Supplemental Online Appendix to "Cause or Effect? Turnout in Hispanic Majority-Minority Districts"

John A. Henderson[*]
Assistant Professor
Dept. of Political Science
Yale University

Jasjeet S. Sekhon[†]
Professor
Travers Dept. of Political Science
Department of Statistics
UC Berkeley

Rocío Titiunik[‡]
Assistant Professor
Dept. of Political Science
University of Michigan

5/12/2016 (09:17)

[*]<john.henderson@yale.edu>, http://jahenderson.com/, Institution for Social and Policy Studies, Yale University, 77 Prospect Street, New Haven, CT 06520-8209

[†]<sekhon@berkeley.edu>, http://sekhon.berkeley.edu/, Center for Causal Inference and Program Evaluation and Institute of Governmental Studies, University of California, 109 Moses Hall, #2370, Berkeley, CA 94720-2370

[‡]<titiunik@umich.edu>, http://www.umich.edu/~titiunik, Center for Political Studies, ISR, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106-1248

1 Introduction

This supplemental appendix to the paper "Cause or Effect? Turnout in Hispanic Majority-Minority Districts" is intended for online publication only. In the following sections, we offer additional descriptive information about the dataset on which all our analysis in the paper is based, particularly information on the specific districts and incumbents included in each step of our analysis. We also present a discussion of the census-block conversion procedure we use to link 2010 to 2000 blocks to build our outcome measures for 2002, 2004 and 2006, and the surname matching methods used by the California Statewide Database (SWDB) to link ethnicity to turnout and registration. We provide a brief discussion of California's experience with racial redistricting over the last three cycles (1990, 2000, and 2010), and give additional selection accounts that may help explain the participation bias we observe emerging from redistricting at each cycle. We then present results from our sensitivity analysis, as well as supplementary statistical analyses and robustness checks to the research design and specifications used in the main body of paper. Lastly, we include additional QQ-Plot figures showing further evidence of the turnout and registration bias we find stemming from the drawing of majority-minority (MM) districts in California.

2 Main Balance Table

Table I: Covariate Balance for Baseline and Matching Analysis

	BASELINE			MATCHED		
	Mean	P-value		Mean	P-value	
	Tr - Co	μ-test	ks-test	Tr - Co	μ-test	ks-test
CONDITIONING SET						
Voting Age Population (VAP)	9.79	0.00	0.00	1.48	0.78	0.79
Black VAP	0.02	0.00	0.00	0.00	0.75	0.99
Hispanic VAP	0.00	1.00	1.00	0.00	0.77	0.16
HH Income 39k	-0.06	0.00	0.00	0.00	0.89	0.10
HH Income 40k to 74k	0.02	0.03	0.00	0.00	0.97	0.04
HH Income $\geq$ 100k	0.02	0.17	0.00	0.00	0.95	0.01
Pop Highschool or Less	-0.01	0.00	0.00	0.01	0.71	0.06
Pop Foreign	0.04	0.00	0.00	0.00	0.90	0.10
ADDITIONAL COVARIATES						
Pop Non-Citizen	0.00	0.06	0.00	0.00	0.76	0.08
Pop Naturalized Citizen	0.04	0.00	0.00	0.00	0.82	0.00
Registration Total 1998	0.03	0.00	0.00	0.02	0.14	0.11
Registration Hispanic 1998	0.03	0.00	0.00	0.00	0.38	0.13
Registration Democrat 1998	-0.01	0.02	0.00	0.02	0.05	0.00
Registration Republican 1998	0.02	0.00	0.00	-0.01	0.60	0.79
Registration Total 2000	0.04	0.00	0.00	0.02	0.08	0.02
Registration Democrat 2000	0.01	0.00	0.00	0.02	0.05	0.00
Registration Republican 2000	0.00	0.97	0.00	0.00	0.68	0.07
Dem. Vote U.S. Senate 1998	-0.13	0.00	0.00	-0.04	0.25	0.41
Dem. Vote Governor 1998	-0.13	0.00	0.00	-0.05	0.17	0.02
Dem. Vote U.S. House 1998	-0.12	0.00	0.00	-0.06	0.08	0.41
Dem. Vote President 2000	0.01	0.41	0.00	0.00	0.92	0.07
Dem. Vote U.S. Senate 2000	-0.01	0.19	0.00	0.00	0.84	0.07
Dem. Vote U.S. House 2000	-0.05	0.00	0.00	-0.01	0.56	0.07
Pop Female	0.02	0.00	0.00	0.01	0.09	0.00
Pop 25 to 44 Years	-0.01	0.00	0.00	-0.02	0.00	0.00
Pop 45 to 59 Years	0.00	0.00	0.00	0.00	0.60	0.57

3 Details About Dataset: Trimming and Matching Choices, Congressional Districts and Incumbents Included in the Analysis

Table II in this Supplemental Appendix gives details about the BASELINE and MATCHED datasets we use in our analyses. (Please refer to Section 4 in the main paper for a more-detailed description of these choices.) In this section, we present additional information about the particular congressional races included in each of these two datasets, with the goal of providing precise information about which races are represented in each stage of our main analysis. In Table III below we show the incumbents who were representing each of the 53 congressional districts in California at the moment of the 2002 general election, the first post-redistricting election that induces the treatment assignment we analyze in the paper. For each incumbent running for reelection in 2002, the third and fourth columns show whether this incumbent was Hispanic or Non-Hispanic White.

Table II: Nested Datasets Used in Analysis at Census Block Level

	BASELINE
Number treated Number control	4,687 census blocks (2000 U.S. Census) 4,687 census blocks (2000 U.S. Census)
Restrictions	Closed seat in congressional district race in 2000 and 2002 Congressional incumbent running in 2000 wins 2000 election
Treatment	Non-Hispanic White congressional incumbent running in 2000 Hispanic congressional incumbent running in 2002)
Control	Non-Hispanic White congressional incumbent running in 2000 and 2002
Matching	Match on *proportion* Hispanic voting age population only.
	MATCHED
Number treated Number control	428 census blocks (2000 U.S. Census) 428 census blocks (2000 U.S. Census)
Restrictions	Same restrictions as in BASELINE
Treatment	Non-Hispanic White congressional incumbent running in 2000 Hispanic congressional incumbent running in 2002
Control	*Same* non-Hispanic White congressional incumbent running in 2000 and 2002
Matching	Define state house, state senate and congressional district “triplet” for every block in 2000. Drop triplets where number of controls less than twice number of treated. Within triplets, trim treatments whose demographics are outside of common support. In addition, (i) match exactly on 2000 triplet; (ii) within each triplet, match on demographics; (iii) then match across triplets, i.e., select matched pairs within each triplet such that overall balance *across* triplets is maximized.

In addition, the last two columns of Table III indicate whether each congressional district is represented in the BASELINE and/or MATCHED datasets. For example, in district 1 incumbent Michael Thompson runs for reelection in 2002; Thompson is Non-Hispanic White and census blocks from his district are included in the BASELINE dataset but not in the MATCHED dataset.

Table III: All incumbents in 2002 California U.S. House Election

District	Incumbent name	Non-Hispanic White	Hispanic	In Baseline	In Matched
1	Thompson, Michael	Yes	No	Yes	No
2	Herger, Walter William (Wally)	Yes	No	Yes	No
3	Ose, Doug	Yes	No	Yes	No
4	Doolittle, John Taylor	Yes	No	Yes	No
5	Matsui, Robert Takeo	No	No	No	No
6	Woolsey, Lynn C.	Yes	No	Yes	No
7	Miller, George	Yes	No	Yes	No
8	Pelosi, Nancy	Yes	No	Yes	No
9	Lee, Barbara	No	No	No	No
10	Tauscher, Ellen O'Kane	Yes	No	Yes	No
11	Pombo, Richard William	Yes	No	Yes	No
12	Lantos, Thomas Peter	Yes	No	Yes	No
13	Stark, Fortney Hillman (Pete), Jr.	Yes	No	Yes	No
14	Eshoo, Anna Georges	Yes	No	Yes	No
15	Honda, Mike	No	No	No	No
16	Lofgren, Zoe	Yes	No	Yes	No
17	Farr, Sam	Yes	No	Yes	No
18	Open Seat	–	–	No	No
19	Radanovich, George P.	Yes	No	Yes	No
20	Dooley, Calvin M.	Yes	No	Yes	No
21	Open Seat	–	–	No	No
22	Thomas, William Marshall	Yes	No	Yes	No
23	Capps, Lois	Yes	No	Yes	No
24	Gallegly, Elton W.	Yes	No	Yes	No
25	Mckeon, Howard P. (Buck)	Yes	No	Yes	No
26	Dreier, David Timothy	Yes	No	Yes	Yes
27	Sherman, Brad	Yes	No	Yes	No
28	Berman, Howard Lawrence	Yes	No	Yes	No
29	Schiff, Adam	Yes	No	Yes	Yes
30	Waxman, Henry Arnold	Yes	No	Yes	No
31	Becerra, Xavier	No	Yes	No	No
32	Solis, Hilda	No	Yes	Yes	Yes
33	Watson, Diane Edith	No	No	No	No
34	Roybal-Allard, Lucille	No	Yes	Yes	No
35	Waters, Maxine	No	No	No	No
36	Harman, Jane L.	Yes	No	Yes	No
37	Millender-Mcdonald, Juanita	No	No	No	No
38	Napolitano, Grace F.	No	Yes	Yes	Yes
39	Open Seat	–	–	No	No
40	Royce, Edward Randall	Yes	No	Yes	Yes
41	Lewis, Charles Jeremy (Jerry)	Yes	No	Yes	Yes
42	Miller, Gary G.	Yes	No	Yes	Yes
43	Baca, Joe	No	Yes	Yes	Yes
44	Calvert, Ken	Yes	No	Yes	No
45	Bono, Mary	Yes	No	Yes	No
46	Rohrabacher, Dana	Yes	No	Yes	No
47	Sanchez, Loretta	No	Yes	Yes	Yes
48	Cox, Charles Christopher	Yes	No	Yes	Yes
49	Issa, Darrell	No	No	No	No
50	Cunningham, Randall (Duke)	Yes	No	Yes	No
51	Filner, Bob	Yes	No	Yes	No
52	Hunter, Duncan Lee	Yes	No	Yes	No
53	Davis, Susan A.	Yes	No	Yes	No

Table IV: Hispanic incumbents in 2002 California U.S. House Election

District	Incumbent name	In Baseline	In Matched
31	Becerra, Xavier	No	No
32	Solis, Hilda	Yes	Yes
34	Roybal-Allard, Lucille	Yes	No
38	Napolitano, Grace F.	Yes	Yes
43	Baca, Joe	Yes	Yes
47	Sanchez, Loretta	Yes	Yes

Over 80% (43 out of 53) of the congressional districts in California as of November 2002 are represented in our main analysis in the paper, either in BASELINE or in both BASELINE and MATCHED. As can be seen in Table III, the only districts that are excluded from our main analysis are those districts where the incumbent running for reelection in 2002 was neither Hispanic nor Non-Hispanic White (districts 5, 9, 15, 33, 35, 37 and 49), and those "open seats" that had no incumbent running in the 2002 election (districts 18, 21 and 39) – see below for analyses that include open seats.

