Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies[*]

Jonathan N. Katz[†] Gary King[‡] Elizabeth Rosenblatt[§]

December 2, 2018

Abstract

We clarify the theoretical foundations of partisan fairness standards for district-based democratic electoral systems, including essential assumptions and definitions that have not been formalized or in some cases even discussed. We pare assumptions down to their minimal essential components and add extensive empirical evidence for those with observable implications. Throughout, we follow a fundamental principle of statistics too often ignored — defining the quantity of interest separately so its measures are vulnerable to being proven wrong, evaluated, and improved. This enables us to prove which approaches — claimed in the literature to be estimators of partisan symmetry, the most widely accepted standard — are statistically appropriate and which are biased, limited, or not measures of symmetry at all. Because real world redistricting involves complicated politics with numerous participants and conflicting goals, measures biased for partisan fairness sometimes still provide useful descriptions of other aspects of electoral systems.

[*] Our thanks to Steve Ansolabehere, Peter Aronow, Robin Best, Shawn Donahue, Moon Duchin, Andrew Gelman, Kosuke Imai, Jonathan Krasno, Daniel Magleby, Michael D. McDonald, Eric McGhee, Nate Persily, Jameson Quinn, Nick Stephanopoulos, Tyler VanderWeele, Sam Wang, Greg Warrington, and Xiang Zhou for helpful comments.

[†] Kay Sugahara Professor of Social Sciences and Statistics, DHSS 228-77, 1200 East California Blvd., Pasadena, CA 91125; jkatz.caltech.edu, jkatz@caltech.edu, (626) 395-4191.

[‡] Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, 1737 Cambridge Street, Harvard University, Cambridge MA 02138; GaryKing.org, King@Harvard.edu, (617) 500-7570.

[§] Affiliate, Institute for Quantitative Social Science, Harvard University, ERosenblatt@alumni.harvard.edu.

1 Introduction

Partisan fairness in modern democracies is defined at the intersection of two grand reresentative institutions – political parties and district-based electoral systems. Whereas parties are mostly defined by voters and candidates, the contiguous geographic districts that collectively tile a political system's landmass constitute the playing field on which the parties compete. This intersection is most obvious during legislative redistricting processes, but it is also crucial for evaluating the fairness of relatively fixed districts, such as for the US Senate and electoral college.

We clarify the theoretical foundations of *partisan symmetry*, the most widely accepted standard of partisan fairness in district-based democratic electoral systems, along with alternative definitions. Although the literature dates back more than a century, doing so requires definitions not fully formalized, essential assumptions not previously discussed, and quantities of interest often left implicitly defined. We also offer empirical evidence from 70,540 district-level elections, in 963 legislature-election years in the US, to shore up, or choose among, assumptions with observable implications. (These data, which we arranged to make public, may be the largest collection of election data ever analyzed at once; see Klarner 2018.) We use this theory and evidence to build on one of the most important principles of statistics — defining the quantities of interest rigorously and separately from the measures used to estimate them. This enables us to use standard statistical approaches to evaluate existing measures. Among measures claimed to be estimators of partisan symmetry, we distinguish between those which are statistically appropriate and those which are in fact biased, limited, or not measures of symmetry at all. We also show how measures biased for partisan fairness can still reveal other interesting features of complicated electoral systems unrelated to fairness.

Section 2 defines the partisan symmetry standard, and Section 3 considers alternatives. Section 4 clarifies assumptions needed for estimating seats-votes curves, and Section 5 evaluates existing measures of partisan fairness. We discuss uncertainty estimates in Section 6 and Section 7 concludes.

2 The Partisan Symmetry Standard

In this section, we describe the partisan symmetry standard for a single member district, where it is easier to understand, and then generalize it to an entire legislature. We also make explicit required assumptions in this approach the literature has left implicit or ignored, and characterize different types of symmetry and asymmetry. The concept of fairness-through-symmetry can be traced to "The Golden Rule" (part of almost every ethical tradition; Blackburn 2003) and the Bible (*Genesis* 13:8-9, *Matthew* 7:12; Wang and Remlinger 2018).

