	16.44	7.12
	White	48.82	66.30	115.13	−17.48	−15.18	74.28
	NA	0.60	0.58	1.18	0.01	1.00	0.76
	Total	77.14	77.86	155.00	−0.71	−0.46	100.00
Comparing the Two Elections	Black	−3.95	0.57	−3.38	−4.52	−12.78	−2.18
	Hispanic	−3.38	−0.16	−3.54	−3.23	−10.46	−2.28
	Other	−2.21	−0.55	−2.76	−1.66	−8.76	−1.78
	White	0.87	9.10	9.97	−8.23	−6.38	6.43
	NA	−0.20	−0.09	−0.29	−0.11	−7.60	−0.19
	Total	−8.87	8.87	0.00	−17.75	−11.45	0.00

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.

Figure 11 highlights the critical methodological result. Despite the total change in votes being zero, the LDA still produces a massive Mean Absolute Error (MAE) of 1.7 million votes. This unallocated error is caused entirely by the LDA's inability to correctly model the non-linear interaction between Rate and Composition. By comparison, the RCVD framework correctly accounts for all votes in the system, yielding an MAE of 0.

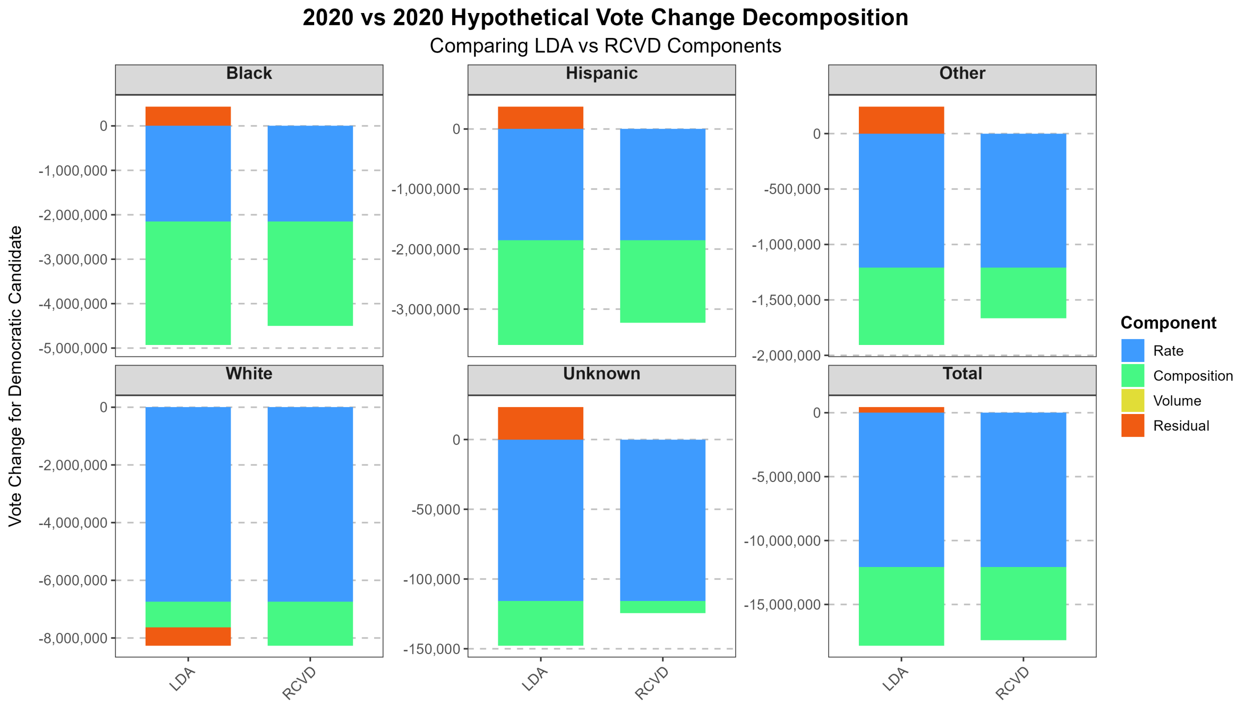

Figure 11: This figure shows that even when Volume is not a factor in evaluating the shifts in an election, the LDA creates significant errors in attribution of vote changes. Here, the total absolute error is 1.7 million votes. The underlying numbers reflected in this figure are in Appendix A.

