



PPIC

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

Technical Appendices

Expanding California's Electorate Will Recent Reforms Increase Voter Turnout?

Eric McGhee

with research support from Daniel Krimm

Contents

Appendix A. Current Population Survey Missing Data Imputation

Appendix B. Detailed Model Results

Supported with funding from the S. D. Bechtel, Jr. Foundation

Appendix A: Current Population Survey Missing Data Imputation

We use the November Supplement to the Current Population Survey (CPS) of the U.S. Census for many of the analyses in this report. The CPS is a survey much like any other public opinion survey, so it has its share of missing data when respondents are not asked a particular question, refuse to answer the question, or profess not to know the answer. Sometimes, missing answers of this sort are interesting in and of themselves because they reveal something about the respondent's point of view. But we use the CPS only to obtain information about which every respondent *ought* to have a true response. Thus, it is preferable to fill in as much of this information as possible.

In its reports of registration and voting, the Census handles the problem of missing data by assuming that every respondent who is missing on those variables was a “no”: i.e., was not registered or did not vote. This is a problematic decision, since it seems highly doubtful that *all* or even *almost all* of these respondents were unregistered nonvoters. The CPS has traditionally avoided overt criticism on this score because the missing data strategy has yielded total turnout figures that match official turnout statistics quite closely.

But the official Census reports are accurate because respondents exaggerate their turnout, and the Census approach to missing data simply adjusts down this overreport. Because the method is not designed to correct for overreporting, its accuracy is simply a happy accident. Indeed, the method is far less successful with registration, where it frequently falls far below official registration rates, even though these official numbers are themselves often deflated because of “dead weight” on the rolls—voters who have died or moved but not been pulled off the registration list. (How serious is dead weight? The official registration numbers exceed the population of eligible voters in some states.) In California, the Census method routinely reports registration rates as much as 10 percentage points below the official numbers from the secretary of state. Even the accuracy of the turnout numbers is no longer guaranteed: their accuracy is likely getting worse as the population of nonrespondents becomes ever more biased toward certain demographic groups (McDonald 2012).

Reflecting these problems, the academic studies that use the CPS typically do not follow the Census's method. Instead, the standard approach is to drop respondents with missing values from the analysis entirely. This is a better approach, but it also has its dangers. It assumes that the missing data are what statisticians call *missing completely at random* (MCAR): that is, it pretends as if respondents drew a number from a hat to decide whether they would answer each question, making the population of nonrespondents equivalent to the population of respondents in all but their decision to answer (King et al. 2001). But those who did not answer the question are actually very different from those who did, so treating them as the same introduces new and different biases into the analysis.

A still better practice is to impute values for the missing data with the data available. The process assumes that the data are *missing at random* (MAR): the probability that a value is missing is related to other values in the data set and can be predicted with reasonable accuracy from them. The imputation model is then a likelihood function, where the variables are assumed to be jointly multivariate normal, and this likelihood is then maximized to produce values for the missing cells.¹ The additional uncertainty that results from this

¹ The actual process of computing the data likelihood and taking random draws from it is more complicated than this implies, since the problem is highly computationally intensive. A variety of shortcuts have been devised to address this problem. See King et al. (2001) for more details.

imputation can be incorporated into calculations of the standard errors, allowing one to temper one's conclusions appropriately.

The imputation model here uses all the demographic and political variables in the data set, and was run in Amelia for R (Honaker et al. 2009). The variables in the model are closely correlated and so are likely to be excellent predictors of one another. When we compared the results of this imputation method to official numbers for registration and voting in California, we obtained results that matched trends over time quite closely (as opposed to the Census method, which implied an absolute decline in turnout and registration that has not in fact occurred), and actually hit the point estimates for California's registration with impressive precision. The voting estimates, as one might expect given respondent tendencies to overstate turnout, were somewhat too high. But the trends were all in line with actual turnout figures. Accordingly, we are careful throughout the report to avoid reporting point estimates, which may be distorted by flawed self-reports, and report only relative effects instead.