2.1 Symmetry in a Single Member District

Although all our results generalize to any number of political parties (as in Ansolabehere and King, 1990; Katz and King, 1999; King, 1990), we use two parties throughout to simplify exposition. We also assume an odd number of voters to eliminate the possibility of a tie (or assume a coin flip in that instance). Then denote the Democratic proportion of the (two party) vote in district d as v_d (for $d = 1, \ldots, L$). In one single member district, denote the *plurality voting rule* as $s(v) = \mathbf{1}(v > 0.5)$, which takes on the value 1 if $v > 0.5$ (meaning the Democratic candidate wins) and 0 otherwise (the Republican wins). In other words, when a political party receives more votes than any other party it wins the seat. The reason this rule is universally judged as fair is because it is symmetric, applying the same way to *any* party, regardless of its name or identity.

We formally express *district level partisan symmetry* (cf. "neutrality" in formal theory; May 1952, p.681–682) as $s(v) = 1 - s(1 - v)$, for all v. In other words, if we swapped the labels on the parties, nothing would change other than who wins the seat. For example, if the Democratic party received 0.55 of the vote in a district, it would win the seat, because $s(0.55) = 1$, and if (instead) the Republican party received 0.55 of the vote, it would receive the seat, because $1 - s(1 - 0.55) = 1$. The plurality voting rule is thus fair with respect to the two parties because it is symmetric.

Deviations from partisan symmetry in a single member district, first-past-the-post electoral system can stem from fraud. For example, if a criminal surreptitiously stuffs the ballot box with an extra 0.1 Democratic proportion of the vote, then the Democratic party will win the seat if it receives more than 0.4 (rather than 0.5) of the votes — that is, $s'(v) = \mathbf{1}(v > 0.4)$, for all v — which is obviously not symmetric. To see this asymmetry formally, consider that a Democratic candidate receiving 0.45 of the vote would win the seat, $s'(0.45) = 1$, but a Republican candidate who (instead) receives the same proportion of the vote would lose: $1 - s'(1 - 0.45) = 0$.

2.2 Symmetry in a Legislature

We now show how partisan symmetry applies to fairness for an entire legislature.

2.2.1 The Seats-Votes Curve

We define here the seats-votes curve from its component parts. Denote the *populace*, $\mathbb{P}$, the set of all individuals living in a state, including systematic patterns in their electoral behavior (or nonbehavior); an *electoral system*, $\mathbb{E}$, all factors that turn the populace's votes into seats, including district boundary lines, district level voting rules (such as plurality voting), and whether the rules are followed (Cox, 1997, p.38); and other *measured exogenous influences* on voter behavior, X, such as demographic variables (e.g., percent African American or immigrant), candidate quality (e.g., incumbency status or uncontestedness), voter behavior (such as lagged vote), and campaign events. Together $\{\mathbb{P}, \mathbb{E}, X\}$ determine a "permutation invariant" joint probability density from which district-level vote proportions are drawn, $p(v_1, \ldots, v_L \mid X)$.[1]

Next, we aggregate the district vote proportions into the statewide average district vote $V = V(v_1, \ldots, v_L) = \text{mean}_d(v_d)$ and the statewide seat proportion $S = S(v_1, \ldots, v_L) = \text{mean}_d[s(v_d)]$, with $s(v_d)$ defined in Section 2.1.[2] Electoral systems $\mathbb{E}$, including changes such as redistricting, are important because sets of district votes that differ, $\{v_1, \ldots, v_L\} \neq$

[1] Because all measures discussed in this paper are invariant to permutations of the district labels, we only need probablity densities specified up to a permutation of its arguments; e.g., $p(v_1, v_2, v_3 \mid X) = p(v_1, v_3, v_2 \mid X)$ (Wimmer, 2010, p.114). This is less restrictive than assuming that individual districts are drawn independently from a univariate density (Gelman and King, 1990; King, 1989).

[2] For set A with cardinality $\#A$, define the mean over i of function $g(i)$ as $\text{mean}_{i\in A}[g(i)] = \frac{1}{\#A}\sum_{i=1}^{\#A} g(i)$. When there is no ambiguity, we simplify notation by letting $\sum_d \equiv \sum_{d=1}^{D}$ and $\text{mean}_d \equiv \text{mean}_{d\in A}$.

