Cause or Effect? Turnout in Hispanic Majority-Minority Districts

John A. Henderson  
Department of Political Science, Institution for Social and Policy Studies, Yale University,  
77 Prospect Street, New Haven, CT 06520-8209

Jasjeet S. Sekhon  
Travers Department of Political Science and Department of Statistics, Institute of Governmental Studies, University of California, 109 Moses Hall, #2370, Berkeley, CA 94720-2370  
e-mail: sekhon@berkeley.edu; URL: http://sekhon.berkeley.edu/ (corresponding author)

Rocio Titiunik  
Department of Political Science, University of Michigan, 5700 Haven Hall,  
505 S. State St, Ann Arbor, MI 48109

Edited by Marisa Abrajano

Legislative redistricting alters the political and electoral context for some voters but not others, thus offering a potentially promising research design to study many questions of interest in political science. We apply this design to study the effect that descriptive representation has on co-ethnic political engagement, focusing on Hispanic participation following California’s 2000 redistricting cycle. We show that when redistrictors draw legislative boundaries in California’s 1990, 2000, and 2010 apportionment cycles, they systematically sort higher-participating Hispanic voters into majority-Hispanic (MH) jurisdictions represented by co-ethnic candidates, biasing subsequent comparisons of Hispanic participation across districts. Similar sorting occurs during redistricting in Florida and Texas, though here the pattern is reversed, with less-participating Hispanic voters redistricted to MH districts. Our study highlights important heterogeneity in redistricting largely unknown or underappreciated in previous research. Ignoring this selection problem could significantly bias estimates of the effect of Hispanic representation, either positively or negatively. After we correct for these biases using a hierarchical genetic matching algorithm, we find that, in California, being moved to a district with an Hispanic incumbent has little impact on Hispanic participation in our data.

1 Introduction

We propose and evaluate a novel research design that exploits redistricting to identify the participatory consequences of changes in representation, building on previous analyses of the incumbency advantage (Ansolabehere, Snyder, and Stewart 2000; Sekhon and Titiunik 2012). Redistricting induces variation across districts and over time, since some voters are moved to different districts while others are not, and people vote both before and after redistricting. The former variation ensures some voters experience different features of incumbency (e.g., party, race, gender), whereas the latter allows us to control for differences in prior participation, enabling our design to study many important electoral effects.

We implement this research design to study the participatory effects of co-ethnic representation, focusing on whether being represented by a Hispanic member of Congress leads to higher voter...
turnout among Hispanic citizens. In spite of considerable research, there is disagreement over whether the shared ethnicity of Hispanic incumbents increases participation among co-ethnic voters. Although some scholars find that co-ethnic representation “empowers” Hispanic voters and leads to higher turnout (Barreto, Segura, and Woods 2004; Barreto 2007), others fail to uncover positive effects (Fraga 2016). Moreover, some scholars argue that creating majority-Hispanic (MH) districts—where most Hispanic incumbents are elected—may lead to unintentional declines in Hispanic participation, since such districts concentrate partisans, dampening both electoral competition and voter mobilization (De la Garza and DeSipio 1992).

Using data from California’s 1990, 2000, and 2010 redistricting cycles, we compare turnout and registration rates of Hispanic voters moved during redistricting to new districts represented by Hispanic incumbents to the rates of those retaining prior Non-Hispanic White (NHW) representatives. A major finding uncovered by our empirical analysis is that policymakers are very effective at sorting voters into districts based on their demographic and political factors, most importantly previous Hispanic turnout and registration. As a consequence, it is very difficult to find voters redistricted to new Hispanic incumbents who are comparable to those retaining NHW representatives, significantly biasing estimates that compare participation across legislative districts. Though we focus on incumbents’ ethnicity, our findings suggest that biases induced during redistricting may affect other electoral analyses of incumbency or representation (Sekhon and Titiunik 2012; Chen and Rodden 2013). Notably, this concern generalizes to other states, including Florida and Texas, where systematic sorting based on previous participation also emerges during redistricting. Ignoring this selection problem could bias estimates of the effect of Hispanic representation (either positively or negatively), and thus researchers must develop empirical strategies to correct for it.