Appendix B: Detailed Model Results

This appendix contains regression results for any analysis in the main text that employed such methods. Some of these models are run on official reports of registration or voting, and others use individual-level data from the voter registration file, as provided to PPIC by Political Data, Inc. However, most of the models are run on a stacked data set of Current Population Survey November Supplements covering the years from 1980 through 2012 or on a subset of those data as appropriate. Because the CPS employs a complex, stratified survey design, any model we ran with those data also clusters on households and weights using the sampling weight provided by the CPS.²

TABLE B1
Model results for analysis of online registration (Dependent variable=registration)

	β	SE
Intercept	-1.514***	0.060
Online registration state	-0.015	0.028
Age	0.017***	0.002
Age ²	0.000***	0.000
Married	0.378***	0.011
Latino	-0.185***	0.021
Asian American/Pacific Islander	-0.775***	0.028
African American	0.393***	0.019
Female	0.161***	0.009
Some high school	0.400***	0.026
High school graduate	0.993***	0.023
Some college	1.732***	0.025
College graduate	2.314***	0.029
Postgraduate	2.665***	0.036
Unemployed	-0.168***	0.027
Moved 1–6 months ago	-0.232***	0.034
Moved 6–11 months ago	-0.196***	0.035
Moved 1–2 years ago	-0.025	0.034
At current address more than two years	0.444***	0.031
Year fixed effects?		Yes
State fixed effects?		Yes
-2 * log likelihood		
N		887,813

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics); The National Conference of State Legislatures (online registration).

NOTE: Cell entries are logit regression coefficients. # p < 0.10, *** p < 0.001.

² For detailed thoughts on regression analysis with CPS data, see Davern et al. (2007).

TABLE B2

Model results for turnout using actual election returns (Table 1, Column 2, Row 2)

	β	SE
Intercept	40.86***	11.19
EDR state	0.29	1.09
Closing date	0.04	0.04
Average age	0.00	0.00
High school graduate (%)	0.06	0.08
College graduate (%)	0.09	0.08
African American (%)	0.20 [#]	0.10
Moved in last six months (%)	-0.20 [#]	0.11
Year fixed effects?		Yes
State fixed effects?		Yes
Adjusted R ²		0.83
Root MSE		2.96
N		450

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics); The Book of the States (closing dates 1994–2012); Mitchell and Wlezien (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration); The United States Elections Project (voter turnout 1980–2012).

NOTE: Cell entries are ordinary least squares regression coefficients. [#] p < 0.10, *** p < 0.001.

TABLE B3

Model results for turnout in Iowa using actual election returns (Table 1, Column 2, Row 10)

	β	SE
Intercept	69.88***	1.49
Iowa under EDR	-0.23	2.31
Year = 2008	-0.29	0.94
State = Minnesota	8.58*	1.89
State = South Dakota	-3.30	1.89
State = Nebraska	-6.84*	1.89
State = Kansas	-7.92*	1.89
State = Missouri	-3.28	1.89
Adjusted R ²		0.93
Root MSE		1.49
N		12

SOURCE: The United States Elections Project (voter turnout 1980–2012).

NOTES: Data include turnout numbers for six states (Iowa, Kansas, Minnesota, Missouri, Nebraska, and South Dakota) and two election years (2004 and 2008). Iowa used EDR in its first presidential election in 2008, and none of the comparison states changed its system during the same two-presidential-election period. Cell entries are ordinary least squares regression coefficients. * p < 0.05, *** p < 0.001.

TABLE B4
Model results for turnout in Montana using actual election returns
(Table 1, Column 2, Row 11)

	β	SE
Intercept	64.43***	1.22
Montana under EDR	3.91	1.92
Year = 2008	-2.03 [#]	0.86
State = Idaho	-3.89 [#]	1.55
State = North Dakota	-3.56	1.55
State = South Dakota	3.02	1.55
State = Wyoming	-3.12	1.55
Adjusted R ²	0.50	
Root MSE	1.22	
N	10	

SOURCE: The United States Elections Project (voter turnout 1980–2012).

NOTES: Data include turnout numbers for five states (Idaho, Montana, North Dakota, South Dakota, and Wyoming) and two election years (2004 and 2008). Montana used EDR in its first presidential election in 2008, and none of the comparison states changed its system during the same two-presidential-election period. Cell entries are ordinary least squares regression coefficients. * p < 0.05, *** p < 0.001.