Tables IV and V present subsets of the information presented in Table III in order to highlight information about our treatment and control groups, respectively. In Table IV, we list only those districts in Table III that are represented by a Hispanic incumbent as of November 2002; as discussed in the main body of the paper, these are the districts from which we selected our treatment group for our main analysis. Table IV shows that there are six Hispanic incumbents in California running for reelection in 2002, five of which are included in BASELINE (Solis, Roybal-Allard, Napolitano, Baca, Sanchez) and four of which are included in MATCHED (Solis, Napolitano, Baca, Sanchez). Thus, our analysis excludes only district 31 represented by Xavier Becerra.

To complement Table IV, Table V lists only those districts in Table III that are represented by a Non-Hispanic White (NHW) incumbent as of November 2002; these are the districts from which we selected our control group for our main analysis in the paper. As shown in Table V, all 37 NHW incumbents who are running for reelection in the 2002 general election are included in our analysis. In particular, all 37 NHW incumbents are included in BASELINE and 6 out these 37 are included in MATCHED. The main reason why a large number of NHW incumbents are not included in MATCHED is because our research design restricts the control group to be originally in the *same* congressional district as our treatment group: those blocks moved from a NHW incumbent to a Hispanic incumbent comprise our treatment group, and our control group are those blocks that before redistricting are in the same congressional district from where the treated blocks are selected. There are many districts represented by a NHW incumbent that do not cede territory to districts represented by a Hispanic incumbent; our research design in MATCHED, since it focuses on treated and control blocks that were represented by the *same* NHW incumbent before redistricting, excludes those districts. However, as mentioned above, all districts represented by a NHW incumbent are included in our analysis because they are included in BASELINE.

Table V: Non-Hispanic-White incumbents in 2002 California U.S. House Election

District	Incumbent name	In Baseline	In Matched
1	Thompson, Michael	Yes	No
2	Herger, Walter William (Wally)	Yes	No
3	Ose, Doug	Yes	No
4	Doolittle, John Taylor	Yes	No
6	Woolsey, Lynn C.	Yes	No
7	Miller, George	Yes	No
8	Pelosi, Nancy	Yes	No
10	Tauscher, Ellen O'Kane	Yes	No
11	Pombo, Richard William	Yes	No
12	Lantos, Thomas Peter	Yes	No
13	Stark, Fortney Hillman (Pete), Jr.	Yes	No
14	Eshoo, Anna Georges	Yes	No
16	Lofgren, Zoe	Yes	No
17	Farr, Sam	Yes	No
19	Radanovich, George P.	Yes	No
20	Dooley, Calvin M.	Yes	No
22	Thomas, William Marshall	Yes	No
23	Capps, Lois	Yes	No
24	Gallegly, Elton W.	Yes	No
25	Mckeon, Howard P. (Buck)	Yes	No
26	Dreier, David Timothy	Yes	Yes
27	Sherman, Brad	Yes	No
28	Berman, Howard Lawrence	Yes	No
29	Schiff, Adam	Yes	Yes
30	Waxman, Henry Arnold	Yes	No
36	Harman, Jane L.	Yes	No
40	Royce, Edward Randall	Yes	Yes
41	Lewis, Charles Jeremy (Jerry)	Yes	Yes
42	Miller, Gary G.	Yes	Yes
44	Calvert, Ken	Yes	No
45	Bono, Mary	Yes	No
46	Rohrabacher, Dana	Yes	No
48	Cox, Charles Christopher	Yes	Yes
50	Cunningham, Randall (Duke)	Yes	No
51	Filner, Bob	Yes	No
52	Hunter, Duncan Lee	Yes	No
53	Davis, Susan A.	Yes	No

Table VI: 2002 and 2004 U.S. House Primaries in California

District	2002 Primary			2004 Primary		
	Open Seat	Dem Margin	Rep Margin	Open Seat	Dem Margin	Rep Margin
1	No	100.0	100.0	No	100.0	100.0
2	No	21.9	80.5	No	13.9	100.0
3	No	100.0	100.0	Yes	100.0	3.1
4	No	100.0	55.0	No	100.0	100.0
5	No	100.0	100.0	No	100.0	100.0
6	No	61.0	100.0	No	68.0	100.0
7	No	100.0	100.0	No	100.0	100.0
8	No	86.2	50.2	No	100.0	100.0
9	No	69.7	23.7	No	100.0	100.0
10	No	66.8	11.1	No	100.0	100.0
11	No	27.9	74.0	No	100.0	100.0
12	No	100.0	100.0	No	53.7	10.9
13	No	100.0	100.0	No	100.0	100.0
14	No	100.0	100.0	No	100.0	100.0
15	No	100.0	100.0	No	100.0	100.0
16	No	100.0	100.0	No	100.0	100.0
17	No	82.2	26.3	No	82.3	49.8
18	Yes	14.6	21.4	No	100.0	37.1
19	No	100.0	100.0	No	100.0	100.0
20	No	100.0	5.1	Yes	46.3	59.3
21	Yes	100.0	4.0	No	100.0	100.0
22	No	100.0	100.0	No	–	100.0
23	No	100.0	24.7	No	100.0	100.0
24	No	100.0	100.0	No	100.0	100.0
25	No	100.0	68.9	No	12.2	100.0
26	No	100.0	100.0	No	23.7	67.1
27	No	100.0	100.0	No	100.0	100.0
28	No	100.0	100.0	No	63.8	100.0
29	No	100.0	100.0	No	100.0	30.3
30	No	79.1	100.0	No	100.0	100.0
31	No	100.0	29.8	No	78.9	100.0
32	No	100.0	100.0	No	100.0	–
33	No	100.0	100.0	No	100.0	–
34	No	100.0	100.0	No	100.0	100.0
35	No	100.0	36.9	No	100.0	100.0
36	No	100.0	14.6	No	100.0	12.0
37	No	55.4	100.0	No	46.2	100.0
38	No	30.0	100.0	No	57.8	–
39	Yes	4.2	40.8	No	100.0	100.0
40	No	100.0	100.0	No	3.4	100.0
41	No	100.0	100.0	No	100.0	100.0
42	No	100.0	100.0	No	100.0	100.0
43	No	100.0	100.0	No	100.0	100.0
44	No	100.0	45.0	No	100.0	71.6
45	No	100.0	100.0	No	41.6	71.9
46	No	100.0	100.0	No	16.0	67.1
47	No	100.0	7.4	No	100.0	18.7
48	No	100.0	82.6	No	100.0	100.0
49	No	100.0	100.0	No	100.0	100.0
50	No	100.0	73.4	No	100.0	100.0
51	No	40.8	5.5	No	53.3	39.4
52	No	100.0	100.0	No	100.0	100.0
53	No	100.0	16.1	No	100.0	100.0

4 Details on Data Conversion and Ethnicity Surname Matching

4.1 Converting 2010 to 2000 Census Blocks

Recall from Section 3.1 of our paper that

> "For 1998 and 2000, the SWDB includes turnout and registration figures for 2000 census blocks. However, for 2002 and 2004, participation data are only available at the 2010 census block, and thus must be converted to 2000 census block. We take a standard approach, using Census Block Relationship Files to convert participation data at the 2010 census-block to the 2000 census-block level."

To measure participation outcomes for 2002 to 2006, we use Census Block Relationship Files provided by the U.S. Census Bureau to convert participation data at the 2010 census-block to the 2000 census-block level.[1] The relationship between 2010 and 2000 blocks may be one-to-one (the 2010 and 2000 blocks are identical) many-to-one (several 2010 blocks are contained in one 2000 block), one-to-many (one 2010 block is split between several 2000 blocks), or many-to-many (parts of a 2010 block are contained in a 2000 block that also contains parts of other 2010 blocks). The first two relationships do not pose problems for converting participation figures, since the 2010 census block is never split. This simply requires leaving the information as is (one-to-one) or aggregating several 2010 blocks into one 2000 block (many-to-one). These procedures introduce no error of any kind, as they only sum the totals of different 2010 blocks to produce total counts for corresponding 2000 blocks.

The latter two relationships pose some difficulty, since these involve portioning the turnout and registration figures of a 2010 block into several 2000 blocks, potentially introducing error. To do this, we split 2010 blocks into 2000 blocks according to the area weights provided in the 2010 Census Block Relationship files. For example, if one-fourth of the area of a 2010 census block belongs to 2000 block A and three-fourths belongs to 2000 block B, we assign 25% of the turnout and registration counts to block A and 75% to block B. This procedure of converting 2010 blocks to 2000 blocks naturally introduces some inaccuracies, but there are two reasons why this dataset is reliable. First, over seventy percent of the 2000 blocks in our dataset are error-free, comprised of one or more

[1] Census Block Relationship files provide detailed information about the linkage between 2000 and 2010 tabulation census blocks, in particular, the (area) proportion of each 2010 block that is contained in each 2000 block.

whole 2010 blocks that do not involve splitting. Second, we identify those 2000 blocks that are obtained from aggregate parts of split 2010 blocks, and do contain error, and reproduce our analysis entirely discarding these 2000 blocks. In doing so we recover very similar (and largely null) results as our main analysis, as seen in Table VII below.

Table VII: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results for Perfectly Converted Blocks

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.009	0.73	-0.049	0.00
Hispanic Registration	0.011	0.24	0.011	0.24
Non Hispanic Turnout	-0.167	0.00	-0.065	0.00
Non Hispanic Registration	-0.022	0.03	-0.003	0.73
MATCHED				
Hispanic Turnout	-0.063	0.59	-0.054	0.39
Hispanic Registration	-0.012	0.77	-0.033	0.41
Non Hispanic Turnout	-0.240	0.03	-0.119	0.03
Non Hispanic Registration	-0.007	0.86	0.015	0.69

Note: Datasets are described in the main paper. Registration is measured as levels in 2000, 2002, or 2004, for each 2000 U.S. census block over Hispanic or Non-Hispanic voting age population. Turnout measures are constructed as 2000, 2002, or 2004 turnout over Hispanic or Non-Hispanic registration in 2000, as reported in the SWDB. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

4.2 U.S. Census Hispanic Surname Matching

The individual-level data we collect from the SWDB, originates from the Registrar of Voters in each of California's fifty-eight counties. These registration files do not contain racial or ethnic information. To determine whether a registered individual is Hispanic, the SWDB matched surnames to the Passel-Word Hispanic Surname List produced by the U.S. Census Bureau. All persons whose surname appeared in this list were classified as Hispanic, and all others were classified as non-Hispanic. The SWDB relies solely on surname matching to construct its Hispanic turnout measure, and it does not incorporate census data to update the ethnic classification of voters.[2]

Measuring minority turnout poses some challenge. Self-declared information on race and ethnicity is not collected during the voter registration process, thus this information must be inferred from the available data (i.e., name and address). There are two common approaches to infer race and ethnicity from this information. The first, "Surname List Method" (SLM), is used by the SWDB, and involves linking individuals' last names to a list of surnames that are typical of a particular ethnic or racial group. When the individual's last name is found in the list, he or she is classified as belonging to that particular race or ethnicity. The second, "Census Geocoding Method" (CGM), links individuals' addresses to census data about their areas of residence, and uses ethnic and racial concentrations in these areas to infer the likelihood that an individual is of a certain race or ethnicity.