We aim to correct for these systematic pretreatment differences, using a matching algorithm to find subsets of voters that are similar on characteristics used during redistricting, but that differ in being assigned to new Hispanic representatives. After matching, we find Hispanic candidates have little impact on Hispanic turnout, but may diminish NHW participation. A major challenge in concluding that these null effects generalize across California is that severe sorting in redistricting necessitates we either trim considerable data or make herculean assumptions that biases in how districts are drawn, including significant differences in prior participation, have no influence on future turnout. Generally, our findings recommend that future studies relying on redistricting-induced variation in candidate- or district-level characteristics should address whether biases emerging from decennial redistricting could confound their empirical results.

2 Redistricting as Research Design

In this application of our redistricting-based design, the unit of analysis is the census block. Let \( T_i^{WH} = 1 \) if block \( i \) is moved from a district represented by a NHW incumbent to one represented by an Hispanic incumbent before election \( t \), and \( T_i^{WH} = 0 \) if block \( i \) stays in a district represented by a NHW incumbent before and after \( t \). Define \( Y_0(i, t) \) to be the potential outcome for \( i \) if \( T_i^{WH} = 0 \), at election \( t \) and \( Y_1(i, t) \) the analogous potential outcome if \( T_i^{WH} = 1 \). We only observe realized outcomes, \( Y(i, t) = Y_0(i, t) \cdot (1 - T_i^{WH}) + Y_1(i, t) \cdot T_i^{WH} \). Our parameter of interest is the average treatment effect on the treated, \( ATT^{WH} = E[Y_1(i, t) - Y_0(i, t)|T_i^{WH} = 1] \).

If redistricting was fully randomized, we could estimate \( ATT^{WH} \) simply differencing the average turnout of voters moved from an NHW to a Hispanic incumbent and voters remaining with an NHW incumbent at election \( t \). Redistricting decisions are not random, as several population characteristics are considered when redrawing district boundaries. Fortunately, we can collect the same characteristics used by redistrictors (i.e., census information, election results, past turnout, registration).

We initially assume that, conditioning on these characteristics, nonredistricted (control) blocks are a valid comparison group to infer the turnout rates that would have occurred after redistricting.

\(^{1}\)Replication data are available online at the Political Analysis Dataverse (Henderson, Sekhon, and Titiunik 2016).
among redistricted (treated) blocks, if the ethnicity of the incumbent had not changed. Letting $X_i$ denote a vector of observable characteristics for block $i$, this selection-on-observables assumption can be formalized as

$$E[Y_0(i, t)|T_i^{WH} = 1, X_i] = E[Y_0(i, t)|T_i^{WH} = 0, X_i],$$

which is sufficient to identify a causal effect, since

$$ATT(X)^{WH} = E[Y_1(i, t) - Y_0(i, t)|T_i^{WH} = 1, X]$$

$$= E[Y(i, t)|T_i^{WH} = 1, X] - E[Y(i, t)|T_i^{WH} = 0, X].$$

We show below that condition (1) may be too strong given evidence in California. Redistricted voters may not be comparable to voters whose previous incumbent remained unchanged, even with similar $X$. Since we observe participation rates before and after redistricting, we can replace (1) with a different assumption that may be more plausible for this application. This assumption requires that, controlling for $X$, the average change in Hispanic turnout among control blocks retaining their old incumbent between election $t-1$ and $t$, is the same change in Hispanic turnout that treated blocks would have experienced had their incumbent remained unchanged. Formally,

$$E[Y_0(i, t) - Y_0(i, t-1)|T_i^{WH} = 1, X] = E[Y_0(i, t) - Y_0(i, t-1)|T_i^{WH} = 0, X].$$