TABLE B5
Model results for Table 1, all CPS respondents

	β	SE
Intercept	-2.302***	0.064
EDR state	0.028	0.062
Closing date	0.003 [#]	0.002
Age	0.030***	0.001
Age ²	0.000***	0.000
Married	0.413***	0.010
Latino	-0.182***	0.018
Asian American/Pacific Islander	-0.761***	0.029
African American	0.470***	0.016
Female	0.141***	0.008
Some high school	0.367***	0.023
High school graduate	0.985***	0.022
Some college	1.685***	0.022
College graduate	2.285***	0.025
Postgraduate	2.627***	0.029
Unemployed	-0.190***	0.023
Moved 1–6 months ago	-0.205***	0.024
Moved 6–11 months ago	-0.150***	0.027
Moved 1–2 years ago	0.064**	0.026
At current address more than two years	0.524***	0.023
Year fixed effects?	Yes	
State fixed effects?	Yes	
-2 * log likelihood	973367	
N	887,813	

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994–2012); Mitchell and Wlezién (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration).

NOTE: Cell entries are logistic regression coefficients. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE B6
Model results for Table 1, CPS respondents without a high school diploma

	β	SE
Intercept	-1.796***	0.077
EDR state	-0.019	0.077
Closing date	0.004 [#]	0.002
Age	0.040***	0.002
Age ²	0.000***	0.000
Married	0.486***	0.013
Latino	-0.349***	0.024
Asian American/Pacific Islander	-0.549***	0.039
African American	0.499***	0.021
Female	0.138***	0.010
Unemployed	-0.200***	0.032
Moved 1–6 months ago	-0.165***	0.029
Moved 6–11 months ago	-0.062	0.043
Moved 1–2 years ago	0.091***	0.030
At current address more than two years	0.555***	0.026
Year fixed effects?	Yes	
State fixed effects?	Yes	
-2 * log likelihood	591418	
N	472,986	

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994–2012); Mitchell and Wleziem (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration).

NOTE: Cell entries are logistic regression coefficients. # p < 0.05, ** p < 0.01, *** p < 0.001.

TABLE B7
Model results for Table 1, CPS respondents who have moved in the last six months

	β	SE
Intercept	-1.924***	0.182
EDR state	0.109	0.162
Closing date	-0.003	0.004
Age	0.018***	0.005
Age ²	0.000	0.000
Married	0.256***	0.028
Latino	-0.267***	0.051
Asian American/Pacific Islander	-0.569***	0.085
African American	0.436***	0.039
Female	0.145***	0.021
Some high school	0.124	0.084
High school graduate	0.746***	0.081
Some college	1.396***	0.088
College graduate	2.007***	0.088
Postgraduate	2.259***	0.096
Unemployed	-0.285***	0.053
Year fixed effects?		Yes
State fixed effects?		Yes
-2 * log likelihood		147234
N		119,052

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994–2012); Mitchell and Wleziem (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration).

NOTE: Cell entries are logistic regression coefficients. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE B8
Model results for Table 1, CPS respondents younger than age 25

	β	SE
Intercept	-2.055***	0.217
EDR	0.104	0.150
Closing date	0.000	0.004
Married	-0.034	0.039
Latino	-0.305***	0.040
Asian American/Pacific Islander	-0.543***	0.070
African American	0.396***	0.040
Female	0.153***	0.025
Some high school	0.848***	0.194
High school graduate	1.275***	0.189
Some college	2.024***	0.186
College graduate	2.746***	0.192
Postgraduate	2.682***	0.261
Unemployed	-0.256***	0.045
Moved 1–6 months ago	-0.040	0.057
Moved 6–11 months ago	-0.005	0.061
Moved 1–2 years ago	0.204**	0.059
At current address more than two years	0.556***	0.061
Year fixed effects?	Yes	
State fixed effects?	Yes	
-2 * log likelihood	151015	
N	118,682	

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994-2012); Mitchell and Wleziem (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration).

NOTE: Cell entries are logistic regression coefficients. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE B9
Model results for Table 1, estimate for Iowa using CPS

	β	SE
Intercept	-1.729***	0.239
EDR	0.067	0.116
Age	-0.008	0.007
Age ²	0.000***	0.000
Married	0.534***	0.055
Female	0.220***	0.038
Some high school	0.644***	0.128
High school graduate	1.310***	0.113
Some college	2.109***	0.118
College graduate	2.931***	0.133
Postgraduate	3.411***	0.180
Unemployed	-0.101	0.120
Moved 1–6 months ago	-0.222	0.182
Moved 6–11 months ago	-0.247	0.167
Moved 1–2 years ago	0.054	0.170
At current address more than two years	0.527**	0.158
Year = 2008	-0.032	0.058
State fixed effects?	Yes	
-2 * log likelihood	19438	
N	20,245	

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994–2012); Mitchell and Wleziem (1996) (closing dates 1980–1992); The National Conference of State Legislatures (election day registration).