Surname list methods work well for ethnicities with distinctive last names, especially Asian and Hispanic groups, but do a poor job distinguishing blacks from NHWs. Census-based methods identify blacks much more effectively, but have less ability to identify Hispanics or Asians. It is possible to combine both sources of information to increase the accuracy of ethnic and racial classification (e.g., Fiscella and Fremont 2006). Compared to SLM, combined approaches produce large increases in the accuracy of classification for blacks and moderate increases for Asians, but the increases for Hispanics are minor.[3] Indeed, the evidence shows that surname matching methods

[2]The improvement in accuracy in Hispanic classification from incorporating geographical data is very small. Additionally, in our statistical analysis, we condition on all census variables that *would have been used* in a combined approach of ethnic classification and therefore we exploit both sources of information (geographic and surname-based) to make inferences. This is important, since the census characteristics that allow for increased accuracy in prediction of Hispanic ethnicity are likely to be correlated with the way in which the redistricting process is implemented.

[3]For instance, Elliott, Fremont, Morrison, Pantoja and Lurie (2008) propose using the racial composition of the census block as the baseline probability of a block-resident belonging to that ethnicity. These prior probabilities are then updated with information from surname lists using Bayes' Theorem to produce posterior probabilities of ethnic or racial

are highly effective for inferring Hispanic ethnicity, and that incorporating geographical information improves classification only slightly. In contrast, using both sources of information is considerably more accurate when classifying blacks than methods that only use residential and demographic data.

5 More Details about Redistricting in California

5.1 Racial Redistricting in California

The passage of the Voting Rights Act (VRA) in 1965 had profound consequences for minority and Hispanic representation. Over the last half-century, the number of Hispanic members of Congress grew seven-fold, from 4 (0.7%) to 31 (7.1%), with 10 of these from California (U.S. Census Bureau 2012). This expansion was bolstered by two legal changes in the 1970s and 1980s that strengthened minority protections during reapportionment. Gains for Hispanic representation were first made possible by Congress amending the VRA in 1975 to include Hispanics as a protected minority group. The most significant changes were brought about by the 1982 amendments, along with the 1985 ruling in *Thornburg v. Gingles*, which barred redistricting plans that diluted minority voting strength even if drawn without an explicitly discriminatory purpose. Following these developments, the VRA was interpreted as mandating MM districts whenever feasible to expand minority representation (Gay 2001).

Following the 1985 *Thornburg* ruling, and a decade of population growth resulting in seven new congressional seats to apportion in the 1990 cycle, events in California conspired to dramatically alter the political landscape of the state. According to contemporary accounts, the Republican party quickly grasped the importance of racial districting as a means to expand party seat shares in Washington and Sacramento (Kousser 1997, 1998; Grofman and Handley 1998). Comparatively, the Democratic party was internally split. Though minority groups within the Democratic party would gain greater representation through racial redistricting, these gains typically would come at the expense of reducing their party's total seat share in the legislature (Bickerstaff 2007; Grofman

membership. In verifying their approach using a dataset of 1,973,362 enrollees in a national health plan, the authors find that using both surname and geocoding data results in a classification correlated with self-reported Hispanic ethnicity at 0.79. In comparison, SLM classification correlates with self-reported Hispanic ethnicity at a factor of 0.77. However, the classification using CGM, is correlated with reported Hispanic membership at only 0.49.

and Reynolds 1996; Hill 1995; Petrocik and Desposato 1998). For their part, the Democrats sought to use their majorities in the state Assembly and Senate to control the extent of racial redistricting, offering new MM seats but limiting the degree to which these would hurt party incumbents.

Republican Pete Wilson was elected Governor, however, just in time to dash Democratic plans. According to Kousser (1997) the new Governor devised an effective redistricting strategy: "Refuse to negotiate...appoint a commission nominally balanced in partisanship...veto all legislative plans, turn the issue over to the (Republican controlled) state Supreme Court...and suggest that the court's Special Masters use the commission's proposal" (Kousser 1997, p 150-51). Thus, the Democrats' efforts to control the redistricting process with a moderate plan was thwarted by the Governor's veto. This veto resulted in a court-led plan, devised by Republican judges, to be adopted. Barring any significant constraint from the legislature, the judges' plan greatly expanded the number of majority-minority districts in California largely in accordance with the contemporary interpretation of the VRA (Kousser 1997; Gay 2001; Grofman and Handley 1998).

By the 2000 redistricting cycle, the legal and political climate had changed considerably. In two major reversals, the Supreme Court decided in *Shaw v. Reno* (1993) and *Miller v. Johnson* (1995) that redistricting plans could be held impermissible if race was the predominant factor in drawing the boundaries of a particular district (Grofman and Handley 1998; Kousser 1997, 1998). Though the 1990 California plan was upheld after higher court review, the ruling left little doubt that future plans to expand MM districts would receive considerable and careful scrutiny (Gay 2001; Kousser 1998). On the other hand, the court retained the provisions in the 1982 VRA amendments prohibiting plans that had the effect of diluting minority vote strength. Given the considerable expansion in the number of MM districts in the 1990 plan, the court did not leave much room for the legislature to make major changes to the racial composition of districts a decade later.

Unified government and large legislative majorities put the Democratic party fully in control of the redistricting process in 2000. Yet, with only one additional congressional seat to apportion, the general belief that the party had "maxed-out" its possible areas of seat-expansion, and the possibility of legal challenge by Republicans and minority Democrats based on *Shaw v. Reno*, the prospects for major party gains looked dim (Barabak 2001; Kousser 1997; McDonald 2004). In response, the party devised a bipartisan 'incumbent protection' plan with the purpose of shoring up existing districts to make them less competitive (McDonald 2004). Thus, the geographies and populations that were moved in 2000, were done so with express end of retaining the balance of partisan and minority control of seats established after 1990.

As a consequence of the legal and political developments across the nation, the 1990 reapportionment cycle was the most racially progressive round of redistricting in American history (Gay 2001; Lublin 1997). In that cycle California led the nation, more than doubling the number of Hispanic and mixed-ethnic majority-minority congressional districts from 6 to 13 in Congress, placing it first amongst states in minority representation. A decade later, despite legal and political changes, the VRA-restriction against diluting minority voting strength and a log-roll coalition of Republicans and minority Democrats, limited the Democratic majority's ability to redraw districts to expand its legislative seat shares. Thus in 2000, California's incumbent protection plan effectively locked in existing MM districts until the 2010 apportionment cycle, which similarly retained MM districts through a non-partisan plan following the adoption of the California Citizens Redistricting Commission (Pierce and Larson 2011).

5.2 Defining Ethnicity

The way in which we defined our treatment of interest and measured ethnicity of voters and incumbents carries important implications for the definition of ethnicity that our analysis is implicitly employing. To date, multiple definitions of ethnicity have been offered, and the lack of agreement has complicated the study of important issues in ethnic politics (see, e.g., Chandra 2006; Chandra and Wilkinson 2008). One definition considers ethnicity as an identity category in which eligibility for membership is determined by descent-based attributes (Chandra 2006). This definition, however, does not explicitly consider the cognitive processes associated with the acquisition and preservation of ethnic identity that has been emphasized by psychology scholars. For example, Bernal, Knight, Garza, Ocampo and Cota (1990) define ethnic identity as primarily concerned with the individual's sense of self as a member of the group, and identify five elements of ethnic identity—ethnic self-identification, ethnic constancy, use of ethnic role behaviors, ethic knowledge and ethnic preference and feelings. The psychological aspects of this definition are important for our study of political behavior, because they suggest that the intensity or the degree to which the individual identifies with his/her ethnic group may have behavioral consequences and thus affect political participation decisions. For example, Spears, Doosje and Ellemers (1997) have shown that individuals who identify very strongly with their group are more likely to use group-level strategies to face threats to social identity.

The treatment we study, shared ethnicity, has forced us to adopt a practical definition of this concept. Our classification of incumbents as Hispanic is based on the Hispanic Americans in Congress website, maintained by the Congressional Research Service. Since this dataset is based on candidates' self-reported characteristics, if a candidate appears listed as Hispanic in the website, this will reflect the candidate's self-identification, and thus capture the important psychological aspect of ethnicity. For voters, the challenger is much greater, because the phenomenon of redistricting occurs at the census block level and thus the census block is our unit of analysis. The only source of data for this unit of analysis is registration files and census counts, both sources that limit the subtlety with which we can measure ethnic identification. As explained above, our turnout and registration data source, the SWDB, classifies a registered voter as Hispanic based on his or her surname. This measure is crude, and can only be considered a proxy for Hispanic descent attributes and psychological markers of Hispanic identity, not a direct measurement.

We note also that because our unit of analysis is aggregate, our design is not well suited to investigate whether individuals whose ethnic identity is stronger are more likely to be affected by having a co-ethnic incumbent than those with a weaker sense of identification. Under the assumptions of our research design, however, the heterogeneity in the strength of ethnic identity within blocks does not affect our inferences, because this heterogeneity should be on average equal for both treated and control blocks. Our inferences at the census block level are informative about the participatory effects of co-ethnic representation, given the average intensity of ethnic identification at the block level in the final sample we consider.

5.3 Candidate-Specific Factors

Another consideration in the way we investigate incumbency using our research design is that representation is a multidimensional 'exposure,' that is, changing incumbents alters many things beyond just ethnicity. In some contexts, we can fix important factors to be constant, for instance partisanship in California, since virtually all Hispanic House representatives in the state are Democrats. Further, we can redefine treatment to investigate other sorts of changes in incumbency, including as these interact, for instance representation by Hispanic women. Though the quality of inferences in using our design could potentially depend on having enough representatives of any particular intersection of interest.

