PUBLIC POLICY INSTITUTE OF CALIFORNIA

Assessing California's Redistricting Commission

Effects on Partisan Fairness and Competitiveness

Technical Appendices

CONTENTS

Eric McGhee

with research support from Lunna Lopes

Appendix A.
Measures of Partisan Gerrymandering

The Efficiency Gap

The efficiency gap (EG) begins with the observation that many votes in single-member district systems are "wasted" in the sense that they do not directly contribute to victory. This includes "lost" votes—all those cast for a losing candidate—and "surplus" votes—those cast for a winner in excess of the number required to fend off all potential rival coalitions. As applies to single-member district elections, surplus votes are defined against the 50 percent threshold (McGhee 2017), even when there are more than two parties. The EG proposes that the wasted votes between any two parties should be roughly equal if the plan is fair. To calculate this comparison, the EG sums the wasted votes in each party, takes the difference between them, and divides by the total votes cast to ease comparison between plans. Working through the algebra, if there are two parties A and B, the EG for Party A reduces to:

$$EG_A = (S_A - 50) - 2 * (V_A - 50) \qquad \text{[A1]}$$

where S_A is seat share for Party A and V_A is vote share for Party A. In words, it says the EG for Party A is the difference between Party A's seat share margin above 50 percent and twice Party A's vote share margin above 50 percent. The valence of the measure is always related to the party referenced by S and V. For example, if S and V refer to the Democratic seat and vote share, then positive values indicate an advantage in favor of Democrats and negative values an advantage in favor of Republicans. This is the way the metric is calculated for this report. The reverse would be true if S and V referred to Republican seat and vote share. The substance of the conclusions is unaffected by this decision.

If turnout is presumed to be roughly equal across districts, or if variation in district turnout is considered unimportant, then V_A in Equation A1 can be calculated with the average district vote share. If n is the total number of districts in the system, v_i^A and v_i^B are the vote counts for Party A and Party B, respectively, in district i, and $t_i = v_i^A + v_i^B$, then

$$V_A = \frac{1}{n}\sum_1^n v_i^A / t_i \qquad \text{[A2]}$$

By contrast, if variation in turnout matters, McGhee (2017) shows that V should instead be calculated by dividing the total raw vote for the party in question by the total raw vote for both parties. Using the same notation as Equation A2:

$$V_A = \sum_1^n v_i^A / \sum_1^n t_i \qquad \text{[A3]}$$

This report uses the version of V_A in Equation A3 for all calculations. Because Democrats tend to win lower-turnout districts, this approach tends to produce results that suggest a larger advantage in favor of Democrats than the version in Equation A2. This is especially true in a state like California, where many districts have large numbers of noncitizens who cannot vote but who still count for the purposes of representation.