$\{v'_1, \ldots, v'_L\}$, but which aggregate into the same average district vote $V(v_1, \ldots, v_L) = V(v'_1, \ldots, v'_L)$, can yield different statewide seat proportions $S(v_1, \ldots, v_L) \neq S(v'_1, \ldots, v'_L)$.[3]

We then define the *seats-votes function* by taking the expected value of the statewide seat proportion $S(v_1, \ldots, v_L)$ over the density $p(v_1, \ldots, v_L \mid X)$, constrained so that $V = \text{mean}(v_d)$:

$$E_p[S(v_1, \ldots, v_L) \mid X, \text{mean}(v_d) = V] = S(V \mid \mathbb{P}, \mathbb{E}, X) \equiv S(V). \tag{1}$$

The seats-votes function is a scalar property of the electoral system computed from random variables $\{v_1, \ldots, v_L\}$ and V, along with fixed characteristics X (King, 1989). A coherent seats-votes function is defined independently of the observed realizations $\{v_1^O, \ldots, v_L^O\}$ (and in turn independently of the observed realization of the average district vote V^O). We call this the *Stable Electoral System Assumption*:

Assumption 1. [SESA: Stable Electoral System] *The probability density of district vote proportions is defined independently of any one set of realized district vote proportions:* $p(v_1, \ldots, v_L \mid X, v_1^O, \ldots, v_L^O) = p(v_1, \ldots, v_L \mid X)$.

Assumption 1 can be thought of as Markov independence, such that an election does not change the electoral system that generates vote proportions (after conditioning on X). However, the assumption will usually be applied to data from one election in isolation, at that one time point, with independence applying over hypothetical replications from the same (stable) electoral system. Violations of this assumption occur when an election prompts a new redistricting controlled by a different party or group, or if an electoral realignment changes the coalitions making up the parties (unless encoded in X). This seats-votes curve would then be incoherent because the electoral system it describes is not stable as it is defined differently depending on the observed vote. A simple numerical example of a violation of SESA, and no single seats-votes curve, is if $S(0.6) = 0.7$ for an election with $V^O = 0.6$ but $S(0.6) = 0.8$ following an election with $V^O = 0.5$.

SESA implies that the seats-votes function is *single-valued*, and not dependent on the election outcome, so that a complete representation of all values of $S(V)$ for populace $\mathbb{P}$

[3] Redistricters often make calculations like these by assuming that individual votes are fixed, at least with respect to redisticting. Although convenient and often not far off, this assumption is unnecessary.

and electoral system $\mathbb{E}$, conditional on X, is the set $\mathcal{S} = \{S(V) : V \in [0, 1]\}$, which we call the *seats-votes curve*. If SESA does not hold, then the seats-votes function is not single-valued and the seats-votes curve is not coherently defined and, as such, concepts like partisan symmetry cannot even be evaluated. Including sufficiently informative variables in X can correct for a violation of this assumption. If SESA holds, then we still need to consider how to estimate it, a subject we address in Section 4. (SESA is related to the *Stable Unit Treatment Value Assumption, SUTVA*, commonly made in the causal inference literature; see Iacus, King, and Porro 2018; Rubin 1991; VanderWeele and Hernan 2012.)

2.2.2 Differential Partisan Turnout Effects

The Supreme Court requires equal population, not equal turnout, across districts (*Baker v. Carr*, 369 U.S. 186 (1962)). As such, when turnout rates differ by party, gerrymanderers can use this fact to their advantage. For example, because turnout is usually lower in Democratic areas (Leighley and Nagler 2013 and Plener Cover 2018, p.1189ff), Republicans can sometimes maintain their majority in meeting a district's population quota by packing in many who prefer the Democrats but are not likely to vote. Similarly, Democrats may settle for a minority of Democratic voters in a district if favorable demographic changes are on the horizon, such as young Hispanic immigrants aging into the electorate or older Republicans dying off.

Differential partisan turnout is represented in the seats-votes curve, as defined in Section 2.2.1. The curve conditions on V — the unweighted *average district vote*, $V = \text{mean}(v_d)$ — and then differential partisan turnout can influence $S(V)$, changing the shape of the curve.

