The Margin and Mix Method:

A Zero-Loss Approach to Decomposing Electoral and Compositional Shifts

Donald Moratz

July 1, 2025

Abstract

Accurately analyzing the impact of demographic changes within and across groups between elections is challenging with existing methods, which often fail to account for total vote shifts. I introduce the Rate, Composition and Volume (RCV) Decomposition, a novel approach that resolves this issue by precisely accounting for both within-group and cross-group compositional changes. Using simulated data, I show RCV not only produces accurate and interpretable results but also provides a framework to standardize the sequencing of the calculation to ensure consistency in analysis. I then show how using RCV allows researchers to identify errors in data estimates by comparing ANES data vs PEW data for recent US elections. Finally, I apply the method to the 2016 and 2020 U.S. elections, showing how the use of RCV would have highlighted shifting support amongst Hispanic voters that presaged the results of the 2024 presidential election.

THE GRIMMER ISSUE IS IN THIS SECTION: Measuring the importance of composition, turnout, and vote choice for conclusions about racial resentment

They hold fixed rate and shift then hold fixed composition and shift and say they offset, but that loses the importance of the cross-partial derivatives

Fraga et al fall guilty of the same issue

2nd argument: we should use grimmer equation 5 not grimmer equation 3 for basically the reasons they argue

Describe the generous position of the assumption is that cross partial derivative impacts are minor, show that they are not necessarily minor

Introduction

Electoral outcomes are shaped not only by individual vote choices but also by the composition of the electorate across demographic and identity groups. Ever since Downs (1957) argued for thinking about voting as an economic choice made by individuals, the focus of the discipline has been on individual-level voter behavior. Despite the justifiable nature of this focus, understanding how group-level dynamics shift over time is essential for explaining election outcomes. Individuals belong to specific groups and it is the composition of the electorate across groups that ultimately determines who wins and loses (Axelrod, 1972). This is particularly relevant to the study of political science since in large elections, candidates must rely on appeals to broad groups rather than tailoring only to individuals. Historically, discussions of group effects were focused on turnout. Conventional wisdom held that conservatives suppress voter turnout amongst the poorest citizens (Piven and Cloward, 1989). To the extent to which that is true, the expectation was that increased turnout would drive more votes to the Democratic party (Lijphart, 1997). However, Citrin et al. (2003) argued that the impacts of changes in turnout could not be so easily identified since the gap between voters and non-voters preferences can change significantly from election to election. As the field turned toward the causal identification revolution, this debate was largely left to lie as research focused more on individual level drivers of voting behavior (Marble et al., 2024).

The discussion of the impact of changes in composition is not restricted to only academic circles. Recent debates on U.S. demographic trends have focused on the influence of racial composition and class-based voting patterns on election outcomes (Klein, 2024b). Some argue that demographic trends favor the Democratic Party

(Klein, 2024a), while others emphasize the need for Democrats to focus on winning working-class, white voters (Judis and Teixeira, 2023). These debates reflect both public curiosity (Fessenden et al., 2020; Kolko and Monkovic, 2020) and a deeper academic interest in understanding vote share shifts across elections (Zingher, 2019; Hill et al., 2021). However, existing approaches to studying this question often fail to adequately account for the complexities of electoral composition, particularly when considering both shifts within groups and changes across groups. Recent work by Marble et al. (2024) argues that correctly accounting for votes requires knowing the group's size and its turnout rate as well as the rate at which the group voted for each candidate. The authors provide an exciting new tool for translating voter survey data, like ANES, into estimations of support for parties by racial groups.

Yet, in their work, as well as in other academic studies, authors fail to correctly account for the impact of the change in composition across groups. For example, in Marble et al. (2024), the authors use what I call a derivative-based approach. They seek to understand the shifting support of Mitt Romney and Donald Trump from the 2012 to the 2016 election. To do so, they first shift the margin of victory for Trump within white voters, fixing the racial composition of the electorate at 2016 levels. They next check the impact of the changing racial composition by fixing the margins within groups at 2016 levels. They stop their analysis here, but following the same logic, fixing the racial composition and the margins within group and simply shifting the total number of votes and summing across the three shifts shows that their method does not accurately capture the total sum of the change in the electorate. Indeed, their approach generates a misspecification of nearly 1.2 million votes.

Currently, there are two main approaches widely used in the literature to study compositional effects: the derivative-based approach and the regression-based approach as in Hill et al. (2021). I argue that neither approach adequately captures the complex dynamics of the 'state space' of electoral change, which refers to how the relationship between demographic variables and time shifts between elections. While both the derivative-based approach (e.g., (Marble et al., 2024)) and regression-based methods (e.g., (Hill et al., 2021)) have contributed to understanding vote shifts, they inadequately model how changes in group composition influence outcomes across elections. To address this gap, I plan to introduce the Rate, Composition and Volume Decomposition (RCV), a zero-loss, non-linear solution that better models the dynamic changes across subgroups. By extending the insights of previous authors, RCV offers a novel solution that fully captures dynamic shifts across both time and subgroups, providing a more precise understanding of electoral outcomes.

