



PPIC

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

Implementing Automated Voter Registration in California

Technical Appendices

CONTENTS

Appendix A. Synthetic Control Analysis Details

Appendix B. Additional Analysis

Eric McGhee, Radhika Mehlotra, and Mindy Romero

Supported with funding from the New Venture Fund

Appendix A. Synthetic Control Analysis Details

The Catalist registration data present both challenges and opportunities. As statewide aggregates they offer a relatively small number of cases for analysis, including just one for California under CNMV. That makes statistical confidence difficult, leading to a greater likelihood of falsely concluding that CNMV had no effect. On the other hand, the data is structured as a panel cross-section, allowing explicit controls for states' pre-treatment characteristics.

In particular, the panel structure allows us to use the synthetic control method. The synthetic control identifies a weighted combination of non-AVR states, to understand what might have happened if California had never adopted CNMV (Abadie, Diamond et al. 2010). The weights are designed to make the synthetic control group's pre-treatment trendline match the trend for California as closely as possible. More formally the synthetic control identifies weights W^* that solve the following constrained minimization problem:

$$W^* = \min_W (X_{CA} - X_S W)' V (X_{CA} - X_S W)$$

where X_{CA} is a $(k \times 1)$ vector of k pre-treatment factors for California and X_S is a similar $(k \times J)$ matrix of factors for all other states. W is a $(J \times 1)$ vector of weights indicating the importance of each untreated state to the synthetic control group and V is a set of regression-style weights indicating the relevance of each factor to the calculation.

If matched well to a long enough pre-treatment series, the synthetic control group should continue to follow the path California would have taken in the absence of CNMV intervention. A difference-in-difference between the synthetic control and California can therefore serve as an estimate of the treatment effect of CNMV.

However, the limited number of post-treatment time periods available for this analysis makes the results especially sensitive to the choice of V . Ordinarily V is chosen to minimize the difference between the pre-treatment outcome variable in the treated and untreated states. If $Y_{CA\sim}$ is the vector of pre-CNMV registration values for California and $Y_{S\sim}$ is a matrix of the same for all other states, the optimal V^* is estimated with

$$V^* = \min_V (Y_{CA\sim} - Y_{S\sim} W^*(V))' (Y_{CA\sim} - Y_{S\sim} W^*(V))$$

However, this does not resolve the question of model specification: that is, which covariates to include in V in the first place. With only a single post-treatment time period available for analysis, a relatively poor model might introduce too much error and bias our findings toward zero. To avoid this problem, we canvass across a wide range of models. For each one, we run a leave-one-out cross-validation process to estimate the out-of-sample mean squared prediction error (MSPE). We then run each of those models on the full data. Finally, we average the treatment estimates for all these models, weighting each by its inverse MSPE from the leave-one-out validation.

The final step in our synthetic control analysis is to estimate our statistical confidence in the results. What is the probability that we would see a result as large as the one observed if CNMV in fact had no effect at all? We use placebo tests to estimate this quantity (Abadie, Diamond et al. 2010): we run the same model separately on each comparison state as if it had adopted CNMV. The share of states with a treatment effect at least as large as the one for California serves as the probability that the observed effect for California is due to random chance. We then average these probabilities by the inverse MSPE, as above.

The models we test and the covariates they include are in Table A1 below. In some cases a particular model would not converge for a particular group, so we left it out of the process. Table A2 displays the treatment effects associated with each model and each subgroup, as well as the probabilities of a treatment due to chance and the MSPE from the leave-one-out validation. Table A3 shows the states with non-zero weights for each

model and subgroup. Some states in the youth data had a large number of cases with missing age values due to reporting problems from the states themselves. Alaska, Hawaii, Mississippi, North Dakota, New Hampshire, Wisconsin, and Wyoming each averaged more than 13% missing, and for some years some had rates over 20%. To account for this problem, we included the percent missing as a matching variable in all the models below run on the youth data.

TABLE A1

Synthetic control models

Variable	All Lags	Lags+ Political	Mixture	Demog.	Full	Lags+ Demog.	Political
Non-Hispanic white			X	X	X	X	
Homeowner				X	X	X	
Unemployed				X	X	X	
Age 18-24				X	X	X	
Age 25-29				X	X	X	
Married				X	X	X	
College graduate				X	X	X	
Moved in last year				X	X	X	
U.S. Senate race in state		X	X		X		X
U.S. Senate race margin		X	X		X		X
Gubernatorial margin		X	X		X		X
CVAP natural log	X	X	X	X	X	X	X
Lag 1	X	X	X		X	X	
Lag 2	X	X	X		X	X	
Lag 3	X						
Lag 4	X						

NOTES: Table shows each of the different models tested in the analysis, along with the variables included in each one.