This finding is definitive: The 1.7 million vote error is not due to the omission of Volume, but is a direct consequence of the LDA's formula. By comparing the LDA and RCVD, it is clear that all 1.7 million unidentified votes come from a misspecification of the impact of Composition and Rate, distributed across the groups in increments of hundreds of thousands of votes.

In summary, this scenario definitively refutes the implicit assumption that the LDA's unexplained variance is merely Volume or general error. It validates the necessity of the zero-loss RCVD framework by demonstrating that even in a highly controlled environment where Volume is eliminated as a potential explanatory variable, the LDA fundamentally fails to correctly partition the vote change between Rate and Composition.

Non-Uniqueness and the Fixed RCVD Path

As noted in the conclusion of Section 2, the non-linear nature of the Rate, Composition, and Volume Decomposition (RCVD) means that the order of calculation is not unique. Mathematically, there are six possible permutations of the Rate (ΔR), Composition (ΔC), and Volume (ΔV) terms that satisfy the zero-loss property of the RCVD framework. This path-dependence is a necessary artifact of correctly modeling the interaction effects that the DBA previously relegated to the unassigned Residual. Therefore, there needs to be a strong theoretical justification for the chosen order ($\Delta R \to \Delta C \to \Delta V$).

Figure 12 illustrates an alternative zero-loss order (Composition → Rate → Volume) when applied to the simple 2016-2020 simulation. While the total MAE remains 0, the calculated contribution of Rate and Composition shifts: the alternative order assigns a slightly larger role to Composition and a slightly smaller role to Rate.