In this paper, I plan to begin by conducting a review of the literature on the topic before contrasting the RCV approach with the derivative-based method, demonstrating its failure to accurately account for the composition shifts it is used to study. I will then highlight how the RCV approach changes our interpretation of election outcomes via simulation, introducing situations where reliance on other methods leads to biased estimations. Next, I will show how the use of the RCV method allows researchers to more easily distinguish problems with underlying data by comparing estimates of the 2016 and 2020 election outcomes based on ANES data versus estimates based on PEW data. Finally, I will conclude by showing a practical impact: how the use of RCV combined with PEW data would have done a better job capturing shifting trends among Latino voters in anticipation of the 2024 presidential election.

The paper proceeds as follows. First, I review existing literature and mathematical approaches, with a focus on the derivative-based method. I then introduce RCV, which more accurately captures compositional shifts compared to the derivative-based approach. In section 3, I highlight how the RCV approach changes our interpretation of election outcomes via simulation, introducing situations where reliance on other methods leads to biased estimations. In section 4, I show how the use of the RCV method allows researchers to more easily distinguish problems with underlying data by comparing estimates of the 2016 and 2020 election outcomes based on ANES data versus estimates based on PEW data. Finally, in section 5, I show a practical impact: how the use of RCV combined with PEW data would have done a better job capturing shifting trends among Latino voters in anticipation of the 2024 presidential election.

Theory

Electoral outcomes are shaped not only by individual vote choices but also by the composition of the electorate across demographic and identity groups. Ever since Downs (1957) argued for thinking about voting as an economic choice made by individuals, the focus of the discipline has been on individual-level voter behavior. Despite the justifiable nature of this focus, understanding how group-level dynamics shift over time is essential for explaining election outcomes. Individuals belong to specific groups and it is the composition of the electorate across groups that ultimately determines who wins and loses (Axelrod, 1972). This is particularly relevant to the study of political science since in large elections, candidates must rely on appeals to broad groups rather than tailoring only to individuals. Historically, discussions of group effects were focused on turnout. Conventional wisdom held that conservatives suppress voter turnout amongst the poorest citizens (Piven and Cloward, 1989). To the extent to which that is true, the expectation was that increased turnout would drive more votes to the Democratic party (Lijphart, 1997). However, Citrin et al. (2003) argued that the impacts of changes in turnout could not be so easily identified since the gap between voters and non-voters preferences can change significantly from election to election. As the field turned toward the causal identification revolution, this debate was largely left to lie as research focused more on individual level drivers of voting behavior (Marble et al., 2024).

Contemporary academic discussion has taken new interest in the topic. Achen and Bartels (2016) argues that political science needs to rethink how we approach studying democracy, moving away from a focus on individuals and towards a focus on identity groups. Sides et al. (2019) argues that the Trump campaign of 2016 was successful because of changes in the coalitions of both the Republican and Democratic party driven by identity politics. Although recent evidence suggests that demographic sorting is not as strong a predictor of vote choice as one might believe (Kim and Zilinsky, 2024), in an era of increasing polarization, building electoral coalitions based on identity and group membership has become a key strategy for candidates seeking broad-based support (Lemi, 2021). While social media has increased the ability of candidates to engage in micro-targeting of their appeals (Hersh and Schaffner, 2013), there are limits to the extent to which campaigns can engage in micro-targeting since data availability can vary drastically from state to state (Hersh, 2015). However, even if candidates could achieve perfect targeting, electoral success relies on a candidate's ability to build coalitions of voters. This is particularly true as voter persuasion declines and turnout-based strategies gain importance (Hill, 2017).

Recent scholarship has increasingly recognized the importance of these compositional effects in shaping election outcomes. A variety of approaches have been developed to disentangle how shifts within and across groups impact overall vote shares. Hill et al. (2021) uses a regression-based approach to attribute changes in vote share to either composition or conversion, while Zingher (2019) focuses on estimating how changes in group size affect party support based on underlying group dynamics.The regression-based approach, though, applies a strict linear specification of the functional form of the equation. A misspecification of the functional form can introduce significant bias in the coefficients from the regression, even if 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. There are also several analyses that rely on simple extrapolation to estimate the effect of shifting compositions. Fraga et al. (2021) looks at how different compositions and turnouts across racial groups harmed Clinton in the 2016 election by using 2012 turnout rates and estimating a new vote share. Carmines et al. (2016) argues that shifts between political coalitions and changes in turnout affect vote share. While these studies contribute to our understanding of compositional effects, they fail to fully capture the dynamic interactions between group shifts and electoral outcomes.

Building on this work, the most extensive attempt to propose a clear methodology to segment compositional effects from rate 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 is the percent of voters within a group who support a given candidate. They reject the linearity assumption imposed by earlier regression based approaches to the problem and adopt a non-linear approach. 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 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 et al. (2023) examine the shift within Latino voters, conducting an analysis of this subgroup, but applying a similar methodology. They find that demographic shifts within the Latino group is a driver for their increased support of Trump in 2020 compared to 2016.