TABLE A2

Synthetic control model results

Model	Treatment	Pr(Treatment).	MSPE
ALL REGISTRANTS			
All lags	-0.011	0.30	0.0001
Mixture	-0.021	0.40	0.0002
Lags + political	-0.006	0.49	0.0003
Demographics	-0.046	0.12	0.0005
Lags + demographics	-0.016	0.10	0.0008
Full	-0.015	0.15	0.0008
Political	-0.024	0.35	0.0024
LATINOS			
All lags	-0.005	0.68	0.0001
Mixture	-0.012	0.62	0.0002
Lags + political	-0.028	0.15	0.0002
Full	-0.012	0.62	0.0002
Lags + demographics	0.012	0.22	0.0003
Demographics	-0.011	0.40	0.0011
ASIAN AMERICANS			
Lags + demographics	0.062	0.20	0.0005
Full	-0.013	0.41	0.0005
Mixture	-0.013	0.41	0.0006
Lags + political	-0.007	0.51	0.0006
All lags	-0.026	0.51	0.0025
Demographics	-0.021	0.43	0.0052
YOUTH			
Full	-0.001	0.95	0.0007
Demographics	-0.009	0.62	0.0010
Political	-0.045	0.41	0.0011
Lags + political	-0.025	0.22	0.0013
Lags + demographics	0.010	0.75	0.0013
Mixture	-0.028	0.21	0.0014
All lags	0.000	0.49	0.0026

SOURCES: IPUMS-USA (Ruggles, Flood et al. 2019) American Community Survey micro-samples (2010–18); Catalyst registration files (2010–18).

NOTES: Table shows the estimated treatment effect ("Treatment") and the probability of an effect that size in the placebo tests ("Pr(Treatment)"). The final column ("MSPE") contains the mean squared prediction error of the leave-one-out cross-validation exercise described in the text above.

TABLE A3

Synthetic control weights by state and model, all registrants

	All Lags	Lags+ Political	Mixture	Demog.	Full	Lags+ Demog.	Political
AL	0.004	0.0010	0	0	0	0	0
AR	0.023	0.0060	0	0	0	0	0
AZ	0.013	0.0050	0	0	0.358	0.002	0
DE	0.006	0.0010	0	0	0	0	0
FL	0.009	0.0030	0	0	0	0.001	0
HI	0.011	0.0020	0	0.19	0.051	0.004	0
IA	0.002	0.0000	0	0	0	0	0
ID	0.012	0.0140	0.001	0	0	0.001	0
IN	0.005	0.0040	0	0	0	0	0
KS	0.008	0.0050	0	0	0	0	0
KY	0.031	0.0320	0.002	0	0	0.001	0
LA	0.005	0.0020	0	0	0	0	0
MA	0.008	0.0050	0	0	0	0	0
MD	0.017	0.4040	0	0	0	0	0
ME	0.012	0.0020	0	0	0	0	0
MI	0.012	0.0040	0	0	0	0	0
MN	0.027	0.0080	0	0.081	0	0.002	0
MO	0.010	0.0040	0	0	0	0	0
MS	0.008	0.0010	0	0	0	0	0
MT	0.014	0.0050	0	0	0	0.002	0
NC	0.017	0.0280	0.653	0.004	0.001	0.004	1
ND	0.004	0.0010	0	0	0	0	0
NE	0.015	0.0160	0	0	0	0	0
NH	0.009	0.0050	0	0.001	0	0.001	0
NJ	0.006	0.0030	0	0	0.002	0.197	0
NM	0.006	0.0040	0	0	0	0	0
NV	0.007	0.0030	0	0	0	0	0
NY	0.010	0.0040	0	0	0	0.001	0
OH	0.005	0.0010	0	0	0	0	0
OK	0.023	0.0050	0	0	0	0.001	0
PA	0.013	0.004	0	0	0	0	0
SC	0.012	0.001	0	0	0	0	0
SD	0.014	0.015	0.001	0	0	0	0
TN	0.407	0.182	0.257	0	0.544	0.518	0
TX	0.009	0.002	0	0	0	0	0
VA	0.105	0.017	0.003	0.269	0.001	0.245	0
WA	0.008	0	0	0	0	0	0
WI	0.009	0.004	0	0	0	0	0
WV	0.081	0.192	0.079	0	0.042	0.015	0
WY	0.008	0.004	0.001	0.453	0	0.002	0

SOURCES: IPUMS-USA (Ruggles, Flood et al. 2019) American Community Survey micro-samples (2010–18); Catalyst registration files (2010–18).