Equation (2) requires preredistricting participation data to estimate causal effects (Abadie 2005). These are obtained through difference-in-differences, comparing the difference between treatment and control in the change in Hispanic turnout before ($t-1$) and after ($t$) redistricting:

$$E[Y(i, t) - Y(i, t-1)|T_i^{WH} = 1, X] - E[Y(i, t) - Y(i, t-1)|T_i^{WH} = 0, X]$$

$$= E[Y_1(i, t) - Y_0(i, t)|T_i^{WH} = 1, X]$$

$$= ATT(X)^{WH}.$$

### 3 Appyling the Design to Hispanic Representation in California

Using our design to study redistricting in California, we uncover evidence of significant sorting on Hispanic participation in the creation and maintenance of districts. During each of the last three cycles (1990, 2000, and 2010), California redistricted higher-participating Hispanic voters into new or existing House districts represented by Hispanic incumbents. This suggests that cross-sectional comparisons of turnout across districts, as done in most prior research on this question, would lead to invalid conclusions.

#### 3.1 Redistricting Data

We utilize data from the 1990, 2000, and 2010 California redistricting plans for U.S. House seats. Results mainly focus on the 2000 redistricting, though additional analyses of 1990 and 2010 show that sorting persists across multiple cycles. Our unit of analysis is the census block, the lowest level of geography in redistricting, allowing us to replicate the redistricting assignment process (Cain, MacDonald, and Hui 2006). Most data come from the California State-Wide Database (SWDB). For 1998 and 2000, the SWDB includes turnout and registration for 2000 census blocks. For 2002

---

2See Section 5.1 of the Supplementary materials for details on California’s majority-minority districting, including possible mechanisms for this phenomenon.
and 2004, participation data are only available at the 2010 census block, which we convert to the 2000 census-block level using Census Block Relationship Files. Finally, 2010 turnout and registration are available from the SWDB at 2010 census blocks, whereas registration data from 1990 are available at the 1990 census block group (King et al. 1997).

The SWDB contains turnout and registration figures in total, and by age, party affiliation and ethnicity, as well as vote returns for congressional, assembly, presidential, and other elections. Hispanic turnout is constructed from block-geocoded and surname-matched individual records in California voter registration files. Details on incumbent ethnicity come from the Hispanic Americans in Congress website, maintained by Congress. To measure important demographic covariates for the matching analysis below, we merged block-level data from the 2000 Census Summary File 3 to the SWDB data set. Every block was assigned its corresponding congressional, assembly, and state senate district according to the redistricting plan effective for each relevant election.

Our primary interest is evaluating the use of redistricting as a research design, though this application to co-ethnic representation raises important conceptual issues about how to appropriately define and measure ethnicity. We discuss such issues at greater length in Section 5.2 of the Supplementary materials.

3.2 Evidence of Bias

Following redistricting in 2000, Hispanic turnout rates are higher in MH districts represented by Hispanic incumbents than in non-MH jurisdictions represented by NHW politicians. After matching to adjust for the proportion of the total voting age population that is Hispanic (P-HVAP), average Hispanic turnout in 2004 is roughly 4.3 percentage points higher in MH than non-MH districts, evidencing a positive association between Hispanic turnout and MH status.

Are these Hispanic participation differences caused by residing in an MH district with a Hispanic representative, or could these arise as a result of pre-existing differences among voters in MH and non-MH districts? The evidence we present strongly suggests the latter. In Figure 1a, we display a QQ-plot of Hispanic turnout in the 2000 election, comparing would-be-treated blocks (those that will be moved from a congressional district represented by an NHW incumbent to a district represented by a Hispanic incumbent) to would-be-control blocks (those that keep the same NHW incumbent before and after redistricting).

We find that average prior Hispanic turnout in 2000 is 4.2 percentage points higher in would-be-treated than in would-be-control blocks. The QQ-plot reveals that this difference persists across the whole distribution (deviations off 45° lines indicate differences in distributions). Given equal P-HVAP and before redistricting occurs, Hispanic voters who will be redistricted into a district with a Hispanic incumbent already vote at systematically higher rates than Hispanic voters who will stay with their old NHW-incumbent district.