However, this derivative based method has two key drawbacks. First, it conflates composition and volume. Composition can be thought of as the proportion of the

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

total electorate comprised by each group; i.e. 60 percent white, 40 percent non-white. Whereas volume is the total size of the electorate. Importantly, the calculation in Marble et al. (2024) allows for changes in group size and turnout rate, but does not capture how group sizes do not change equally for all groups. This means that we cannot identify what component of the change is related to a simple increase in the size of the electorate that would have no effect on the share of votes a candidate receives – only the total – and what component of the change is a shift across groups that might have an impact on the outcome. By way of example, consider turnout in the 2008 presidential election. Turnout in that election was at a 40 year high (Woolley and Peters, 2021), but increased turnout was only a part of the story. In particular, turnout amongst black voters rose by nearly 5%, whereas turnout amongst white voters declined by just over 1% (Lopez and Taylor, 2009). While the formulation from Marble et al. (2024) would show a positive effect from turnout for black voters and a negative one for white voters, it would understate the importance of these effects because black voters became a relatively more larger proportion of the electorate compared to 2008.

The second drawback of the derivative based approach is that – having already calculated the effect of the shift in vote share within a group – by calculating their composition effects using the previous elections rate combining the two becomes difficult. The crux of the issue is that the sum of the two first derivatives will either overstate the impact of the shifts (if the state-space is concave) or understate the impact of the shifts (if the state-space is convex). This is because the correct calculation of the impact of both rate and composition requires also adding the cross-partial second derivative. To show this, consider equation 1 for a single group. We can therefore state the variables as a function only of time t and not of group x. For clarity, I will use $r(t)$ to mean [Vote Share$(t,$ Republican$)$ − Vote Share$(t,$ Democrat$)$], (t) to mean Turnout$(t)x$Group Size(t). Finally, define Diff Net$(t, t-1)$ as $\Delta z(t)$ Then we can restate equation 1 as:

$$\Delta z(t) = r(t)c(t) \tag{2}$$

Which is similar to equation 5 in Marble et al. (2024)[2]. Now consider that $\Delta z(t) = z(t_2) - z(t_1)$. Then we can solve for the role of the change in rate, $r(t)$, and composition, $c(t)$ as follows:

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

Therefore, without accounting for the cross-partial derivative, the resulting combination of the two first derivatives will misrepresent the impact of the shifts[3], understating their impact when the shifts move in the same direction and overstating their impact when they move in opposite directions. Of course, one might argue that therefore the correct approach is to not sum the pieces, but the question of attribution remains. This becomes even more complicated when the equation uses three variables, as in Marble et al. (2024). The correct equation becomes the sum of the first, second, and third cross-partial derivatives. Suppose the third variable is $v(t)$, then the correct specification is:

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

[2] $C_t(x) = \text{Vote Share}_t(x, C) x \text{Share Voters}_t(x) x \text{Overall Turnout}_t$

[3] Except in the trivial case where the cross-partial derivative is zero

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

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

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

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

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

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

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

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

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

As can be seen, to accurately capture the shifts requires including multiple cross-partial derivatives and tracing through the expected disconnect generated by failing to account for these shifts becomes a significant challenge. As a result, methods that rely on only the simple specification of the component parts do not take into account the simultaneous nature of the problem, which will ultimately result in a misstatement of the total effects of the shifts in rate and composition. 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 more fully developed later but intuitively, by calculating the effects of composition using prior rates, the derivative-based approach 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 derivative-based approach 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. In order 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 party of a 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:

$$z(t) = r(t)c(t)v(t) \tag{3}$$

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 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. Given that this equation is a composite function, we can apply the derivative multiplication rule to find the derivative:

$$z'(t) = r'(t)c(t)v(t) + r(t)c'(t)v(t) + r(t)c(t)v'(t) \tag{4}$$

And indeed this derivative suggests using a formula essentially identical to the one utilized by Marble et al. (2024), whereby the first part of the expression is the impact of the change in rate, the second part the impact of the change in composition, and the third part the impact of the change in volume. However, as shown above, this formula is misleading. As defined, $z(t)$ is a fundamentally non-linear equation in a four dimensional state-space. Therefore, the slope is not constant, and, as a result, adequately describing the changes requires the use of cross-partial derivatives. Instead, to calculate the impact of rate, composition, and volume separately, I propose employing a sequential gradient ascent method to explore the state-space:

$$\begin{aligned}\Delta z(t) =& [r(t_2) - r(t_1)]c(t_1)v(t_1) \\ &+ r(t_2)[c(t_2) - c(t_1)]v(t_1) \\ &+ r(t_2)c(t_2)[v(t_2) - v(t_1)]\end{aligned} \tag{5}$$

While initially counter intuitive, the proof that this is equal to $z(t_2) - z(t_1)$ is presented below in equation 6. I call this the Rate, Composition, and Volume (RCV) approach.

$$
\begin{aligned}
\Delta z(t) &= z(t_2) - z(t_1) \\
z(t_2) - z(t_1) &= r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) \\
&= r(t_2)[c(t_1)v(t_1) - c(t_1)v(t_1) + c(t_2)v(t_1) - c(t_2)v(t_1) + c(t_2)v(t_2)] - \\
&r(t_1)c(t_1)v(t_1) \\
&= r(t_2)c(t_1)v(t_1) - r(t_1)c(t_1)v(t_1) + r(t_2)c(t_2)v(t_1) - \\
&r(t_2)c(t_1)v(t_1) + r(t_2)c(t_2)v(t_2) - r(t_2)c(t_2)v(t_1) \qquad (6) \\
r(t_2)c(t_2)v(t_2) - r(t_1)c(t_1)v(t_1) &= [r(t_2) - r(t_1)]c(t_1)v(t_1) + \\
&r(t_2)[c(t_2) - c(t_1)]v(t_1) + \\
&r(t_2)c(t_2)[v(t_2) - v(t_1)]
\end{aligned}
$$