For academic purposes, researchers may *also* be interested in the counterfactual seats-votes curve we would see if turnout were equalized across districts, a "controlled direct effect" (Acharya, Blackwell, and M. Sen, 2016). To construct this counterfactual curve, we switch from the average district vote to the total *statewide vote*, the weighted average of district vote proportions: $U = \sum_d n_d v_d / \sum_d n_d$, with n_d, the number of voters in district d, as weights. The two quantities coincide (i.e., $U = V$) when the turnout and votes are uncorrelated. To see this, let $n_d = \bar{n} + t_d$, where $\bar{n} = \text{mean}_d(n_d)$ and $t_d = n_d - \bar{n}$. Then,

$$U = \frac{\sum_d n_d v_d}{\sum_d n_d} = \frac{\sum_d \bar{n} v_d + \sum_d t_d v_d}{\sum_d n_d} = V + \frac{\sum_d t_d v_d}{\sum_d n_d}. \tag{2}$$

The last term of the last equality vanishes when $\text{Cov}(t_d, v_d) = \text{Cov}(n_d, v_d) = 0$.

It may seem paradoxical that weighting by turnout in the vote calculation controls away the effect of turnout on the seats-votes curve, while ignoring turnout enables its effect on $S(V)$ to be seen. Yet, turnout is in part a consequence of the electoral system $\mathbb{E}$ and therefore post-treatment. The quantity $S(V)$, conditional as it is on V, already has differential partisan turnout accounted for in its effect on seats (Ansolabehere, Brady, and Fiorina 1988; Grofman, Koetzle, and Brunell 1997; Gudgin and Taylor 2012, p.56). Researchers who want to measure all effects of redistricting including turnout use V and avoid U or they risk post-treatment bias (King and Zeng, 2006, §3.4).

Using U has an unrelated difficulty because of severe measurement error from total turnout often not being reported in uncontested districts and, even when it is, voters often skip casting ballots in these pointless “races”. Unfortunately, uncontestedness itself is quite prevalent in many state legislatures, in part a consequence of redistricting, and thus another important tool of gerrymanderers that should not be controlled away (*LULAC v. Perry*, 548 U.S. 399 (2006)). As such, this measurement error is post-treatment and may induce even more post-treatment bias in U. (Uncontestedness also affects V, but its effects are comparatively minor for most applications.)

Thus, although U and $S(U)$ are not of interest for evaluating the total effects of electoral systems or legislative redistricting maps from the point of view of democratic representation, they are sometimes important for academic purposes. See Campbell (1996).

2.2.3 Characteristics of Seats-Votes Curves

The most commonly accepted standard for fairness of voting in a legislature is statewide *partisan symmetry* (King and Browning, 1987) which we define formally as:

Definition 1 (Partisan Symmetry)**.** *An electoral system satisfies the partisan symmetry standard if* $S(V) = 1 - S(1 - V)$ *for all* $V \in [0, 1]$

(See Section 5.3 for an alternative representation.) Because of the impact of districting, even if $s(v) = 1 - s(1 - v)$ holds for every individual district, statewide partisan symmetry may not hold.

Any deviation from partisan symmetry is known as the degree and direction of *partisan bias*, which we define formally as follows:

Definition 2 (Partisan Bias)**.** *Partisan bias is the deviation from partisan symmetry:* $\beta(V) = \{S(V) - [1 - S(1 - V)]\}/2$*, for any* $V \in [0, 1]$.

The quantity $\beta(V)$ is the (perhaps negative) proportion of seats that should be taken from the Democrats (and thus given to the Republicans) to make the system fair. (The division by 2 makes $\beta(V)$ the distance from each party to symmetry, as desired, rather than to each other.) Thus, special cases of partisan bias include (a) partisan symmetry, where $\beta(V) = 0$; (a) Democratic bias, where $\beta(V) > 0$; and (c) Republican bias, $\beta(V) < 0$. Although $\beta(V)$ is defined for any $V \in [0, 1]$, only half this range is needed, say $V \in [0.5, 1]$, because $\beta(V) = \beta(1 - V)$. (Partisan bias is unrelated to statistical bias, where the expected value of an estimator is not equal to the population quantity of interest.)