We replicate our analysis for 1990, when most of California’s MH districts were created, and 2010, when a nonpartisan citizens commission maintained these districts, and find that bias also persists in these redistricting cycles. Figure 1b and Figure 1c present QQ-plots for Hispanic registration over Hispanic voting-age population (HVAP) for the 1990 and 2010 redistricting, respectively. Plots show that blocks moved to districts represented by Hispanic politicians consistently have higher rates of previous Hispanic registration (by 4–5%) than those remaining with NHW incumbents in both 1990 and 2010. Combined, these provide compelling evidence that California’s redistricting process induces significant differences in prior participation across MH and non-MH
districts. This bias appears regardless of whether redistricting is driven by Republican (1990), Democratic (2000), or nonpartisan (2010) efforts.

An analogous examination of Florida and Texas reveals that systematic differences in Hispanic participation induced by redistricting generalize beyond California. Shown in Figure 2, after matching on P-HVAP, blocks that will be redistricted from an NHW to a Hispanic incumbent have systematically lower rates of Hispanic registration than blocks represented by a NHW incumbent before and after. Interestingly, these biases move in the opposite direction of California, suggesting that different forces drive redistricting across these states. Yet, these highlight the same concern that systematic biases (in any direction) across districts induced at their genesis must be addressed when studying the electoral consequences of redistricting in representation.

4 Matching Results for Redistricting in 2000

Given this clear bias, we implement trimming and matching procedures to reduce demographic, socioeconomic, and electoral dissimilarities across redistricted geographies for the 2000 redistricting cycle. We construct two data sets—BASELINE and MATCHED—from California’s more than 300,000

Fig. 1 Bias in California’s Redistricting Process: QQ-plots of (a) Hispanic turnout in 2000, and Hispanic registration in (b) 1990 and (c) 2010, illustrating substantial differences across blocks to be moved from NHW incumbents to Hispanic incumbents and those with the same NHW incumbent after redistricting in 1990, 2000 or 2010, matching on P-HVAP.
blocks.\(^8\) We always define treatment blocks as having a NHW incumbent in the 2000 congressional election and a Hispanic incumbent in 2002 and 2004 elections. For BASELINE, we exclude open-seat House races in 2000 or 2002, and races with defeated incumbents in 2000, and define control blocks as having a NHW congressional incumbent in elections from 2000 to 2004. We match each treated block to a control with the most similar P-HVAP, resulting in 4687 matched-pairs.\(^9\)

Before adjustment, blocks differ greatly in Hispanic population, with 49% P-HVAP in treatment compared to 29% in control. Although matching on P-HVAP eliminated this difference, treatment and control groups in BASELINE differ in nearly every observable characteristic. Treatment blocks consistently have higher income, greater proportions of naturalized citizens, and significantly higher rates of previous Hispanic registration, which all positively correlate with Hispanic participation.

To reduce remaining imbalances, we construct MATCHED, imposing additional restrictions and conducting matching. We restrict controls to have the same NHW incumbent from 2000 to 2004. We drop blocks in each 2000 assembly-senate-congress triplet, where the number of control blocks was less than twice the number of treated blocks.\(^{10}\) To further reduce important covariate imbalances, we drop treated blocks with pre-redistricting variation outside the covariate support of control units within each triplet, shrinking the number of treatments from 4687 in BASELINE to 428 in MATCHED. These 4259 blocks were discarded because the available controls were too dissimilar in their pretreatment characteristics to provide quality matches.

Importantly, these choices involve only pretreatment stratifications. Since no Hispanic incumbent retires or is defeated in our sample, we never drop blocks enforcing that voters have the same (new) Hispanic incumbent in 2002 and 2004. Our restrictions on controls include removing blocks where sitting incumbents did not contest general elections in 2000 or 2002. Though a few incumbents lost races or retired in 2000, no voters in these districts were to be incorporated into Hispanic-represented districts. Similarly, no incumbents lost or retired in 2004 who previously represented any voters moved to new Hispanic districts in the 2000 redistricting.

\(^8\)Detailed data descriptions are presented in Table II in the Supplementary materials.