NOTES: Table contains state weights in the synthetic control analysis of all registrants, separately for each model.

TABLE A4

Synthetic control weights by state and model, Latino registrants

	All Lags	Lags+ Political	Mixture	Demog.	Full	Lags+ Demog.
AL	0	0	0.001	0	0.001	0.015
AR	0	0	0	0	0	0.017
AZ	0	0	0	0	0	0.020
DE	0	0	0.002	0	0.002	0.039
FL	0	0.129	0.011	0	0.011	0.019
HI	0.194	0.148	0.001	0	0.001	0.019
IA	0	0	0.002	0	0.002	0.017
ID	0	0	0.001	0	0.001	0.017
IN	0	0	0.001	0	0.001	0.019
KS	0	0	0.001	0	0.001	0.018
KY	0	0	0.001	0	0.001	0.038
LA	0	0.006	0.001	0	0.001	0.019
MA	0	0	0.001	0	0.001	0.020
MD	0	0	0.020	0	0.020	0.025
ME	0	0.045	0.001	0	0.001	0.023
MI	0	0	0.003	0	0.003	0.032
MN	0	0	0.001	0	0.001	0.015
MO	0	0	0.019	0	0.019	0.016
MS	0	0	0	0	0	0.015
MT	0	0.203	0.238	0	0.238	0.114
NC	0	0	0.247	0.931	0.247	0.021
ND	0	0	0.001	0	0.001	0.02
NE	0	0	0.001	0	0.001	0.022
NH	0	0	0.001	0	0.001	0.026
NJ	0	0	0.032	0	0.032	0.017
NM	0	0	0.001	0	0.001	0.05
NV	0	0	0.404	0	0.404	0.019
NY	0	0	0.001	0	0.001	0.033
OH	0	0	0.001	0	0.001	0.021
OK	0	0	0.001	0	0.001	0.023
PA	0	0	0.001	0	0.001	0.024
SC	0	0	0	0	0	0.017
SD	0.290	0.465	0.001	0.069	0.001	0.022
TN	0	0	0.001	0	0.001	0.025
TX	0	0	0.001	0	0.001	0.019
VA	0	0	0.001	0	0.001	0.027
WA	0.147	0	0	0	0	0.020
WI	0	0.002	0.001	0	0.001	0.023
WV	0	0	0.001	0	0.001	0.032
WY	0	0	0	0	0	0.021

SOURCES: IPUMS-USA (Ruggles, Flood et al. 2019) American Community Survey micro-samples (2010–18); Catalyst registration files (2010–18).

NOTES: Table contains state weights in the synthetic control analysis of Latino registrants, separately for each model.

TABLE A5

Synthetic control weights by state and model, Asian American registrants

	All Lags	Lags+ Political	Mixture	Demog.	Full	Lags+ Demog.
AL	0	0	0	0	0	0.005
AR	0	0	0	0	0	0.011
AZ	0.001	0	0	0	0	0.012
DE	0	0	0	0	0	0.016
FL	0	0.005	0	0	0	0.013
HI	0.281	0.251	0	0.999	0	0.012
IA	0	0	0	0	0	0.004
ID	0.001	0	0	0	0	0.009
IN	0	0	0	0	0	0.007
KS	0	0	0	0	0	0.009
KY	0	0	0	0	0	0.004
LA	0	0	0	0	0	0.005
MA	0	0	0	0	0	0.006
MD	0	0	0	0	0	0.011
ME	0	0	0	0	0	0.021
MI	0	0	0	0	0	0.014
MN	0.027	0	0	0	0	0.070
MO	0	0	0	0	0	0.007
MS	0	0	0	0	0	0.066
MT	0	0	0	0	0	0.013
NC	0.136	0	0.486	0	0.486	0.010
ND	0	0	0	0	0	0.006
NE	0	0	0	0	0	0.009
NH	0	0	0	0	0	0.054
NJ	0	0.001	0	0	0	0.306
NM	0	0	0	0	0	0.011
NV	0.002	0	0	0	0	0.012
NY	0	0	0	0	0	0.022
OH	0	0	0	0	0	0.008
OK	0	0	0	0	0	0.009
PA	0	0.186	0.434	0	0.434	0.053
SC	0	0.074	0	0	0	0.007
SD	0	0	0	0	0	0.009
TN	0	0	0	0	0	0.015
TX	0.003	0	0	0	0	0.005
VA	0.223	0.161	0.078	0	0.078	0.026
WA	0	0	0	0	0	0.004
WI	0.009	0	0	0	0	0.013
WV	0	0.033	0	0	0	0.085
WY	0.314	0.289	0.001	0	0.001	0.018