The chosen value of V in a seats-votes function must be a *possible* result of the electoral system so that there is a defined value of $S(V) \in [0, 1]$. For example, if one party would not tolerate the other party winning, so that war would break out and end the democracy if say $V > 0.5$, then $S(V)$ would be undefined for $V > 0.5$. Similarly, a party system defined based on fixed ethnic or racial divisions would mean that only slight variations in V from V^O would be possible (due to changes in turnout or demographic change). This assumption does not require that any outcome be likely. For example, presently, the state houses in Massachusetts and Utah are 77% and 17% Democratic, respectively. Given what we know about electoral politics, the probability of either one being controlled by the opposition party in the near future is very small, but certainly not zero. The election of an African American as president was seen as highly unlikely only a few years before the election of Barack Obama, as was the election of Donald Trump before 2016; each was improbable for some researchers, but not impossible.

The assumption we need formalizes the venerable concept of rotation in office which “was a political principle put into the design of new political systems in order to prevent the corruption of elected officials, check government tyranny, guarantee liberty, enhance the quality of political representation, and promote widespread service in government, among other values” (Petracca, 1996). The rotation in office principle says that it is conceivable for both parties to win office, if enough elections are run under the same electoral system. We formalize this assumption as follows:

Assumption 2. [Rotation in Office] *For a given electoral system and “average district vote victory size” parameter $\eta \in [0, 0.5]$ chosen by the researcher, the range of possible values for the average district vote is no smaller than $V \in [0.5 - \eta, 0.5 + \eta]$.*

This assumption allows the range of possible vote proportions to be asymmetric, so long as it has as a subset a smaller symmetric range (e.g., $[0.4, 0.8]$ includes $[0.4, 0.6]$, so that $\eta = 0.1$). With the possible victory size parameter set to its maximum, $\eta = 0.5$, any value of $V \in [0, 1]$ may be used with $S(V)$ so that for example the full version of partisan bias in Definition 2 can be used. We allow η to take smaller values so that special cases of the partisan symmetry standard can apply in electoral systems where certain lopsided outcome sizes are inconceivable as long as a symmetric range exists. For example, for $\beta(0.5)$, we can use $\eta = 0$. In all cases, the range of conceivable values of V may be larger than $[0.5 - \eta, 0.5 + \eta]$. Although Assumption 2 is defined in terms of possible electoral outcomes, those that are exceedingly unlikely, such as Washington DC voting overwhelming Republican, do not violate this assumption but may generate model dependence in estimation (see Section 4).

2.2.4 Summaries

Partisan bias is sometimes summarized at (a) bias at 0.5, $\beta(0.5) = S(0.5) - 0.5$; (b) bias at another point such as $\beta(0.55) = \{S(0.55) - [1 - S(1 - 0.55)]\}/2 = \beta(0.45)$; (c) an average over a range of vote values, such as $E[\beta(V)] = \int_{0.5}^{0.55} \beta(V)p(V)dV$, where $p(V)$ is the predictive density of likely votes or a uniform with range based on plausible average district vote values (Gelman and King, 1994a); or (d) an indicator as in for whether $\mathbf{1}(V > 0.5) = \mathbf{1}[S(V) > 0.5]$ (Best, Donahue, Krasno, Magleby, and McDonald, 2018).

These summaries are easier to estimate than the entire curve but are only useful if they accurately represent partisan bias for all empirically likely values of V. If a summary differs from the value of partisan bias for other empirically reasonable values of V, then an electoral system judged to be fair by the summary can instead turn out to be biased in a real election. This pattern may even be intended by gerrymanderers who sometimes misjudge their likely average district vote and instead of having an electoral system biased in their favor, such as by winning a large number of districts by a small amount, they have one massively biased against them, by losing them all by a small amount.

For competitive electoral systems, (c) can be a reasonable summary if the values of V we are likely to observe are included in the specified range. In contrast, (a) is best used with another assumption because, even when $\beta(0.5) = 0$, $\beta(V)$ may be far from 0 for any other value of V. Summary (c) will normally be the most statistically stable of the three. These warnings do not mean that summaries should not be used, only that they come with an assumption that needs to be understood.