There are two noteworthy constraints to 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 both the percentage of the total voters that a group comprises is a function of both the proportion of the population comprised of the that group and the turnout rate of that group as in equation 1. 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 caveat is that this sequence is not unique. The gradient ascent can be calculated by moving any order-combination of rate, composition, and volume. Unless the state-space is defined by a regular cuboid, a shift in the sequencing of the calculation will by necessity change the interpretation of the estimated coefficients. In order to address this, In the next section I highlight how changing the sequencing affects the interpretation of a simulated election result. There are two key arguments for why this proposed ordering is the most justifiable. 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. 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 stretching or shrinking of the newly specified outcome.

Section 3 - simulation, show how the order changes interpretation, Show how we think about composition and volume shifts in terms of election impacts Section 4 - ANES vs PEW Section 5 - Hispanic Votes

The Method - Simulated Data

Simple Case

In this section, I will introduce a basic example of the RCV math introduced in the previous sections. Table 1 shows a simple comparison of data from the 2016 and 2020 US presidential elections. The data is based on the blocs package introduced by Marble et al. (2024), with racial groups collapsed to the simple case of white and non-white.

For illustrative purposes, I have introduced rounding to the closest 500,000^{ths} place. Table 1 shows a result favoring Clinton in the 2016 election by 8 million votes, with margin and margin rate calculated from the perspective of the democratic candidate.[4] Comparably, Biden performed significantly better in 2020, winning by 17 million votes, for a change in Non-White margin of 4 million votes and a change in White margin of 5 million votes. This is equivalent to a 7.5% decline in margin within the Non-White voting block and an improvement of 6.6% in margin within the White community. There was also a 3.6% shift towards non-white voters and away from white voters. Total votes in the system increased by 26 million, which explains why Biden improved his margin of victory in both groups.

Table 1: Comparing Two Elections

Election	Group	Candidate 1	Candidate 2	Total	Margin	Margin Rate (%)	Mix (%)
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.

Table 2 demonstrates the importance of shifting away from the calculation introduced in previous work. While none of the other papers mentioned in this study calculated the impact of volume, I have produced the logical calculation that would follow from the derivative-based approach discussed above. By leaving volume unspecified, previous work in essence shifted any unexplained votes to an error term that was comprised not just of volume, but to incorrectly calculated effects of mix. The upper third of Table 2 shows that using the derivative-based approach leaves a Mean Absolute Error (MAE) of just over 1.4 million votes misattributed. However, the RCV approach correctly specifies the proper placement of the votes into their appropriate

[4] As best as I can tell, I have replicated the process used by Marble et al. (2024) in creating this data. The difference in their reported votes from reality seems to be related to their estimation of voter support for candidates based on the ANES.

categories, leaving a MAE of 0 votes, precisely explaining the change within the system. While a large portion of the mispecification indeed does belong in the previously undefined category of volume, in the final third of the table I show that the benefit that Biden received from the change in the racial composition of the electorate is understated by 650,000 votes using the derivative method. The RCV approach also has the appealing property of explaining rate in the exact same way as the derivative-based approach, leaving our interpretation of previous work that calculated the impact of rate unchanged.

Table 2: Concept Demonstration

Method	Group	Rate	Mix	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.

In summary, the RCV approach has the desirable property of leaving every vote in the system precisely specified, delivering a zero-loss summary of the changes in the system, and this result is robust to further partitions of the underlying data[5]. The RCV approach also has the appealing property of explaining rate in an identical fashion to previous work. In future work, I intend to examine the robustness of the RCV approach to subpartions of the data.

Alternative Orders

Thinking about Rate vs Composition Changes

In this section, I highlight the importance in understanding and correctly specifying the roles of rate and mix within elections. To do so, I will examine hypothetical scenarios that would have led to Trump winning the popular vote for the 2020 presidential

[5]See Appendix for the robustness of the method to partitions

election and compare those to the actual results implied by the blocs package. I will do so only for rate and for mix, since any shift in volume would be equivalent to applying using an eigenvector to achieve a linear transformation of the data, perfectly preserving the underlying relationships. To begin, Table 5 shows how rates would have had to have shifted in order to guarantee a win for Trump in 2020 without any mix 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.

Table 3: Hypothetical Data, Only Rate Changes

Election	Group	Candidate 1	Candidate 2	Total	Gross Margin	Margin Rate (%)	Mix (%)
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.

Conversely, Table 6 demonstrates how mix would have had to shift, without any changes in the underlying rates within groups, in order to deliver a Trump 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 mix 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.

Table 4: Hypothetical Data, Only Mix Changes

Election	**Group**	**Candidate 1**	**Candidate 2**	**Total**	**Gross Margin**	**Margin Rate (%)**	**Mix (%)**
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.