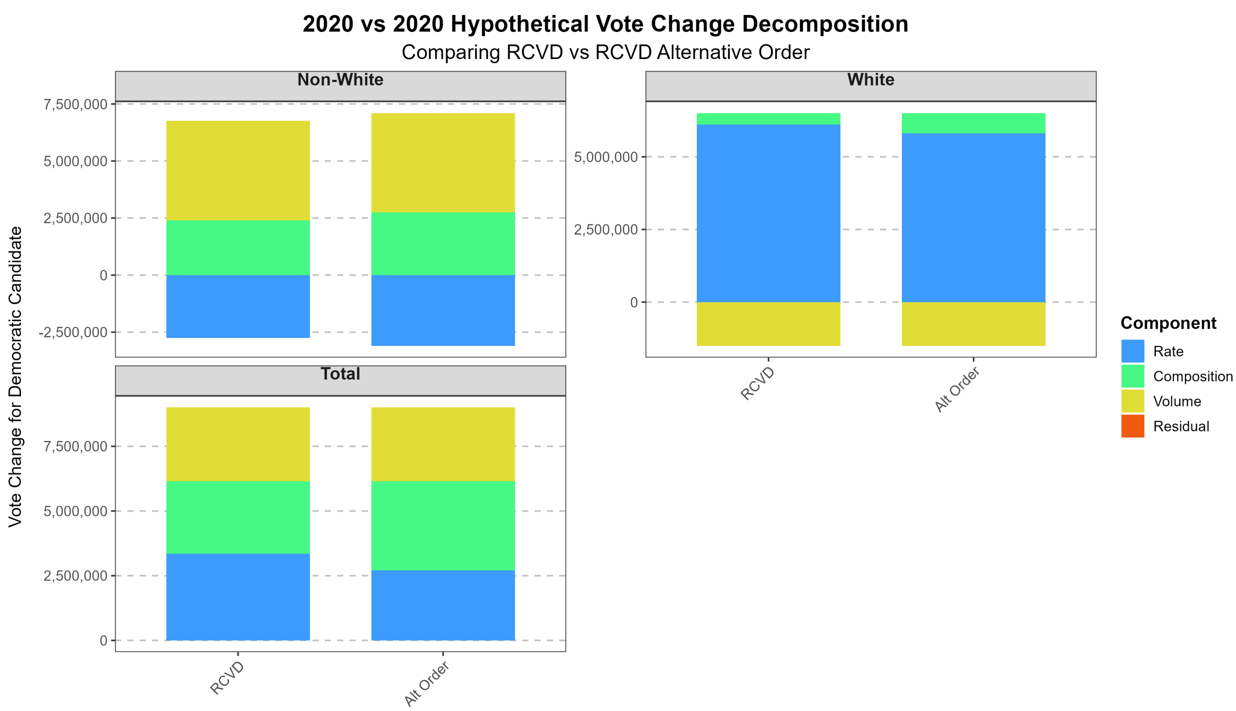

Figure 12: This figure shows that shifting the order in which RCVD is calculated yields no error term, but can shift the interpretation of the impacts of both Rate and Composition.

Despite the relative closeness of these results in the simple case, I adhere to the

$\Delta R \rightarrow \Delta C \rightarrow \Delta V$ order for two core reasons related to the underlying political theory. First, this order preserves Rate's causal primacy. The convention in electoral studies has always been to calculate the change in group loyalty (Rate) based on the prior period's data. Maintaining this sequence ensures that the ΔR term represents the clean, unadulterated shift in group preference, free from the simultaneous influence of demographic change. This allows our ΔR term to retain the exact same highly intuitive value calculated by the DBA, validating the existing literature while correcting the other terms.

Second, this order maintains Volume as an eigenvalue. By sequencing Volume last, we ensure it serves as an eigenvalue—a perfect proportional shift that occurs after all components affecting the margin have been calculated. Intuitively, shifting Composition first (e.g., more non-White voters) will ultimately change the final national aggregate Rate, even if the Rates within each group do not change. By placing ΔC before ΔV, we ensure that Volume is purely an increase or decrease in the system's size, distributed proportionally across the new electoral configuration set by Rate and Composition. This preserves the most theoretically clean interpretation of Volume.

In sum, the $\Delta R \rightarrow \Delta C \rightarrow \Delta V$ ordering is not arbitrary, but a deliberate choice that preserves the most theoretically clean and consistent interpretation of Rate against existing literature, while ensuring Composition and Volume are correctly specified.

Rate Only and Composition Only

Table A3 shows how Rates would have had to have shifted in order to guarantee a win for Trump in 2020 without any Composition shifts. To achieve this effect, I multiplied Biden's margin Rate within each group by a factor of .9, preserving the relative Rates at which each group voted for Biden, but reducing his vote totals. As can be seen, this shift is sufficient to shift just over 17.2 million votes to Trump versus the actual reported result.

Conversely, Table A4 demonstrates how Composition would have had to shift, without any changes in the underlying Rates within groups, in order to deliver a Trump

Table A3: Hypothetical Data, Only Rate Changes

Election	Group	Candidate 1	Candidate 2	Total	Gross Margin	Margin Rate (%)	Composition (%)
1st Election (2020)	Black	15.42	1.47	16.89	13.95	82.60	10.89
	Hispanic	13.21	4.47	17.68	8.73	49.40	11.41
	Other	8.64	5.16	13.80	3.48	25.20	8.90
	White	47.95	57.21	105.16	−9.25	−8.80	67.84
	NA	0.80	0.67	1.47	0.13	8.60	0.95
	Total	86.02	68.98	155.00	17.03	10.99	100.00
2nd Election (2020)	Black	13.88	3.01	16.89	10.86	64.34	10.89
	Hispanic	11.89	5.79	17.68	6.09	34.46	11.41
	Other	7.78	6.03	13.80	1.75	12.68	8.90
	White	43.16	62.00	105.16	−18.84	−17.92	67.84
	NA	0.72	0.75	1.47	−0.03	−2.26	0.95
	Total	77.41	77.59	155.00	−0.17	−0.11	100.00
Comparing the Two Elections	Black	−1.54	1.54		−3.08	−18.26	0.00
	Hispanic	−1.32	1.32		−2.64	−14.94	0.00
	Other	−0.86	0.86		−1.73	−12.52	0.00
	White	−4.80	4.80		−9.59	−9.12	0.00
	NA	−0.08	0.08		−0.16	−10.86	0.00
	Total	−8.60	8.60		−17.20	−11.10	0.00