2.2.5 Types of Symmetry and Asymmetry

Partisan symmetry is a minimal and thus flexible standard of fairness which many different types of electoral systems satisfy. We first clarify the range of variation of symmetric electoral systems and then characterize types of biased electoral systems. We order electoral systems meeting the partisan symmetry standard by the size of the bonus going to the statewide majority vote winner or, in other words, by the degree of *electoral responsiveness*, of $S(V)$ to changes in votes V, as follows:

Definition 3. [Electoral Responsiveness] *Electoral responsiveness, which quantifies how much the statewide seat proportion is altered by a change in the average district vote, is* $\rho(V) = \partial S(V)/\partial V$.

Because the number of legislative seats is discrete, seats-votes curves are inherently discrete, and $\rho(V)$ is not uniformly continuous. Thus, in practice, the curve is summarized by smoothing via a discrete derivative $\rho(V, V') = [S(V) - S(V')]/(V' - V)$, given chosen values V and V'. We will use the shorthand $\rho(V)$ to refer to both the theoretical continuous quantity and the discrete estimator.

Electoral responsiveness is commonly summarized at (a) $\rho(0.5)$; (b) an empirically reasonable value such as $\rho(V^O)$, where V^O is the observed average district vote for a real election; or (c) an empirically reasonable range, such as $\rho(0.45, 0.55)$.

We first use Definition 3 to define a minimal standard for a fair democratic electoral system, which we call *symmetric democracy*:

Definition 4 (Symmetric Democracy)**.** *An electoral system characterized by symmetric democracy satisfies (a)* partisan symmetry *(Definition 1), (b)* nonnegative responsiveness*,* $\rho(V) \geq 0$ *for all* V*, and (c)* unanimity*,* $S(0) = 0$*.*

Conditions (a) and (c) imply also that $S(1) = 1$. Conditions (b) and (c) imply, for at least one point in $V \in [0, 1]$, that $\rho(V) > 0$. Condition (c) is referred to as "unanimity" or the "Pareto principle" in social choice theory (A. Sen, 1976). (We suggest a modification of condition (c) in Section 3.2 when one party is unlikely to ever win a majority of votes.)

Four ranges of electoral responsiveness that satisfy Definition 4 are often discussed, each of which we illustrate with a fair seats-votes curve in the left panel of Figure 1. First, *proportional representation* meets the partisan symmetric standard because $S(V) = V$ and $1 - S(1 - V) = V$, or in other words $\rho(V) = 1$ and $\beta(V) = 0$ for all V (green line in the figure). Legislatures with single member, plurality voting systems are not guaranteed to be proportional by law and tend to be *majoritarian* by empirical pattern, which means that they usually give a bonus to the party winning a majority of votes statewide, with $1 < \rho(V) < \infty$ (see blue line). For example, suppose the Democrats receive $V = 0.55$ proportion of the average district vote statewide and, because of how the district lines are drawn, receive $S(0.55) = 0.75$ proportion of the seats. This is not proportional, but it would be fair according to partisan symmetry if we knew the Republicans, if they had received $1 - V = 0.55$ proportion of the vote, would also receive $1 - S(1 - 0.55) = 0.75$ proportion of the seats. Third, a more extreme type of electoral system still meeting partisan symmetry is *winner-take-all* (with $\rho \to \infty$), where the majority vote winner receives all of the seats (solid black line in left panel of Figure 1). A final type of system that meets partisan symmetry is where the party winning a majority of votes receives a *negative bonus* ($0 < \rho < 1$), such as if $S(0.65) = 0.55$ and $1 - S(1 - 0.65) = 0.55$ (red line).