Yet, an important limitation to the study of representation generally (including in using our design), is that it is very difficult to isolate the effects of idiosyncratic or candidate-specific factors that might influence political engagement or turnout. Such factors could be theoretically or empirically important, but would be exceedingly challenging to empirically measure since multiple factors would all cluster together in one particular case. And this is a problem general to both observational and experimental research. Such candidate-specific factors could also yield bias in using our design to study the participation effects of Hispanic representation, if such factors tend to correlate with Hispanic incumbency. If these factors are independent of incumbent's ethnicity, these would not bias estimates, though could increase variance in turnout not explained by co-ethnicity. One way to assess this concern that candidate-level factors might add noise or bias to findings is to include candidate fixed-effects when estimating *difference-in-differences*. Importantly, doing so does not alter our results.

6 Additional Bias Generating Selection Accounts of Redistricting

A number of accounts that may explain the bias induced during redistricting rely on redistrictors having direct or indirect access to turnout or registration figures when drawing district boundaries at some stage in the apportionment process. From conversations with practitioners, we learned that there is some variation in the prior participation information that is typically used to draw district boundaries. Some of the clearest evidence that certain states have used turnout data comes from Texas' 2010 redistricting of the 23rd Congressional district that failed to obtain preclearance in a 2012 ruling. Unlike in California, Texas Republicans in 2010 redistricted Hispanic voters to weaken Democratic incumbents through minority vote dilution. The D.C. circuit court cited clear evidence of these efforts by "mapdrawers (who) consciously replaced many of the district's active Hispanic voters with low-turnout Hispanic voters in an effort to strengthen the voting power of CD 23's Anglo citizens." In other words, they sought to reduce Hispanic voters' ability to elect without making it look like anything in CD 23 had changed" (Texas v. United States 2012, p. 32). The court uncovered communications between lawmakers and map-drawers directing the latter to create maps that sustained previous Hispanic population figures, but lowered Hispanic and minority turnout (Texas v. United States 2012). Legislators also provided sample maps to redistrictors that considered "'voting patterns and ethnicity' to see what could be done 'to change the district"' (Texas v. United States 2012). Although this evidence may not generalize to other states, minority turnout (which is not collected during voter registration in Texas) was clearly used to alter district boundaries in a way that was likely to attract scrutiny.

Most states, including California, do not require voters to provide race or ethnicity during voter registration.[4] Yet, turnout (abstention) by ethnicity data may be used in California by the parties or prominent analysts that influence the redistricting proposals that get considered. Since these data are not universally collected in registration forms, parties and analysts typically use ecological inference methods to estimate turnout rates by race or ethnicity. The most common method used is double regression, which is a two step estimator (Lichtman 1991). More involved methods make use of the full $R \times C$ contingency table with an abstention column alongside party vote share (e.g., Cho and Judge 2009; Greiner and Quinn 2009). Many people are involved in redistricting, yet only some must use turnout estimates (or its proxies) for turnout to influence the outcome.[5]

[4]As of 2004, eleven states collected race or ethnicity data during voter registration, and in only three of those states (Alabama, North Carolina, and South Carolina) was providing race information mandatory. California did not collect such data until 2003, when a bill passed by the state legislature called for registration forms to have a space for providing information on race and ethnicity. Reporting this information, however, is optional. See Cruz and Hayes (2009).

[5]A sharp example of only some actors having to use turnout data is offered by California in 2010. Although the full story of the 2010 redistricting has yet to be written, many practitioners claim that, though the non-partisan redistricting commission did not use turnout and partisan registration data, many of the sample maps that they based the final plan on did. Hence turnout and partisan registration information was influential in how the final plan was drawn. For example, see Pierce and Larson (2011).

Practitioners have also indicated to us that Democratic incumbents often lobby to place certain Hispanic precincts into the MM districts they represent, due to their party's prior experience conducting get-out-the-vote (GOTV) campaigns in those communities. This may make it easier for these incumbents to mobilize friendly voters in future elections, or politicians may simply be more familiar with these political geographies given their party's past electoral activity there. Additionally, Hispanic community leaders may prefer their areas to be moved into an Hispanic-majority district, and may lobby to influence the particular district maps to make this happen. Having been targeted before, voters in these areas may be more likely to vote again, suggesting that prior electoral mobilization may account for these differences in minority turnout without redistrictors explicitly utilizing turnout data.

Spurious participation differences also could emerge indirectly, as an artifact of the geographic or political constraints on redistricting plans. The most relevant of these limits include compactness, contiguity, and respect for natural boundaries and communities criteria that tend to prohibit drawing 'expansive', 'unnatural' or 'unusual' districts (Cain, MacDonald and Hui 2006). To redistrict, legislatures must take geographies and populations as they naturally occur when redrawing district boundaries. This may lead to plans that select particular demographic characteristics that predict voting, inducing spurious participation differences across MM and non-MM districts. Urban and suburban minority populations reside in relatively concentrated areas, making them more difficult to break apart without violating community-of-interest or non-dilution rules (Cain, MacDonald and Hui 2006). Conversely, rural minority populations are widely dispersed, dimming the prospects for contiguous rural MM districts that satisfy the 'one-person, one-vote' constitutional requirement (U.S. Census Bureau 2012). Thus, it may be prohibitively difficult to create minority opportunity seats outside of urban and suburban areas. Yet, the baseline rate of minority participation is much higher in urban and especially suburban areas, compared to rural geographies (U.S. Census Bureau 2008). The net result may be a redistricting process that is limited to creating MM districts from non-rural minority populations, more likely to register and vote by virtue of their demographic and residential characteristics.

7 Robustness Analysis: Sensitivity Tests, Open Seat Analysis and Supplementary Results

7.1 Sensitivity Analysis and Results

In a prior version of our paper we wrote:

> "In observational work on minority empowerment, researchers assume that units with similar values of predetermined characteristics X, have the same probability of being represented by His- panic incumbents, justifying a comparison of treated and control blocks with similar X. Since we observe much of the redistricting process, including the data used by redistrictors, we think this assumption is plausible here. However, it may not be possible to find such similar units in a particular sample. Indeed, as we show for California, it is quite difficult to find redistricted and non-redistricted blocks with similar values in their pre-redistricting covariates, even when these share the same prior incumbent and legislative-district histories. This lack of common support across treated and control groups can lead to biased inferences."

There are three ways in which a "selection on observables" assumption can be violated and bias could be introduced in an observational study.[6] After controlling for X, unobserved factors U affect both treatment assignment and the outcome of interest, generating "selection" or "endogeneity" that biases inference. Next, the range of values of X taken by treated units may be different from the range of values of X taken by the control units, a condition that is commonly referred to as *lack of common support*. Finally, even if treatment and control units take values of X in the same range, the distribution of X can be different between the groups (for example, the mean and median of X in the treatment group may be very different from the mean and median of X in the control group even in the common support of X). The first type of bias, the one that depends on unobservables, is referred to as *hidden* bias, while the last two are referred to as *overt* bias.

A typical sensitivity analysis assumes that overt bias has been removed by conditioning on X covariates, but that covert bias due to imbalances on U may remain. Under specific assumptions, the test provides researchers with information about how robust their findings are to selection into treatment based on an unobservable variable (Rosenbaum 2002). Formally, for a treated-control matched pair $\{i, j\}$ with covariates X_i and X_j, respectively, define the probability of treatment for

[6] See Heckman, Ichimura and Todd (1998) for a formal derivation

unit i to be $\pi_i = Pr(T_i^{WH} = 1)$, and the odds of treatment $O_i = \pi_i/(1 - \pi_i)$. Define π_j and O_j analogously for j. A study is free from hidden bias if $O_i/O_j = 1$ for all $\{i, j\}$ pairs. In an experiment, this occurs if both units in each pair are assigned treatment with equal probability. In an observational study, this amounts to assuming that $\pi_i = \pi_j$ whenever $X_i = X_j$, that is, two units with identical observed characteristics will have the same probability of receiving treatment

In a study with hidden bias, two units with identical values of the covariates have different probabilities of receiving treatment. In other words, $X_i = X_j$, but $\pi_i \neq \pi_j$. These probabilities are assumed to differ by no more than some factor Γ, where $\frac{1}{\Gamma} \leq (O_i/O_j) \leq \Gamma$. A logit model for the odds of receiving treatment is typically assumed: $O_i = \exp\{f(X_i) + \gamma U_i\}$. Under this model,

$$\frac{1}{\Gamma} \leq \exp\{(f(X_i) + \gamma U_i) - (f(X_j) + \gamma U_j)\} \leq \Gamma.$$

$$\frac{1}{\Gamma} \leq \exp\{\gamma(U_i - U_j)\} \leq \Gamma$$

when $X_i = X_j$, so Γ may be directly interpreted as an upper and lower bound for the level of association between treatment and the unobserved covariate (Rosenbaum 2002). Since the parameter Γ is a measure of departure from the no hidden bias assumption, researchers look for the the minimum value of Γ that would alter the results of the study, and then asses whether that value is large.

To quantify the degree of overt bias in our study, we adapt the classic sensitivity models. We start by assuming that (as we see in our data and as occurs in many other observational studies), treated-control pairs do not have the same values of observed characteristics, i.e. $X_i \neq X_j$. Instead of assessing robustness at arbitrary values of Γ, we use the actual imbalance in observable characteristics (before or after matching on X) to find empirically-relevant values for Γ. To do this, we consider the probability of receiving treatment given the observed covariates or *propensity score*, $q_i = P(T_i^{WH} = 1|X_i)$ for all units. We define $m_1 = median\{q_i|T_i^{WH} = 1\}$ and $m_0 = median\{q_j|T_j^{WH} = 0\}$, to be the median of the distribution of propensity scores for all i treated and j control units, respectively. We then define sensitivity values as:

$$\Gamma = \frac{m_1/(1 - m_1)}{m_0/(1 - m_0)}.$$

In words, Γ is defined as the ratio of the estimated odds of receiving treatment for the median treated and control units, using X to model T^{WH}. We then assess whether Γ alters our findings.[7]

Table VIII: Observed Sensitivity Analysis: 2004-2000 and 2002-2000 Difference Results

	2004		2002	
	Γ	P-value	Γ	P-value
BASELINE				
Hispanic Turnout	2.10	1.00	2.10	1.00
Hispanic Registration	2.10	1.00	2.10	1.00
Non Hispanic Turnout	2.10	1.00	2.10	1.00
Non Hispanic Registration	2.10	1.00	2.10	1.00
MATCHED				
Hispanic Turnout	1.24	0.92	1.24	0.90
Hispanic Registration	1.24	0.30	1.24	0.93
Non Hispanic Turnout	1.24	0.42	1.24	0.21
Non Hispanic Registration	1.24	0.89	1.24	0.96

Note: Γ is the ratio of the odds of treatment for treated and control groups, estimated from the median propensity score for each. P-values are Wilcoxon p-values of the null hypothesis of no effect when Γ takes a particular value.