I have not produced a table comparing the RCV approach and the derivative-based approach for either Table 5 or Table 6, as either specification produces identical results. However, in Table 7 and 8 I utilize similar hypothetical data to highlight the shortcomings of the derivative-based approach. In Table 7, as in Table 5 and Table 6, I shift 2020 data in order to turn a Biden victory into a narrow Trump win. To do so, I multiply the rate each group voted for Biden by a factor of .93 and the racial composition of all non-white groups by a factor of .8. This is still a strong mix shift towards white-voters, but the size of this shift is unimportant. More important is that in this specification, there is no volume at play in the system. Therefore, this is the closest possible comparison to what previous work has attempted to explain, without assuming the calculation that prior authors would have used to account for volume. Table 7 presents the resulting shift in votes, with the value at the bottom of column 3 highlighing that the total votes in the system has remained identical. In this synthetic example, the only two possible buckets for explaining votes are rate within groups and mix between them.

Table 8 highlights that while both the derivative-based approach and the RCV approach correctly specify that 0 votes were contributed to either candidate due to volume changes, the derivative-based approach has an MAE of 1.7 million votes. By

Table 5: Hypothetical Data, Both Rate and Mix

Election	Group	Candidate 1	Candidate 2	Total	Gross Margin	Trad Mar (%)	Trad Mix (%)
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.

comparing this total to the RCV approach, it is clear all of these unidetified votes come from a misspecification of the impact of mix, in almost all cases by hundreds of thousands of votes.

In summary, this section has shown that changes in both rate and mix can be used to engender different outcomes in elections. Crucially, it has also shown that even when volume is not at issue, the derivative-based approach fails to correctly specify the votes explained in the system. The unspoken assumption of previous work that any unexplained variance belonged in the volume category (or in unexplainable error), is therefore insufficient.

Table 6: Hypothetical Data, Both Rate and Mix

Method	Group	Rate	Mix	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.

1 Evaluating Real Data - ANES vs PEW and Hispanic Voters

Need to show table with the voting info based on ANES and voting info based on PEW Do the analysis using both data, discuss the differences, how thinking about votes this way let's us catch the issue

ANES vs PEW

In this section I apply the RCV approach to the fully specified data available in the blocs package. I again show data from the 2016 and 2020 US presidential elections, but this time with the data split into 5 racial categories (as defined in the blocs package). The blocs package estimates a percentage of votes of the total from a given racial group. While the blocs package allows for an error band around this estimation, I have taken the mean estimates and multiplied them by the total votes from the two US presidential elections. An important note about the RCV approach is that, as a zero-loss calculation, there is no MAE in the calculation, but if the underlying data has a confidence interval (as it would if it is itself an estimation), then the RCV approach can be used to estimate the bins for any specified estimate taken from the underlying data.

Unlike in the simple example in the previous section, I have utilized no rounding in the individual categories, but have rounding the total votes in the system to the nearest millionth place. As can be seen by comparing these vote totals to those actually reported in the election (Wikipedia, 2024), the blocs package estimations, which are based upon ANES exit surveys (Marble et al., 2024), strongly overestimate support for Biden. Nevertheless, these estimates are the best available for racial group support for presidential candidates and these issues are not relevant for the purposes of this paper. Table 3 shows the racial categories and their estimated vote totals for both presidential elections. Again, margin and margin rate are calculated from the perspective of the democratic candidate.

Table 7: Comparing Two Elections, Real Data from Marble et al. (2024)

Election	Group	Candidate 1	Candidate 2	Total	Margin	Margin Rate (%)	Mix (%)
1st Election (2016)	Black	13.25	0.89	14.14	12.36	87.40	10.96
	Hispanic	9.87	3.17	13.03	6.70	51.40	10.10
	Other	5.84	3.16	8.99	2.68	29.80	6.97
	White	39.17	52.78	91.95	−13.61	−14.80	71.28
	NA	0.49	0.39	0.88	0.10	10.80	0.68
	Total	68.61	60.39	129.00	8.22	6.38	100.00
2nd 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
Comparing the Two Elections	Black	2.17	0.58	2.75	1.59	−4.80	−0.07
	Hispanic	3.34	1.31	4.65	2.03	−2.00	1.30
	Other	2.80	2.01	4.81	0.80	−4.60	1.93
	White	8.78	4.43	13.21	4.35	6.00	−3.44
	NA	0.31	0.28	0.59	0.03	−2.20	0.27
	Total	17.40	8.60	26.00	8.81	4.61	0.00

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

Table 4 compares the results of the derivative-based approach to the RCV approach. I show that, again, the calculations are identical with respect to rate, but that they differ both in respect to mix and volume. The total MAE of the derivative-based approach in this specification is 1.2 million, versus the MAE of 0 for the RCV approach. This comparison allows for an opportunity to highlight another appealing property of RCV over the derivative-based approach. Appendix Table A3 shows the same calculation from the perspective of the Republican candidate (in both cases, Trump). A careful comparison between Table A3 and Table 4 shows that while each vote gained (lost) for Biden compared to Clinton from a rate perspective is a vote lost (gained) for Trump, this one-to-one trade off completely collapses when calculating the impact of mix and of volume under the derivative-based approach. The total MAE when the impact of the three bins is calculated from the perspective of Trump is 4.2 million. Comparatively, in all three bins using the RCV approach, a vote gained or lost by a candidate in either party is perfectly offset by a vote gained or lost by the candidate of the opposing party.