\(^9\)Pre-matching restrictions discard about 100,000 blocks, retaining 4802 treated and 197,354 controls. Matching on P-HVAP produces BASELINE and matching on the full conditioning-set produces MATCHED. We lose 115 pairs from BASELINE and 16 pairs from MATCHED converting 2010 to 2000 blocks, with final samples of 9374 and 856 blocks.

\(^{10}\)For example, some blocks reside in Congressional district 45, Assembly district 68, and State Senate district 34. These would be designated as members of triplet “cd45-ad68-sd34,” and could only be matched to each other. Only triplets with twice as many controls as treated are retained.
Finally, we match these 428 treated blocks to comparable controls in MATCHED. Treatment and controls were matched exactly on triplet indicator, ensuring that features of lower districts, including incumbent ethnicity, cannot affect our inferences. Treatment and controls within each triplet then were matched on important characteristics. Finally, we use a second genetic matching step implemented hierarchically to select matched-pairs within triplets that maximize balance across triplets—details are provided in Section 8 in the Supplementary materials. After matching, treated and control blocks in MATCHED are much more similar in their characteristics used during redistricting. All mean differences on the conditioning set are statistically insignificant, and the entire distributions are more similar than in BASELINE.11

4.1 Placebo Analysis

Before redistricting, treatment and control units have the same House, assembly, and state senate incumbents. If condition (1) holds, would-be-treatment and would-be-control blocks should be very similar in pre-redistricting Hispanic participation, controlling for characteristics used during redistricting. This observation motivates a “placebo” analysis to test whether controlling for pretreatment covariates recovers true zero effects on prior participation—required for valid cross-sectional comparisons—or if estimating difference-in-differences under condition (2) appears more appropriate.

Table 1 presents placebo results. Controlling for P-HVAP in Baseline, significant differences remain in 2000 Hispanic turnout and registration before redistricting occurs. This is in spite of having the same House incumbent in 2000 and similar proportions of Hispanic adults, suggesting that redistrictors are exceedingly effective at sorting already-engaged Hispanic voters into MH districts. Pre-existing differences are greatly reduced after trimming and genetic matching. In Table 1 for MATCHED, average Hispanic turnout in 2000 is 69% among treatments and 68% among controls, statistically indistinguishable. Average differences in Hispanic registration also shrink, although still statistically significant. Non-Hispanic registration and turnout differences are now indistinguishable from zero. The QQ-plot in Figure 3 presents placebo results for 2000 Hispanic turnout, comparing would-be-treatment and would-be-controls in MATCHED. Although turnout quantiles are more similar, some differences remain: treated blocks are slightly larger than controls in lower quantiles.

4.2 Main Results

Due to the small differences in baseline Hispanic turnout and registration, we estimate difference-in-differences defined above. Table 2 shows the difference in participation for treatments and controls, differencing before and after redistricting. As seen, the 2004–2000 differences and the 2002–2000 differences in Hispanic turnout are small or negative across every analysis. In BASELINE, the 2002–2000 differences actually show a greater drop-off in Hispanic turnout in treated than in control blocks in the midterm election following redistricting. After conditioning on important covariates in MATCHED, all turnout differences become statistically indistinguishable between treatment and control for both the 2004 general and 2002 midterm elections.

This evidence suggests that Hispanic citizens in comparable blocks are no more likely to vote by virtue of being moved to districts represented by co-ethnic politicians. Additionally, we find no co-ethnic representation effect on Hispanic registration for either the 2002–2000 or the 2004–2000 differences. The most consistent finding is that being moved to a Hispanic representative has a small negative effect on non-Hispanic turnout. The 2002–2000 differences in non-Hispanic turnout across treated and control blocks are −6.2% in BASELINE and −5.6% in MATCHED. Similar results are found for 2004–2000 differences.

To address important concerns about generalizability, we replicate analysis for Hispanic candidates running in open-seat competitive races in 2002 (CA-21 and CA-39). We also conduct sensitivity analysis to assess if remaining pre-redistricting covariate imbalances could account for post-redistricting differences observed in Hispanic turnout or registration without trimming any blocks.

11See Table I in the Supplementary materials.
Both analyses confirm our main findings. Nevertheless, severe sorting in redistricting makes it difficult to conclude that null effects generalize across California, necessitating trade-offs between trimming considerable data or making herculean assumptions about district formation.