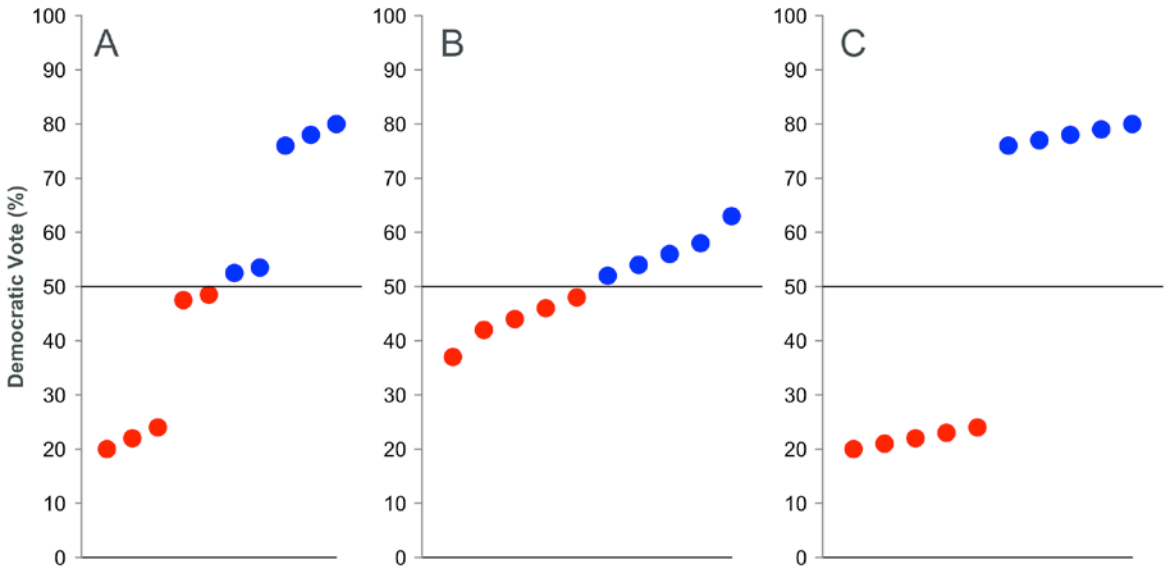
Declination

Declination (Warrington 2017) starts from the assumption that a plan drawn with the intent to advantage one party will arrange the distribution of district vote shares in a way that treats the 50 percent threshold for victory differently than other vote values. In the absence of partisan intent, if all the districts in a plan are lined up from the least Democratic to the most Democratic, the line formed by one party's seats should be about as far from 50

percent on average as the other party's. Some examples of distributions that are balanced in this sense can be seen in Figure A1.

FIGURE A1

Hypothetical vote distributions that would be fair according to declination

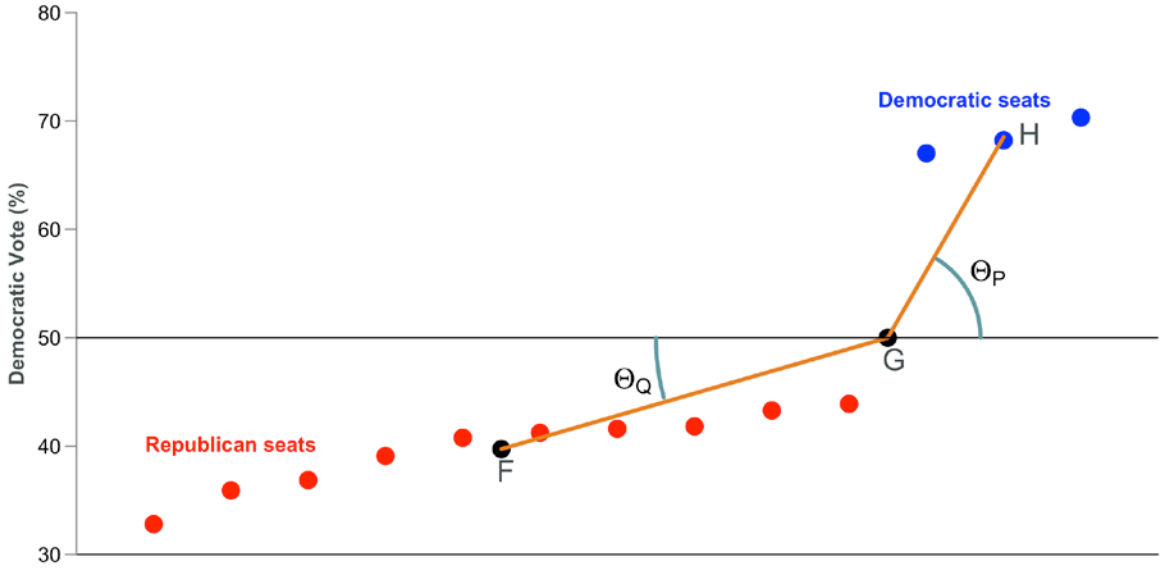


SOURCE: Adapted from examples provided by Greg Warrington, University of Vermont (2017)

NOTE: Graphs show hypothetical distributions of votes that are fair according to the declination measure.

FIGURE A2

Example of a vote distribution with an unfair declination



SOURCE: Congressional Quarterly Voting and Elections Collection (congressional vote 2002-2016); adapted from an example provided by Greg Warrington, University of Vermont (2017).

NOTE: Graph shows an unfair distribution of votes, where the points below the 0.5 vote share line are much closer to that line and show signs of deliberately avoiding that line to ensure more wins for that party.

When this condition is not met, the distribution might look more like the one in Figure A2, which comes from the actual 2016 congressional results in North Carolina and replicates Figure 1 from the main text. The districts won by Republicans (in red) are much closer to 50 percent than the ones won by Democrats (in blue). This distribution appears designed to ensure that Republican seats do not cross the 50 percent line, in part by packing the Democratic seats.