7.2 Open Seat Robustness Analysis: Results, Data Description, Placebo Tests and Covariate Balance

In our main analysis, the treatment condition is defined as movement from a NHW incumbent before redistricting to an Hispanic incumbent after redistricting, and the control as staying with a NHW incumbent before and after redistricting. This measurement explicitly focuses on incumbent races to evaluate whether the presence of a co-ethnic *incumbent representative* on the ballot has any effect on minority turnout. However, elections contested by incumbents are often much less competitive than open seat races. Minority voting in these incumbent-held elections might also

[7] To implement the sensitivity test, we estimate the propensity score using a standard linear logit link function, so that $f(X_i) = X_i^T\beta$, though in principle, any functional relationship may be assumed. From here, the implementation is straightforward: we compute Γ as the ratio of the odds of receiving treatment for treated and control blocks, which are estimated for each dataset regressing T^{WH} on X. Then we conduct permutation inference analysis as outlined in Rosenbaum (2002), using the estimated Γ statistics as the basis for hypothesis testing.

be dampened since the Hispanic incumbent is very likely to be reelected. If so, an analysis that focuses exclusively on incumbent races might be more likely to find null turnout effects than one that included open seats (races where no incumbent candidate is running), since the opportunity to elect an Hispanic *non-incumbent candidate* could have larger effects on co-ethnic turnout.

To address this concern about generalizability, we include here an analysis of races where no incumbent is running. We focus on Congressional Districts (CD) 21 and 39, the two majority-Hispanic districts where Hispanic candidates ran in open-seat races in the 2002 general election.[8] In this analysis, the treatment group is comprised of blocks moved from districts represented by a NHW incumbent in 2000 to either CD-21 or CD-39 after redistricting. The control group is constructed in exactly the same manner as control blocks in the main analysis. For OPEN-BASELINE, controls are blocks represented by a NHW running in 2000 and 2002, matched to comparable treated blocks using P-HVAP. In OPEN-MATCHED, controls are represented by the same NHW incumbent running in 2000 and 2002, and are contained in the same triplets as treated blocks. Trimming was done exactly as before to ensure the common support of the covariates, and matching was done using the hierarchical algorithm exactly as in the main analysis. After trimming and matching, the final open seat sample has 8,403 pairs in OPEN-BASELINE, and 454 matched pairs in OPEN-MATCHED. This expands our effective final sample size to 13,090 total pairs (26,180 total blocks) in our baseline results, and 882 matched pairs (1,764 total blocks) in our final matched dataset.

The open seat analysis confirms the above findings, and provides little evidence that the opportunity to elect new Hispanic candidates in open races increases Hispanic turnout or registration. The open seat results are displayed in Table IX below. As seen in OPEN-BASELINE, Hispanic turnout (-2.5%) *and* Hispanic registration (-1.2%) actually decline in blocks moved to open seat MM districts in 2002, relative to blocks that remain with NHW incumbents (with comparable P-HVAP). This decline may be especially informative, since the 2002 election *is* the race in which no incumbent is running, and thus where we would expect the greatest co-ethnic turnout boost if open seat

[8]There were three open-seat races in California in the 2002 general election, but we only include in our analysis the two races that occurred in majority-Hispanic districts: CD-21 and CD-39. The 2002 general election was won in CD-21 by Hispanic Republican candidate Devin Nunes, and in CD-39 by Hispanic Democratic candidate Linda Sánchez. Note also that CD-21 was a newly created majority-Hispanic district, thus providing additional generality for our findings to the process of creating new MM seats rather than just maintaining existing ones, albeit in a district that is likely to elect a Republican Hispanic representative.

races with new Hispanic candidates have the largest empowerment effects. By 2004, the difference-in-difference estimates for Hispanic turnout become positive (3.5%), while Hispanic registration differences remain negative (-1.0%), though no longer statistically significant. Balance and placebo tests, however, show that treated and control blocks in OPEN-BASELINE are substantially different on both prior Hispanic turnout and registration, but are indistinguishable in OPEN-MATCHED.[9] Thus, turning to the analysis in OPEN-MATCHED, we again find negative differences in Hispanic turnout for both 2002-2000 (-3.3%) and 2004-2000 (-4.3%) differences, though neither estimate is significantly different from zero (p-values of 0.39 and 0.38, respectively). Differences in Hispanic registration across treated and control blocks are also indistinguishable from zero in OPEN-MATCHED. Overall, the results indicate little support for a positive turnout or registration effect in MM seats, whether contested by Hispanic candidates in open races or by relatively safe Hispanic incumbents.[10]

[9] Open seat balance and placebo tests are reported in Table XII and Table XI of the Supplemental Appendix.

[10] Pooling over both the main and open seat analyses generally results in small, negative, but statistically insignificant differences across redistricted and not-redistricted blocks in terms of subsequent Hispanic turnout and registration.

Table IX: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results for Open Seats

	2004		2002	
	Tr - Co	P-value	Tr - Co	P-value
OPEN-BASELINE				
Hispanic Turnout	0.035	0.07	-0.025	0.03
Hispanic Registration	-0.010	0.13	-0.012	0.07
Non-Hispanic Turnout	0.203	0.00	0.041	0.10
Non-Hispanic Registration	0.000	0.96	-0.009	0.15
OPEN-MATCHED				
Hispanic Turnout	-0.043	0.38	-0.033	0.39
Hispanic Registration	0.021	0.49	0.008	0.78
Non-Hispanic Turnout	0.051	0.36	0.029	0.57
Non-Hispanic Registration	0.022	0.27	0.004	0.86

Note: Registration is measured as registration levels in 2000, 2002, or 2004, for each 2000 U.S. census levels in 2000, 2002, or 2004, for each 2000 U.S. census block over Hispanic or Non-Hispanic voting age population. Turnout measures are constructed as 2000, 2002, or 2004 turnout over Hispanic or Non-Hispanic registration in 2000, as reported in the SWDB. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

Table X: Nested Open Seat Datasets Used in Analysis at Census Block Level

	OPEN-BASELINE
Number treated Number control	8,403 census blocks (2000 U.S. Census) 8,403 census blocks (2000 U.S. Census)
Restrictions	Closed seat in congressional district race in 2000 and 2002 Congressional incumbent running in 2000 wins 2000 election
Treatment	Non-Hispanic White congressional incumbent running in 2000 Open seat election with an Hispanic congressional candidate running in 2002 (CD-21 or CD-39)
Control	Non-Hispanic White congressional incumbent running in 2000 and 2002
Matching	Match on *proportion* Hispanic voting age population only.
	OPEN-MATCHED
Number treated Number control	454 census blocks (2000 U.S. Census) 454 census blocks (2000 U.S. Census)
Restrictions	Same restrictions as in OPEN-BASELINE
Treatment	Non-Hispanic White congressional incumbent running in 2000 Open seat election with an Hispanic congressional candidate running in 2002 (CD-21 or CD-39)
Control	*Same* non-Hispanic White congressional incumbent running in 2000 and 2002
Matching	Define state house, state senate and congressional district “triplet” for every block in 2000. Drop triplets where number of controls less than twice number of treated. Within triplets, trim treatments whose demographics are outside of common support. In addition, (i) match exactly on 2000 triplet; (ii) within each triplet, match on demographics; (iii) then match across triplets, i.e., select matched pairs within each triplet such that overall balance *across* triplets is maximized.

Table XI: Intermediate Analysis: 2000 Cross-Sectional Results for Open Seats Analysis

	2000		
	Mean		P-value
	Tr	Co	μ-test
OPEN-BASELINE			
Hispanic Turnout	0.621	0.579	0.00
Hispanic Registration	0.384	0.445	0.00
Non Hispanic Turnout	0.732	0.667	0.00
Non Hispanic Registration	0.635	0.697	0.00
OPEN-MATCHED			
Hispanic Turnout	0.624	0.621	0.88
Hispanic Registration	0.459	0.485	0.21
Non Hispanic Turnout	0.763	0.733	0.01
Non Hispanic Registration	0.700	0.735	0.01

Note: Datasets are described in Table X (in the Supplemental Appendix). Registration is measured as levels in 2000 for each 2000 U.S. census block over Hispanic or Non-Hispanic voting age population. Turnout measures are constructed as 2000 turnout over Hispanic or Non-Hispanic registration in 2000, as reported in the SWDB. P-values are standard t-test probabilities.

Table XII: Covariate Balance Tests for Census-Block Level Open Seat Analysis

	OPEN-BASELINE				OPEN-MATCHED			
	Mean		P-value		Mean		P-value	
	Tr	Co	μ-test	ks-test	Tr	Co	μ-test	ks-test
CONDITIONING SET								
Voting Age Population (VAP)	48.94	49.92	0.33	0.00	48.86	47.49	0.57	0.18
Black VAP	0.01	0.05	0.00	0.00	0.01	0.01	0.89	1.00
Hispanic VAP	0.35	0.36	0.35	0.23	0.23	0.23	0.80	0.91
HH Income 39k	0.53	0.55	0.00	0.00	0.49	0.49	0.82	0.10
HH Income 40k to 74k	0.30	0.27	0.00	0.00	0.30	0.31	0.30	0.10
HH Income $\geq$ 100k	0.11	0.12	0.01	0.00	0.10	0.10	0.99	0.10
Pop Highschool or Less	0.29	0.28	0.00	0.00	0.28	0.28	0.30	0.10
Pop Foreign	0.19	0.17	0.00	0.00	0.15	0.15	0.75	0.45
ADDITIONAL COVARIATES								
Pop Non-Citizen	0.13	0.12	0.00	0.00	0.10	0.09	0.24	0.02
Pop Naturalized Citizen	0.06	0.06	0.06	0.00	0.05	0.06	0.04	0.00
Registration Total 1998	0.61	0.69	0.00	0.00	0.66	0.77	0.00	0.00
Registration Hispanic 1998	0.25	0.27	0.00	0.00	0.16	0.18	0.06	0.24
Registration Democrat 1998	0.42	0.48	0.00	0.00	0.41	0.45	0.00	0.01
Registration Republican 1998	0.44	0.40	0.00	0.00	0.47	0.44	0.07	0.07
Registration Total 2000	0.63	0.69	0.00	0.00	0.69	0.75	0.00	0.00
Registration Democrat 2000	0.40	0.44	0.00	0.00	0.39	0.42	0.01	0.00
Registration Republican 2000	0.47	0.43	0.00	0.00	0.48	0.47	0.19	0.02
Dem. Vote U.S. Senate 1998	0.62	0.64	0.45	0.00	0.46	0.45	0.88	0.00
Dem. Vote Governor 1998	0.65	0.67	0.20	0.00	0.47	0.48	0.81	0.01
Dem. Vote U.S. House 1998	0.26	0.34	0.00	0.00	0.09	0.10	0.85	0.91
Dem. Vote President 2000	0.47	0.54	0.00	0.00	0.41	0.43	0.70	0.00
Dem. Vote U.S. Senate 2000	0.59	0.66	0.00	0.00	0.62	0.55	0.11	0.00
Dem. Vote U.S. House 2000	0.43	0.50	0.00	0.00	0.38	0.37	0.95	0.00
Pop Female	0.50	0.51	0.00	0.00	0.50	0.52	0.01	0.00
Pop 25 to 44 Years	0.27	0.27	0.49	0.05	0.25	0.25	0.43	0.01
Pop 45 to 59 Years	0.18	0.18	0.11	0.00	0.20	0.18	0.15	0.27

Note: Datasets are described in Table X (in the Supplemental Appendix). Total registration reported as share of total voting age population. All other registration covariates reported as share of total registration. All demographic variables from 2000 U.S. Census.