Table 8: Concept Demonstration, Real Data from Marble et al. (2024)

Method	**Group**	**Rate**	**Mix**	**Volume**	**Calc Mar Chng**	**Var to Act**	**Variance %**
Derivative Calculation	Black	−0.68	−0.08	2.49	1.74	0.15	9.3
	Hispanic	−0.26	0.86	1.35	1.95	−0.08	−4.0
	Other	−0.41	0.74	0.54	0.87	0.07	8.9
	White	5.52	0.66	−2.74	3.43	−0.92	−21.2
	NA	−0.02	0.04	0.02	0.04	0.01	17.6
	Total	4.14	2.22	1.66	8.03	−0.78	−8.9
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		0.00	0.15	0.15		
	Hispanic		0.03	−0.11	−0.08		
	Other		0.11	−0.04	0.07		
	White		0.27	−1.19	−0.92		
	NA		0.01	0.00	0.01		
	Total		0.42	−1.20	−0.78		

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

In summary, the RCV approach is more consistent then the derivative-based approach, correctly specifying the appropriate bins of vote changes, leaving an MAE of 0 when applied to real-world data. In addition, it has the appealing property of highlighting the zero-sum nature of elections- any vote won by one party, is a vote lost for the opposition.

Interpretation

Write about how with the PEW data and the correct method, we anticipate the shift of hispanic voters towards Trump in 2024. Show some 2020 vs 2024 data if it exists

Conclusion

As demographic shifts reshape the U.S. voting landscape, understanding how these changes between groups affect election outcomes is crucial for both scholars and policymakers. Previous methods for explaining these shifts have often introduced significant errors, oftentimes amounting to millions of votes. The Rate, Composition and Volume (RCV) approach, as defined in this paper, eliminates such errors, providing a zero-loss approach that precisely specifies how changes in rate, mix, and volume influence election outcomes. Importantly, the RCV approach is broadly applicable, extending beyond U.S. elections. It can be used in any context where subdividing an electorate or population into distinct groups— whether based on ethnicity, geography, age, education, or income— helps explain outcomes. The method is also valuable for analyzing other time-series cross-sectional data where participation changes, as it accurately accounts for shifts in group composition over time.

Future work will explore the robustness of the RCV approach in two key areas. First, I will investigate its ability to account for nested subgroups, demonstrating how shifts within smaller partitions contribute to broader electoral changes, as explored by Fraga et al. (2024). Second, I will extend the analysis across multiple time periods to capture long-term trends, showing how the RCV approach can reveal the evolving dynamics of group shifts over time. These studies will further validate RCV as a tool for analyzing both historical shifts in political support and provide a basis for extending our understanding of electorate dynamics for predicting future electoral outcomes.

References

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

[2] Robert Axelrod. "Where the Votes Come From: An Analysis of Electoral Coalitions, 1952–1968". en. In: *American Political Science Review* 66.1 (Mar. 1972), 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).

[3] Edward G. Carmines, Michael J. Ensley, and Michael W. Wagner. "Ideological Heterogeneity and the Rise of Donald Trump". en. In: *The Forum* 14.4 (Dec. 2016), 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).

[4] Jack Citrin, Eric Schickler, and John Sides. "What if Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections". en. In: *American Journal of Political Science* 47.1 (Jan. 2003), 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).

[5] Anthony Downs. "An Economic Theory of Political Action in a Democracy". en. In: *Journal of Political Economy* 65.2 (Apr. 1957), pp. 135–150. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/257897. URL: https://www.journals.uchicago.edu/doi/10.1086/257897 (visited on 10/13/2024).

[6] Andrew M. Engelhardt. "Trumped by Race: Explanations for Race's Influence on Whites' Votes in 2016". In: *Quarterly Journal of Political Sci-*

ence 14.3 (July 2019), pp. 313–328. ISSN: 15540634. DOI: 10.1561/100.00018068. URL: http://www.nowpublishers.com/article/Details/QJPS-18068 (visited on 03/12/2024).

[7] Ford Fessenden, Lazaro Gamio, and Rich Harris. "Even in Defeat, Trump Found New Voters Across the U.S." en-US. In: *The New York Times* (Nov. 2020). ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2020/11/16/us/politics/election-turnout.html (visited on 02/18/2024).

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

[9] Bernard L. Fraga, Yamil R. Velez, and Emily A. West. "Reversion to the Mean, or Their Version of the Dream? Latino Voting in an Age of Populism". en. In: *American Political Science Review* (May 2024), 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).

[10] Bernard L. Fraga et al. "Analysis | Why did Trump win? More whites — and fewer blacks — actually voted." en-US. In: *Washington Post* (Dec. 2021). 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).