SOURCES: IPUMS-USA (Ruggles, Flood et al. 2019) American Community Survey micro-samples (2010–18); Catalyst registration files (2010–18).

NOTES: Table contains state weights in the synthetic control analysis of Asian American registrants, separately for each model.

TABLE A6

Synthetic control weights by state and model, young (18-24) registrants

	All Lags	Lags+ Political	Mixture	Demog.	Full	Lags+ Demog.	Political
AL	0.001	0.001	0	0	0	0	0
AR	0.003	0.011	0	0	0	0	0
AZ	0.001	0.001	0	0	0	0	0
DE	0.02	0	0	0	0	0	0
FL	0.002	0.001	0	0	0	0	0
HI	0.003	0.053	0	0.345	0.176	0.101	0
IA	0.002	0.001	0	0	0	0	0
ID	0.446	0.055	0.054	0	0.311	0.408	0.049
IN	0.005	0.001	0	0	0	0	0
KS	0.002	0.001	0	0	0	0	0
KY	0.002	0.001	0	0	0.001	0	0.016
LA	0.001	0.001	0	0	0	0	0
MA	0.002	0.001	0	0	0	0	0
MD	0.002	0.001	0	0	0	0	0
ME	0.002	0.001	0	0	0	0	0
MI	0.003	0.001	0	0	0	0	0
MN	0.012	0.001	0	0	0	0	0
MO	0.002	0.001	0	0	0	0	0
MS	0.005	0.001	0	0	0	0	0
MT	0.283	0.387	0.398	0	0.164	0.216	0
NC	0.005	0.062	0.077	0.076	0.126	0.033	0.933
ND	0.002	0	0	0	0	0	0
NE	0.001	0.001	0	0	0	0	0
NH	0.001	0.001	0	0	0	0	0
NJ	0.002	0.001	0	0	0.001	0.1	0
NM	0.001	0.001	0	0	0	0	0
NV	0.001	0.001	0	0	0	0	0
NY	0.002	0.001	0	0	0.003	0	0
OH	0.001	0.001	0	0	0	0	0
OK	0.026	0.406	0.387	0	0.154	0.142	0
PA	0.002	0.001	0	0	0	0	0
SC	0.002	0.001	0	0	0	0	0
SD	0.002	0.001	0	0	0.001	0	0
TN	0.011	0.001	0	0	0	0	0
TX	0.002	0.001	0	0	0	0	0
VA	0.004	0.001	0	0.33	0.002	0	0
WA	0.002	0.001	0.074	0.129	0	0	0
WI	0.001	0.001	0	0	0	0	0
WV	0.003	0	0	0	0.058	0	0
WY	0.131	0.001	0.009	0.119	0	0	0