7.3 Supplementary Results

In this section, we present some additional results and robustness checks on our main analysis. Table XIII presents cross-sectional analysis of the main results, making just the Selection on Observables (SOA) assumption in Eq. (1) of the main paper. As can be seen, we recover consistently null results in MATCHED under the cross-sectional SOA assumption, even before we estimate *difference-in-differences* under Eq. (2).

Table XIV presents the *difference-in-differences* results using 2000 census HVAP and NHVAP as the denominator for Hispanic and non-Hispanic turnout, rather than using 2000 registration as the denominator as done in the main paper results. (HVAP and NHVAP are always used as the denominator for Hispanic and non-Hispanic registration outcomes.) Again we recover null results for Hispanic and non-Hispanic turnout.

Due to measurement error in the way (N)HVAP is measured by the census, and the way turnout and registration is measured by the SWDB, there are some blocks with 'overruns' in using turnout over HVAP. An overrun is when the numerator is greater than the denominator in measuring participation rates due to the different ways U.S. Census population and SWDB participation data are collected at the block-level. (Note, the presence of these overruns when measuring turnout over HVAP is one of the original justifications for using turnout over registration, since the latter does not produce overruns being measured through the same data and collection process.) Table XV presents a replication of our main analysis allowing overrunning blocks to 'float', that is take on proportions greater than one. Retaining overruns, we find consistently small or null effects for Hispanic participation, though recover negative effects for non-Hispanic participation.

Finally, Table XVI and Table XVII explore the robustness of our results to using different registration years in the denominator. Table XVI uses registration measured in 1998, and XVII uses the concurrent (post-treatment) registration measure, e.g., 2002 turnout divided by 2002 registration. Again each of these checks replicates the null results found in the main analysis.

7.3.1 Cross-Sectional Results – Selection on Observables Assumption

Table XIII: Cross Section Estimates: 2004 and 2002 Main Results

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.088	0.00	0.008	0.43
Hispanic Registration	0.105	0.00	0.103	0.00
Non Hispanic Turnout	-0.133	0.00	-0.050	0.00
Non Hispanic Registration	0.016	0.00	0.033	0.00
MATCHED				
Hispanic Turnout	0.024	0.74	0.027	0.47
Hispanic Registration	0.070	0.10	0.030	0.10
Non Hispanic Turnout	-0.072	0.29	-0.047	0.14
Non Hispanic Registration	0.015	0.33	0.020	0.22

Note: Datasets are described in the main paper. Registration is measured as levels in 2000, 2002, or 2004, for each 2000 U.S. census block over Hispanic or Non-Hispanic voting age population. Turnout measures are constructed as 2000, 2002, or 2004 turnout over Hispanic or Non-Hispanic registration in 2000, as reported in the SWDB. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

7.3.2 Difference Results with (N)HVAP in the Denominator

Table XIV: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results & (N)HVAP in the Denominator

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.015	0.01	-0.044	0.00
Hispanic Registration	0.012	0.10	0.009	0.20
Non Hispanic Turnout	-0.015	0.03	-0.031	0.00
Non Hispanic Registration	-0.013	0.07	0.003	0.63
MATCHED				
Hispanic Turnout	0.003	0.91	0.004	0.86
Hispanic Registration	0.032	0.21	-0.008	0.76
Non Hispanic Turnout	0.014	0.52	0.011	0.57
Non Hispanic Registration	0.033	0.15	0.037	0.11

Note: Datasets are described in the main paper. All registration and turnout are constructed, respectively, as registration and turnout levels in 2000, 2002, and 2004 for each 2000 census block over Hispanic or Non-Hispanic voting age population in each block, as reported by the U.S. 2000 Census. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

7.3.3 Difference Results with (N)HVAP in the Denominator – Allowing Overruns

Table XV: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results & (N)HVAP in the Denominator Allowing Overruns

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.023	0.11	-0.047	0.00
Hispanic Registration	0.013	0.51	0.028	0.03
Non Hispanic Turnout	-0.158	0.01	-0.034	0.01
Non Hispanic Registration	-0.206	0.00	-0.035	0.10
MATCHED				
Hispanic Turnout	0.052	0.29	-0.002	0.95
Hispanic Registration	0.093	0.17	0.078	0.23
Non Hispanic Turnout	-0.077	0.03	-0.055	0.03
Non Hispanic Registration	-0.107	0.02	-0.106	0.02

Note: Datasets are described in the main paper. All registration and turnout are constructed, respectively, as registration and turnout levels in 2000, 2002, and 2004 for each 2000 census block over Hispanic or Non-Hispanic voting age population in each block, as reported by the U.S. 2000 Census. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

7.3.4 Difference Results with 1998 Registration in the Denominator

Table XVI: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results & 1998 Registration in the Denominator

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.050	0.09	-0.055	0.00
Hispanic Registration	0.031	0.47	0.063	0.09
Non Hispanic Turnout	-0.103	0.05	-0.046	0.29
Non Hispanic Registration	-0.138	0.10	-0.052	0.50
MATCHED				
Hispanic Turnout	-0.100	0.39	-0.030	0.65
Hispanic Registration	-0.163	0.33	-0.085	0.49
Non Hispanic Turnout	-0.058	0.41	-0.020	0.68
Non Hispanic Registration	-0.096	0.35	-0.099	0.30

Note: Datasets are described in the main paper. Registration is measured as levels in 2000, 2002, or 2004, for each 2000 U.S. census block over Hispanic or Non-Hispanic registration in 1998. Turnout measures are constructed as 2000, 2002, or 2004 turnout over Hispanic or Non-Hispanic registration in 1998, as reported in the SWDB. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

7.3.5 Difference Results with Concurrent Registration Year in the Denominator

Table XVII: Difference-in-Difference Estimates: 2004-2000 and 2002-2000 Difference Results & Concurrent Registration in the Denominator

	2004		2002	
	Mean Tr - Co	P-value μ-test	Mean Tr - Co	P-value μ-test
BASELINE				
Hispanic Turnout	0.008	0.25	-0.053	0.00
Hispanic Registration	0.012	0.10	0.009	0.20
Non Hispanic Turnout	0.005	0.44	-0.026	0.00
Non Hispanic Registration	-0.013	0.07	0.003	0.63
MATCHED				
Hispanic Turnout	-0.033	0.12	0.010	0.64
Hispanic Registration	0.032	0.21	-0.008	0.76
Non Hispanic Turnout	-0.010	0.50	0.003	0.82
Non Hispanic Registration	0.033	0.15	0.037	0.11

Note: Datasets are described in the main paper. Registration is measured as levels in 2000, 2002, or 2004, for each 2000 U.S. census block over Hispanic or Non-Hispanic registration concurrently. Turnout measures are constructed as 2000, 2002, or 2004 turnout over Hispanic or Non-Hispanic concurrent registration, as reported in the SWDB. P-values are OLS with Huber-White standard errors and interacted time fixed effects.

7.4 QQ-Plots and Additional Figures

In this section, we present additional QQ-Plot figures. These figures provide further evidence that racial redistricting in California creates and maintains MM districts by moving more participatory Hispanic populations into minority districts, and retaining lower participating populations in districts represented by non-Hispanic White (NHW) incumbents. We also show that this bias is pronounced not only for U.S. House districts, but also for overlapping jurisdictions contained in three MM districts at the House, state senate and assembly levels.

Figure I presents a QQ-Plot of 2004 Hispanic turnout in blocks contained in three overlapping jurisdictions, compared to Hispanic turnout in blocks contained in zero MM districts. Figure II presents 2000 Hispanic turnout for blocks *to be moved to* three overlapping MM districts, to those *to be left in* three districts represented by three NHW incumbents. Figure II(a) shows this measure matching on proportion HVAP, and Figure II(b) shows this matching on the full conditioning set from the main paper. As can be seen, there remains considerable bias stemming from the redistricting process across blocks moved to three overlapping MM districts, and that matching reduces most, but not all of this bias.

Figures I - II present these additional results using turnout rates that are the ratio of Hispanic turnout divided by Hispanic registration in 2000. Figures III - V replicate and extend this analysis using turnout measures that are Hispanic turnout divided by HVAP in 2000. As can be seen, using HVAP in the denominator generally makes the bias 'worse'. Blocks in or to be moved to MM districts appear to be much more participatory that blocks in or retained in non-MM districts. Figure III shows these differences in 2004 Hispanic turnout for III(a) blocks in House MM districts, and for III(b) blocks in three overlapping MM legislative districts.

Figure IV shows these differences in 2004 are likely to be spurious. Figure IV(a) shows 2000 Hispanic turnout in blocks *to be moved to* a U.S. House MM district, and Figure IV(b) shows blocks *to be moved to* three overlapping MM legislative districts. Again these figures illustrate the considerable bias that emerges in the way MM districts are drawn, not only at the House level, but also at the state assembly and senate levels, regardless of whether we use registration or HVAP to normalize our turnout measures. Figure V shows that this bias is greatly reduced through our trimming and matching procedures for V(a) blocks to be moved to House MM districts and V(b) blocks to be moved to three overlapping MM districts.

Finally, Figure VI shows that this spurious bias persists not only for the Democratic-dominated, incumbent protection plan implemented in the 2000 redistricting, but also in the citizens' commissions maps drawn in 2010. Here we show QQ-Plots for 2010 Hispanic turnout in blocks *to be moved to* U.S. House MM districts in the 2010 redistricting cycle. Figure VI(a) shows this bias using 2010 Hispanic turnout divided by 2010 HVAP as the participation measure, and Figure VI(b) shows this bias using 2010 Hispanic registration in the denominator, after just matching on proportion of HVAP across blocks. Figure VI(a) shows clear, positive bias in Hispanic turnout emerging from the 2010 redistricting when using HVAP in the the denominator. However, Figure VI(b) shows bias that is positive in blocks with relatively fewer Hispanic populations, but negative in blocks with relatively greater Hispanic populations, when using 2010 registration in the denominator. This finding may be due to the spurious differences also being induced in Hispanic registration as illustrated in Figure 1(c) from the main paper, which could be artificially inflating Hispanic registration relative to Hispanic turnout in blocks with greater Hispanic populations.