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

[12] Eitan D. Hersh and Brian F. Schaffner. "Targeted Campaign Appeals and the Value of Ambiguity". en. In: *The Journal of Politics* 75.2 (Apr. 2013), pp. 520–534. ISSN: 0022-3816, 1468-2508. DOI: 10.1017/S0022381613000182. URL: https://www.journals.uchicago.edu/doi/10.1017/S0022381613000182 (visited on 09/30/2024).

[13] Seth J. Hill. "Changing votes or changing voters? How candidates and election context swing voters and mobilize the base". en. In: *Electoral Studies* 48 (Aug. 2017), pp. 131–148. ISSN: 02613794. DOI: 10.1016/j.electstud.2017.06.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0261379416301743 (visited on 03/12/2024).

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

[15] John B. Judis and Ruy A. Teixeira. *Where have all the Democrats gone? the soul of the party in the age of extremes*. First edition. New York: Henry Holt and Company, 2023. ISBN: 9781250877505.

[16] Seo-young Silvia Kim and Jan Zilinsky. "Division Does Not Imply Predictability: Demographics Continue to Reveal Little About Voting and Partisanship". en. In: *Political Behavior* 46.1 (Mar. 2024), pp. 67–87. ISSN: 0190-9320, 1573-6687. DOI: 10.1007/s11109-022-09816-z. URL: https://link.springer.com/10.1007/s11109-022-09816-z (visited on 10/13/2024).

[17] Ezra Klein. *The Ezra Klein Show: 'The Strongest Democratic Party That Any of Us Have Ever Seen' on Apple Podcasts*. en-US. Jan. 2024. 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).

[18] Ezra Klein. *The Ezra Klein Show: 'Why Haven't the Democrats Completely Cleaned the Republicans' Clock?' on Apple Podcasts*. en-US. Feb. 2024. URL: https://podcasts.apple.com/us/podcast/why-havent-the-democrats-completely-cleaned/id1548604447?i=1000643778108 (visited on 02/18/2024).

[19] Jed Kolko and Toni Monkovic. "The Places That Had the Biggest Swings Toward and Against Trump". en-US. In: *The New York Times* (Dec. 2020). ISSN: 0362-4331. URL: https://www.nytimes.com/2020/12/07/upshot/trump-election-vote-shift.html (visited on 02/18/2024).

[20] Danielle Casarez Lemi. "Do Voters Prefer Just Any Descriptive Representative? The Case of Multiracial Candidates". en. In: *Perspectives on Politics* 19.4 (Dec. 2021), 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).

[21] Arend Lijphart. "Unequal Participation: Democracy's Unresolved Dilemma Presidential Address, American Political Science Association, 1996". en. In: *American Political Science Review* 91.1 (Mar. 1997), pp. 1–14. ISSN: 0003-0554, 1537-5943. DOI: 10.2307/2952255. URL: https://www.cambridge.org/core/product/identifier/S0003055400230980/type/journal_article (visited on 10/13/2024).

[22] Mark Hugo Lopez and Paul Taylor. *Dissecting the 2008 Electorate: Most Diverse in U.S. History*. Tech. rep. Pew Research Center, 2009. 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).

[23] William Marble, Justin Grimmer, and Cole Tanigawa-Lau. "Measuring the Contribution of Voting Blocs to Election Outcomes". en. In: *The Journal of Politics* (Sept. 2024), 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).

[24] Frances Fox Piven and Richard A. Cloward. *Why Americans don't vote*. eng. 1. paperback ed., updated with a post-election pref. by the authors. New York, NY: Pantheon Books, 1989. ISBN: 9780679723189.

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

[26] Wikipedia. *2020 United States presidential election — Wikipedia, The Free Encyclopedia*. 2024. URL: http://en.wikipedia.org/w/index.php?title=2020%20United%20States%20presidential%20election&oldid=1248504309.

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

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

[29] Joshua N. Zingher. "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 (Aug. 2019), 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

Proof of Equation 4

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

Non-Linearity, An Illustration

Linear estimation is the most commonly used econometric tool in evaluating relationships between data within political science. As a result, any argument for moving away from a linear model needs careful justification. While the derivative-based approach of Marble et al. (2024) is also non-linear, Figure 1 provides a clear illustration of why this particular problem requires a non-linear approach. Consider Figure 1a. In this figure, I present an electorate divided into 8 equal parts. On the left hand side, the politician of interest receives 3 votes from each of 4 of those parts of the electorate, and 0 from the remaining four parts in the first election. This yields 12 total votes for the politician. The right hand side displays the results of a second election. In this second election they won 32.4 votes. This shift in the total votes for the politician comes from three sources: rate (the extent to which a group supports the politician), mix (the relative mix of groups that both support and do not support the politician) and volume (the total number of voters).

For this figure, the changes are as follows. Since the first election, the politician has increased their total support within the groups that they had previously received votes from to the tune of 50%, yielding 4.5 votes from each of those groups instead of 3. Importantly, this change is without any change in the underlying size of the electorate- this comes from vote switching, not from turnout or population dynamics. Additionally, the politician manages to capture an equal number of votes in a 5th, previously apathetic group of voters, yielding a 25% increased share over the previous set of 4 groups. Finally, the electorate grows, either through immigration, through an increase in turnout, or simply from population growth. This change in the population is equal to 44% in this example. While these numbers reflect changes of significant magnitude far and beyond what would ever be expected to happen in a real election, they provide clear examples of the impacts of the shift.