Table 1  Intermediate analysis: 2000 cross-sectional results

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>BASELINE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic turnout</td>
<td>0.042</td>
<td>0.00</td>
</tr>
<tr>
<td>Hispanic registration</td>
<td>0.094</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-Hispanic turnout</td>
<td>0.012</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-Hispanic registration</td>
<td>0.030</td>
<td>0.00</td>
</tr>
<tr>
<td>MATCHED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic turnout</td>
<td>0.011</td>
<td>0.52</td>
</tr>
<tr>
<td>Hispanic registration</td>
<td>0.038</td>
<td>0.04</td>
</tr>
<tr>
<td>Non-Hispanic turnout</td>
<td>0.009</td>
<td>0.41</td>
</tr>
<tr>
<td>Non-Hispanic registration</td>
<td>-0.017</td>
<td>0.29</td>
</tr>
</tbody>
</table>

P-values are standard t-test probabilities.

Fig. 3  Reductions in 2000 turnout bias after matching in California: QQ-plot of Hispanic turnout in 2000 comparing blocks moved from NHW incumbents to Hispanic incumbents and those with the same NHW incumbent after redistricting in 2000, genetic matching on the full conditioning set in MATCHED.

Both analyses confirm our main findings. Nevertheless, severe sorting in redistricting makes it difficult to conclude that null effects generalize across California, necessitating trade-offs between trimming considerable data or making herculean assumptions about district formation.

See Sections 3, 7.1, and 7.2 in the Supplementary materials.
Conclusion

We develop a new research design to study the electoral effects of variation in representation, implemented here to investigate whether Hispanic incumbency impacts co-ethnic participation. A feature of the design is that we can isolate biases emerging from how districts are built during estimation. One major finding is that redistrictors in California consistently sort more engaged Hispanic voters into MH rather than non-MH districts, though this pattern is reversed in other states, highlighting important heterogeneity in redistricting largely unknown or underappreciated in previous research. Future work should investigate possible explanations for this heterogeneity in redistricting, looking more broadly at turnout by historically underpresented groups across the rest of the country. We recommend that scholars consider whether similar biases could affect their inferences about the effects of Hispanic representation. The biases are present in the data even if researchers do not explicitly use redistricting as a research design.

References


Table 2  Difference-in-difference estimates: 2004–2000 and 2002–2000 difference results

<table>
<thead>
<tr>
<th></th>
<th>2004</th>
<th>p-value</th>
<th>2002</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tr-Co</td>
<td></td>
<td>Tr-Co</td>
<td></td>
</tr>
<tr>
<td>BASELINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic turnout</td>
<td>0.046</td>
<td>0.02</td>
<td>−0.034</td>
<td>0.00</td>
</tr>
<tr>
<td>Hispanic registration</td>
<td>0.012</td>
<td>0.10</td>
<td>0.009</td>
<td>0.20</td>
</tr>
<tr>
<td>Non-Hispanic turnout</td>
<td>−0.145</td>
<td>0.00</td>
<td>−0.062</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-Hispanic registration</td>
<td>−0.013</td>
<td>0.07</td>
<td>0.003</td>
<td>0.63</td>
</tr>
<tr>
<td>MATCHED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic turnout</td>
<td>0.014</td>
<td>0.84</td>
<td>0.013</td>
<td>0.75</td>
</tr>
<tr>
<td>Hispanic registration</td>
<td>0.032</td>
<td>0.21</td>
<td>−0.008</td>
<td>0.76</td>
</tr>
<tr>
<td>Non-Hispanic turnout</td>
<td>−0.082</td>
<td>0.24</td>
<td>−0.056</td>
<td>0.10</td>
</tr>
<tr>
<td>Non-Hispanic registration</td>
<td>0.033</td>
<td>0.15</td>
<td>0.037</td>
<td>0.11</td>
</tr>
</tbody>
</table>

P-values are OLS with Huber-White standard errors and interacted time fixed effects.