Figure I: QQ Plots of 2004 Hispanic Turnout Rates for California Census Blocks In and Not In Hispanic Majority-Minority Districts, Matched on Proportion Hispanic Voting Age Population

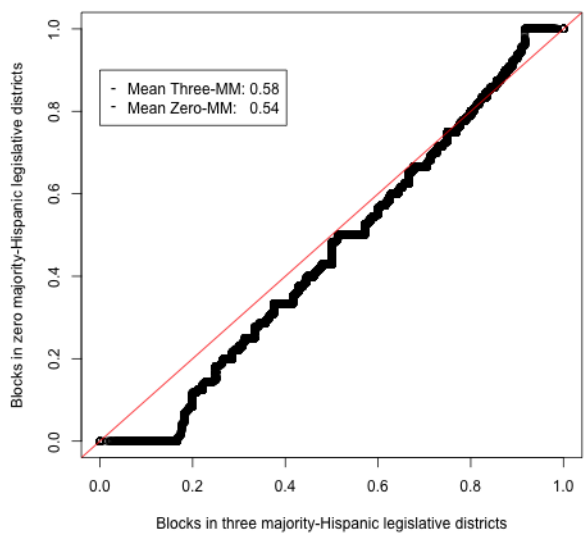
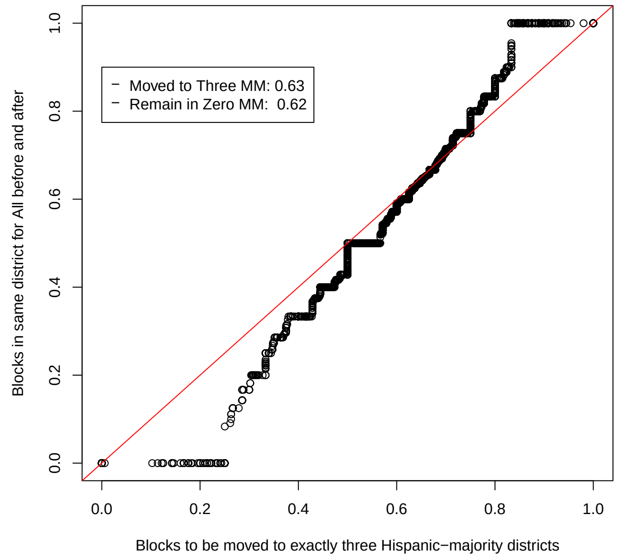
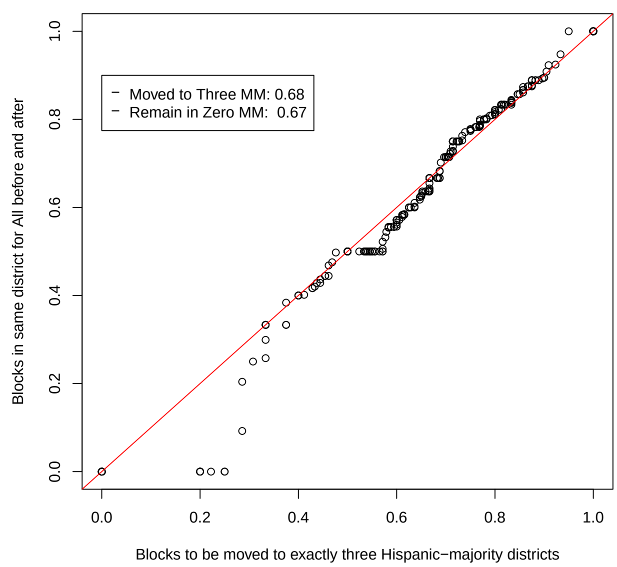


Figure II: QQ Plots of 2000 Hispanic Turnout Rates for California Census Blocks To Be Moved From NHW Incumbent To Hispanic Incumbent After Redistricting

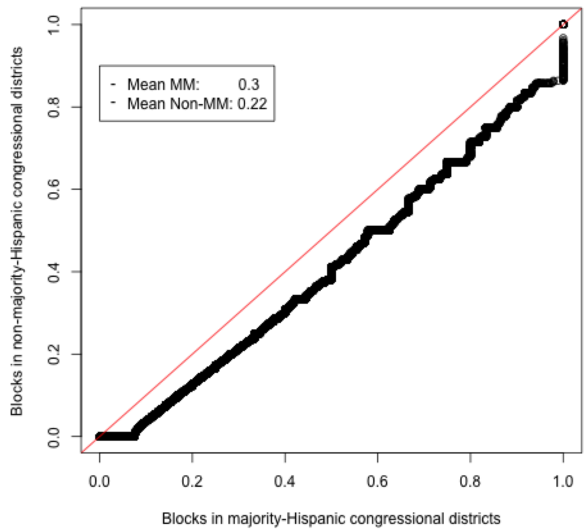


(a) Matched on Proportion Hispanic VAP

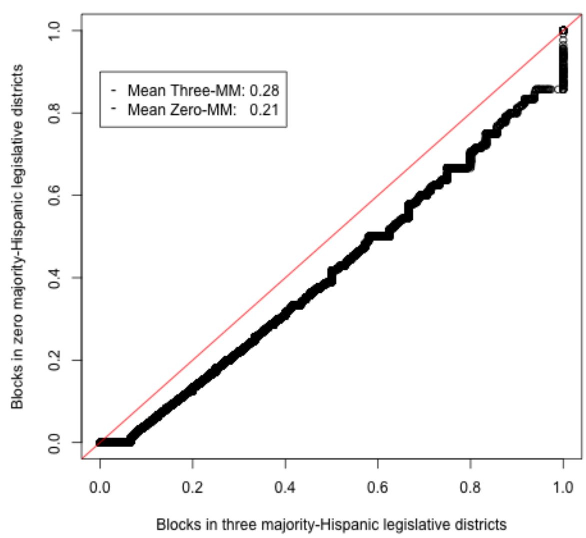


(b) Matched on Conditioning Set

Figure III: QQ Plots of 2004 Hispanic Turnout Rates for California Census Blocks In and Not In Hispanic Majority-Minority Districts, Matched on Proportion Hispanic Voting Age Population

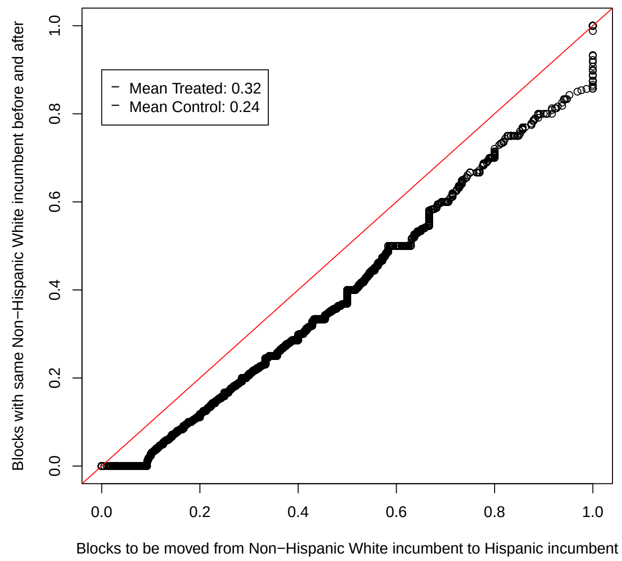


(a) CONGRESSIONAL DISTRICTS

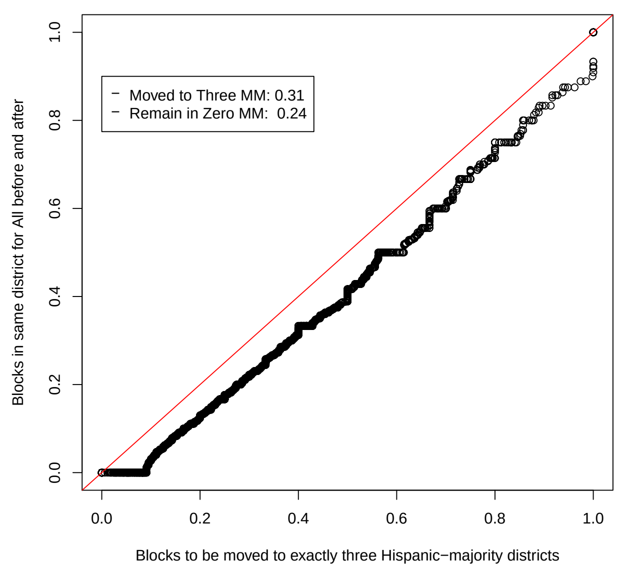


(b) OVERLAPPING ASSEMBLY, SENATE AND CONGRESSIONAL DISTRICTS

Figure IV: QQ Plots of 2000 Hispanic Turnout Rates for California Census Blocks To Be Moved From NHW Incumbent To Hispanic Incumbent After Redistricting, Matched on Proportion Hispanic Voting Age Population

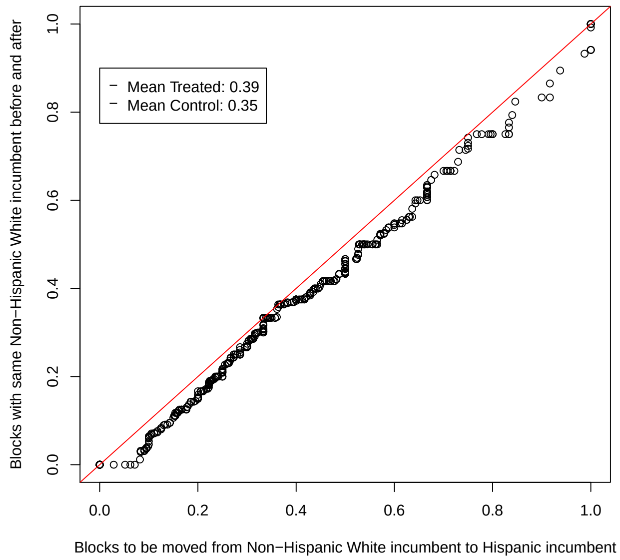


(a) CONGRESSIONAL DISTRICTS

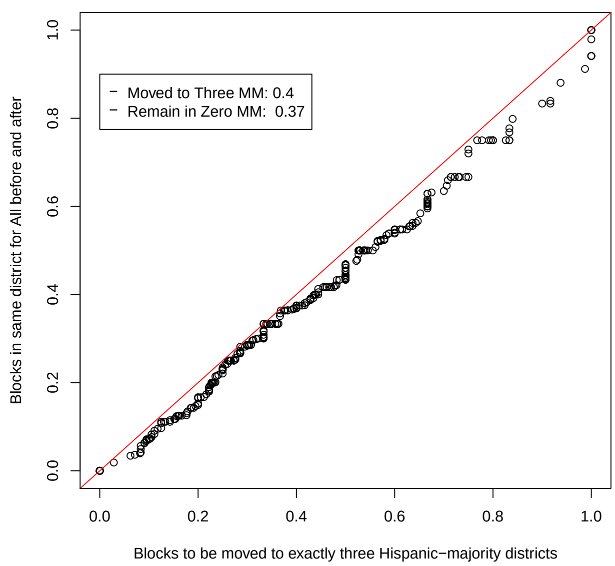


(b) OVERLAPPING ASSEMBLY, SENATE AND CONGRESSIONAL DISTRICTS

Figure V: QQ Plots of 2000 Hispanic Turnout Rates for California Census Blocks To Be Moved From NHW Incumbent To Hispanic Incumbent After Redistricting, Matched on Conditioning Set