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.

popular vote win. In this specification, to achieve the desired result, I had to multiply the size of each of the non-White voting blocs by a factor of .44, essentially reducing them by more than half. This leads to a shift in the Composition of the electorate from just under 68% White to nearly 86% White, an enormous shift. The impossibility of this shift highlights the importance that Republicans will need to place on changing Rates within minority groups in future elections, especially as demographic changes continue to accelerate in the future. However, as will be shown in Section 4, there is significant reason to believe that Republicans have succeeded at changing Rates in recent elections, suggesting that demographics driving elections towards Democrats is not an assured outcome.

I have not produced a table comparing the RCVD approach and the DBA for either Table A3 or Table A4, as either specification produces identical results.

Rate and Composition Shift Data

Robustness

To demonstrate the robustness of the method, I present Appendix Tables 1 and 2. Table A1 splits votes out from the Non-White category into Black and All Others, leaving the White category untouched. Table A2 shows that both the DBA as well as

Table A4: Hypothetical Data, Only Composition Changes

Election	Group	Candidate 1	Candidate 2	Total	Gross Margin	Margin Rate (%)	Composition (%)
1st Election (2020)	Black	15.42	1.47	16.89	13.95	82.60	10.89
	Hispanic	13.21	4.47	17.68	8.73	49.40	11.41
	Other	8.64	5.16	13.80	3.48	25.20	8.90
	White	47.95	57.21	105.16	−9.25	−8.80	67.84
	NA	0.80	0.67	1.47	0.13	8.60	0.95
	Total	86.02	68.98	155.00	17.03	10.99	100.00
2nd Election (2020)	Black	6.78	0.65	7.43	6.14	82.60	4.79
	Hispanic	5.81	1.97	7.78	3.84	49.40	5.02
	Other	3.80	2.27	6.07	1.53	25.20	3.92
	White	60.68	72.39	133.07	−11.71	−8.80	85.85
	NA	0.35	0.30	0.65	0.06	8.60	0.42
	Total	77.43	77.57	155.00	−0.14	−0.09	100.00
Comparing the Two Elections	Black	−8.63	−0.82	−9.46	−7.81	0.00	−6.10
	Hispanic	−7.40	−2.50	−9.90	−4.89	0.00	−6.39
	Other	−4.84	−2.89	−7.73	−1.95	0.00	−4.99
	White	12.73	15.18	27.91	−2.46	0.00	18.01
	NA	−0.45	−0.38	−0.83	−0.07	0.00	−0.53
	Total	−8.59	8.59	0.00	−17.18	−11.08	0.00

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.

the RCVD approach correctly leave the specification of the impact of Rate, Composition, and Volume to White voters (the unchanged category) unchanged from Table 2. This highlights the robustness to irrelevant alternatives of the RCVD approach, showing how it behaves comparably to the DBA.

Table A5: Hypothetical Data, Both Rate and Composition

Method	Group	Rate	Composition	Volume	Calc Mar Chng	Var to Act	Variance %
Derivative Calculation	Black	−2.16	−2.79	0	−4.95	−0.43	9.6
	Hispanic	−1.85	−1.75	0	−3.60	−0.37	11.5
	Other	−1.21	−0.70	0	−1.91	−0.24	14.5
	White	−6.71	−0.88	0	−7.59	0.64	−7.7
	NA	−0.11	−0.03	0	−0.14	−0.02	19.5
	Total	−12.04	−6.13	0	−18.18	−0.43	2.4
RCV Calculation	Black	−2.16	−2.36	0	−4.52	0.00	0.0
	Hispanic	−1.85	−1.38	0	−3.23	0.00	0.0
	Other	−1.21	−0.45	0	−1.66	0.00	0.0
	White	−6.71	−1.51	0	−8.23	0.00	0.0
	NA	−0.11	0.00	0	−0.11	0.00	0.0
	Total	−12.04	−5.71	0	−17.75	0.00	0.0
Comparing the Two Approaches	Black		−0.43	0	−0.43		
	Hispanic		−0.37	0	−0.37		
	Other		−0.24	0	−0.24		
	White		0.64	0	0.64		
	NA		−0.02	0	−0.02		
	Total		−0.43	0	−0.43		

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.