NOTES: Data have been limited to Iowa and similar neighboring states (Minnesota, South Dakota, Nebraska, Kansas, and Missouri), as well as to the presidential election just before (2004) and just after (2008) Iowa changed to EDR. Cell entries are logistic regression coefficients. # p < 0.05, ** p < 0.01, *** p < 0.001.

TABLE B10
Model results for Table 1, estimate for Montana using CPS

	β	SE
Intercept	-1.612***	0.338
EDR	0.021	0.154
Age	-0.004	0.009
Age ²	0.000*	0.000
Married	0.515***	0.063
Female	0.146***	0.046
Some high school	0.330 [#]	0.172
High school graduate	1.029***	0.153
Some college	1.687***	0.159
College graduate	2.503***	0.171
Postgraduate	2.795***	0.248
Unemployed	-0.376*	0.146
Moved 1–6 months ago	-0.123	0.225
Moved 6–11 months ago	-0.189	0.246
Moved 1–2 years ago	-0.043	0.208
At current address more than two years	0.589**	0.197
Year = 2008	-0.097	0.069
State fixed effects?	Yes	
-2 * log likelihood	12752	
N	12,368	

SOURCES: U.S. Census Current Population Survey, November Supplement, 1980–2012 (demographics and turnout); The Book of the States (closing dates 1994–2012); The National Conference of State Legislatures (election day registration).

NOTE: Data have been limited to Montana and similar neighboring states (Idaho, North Dakota, South Dakota, and Wyoming), as well as to the presidential election just before (2004) and just after (2008) Montana changed to EDR.

TABLE B11
Late ballot statistics from Table 3 by county

	Total number of late ballots	Share of VBM ballots cast	Share of uncounted ballots
Amador	29	0.27%	24%
Butte	123	0.22	7
Calaveras	52	0.00	33
Colusa	25	0.85	81
Contra Costa	866	0.38	33
Fresno	545	0.41	68
Glenn	15	0.28	79
Humboldt	101	0.35	45
Inyo	17	0.35	16
Lassen	18	0.30	82
Los Angeles	5,576	0.52	66
Madera	55	0.22	19
Mariposa	14	0.26	19
Merced	191	0.69	74
Mono	19	0.96	15
Napa	138	0.39	27
Orange	2,343	0.39	74
Placer	330	0.29	46
Plumas	47	0.69	60
Sacramento	1,242	0.40	40
San Benito	52	0.47	25
San Diego	1,677	0.24	55
San Francisco	1,327	0.70	41
Santa Clara	1,838	0.39	36
Shasta	88	0.19	20
Sierra	19	1.03	24
Siskiyou	94	0.94	44
Sonoma	307	0.21	36
Sutter	41	0.23	26
Tehama	21	0.14	16
Ventura	668	0.41	29

SOURCES: Political Data, Inc., CACEO (total late and VBM ballots), California Secretary of State (total uncounted VBM ballots by county).

NOTES: Political Data, Inc. provided data for Calaveras, Madera, Mariposa, San Francisco, Sierra, Sonoma, and Tehama counties. CACEO provided data for Amador, Fresno, Humboldt, Inyo, Los Angeles, Mono, Orange, Plumas, Sacramento, San Benito, San Diego, Santa Clara, and Siskiyou counties. Data for the remaining counties were available from both sources. Where the sources did not match, the higher number is reported.

TABLE B12
Model results for late ballot estimates, VBM voters in late ballot counties only
(Table 4)

	β	SE
Intercept	-5.529***	0.056
New registrant since 2010	-0.127**	0.041
Voted in 2010	-1.017***	0.041
Age < 25	1.484***	0.039
Democrat	0.342***	0.044
Independent	0.587***	0.045
Latino	-0.290***	0.050
Asian American/Pacific Islander	0.346**	0.051
County fixed effects?	Yes	
-2 * log likelihood	51090.38	
N	1,230,536	

SOURCE: Political Data, Inc.