Figure A2 imagines a point G that lies on the 50 percent line, directly between the mass of points representing seats won by each party. It also presents two lines connecting G to the center of mass of each party's set of seats, designated as F and H. These lines form the angles θ_Q and θ_P, which represent the deviation of each line from 50 percent. Declination suggests that when a plan is not deliberately drawn to favor one party, these two angles will be roughly equal, as they would be in any of the graphs in Figure A1. When they deviate from each other, the smaller angle will generally identify the favored party. To capture this idea, declination takes the difference between those two angles and divides by $\pi/2$ to convert the result from radians to fractions of 90 degrees:

$$DECLINATION = 2(\theta_Q - \theta_P)/\pi \qquad [A4]$$

This produces a number between -1 and 1. As calculated here, positive values favor Democrats and negative values favor Republicans.

Comparing the Measures

Declination is similar to the EG to the extent that both capture differences between the two parties in their margins of victory. In fact, the two are strongly correlated: 0.87 for Congress and 0.84 for state legislatures. To the extent that there are differences, declination is arguably a better measure of intent. It seems unlikely—though not impossible—that a large declination could occur without some desire to produce it. The EG cannot provide the same evidence of intent by itself, since a large EG could conceivably happen by accident if one party manages to claim a few key seats by very narrow margins.

However, the EG probably does a better job of measuring the core goal of a gerrymander, which is the additional seats a gerrymandering party wins as a product of the redistricting plan. The EG is seat-denominated: if a party wins more seats without winning more votes, the EG will always show that party to be more advantaged than before. By contrast, the units for declination are less intuitive, and it is possible for a party to gain seats without gaining votes and still appear no better or even worse off. This is because declination uses the ratio of each party's vote margin in the seats it holds to the number of seats it holds, so information about seat share is lost in the calculation.[1]

Thus, the approach in this report is to use both measures. Declination is used primarily as an indicator of intent, while the EG indicates the gain in seat share that a party obtains through that intent.

[1] The angles in Equation A4 come from calculating the corresponding arctangent, which ends up embedding this ratio into the formula.

Appendix B.
Imputation Models for Uncontested Seats

A redistricting plan is drawn as a whole that includes all the districts, so any intent to produce more seats for one party or to increase or decrease competition is implemented through the entire plan. If there are any seats that are not contested—either because one major party fails to run a candidate, or because the seat is not up for election in a particular cycle—it is important to offer reasonable guesses as to how voters in those seats would have voted if they had faced a contested race in that cycle. This includes, for example, all same-party races in California under the Top Two primary, and the 20 state Senate seats that are not contested in any given cycle.

To generate these imputed values, I first estimate a hierarchical model of vote shares using the votes in contested races, and then predict values for the uncontested races using this model. This necessarily changes the quantity being measured from the actual number of competitive races or seats won by each party to the number that would have obtained if all seats in the redistricting plan had hosted contests.

TABLE B1

Imputation model: Vote share

	State Legislatures	Congress
Intercept	50.24	51.82
	(0.89)	(1.10)
District presidential vote (mean-deviated by year)	0.77	0.76
	(0.00)	(0.01)
Incumbency	6.67	8.00
	(0.06)	(0.15)
Chamber	-0.58	--
	(0.10)	
Random effects (St. Dev.)		
State	2.71	2.41
Year	2.48	2.93
Residual error	6.50	5.56
Deviance	159662.7	18985.3
N	24,221	3007

SOURCES: Steven Rogers, University of St. Louis (presidential vote by state legislative district, 2004 and 2008); Carl Klarner, University of Florida (state legislative vote 2002-2016); Daily Kos (presidential vote by state legislative district, 2012 and 2016); Polidata (presidential vote by congressional district 2002-2016); Congressional Quarterly Voting and Elections Collection (congressional vote 2002-2016).

NOTES: Models were fit using the lme4 package (https://cran.r-project.org/web/packages/lme4/index.html) in R 3.4.0.