Table A6: Comparing Two Elections, Robustness to Irrelevant Alternatives

Election	Group	Candidate 1	Candidate 2	Total	Margin	Margin Rate (%)	Composition (%)
1st Election (2016)	Black	13.25	0.90	14.15	12.35	87.3	11.0
	All Others	16.25	6.60	22.85	9.65	42.2	17.7
	White	39.00	53.00	92.00	−14.00	−15.2	71.3
	Total	68.50	60.50	129.00	8.00	6.2	100.0
2nd Election (2020)	Black	15.40	1.50	16.90	13.90	82.2	10.9
	All Others	22.60	10.50	33.10	12.10	36.6	21.4
	White	48.00	57.00	105.00	−9.00	−8.6	67.7
	Total	86.00	69.00	155.00	17.00	11.0	100.0
Comparing the Two Elections	Black	2.15	0.60	2.75	1.55	−5.0	−0.1
	All Others	6.35	3.90	10.25	2.45	−5.7	3.6
	White	9.00	4.00	13.00	5.00	6.6	−3.6
	Total	17.50	8.50	26.00	9.00	4.8	0.0

Note: Votes are reported in millions. Candidate 1 refers to Clinton (2016) and Biden (2020); Candidate 2 refers to Trump.

Table A7: Concept Demonstration, Robustness to Irrelevant Alternatives

Method	Group	Rate	Composition	Volume	Calc Mar Chng	Var to Act	Variance %
Derivative Calculation	Black	−0.71	−0.07	2.49	1.70	0.15	9.9
	All Others	−1.30	1.98	1.94	2.63	0.18	7.4
	White	6.11	0.70	−2.82	3.99	−1.01	−20.1
	Total	4.11	2.61	1.61	8.33	−0.67	−7.4
RCV Calculation	Black	−0.71	−0.07	2.33	1.55	0.00	0.0
	All Others	−1.30	1.72	2.03	2.45	0.00	0.0
	White	6.11	0.40	−1.51	5.00	0.00	0.0
	Total	4.11	2.04	2.85	9.00	0.00	0.0
Comparing the Two Approaches	Black		0.00	0.16	0.15		
	All Others		0.27	−0.08	0.18		
	White		0.31	−1.31	−1.01		
	Total		0.57	−1.24	−0.67		

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.

Table A8: Concept Demonstration for Trump, Real Data from Marble et al. (2024)

Method	Group	Rate	Composition	Volume	Calc Mar Chng	Var to Act	Variance %
Derivative Calculation	Black	0.68	0.04	−0.82	−0.11	1.48	−93.1
	Hispanic	0.26	−0.40	−0.45	−0.59	1.44	−71.0
	Other	0.41	−0.35	−0.18	−0.11	0.69	−85.8
	White	−5.52	−0.31	0.91	−4.92	−0.56	12.9
	NA	0.02	−0.02	−0.01	0.00	0.03	−86.4
	Total	−4.14	−1.04	−0.55	−5.73	3.07	−34.9
RCV Calculation	Black	0.68	0.07	−2.34	−1.59	0.00	0.0
	Hispanic	0.26	−0.83	−1.46	−2.03	0.00	0.0
	Other	0.41	−0.63	−0.58	−0.80	0.00	0.0
	White	−5.52	−0.39	1.55	−4.35	0.00	0.0
	NA	0.02	−0.03	−0.02	−0.03	0.00	0.0
	Total	−4.14	−1.81	−2.86	−8.81	0.00	0.0
Comparing the Two Approaches	Black		−0.04	1.52	1.48		
	Hispanic		0.43	1.02	1.44		
	Other		0.28	0.40	0.69		
	White		0.08	−0.65	−0.56		
	NA		0.01	0.01	0.03		
	Total		0.77	2.31	3.07		

Note: Votes are reported in millions. Variance percentages are shown with one decimal place.