NOTE: Analysis includes data from 18 counties provided by Political Data, Inc.: Butte, Calaveras, Colusa, Contra Costa, Glenn, Lassen, Madera, Mariposa, Merced, Napa, Placer, San Francisco, Shasta, Sierra, Sonoma, Sutter, Tehama, and Ventura.

TABLE B13

Model results for USPS consolidation analysis, Dependent variable=late ballots, VBM voters in counties with late ballot data

	Consolidated = closed <u>and</u> merged		Consolidated = closed <u>or</u> merged	
	β	SE	β	SE
Intercept	-4.926***	0.059	-5.515***	0.100
Consolidated zip code	-2.330***	0.091	-0.014	0.083
New registrant since 2010	-0.161***	0.041	-0.127***	0.041
Voted in 2010	-1.022***	0.041	-1.017***	0.041
Age < 25	1.553***	0.039	1.484***	0.039
Democrat	0.196***	0.045	0.342***	0.044
Independent	0.425***	0.046	0.587***	0.045
Latino	-0.187***	0.050	-0.291***	0.050
Asian American/Pacific Islander	0.174***	0.052	0.346***	0.051
County fixed effects?	Yes		Yes	
-2 * log likelihood	50067.43		51090.35	
N	1,230,536		1,230,536	

SOURCE: Political Data, Inc.

NOTE: Analysis includes data from 18 counties provided by Political Data, Inc.: Butte, Calaveras, Colusa, Contra Costa, Glenn, Lassen, Madera, Mariposa, Merced, Napa, Placer, San Francisco, Shasta, Sierra, Sonoma, Sutter, Tehama, and Ventura.

TABLE B14

Model results for USPS consolidation analysis, Dependent variable=counted ballots, all counties

	Consolidated = closed <u>and</u> merged		Consolidated = closed <u>or</u> merged	
	β	SE	β	SE
Intercept	-1.388***	0.002	-1.413***	0.002
VBM in consolidated zip code	0.018***	0.004	-0.012***	0.003
VBM	0.489***	0.001	0.493***	0.002
Consolidated zip code	-0.008**	0.003	0.059***	0.002
New registrant since 2010	1.062***	0.002	1.062***	0.002
Asian American/Pacific Islander	-0.379***	0.002	-0.381***	0.002
Latino	-0.230***	0.001	-0.221***	0.001
Foreign-born	-0.083***	0.002	-0.086***	0.002
Age	0.013***	0.000	0.014***	0.000
Homeowner	0.669***	0.001	0.668***	0.001
Democrat	0.323***	0.001	0.322***	0.001
Republican	0.207***	0.002	0.211***	0.002
Voted in 2010	2.034***	0.001	2.033***	0.001
-2 * log likelihood	16482375		16480555	
N	17,795,670		17,795,670	

SOURCE: Political Data, Inc.

NOTE: Analysis includes data from all counties in 2012.

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research on major economic, social, and political issues. The institute's goal is to raise public awareness and to give elected representatives and other decisionmakers a more informed basis for developing policies and programs.

The institute's research focuses on the underlying forces shaping California's future, cutting across a wide range of public policy concerns, including economic development, education, environment and resources, governance, population, public finance, and social and health policy.

PPIC is a public charity. It does not take or support positions on any ballot measures or on any local, state, or federal legislation, nor does it endorse, support, or oppose any political parties or candidates for public office. PPIC was established in 1994 with an endowment from William R. Hewlett.

Mark Baldassare is President and Chief Executive Officer of PPIC.
Donna Lucas is Chair of the Board of Directors.

Short sections of text, not to exceed three paragraphs, may be quoted without written permission provided that full attribution is given to the source.

Research publications reflect the views of the authors and do not necessarily reflect the views of the staff, officers, or Board of Directors of the Public Policy Institute of California.

Copyright © 2014 Public Policy Institute of California
All rights reserved.
San Francisco, CA

PUBLIC POLICY INSTITUTE OF CALIFORNIA
500 Washington Street, Suite 600
San Francisco, California 94111
phone: 415.291.4400
fax: 415.291.4401
www.ppic.org

PPIC SACRAMENTO CENTER
Senator Office Building
1121 L Street, Suite 801
Sacramento, California 95814
phone: 916.440.1120
fax: 916.440.1121