TABLE B2

Imputation model: Turnout

	State Legislatures	Congress
Intercept	1.97	4.95
	(36.07)	(5.28)
Total district presidential vote (raw)	0.92	0.95
	(0.00)	(0.01)
Midterm	-51.55	-31.96
	(42.95)	(6.47)
Presidential vote X Midterm	-0.30	-0.20
	(0.00)	(0.01)
Chamber	6.66	--
	(0.94)	
Random effects (St. Dev.)		
State	21.13	13.77
Year	71.80	7.68
Residual error	46.36	17.42
Deviance	254871.6	25908.9
N	24,221	3007

SOURCES: Steven Rogers, University of St. Louis (presidential vote by state legislative district, 2004 and 2008); Carl Klarner, University of Florida (state legislative vote 2002-2016); Daily Kos (presidential vote by state legislative district, 2012 and 2016); Polidata (presidential vote by congressional district 2002-2016); Congressional Quarterly Voting and Elections Collection (congressional vote 2002-2016).

NOTES: Models were fit using the lme4 package (https://cran.r-project.org/web/packages/lme4/index.html) in R 3.4.0. For state legislature, the raw outcome votes and presidential votes by district were divided by 100 to ensure the covariates were closer to the same scale. The same numbers were divided by 1000 for the congressional model.

Because the number of wasted votes for each party can depend on variations in turnout across districts, this imputation process uses two models: one to estimate turnout in the uncontested races, and one to estimate the two-party vote share. The turnout model consists of the raw total presidential vote in each seat, an indicator for midterm elections (when turnout is typically lower), and an interaction between the two to capture the fact that raw midterm turnout will fall more in larger districts (while likely being about the same share of all voters in the district). The vote share model consists of the two-party Democratic presidential vote share in each seat and a variable for the incumbency status in the race (-1 Republican incumbent, 0 open, 1 Democratic incumbent). Both models include random terms for states and years, and the versions of each model for state legislative races also include an indicator variable for the chamber of the legislature (0 lower, 1 upper). I mean-deviate the presidential vote by year to ensure that it captures only cross-sectional variation in district outcomes. Nebraska's unicameral legislature was included as a lower chamber.

After estimating these models, I then use the coefficients and standard errors to simulate 200 vectors of vote shares and turnout levels, multiply these two together to get estimated raw Democratic vote, apply these numbers to the uncontested races, calculate 200 versions of each metric, and collapse these simulated metrics to a mean and a standard deviation for each.

I tested this approach with a k-fold cross-validation where k=10. I divided all contested seats into ten random subsets. For each 90%/10% split, I ran the model on the 90% portion and used the results to predict the remaining 10%. For each of these 10 subsets I calculated the average bias of the estimate (the amount the predictions missed the actual result on average, in standard deviations) and the proportion missing the actual result by more than the 95% confidence interval. I then averaged these metrics across all 10 subsets. A high-quality model will have an average bias very close to zero, with about 5 percent of the predictions falling outside the 95% confidence interval.

TABLE B3

Imputation model 10-fold cross-validation

	State Legislatures		Congress	
	Turnout	Vote	Turnout	Vote
Average bias (st.devs.)	0.01	0.00	0.00	0.00
Proportion > 95% confidence	0.03	0.07	0.05	0.05

SOURCES: Steven Rogers, University of St. Louis (presidential vote by state legislative district, 2004 and 2008); Carl Klarner, University of Florida (state legislative vote 2002-2016); Daily Kos (presidential vote by state legislative district, 2012 and 2016); Polidata (presidential vote by congressional district 2002-2016); Congressional Quarterly Voting and Elections Collection (congressional vote 2002-2016).

NOTES: Models were fit using the lme4 package (https://cran.r-project.org/web/packages/lme4/index.html) in R 3.4.0. For state legislature, the raw outcome votes and presidential votes by district were divided by 100 to ensure the covariates were closer to the same scale. The same numbers were divided by 1000 for the congressional model.

The results of this exercise are reported in Table B3. The calibration of the model for Congress is excellent: the average bias is zero and exactly 5% of predictions fall outside the 95% confidence window for each dependent variable. The model quality for state legislatures is somewhat weaker. Three percent of turnout predictions and 7 percent of vote share predictions miss the actual outcome by more than the 95% window, suggesting a model that is too conservative for turnout and too confident for vote share. But these deviations from expectation are not large and the average bias is still small in each case. On the whole, the model is accurate and well-calibrated.

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