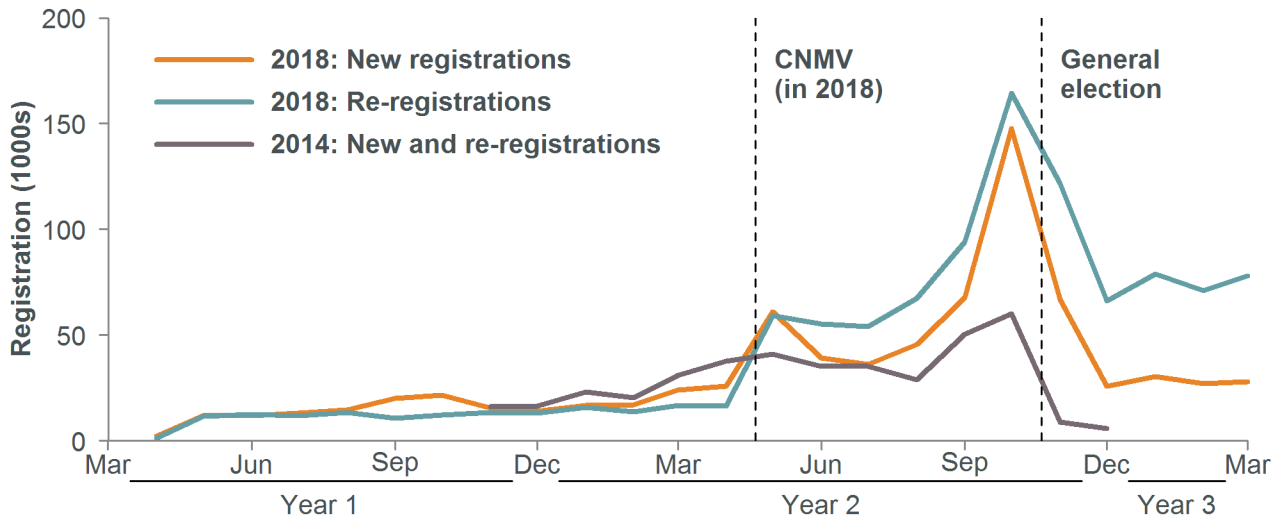
SOURCES: IPUMS-USA (Ruggles, Flood et al. 2019) American Community Survey micro-samples (2010–18); Catalyst registration files (2010–18).

NOTES: Table contains state weights in the synthetic control analysis of young registrants, separately for each model.

Appendix B. Additional analysis

FIGURE B1

New and re-registrations: Latinos

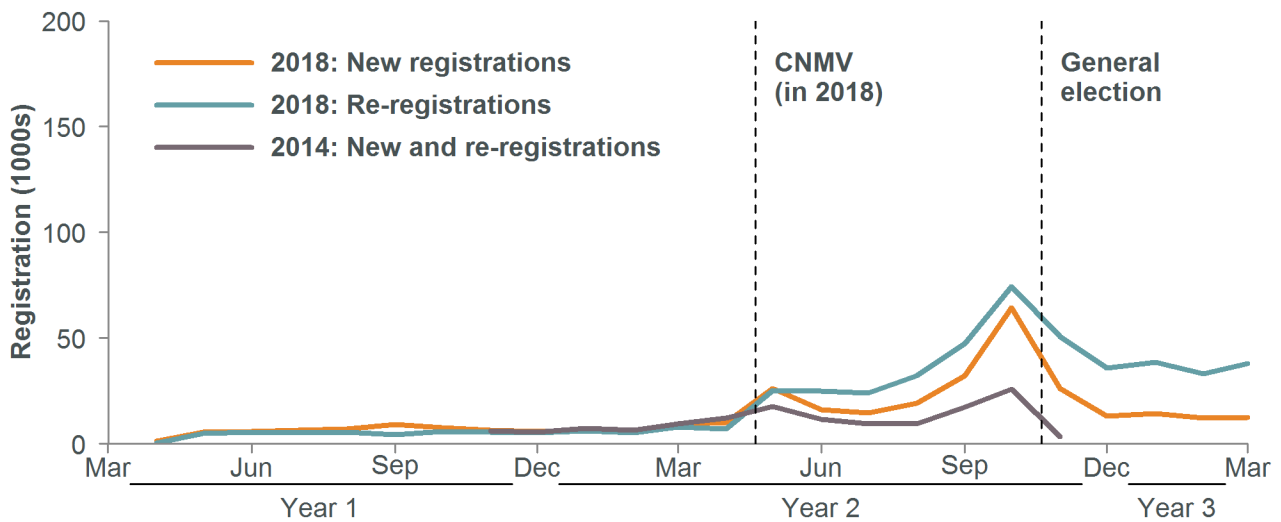


SOURCE: Political Data, Inc.

NOTE: Figure shows the number of Latinos who registered by month of registration.