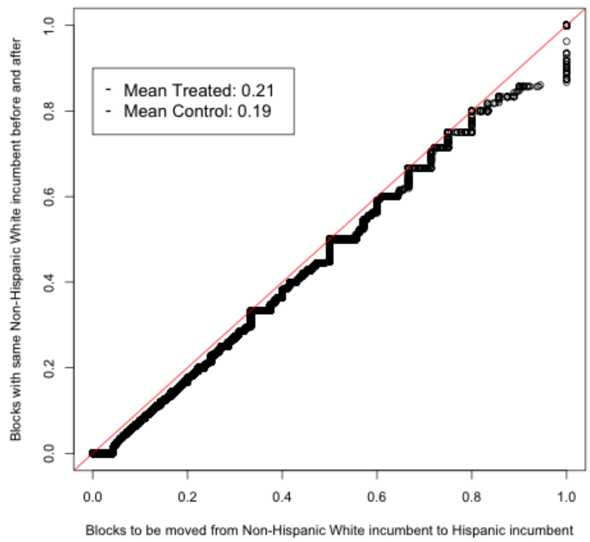


(a) CONGRESSIONAL DISTRICTS

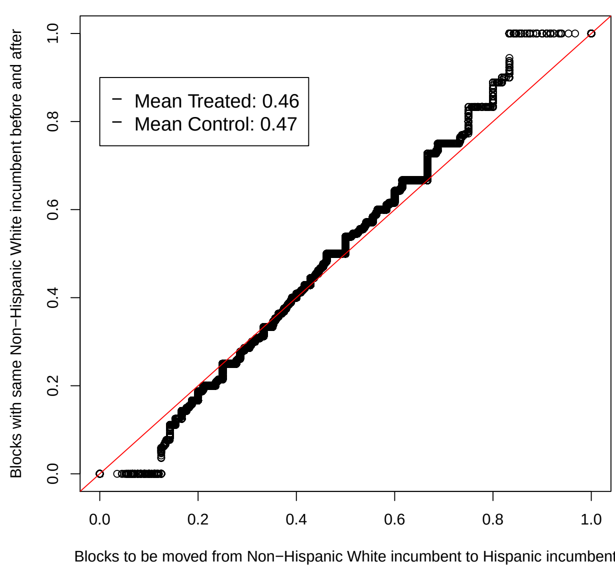


(b) OVERLAPPING ASSEMBLY, SENATE AND CONGRESSIONAL DISTRICTS

Figure VI: QQ Plots of 2010 Hispanic Turnout Rates for California Census Blocks To Be Moved From NHW Incumbent To Hispanic Incumbent After Redistricting, Matched on P-HVAP or Registration



(a) TURNOUT WITH HVAP IN THE DENOMINATOR, MATCHED ON P-HVAP



(b) TURNOUT WITH REGISTRATION IN THE DENOMINATOR, MATCHED ON P-HVAP

8 Hierarchical Genetic Matching

Legislatures simultaneously redistrict overlapping congressional, assembly and state senate jurisdictions. We focus analysis on movements of blocks between congressional districts, since we expect these elections to be most salient for voters. However, we only compare treated and control voters within the same (pre-redistricting) triplet of congressional, assembly and senate districts, as defined in the 1990 redistricting plan. We limit our analysis in this way because, as Barreto, Segura and Woods (2004) show, areas represented by Hispanic incumbents at all three levels are systematically different from areas represented by fewer or no Hispanic incumbents. This restriction ensures that variation in features of prior state and House legislative districts does not affect our inferences.

Due to the hierarchical nature of redistricting, we develop a multilevel algorithm (*hierarchical matching*) that uses a genetic optimizer to match blocks moved from a NHW to an Hispanic congressional incumbent to unmoved blocks based on the baseline covariates used during redistricting (Diamond and Sekhon 2012; Sekhon 2011). Genetic matching utilizes a nearest-neighbor algorithm to create treated-control pairs that minimize distances in multivariate space. The algorithm prioritizes matches that reduce differences across matched-units on the most dissimilar characteristics, gradually increasing the similarities (*balance*) on each covariate in X over successive generations. Hierarchical genetic matching uses a constrained optimization that limits the available matches to be within the same legislative district "triplet", minimizing within-triplet distances. Matched pairs *within* each triplet are selected to maximize balance *across* all triplets.

Our study of minority turnout is the first use of this hierarchical matching algorithm in empirical research. Though novel in application, hierarchical matching, in fact, is a special case of traditional genetic matching, currently used widely in political science, economics and the other social sciences. The matches recovered through the hierarchical algorithm could also be recovered via 'GenMatch', or other similar genetic evolutionary matching approaches. The main benefit of hierarchical genetic matching is that it is much more efficient at obtaining balance in clustered data, since it utilizes a top-level optimizer to minimize imbalances across clusters. The evolutionary optimization feature of both forms of genetic matching, as discussed in Diamond and Sekhon (2012), ensure both eventually converge on the best-balanced dataset. Yet, a clustered optimization gets there faster.

References

Barabak, Mark Z. 2001. "California to Gain Only One Congressional Seat.". Los Angeles Times.

Barreto, Matt A., Gary M. Segura and Nathan D. Woods. 2004. "The Mobilizing Effect of Majority-Minority Districts on Latino Turnout." *American Political Science Review* 98(1):65–75.

Bernal, Martha E, George P Knight, Camille A Garza, Katheryn A Ocampo and Marya K Cota. 1990. "The Development of Ethnic Identity in Mexican-American Children." *Hispanic Journal of Behavioral Sciences* 12(1):3–24.

Bickerstaff, Steve. 2007. *Lines in the Sand*. Austin, Texas: University of Texas Press.

Cain, Bruce, Karin MacDonald and Iris Hui. 2006. "Competition and Redistricting in California: Lessons for Reform." Institute of Governmental Studies.

Chandra, Kanchan. 2006. "What is ethnic identity and does it matter?" *Annu. Rev. Polit. Sci.* 9:397–424.

Chandra, Kanchan and Steven Wilkinson. 2008. "Measuring the effect of ethnicity." *Comparative Political Studies* 41(4-5):515–563.

Cho, Wendy K. Tam and George G. Judge. 2009. "Recovering Vote Choice from Partial Incomplete Data." *Journal of Data Science* 6:155–171.

Cruz, Jose E. and Jackie Hayes. 2009. "Adding Race and Ethnicity: Electoral Data Collection Practice and Prospects for New York State." `http://www.nylarnet.org/reports/`. NY-LARNet, State University of New York.

Diamond, Alexis and Jasjeet S. Sekhon. 2012. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* . Forthcoming.

Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja and Nicole Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43(5):1722–1736.

Fiscella, Kevin and Allen M. Fremont. 2006. "Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity." *Health Services Research* 41(4):1482–1500.

Gay, Claudine. 2001. "The Effect of Minority Districts and Minority Representation on Political Participation in California." Public Policy Institute of California.

Greiner, James and Kevin Quinn. 2009. "$R \times C$ Ecological Inference: Bounds, Correlations, Flexibility and Transparency of Assumptions." *Journal of the Royal Statistical Society* 171(6):67–81.

Grofman, Bernard and Andrew Reynolds. 1996. "Modeling the Dropoff Between Minority Population Share and the Size of the Minority Electorate in Situations of Differential Voter Eligibility Across Groups." *Electoral Studies* 15(2):255–261.

Grofman, Bernard and Lisa Handley. 1998. Voting Rights in the 1990s: An Overview. In *Race and Redistricting in the 1990s*, ed. Bernard Grofman. New York: Agathon Press.

Heckman, James J., Hidehiko Ichimura and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65(2):261–294.

Hill, Kevin A. 1995. "Does the Creation of Majority Black Districts Aid Republicans? An Analysis of the 1992 Congressional Elections in Eight Southern States." *Journal of Politics* 57(2):384–401.

Kousser, J. Morgan. 1997. Redistricting California 1971 - 2001. In *Governing California: Politics, Government, and Public Policy in the Golden State*, ed. Gerald C. Lubenow and Bruce E. Cain. Berkeley: Institute of Governmental Studies Press.

Kousser, J. Morgan. 1998. Reapportionment Wars: Party, Race, and Redistricting in California, 1971 - 1992. In *Race and Redistricting in the 1990s*, ed. Bernard Grofman. New York: Agathon Press.

Lichtman, Allan J. 1991. "Passing the Test: Ecological Regression Analysis in the Los Angeles County Case and Beyond." *Evaluation Review* 15(6):770–799.

Lublin, David. 1997. *The Paradox of Representation*. New Jersey: Princeton University Press.

McDonald, Michael P. 2004. "A Comparative Analysis of Redistricting Institutions in the United States, 2001-02." *State Politics & Policy Quarterly* 4(4):371–395.

Petrocik, John R. and Scott W. Desposato. 1998. "The Partisan Consequences of Majority-Minority Redistricting in the South, 1992 and 1994." *Journal of Politics* 60(3):613–633.

Pierce, Olga and Jeff Larson. 2011. "How Democrats Fooled Californias Redistricting Commission.". ProPublia.

Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer-Verlag.

Sekhon, Jasjeet S. 2011. "Matching: Multivariate and Propensity Score Matching with Automated Balance Search." *Journal of Statistical Software* 42(7):1–52. Computer program available at `http://sekhon.berkeley.edu/matching/`.

Spears, Russell, Bertjan Doosje and Naomi Ellemers. 1997. "Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification." *Personality and Social Psychology Bulletin* 23(5):538–553.

Texas v. United States, Civil Action No. 11-1303. 2012.

U.S. Census Bureau. 2008. "Current Population Survey, November Voting and Registration Supplement." `http://www.nber.org/data/current-population-survey-data.html`. Washington, DC.

U.S. Census Bureau. 2012. "Statistical Abstract of the United States." `http://www.census.gov/statab/www/`. Washington, DC.