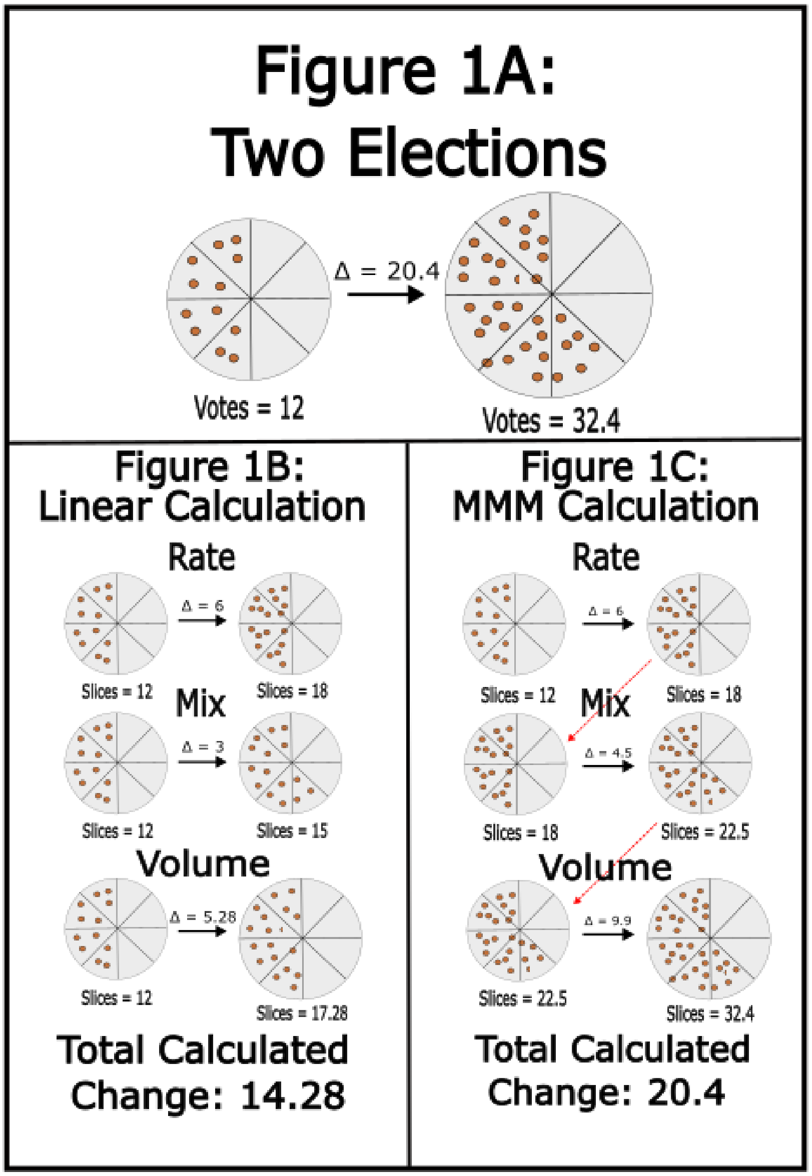

Figure 1: An Illustration of why the RCV Approach is necessary.

Now consider Figure 1b. In this figure, I utilize the derivative-based approach. I first calculate the impact of the 50% change in rate, then the impact of the 25% change in the mix and finally, the impact of the 44% change in total volume. As can be seen, the sum of these changes is 14.28. Since our total difference is 20.4, this approach leaves just over 30% of the vote in the unspecified error term, a value greater than any of the calculated component parts. However, in Figure 1c, I apply the method I propose in this paper, where the impact is calculated sequentially. Here, I show that while the impact of rate is identical in both methods (a shift of 6 total votes), mix is underestimated in Figure 1b by 50% and the impact of volume is underestimated by just over 53%. Even in this simple example, I have shown that if we truly want to understand the impact of population dynamics and turnout, which plays into both mix and volume, the derivative-based approach and the linear regression approach simply will not suffice. Nevertheless, this artificial example bears little resemblance to the real world, so in the next section I will examine the impact of this approach with data much closer to a real world example.

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 derivative-based approach as well as the RCV approach correctly leave the specification of the impact of rate, mix, and volume to white voters (the unchanged category) unchanged from Table 2. This highlights the robustness to irrelevant alternatives of the RCV approach, showing how it behaves comparably to the derivative-based approach.

Table A1: Comparing Two Elections, Robustness to Irrelevant Alternatives

Election	Group	Candidate 1	Candidate 2	Total	Margin	Margin Rate (%)	Mix (%)
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.

On Overfitting

While zero-error terms in regression models often raise concerns about overfitting (Wooldridge, 2010), the RCV approach is not subject to these risks. As it is not

Table A2: Concept Demonstration, Robustness to Irrelevant Alternatives

Method	Group	Rate	Mix	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 A3: Concept Demonstration for Trump, Real Data from Marble et al. (2024)

Method	Group	Rate	Mix	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.

a regression or estimation technique, but rather an accounting framework, zero-error reflects the accurate decomposition of shifts rather than problematic estimation.