FIGURE B2

New and re-registrations: Asian-Americans

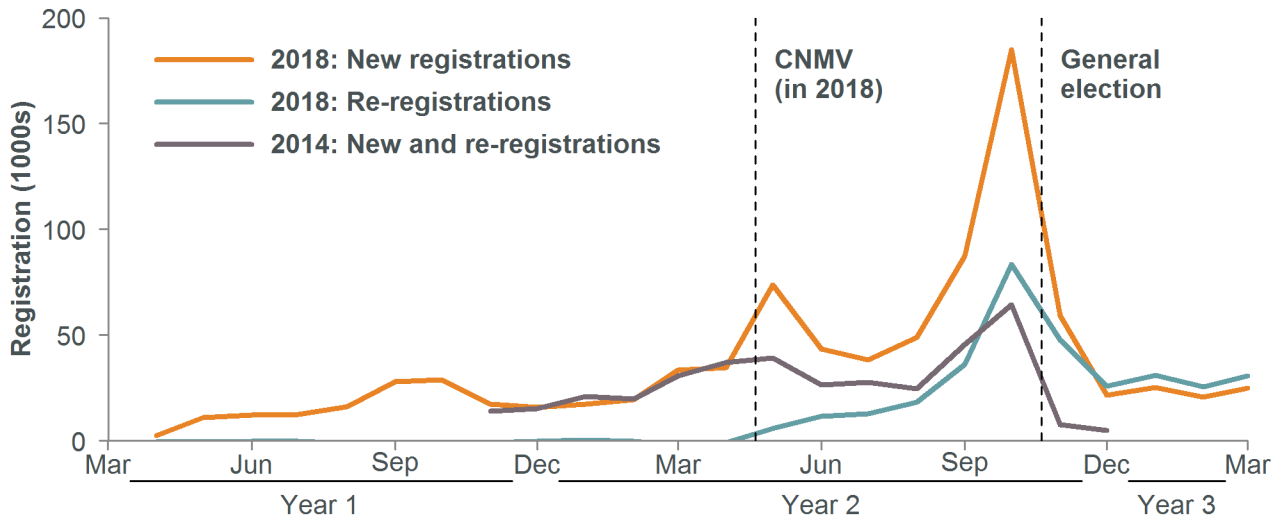


SOURCE: Political Data, Inc.

NOTE: Figure shows the number of Asian Americans who registered by month of registration.

FIGURE B3

New and re-registrations: Young people (18-24)



SOURCE: Political Data, Inc.

NOTE: Figure shows the number of young people who registered by month of registration.

TABLE B1

DMV registration numbers by county, before and after CNMV

County	Pre-CNMV Average	Post-CNMV Average	Change (%)
Alameda	6311	16611	163
Alpine	10	10	0
Amador	199	422	112
Butte	995	2210	122
Calaveras	243	571	135
Colusa	43	163	279
Contra Costa	4513	12164	170
Del Norte	81	255	215
El Dorado	969	2226	130
Fresno	2970	7877	165
Glenn	62	239	285
Humboldt	583	1374	136
Imperial	718	1427	99
Inyo	53	156	194
Kern	2689	7086	164
Kings	339	952	181
Lake	264	605	129
Lassen	80	264	230
Los Angeles	36698	94705	158
Madera	405	1229	203
Marin	1046	2975	184
Mariposa	83	213	157
Mendocino	368	781	112
Merced	569	1987	249
Modoc	38	91	139
Mono	38	127	234
Monterey	1244	3522	183
Napa	624	1463	134
Nevada	551	1388	152
Orange	12664	31974	152
Placer	1803	4395	144
Plumas	75	339	352
Riverside	9082	22413	147
Sacramento	6149	15247	148
San Benito	231	731	216
San Bernardino	7259	19984	175
San Diego	13825	34530	150
San Francisco	3901	9386	141
San Joaquin	2712	6656	145
San Luis Obispo	1407	3203	128
San Mateo	2918	8076	177

County	Pre-CNMV Average	Post-CNMV Average	Change (%)
Santa Barbara	1567	4094	161
Santa Clara	7736	19140	147
Santa Cruz	982	2900	195
Shasta	911	1991	119
Sierra	11	53	382
Siskiyou	205	491	140
Solano	1839	4622	151
Sonoma	2141	5073	137
Stanislaus	1817	4605	153
Sutter	267	896	236
Tehama	203	675	233
Trinity	58	132	128
Tulare	1224	3502	186
Tuolumne	256	610	138
Ventura	3315	8653	161
Yolo	827	2040	147
Yuba	236	625	165

SOURCES: California Secretary of State

NOTES: Pre-CNMV Average is the average monthly DMV registration total from April 2015 through April 2018. Post-CNMV Average is the average monthly DMV registration total from May 2018 through December 2018.



PPIC

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.

Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
T: 415.291.4400
F: 415.291.4401
PPIC.ORG

PPIC Sacramento Center
Senator Office Building
1121 L Street, Suite 801
Sacramento, CA 95814
T: 916.440.1120
F: 916.440.1121