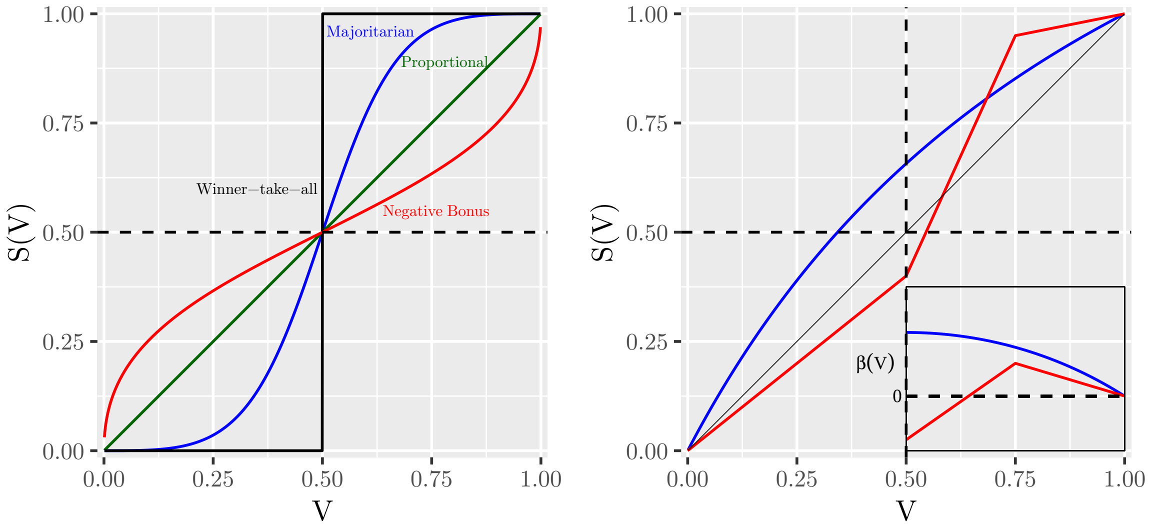

Figure 1: Types of Seats-Votes Curves. Left panel: Symmetric (fair) curves with differing levels of electoral responsiveness. Right panel: Asymmetric (biased) curves, including one consistently biased toward the Democrats (blue) and one with biases favoring different parties depending on V (red); the inset graph is for $\beta(V)$ for $V \in [0.5, 1]$ with the vertical axis scaled to be the same as the main plot, and lines color coded to the seats-votes curves.

Although partisan symmetry is widely viewed as a required standard of any minimally fair electoral system, different levels of electoral responsiveness may reasonably be chosen as preferable or appropriate for and by different people and governments. Many would prefer that their electoral system meet partisan symmetry but not be proportional, winner-take-all, or negative bonus, and so would impose the restrictions of an unbiased ($\beta(V) = 0$) majoritarian ($1 < \rho < \infty$) electoral system. Similarly, although no US state constitution rejects partisan symmetry, the constitutions differ in their requirements regarding electoral responsiveness. Some state constitutions require their redistricters to draw highly responsive districts, in order to encourage competitive elections and party change in office, whereas others encourage their redistricters to draw minimally responsive districts, which protects their incumbents, perhaps to help them gain experience or seniority and thus power on congressional committees. Brunell (2010) even argues that less responsiveness (and thus less competitiveness) produces happier constituents; see also (Gerber and Lewis, 2004, p.1378).

We also distinguish between two types of electoral systems that deviate from partisan symmetry — (a) those biased consistently in favor of one party and (b) those that switch from biased in favor of one party to the other as V changes. The right panel of Figure 1 gives one example of each of these seats-votes curves, in along with an inset graph at the lower right, with $\beta(V)$ plotted by V and color-coded to the corresponding seats-votes curve. The blue seats-votes curve is biased in favor of the Democratic party for every value of V, although by different amounts. We can see this by the corresponding blue line in the inset graph. For example, at $V = 0.5$, $S(V) = 0.66$, and so $\beta(0.5) = (0.66 - 0.5)/2 = 0.08$, which is also the height of the left end of the blue line on the inset graph (although numbers on the vertical axis of the inset graph have been removed to reduce clutter, distances from zero are the same as for the main graph). Whereas the blue $\beta(V)$ line in the inset graph is always above zero, indicating consistent bias toward the Democrats for all V, the red line indicates bias toward the Republicans for $V < 0.125$ and toward Democrats for larger average district vote values.

Partisan bias that switches parties with V is important to consider when using summary measures of bias to represent the entire seats-votes relationship. This type of seats-votes curve can also be the result of a gerrymandering strategy where the party in control draws district maps biased against it, at values of V it sees as unlikely, so long as the same map has more bias in its favor at values of V in future elections it sees as likely.

2.2.6 Seat- vs Vote-Denominated Partisan Bias

The seats-votes curve represents seats as function of votes, $S(V)$, reflecting how electoral systems work, with partisan bias *seat-denominated*. A simple case can be seen in the right panel of Figure 1 as the vertical distance from where the two dashed lines cross (at $S(V) = 0.5, V = 0.5$) to where the red line crosses the ($V = 0.5$) vertical dashed line. This vertical distance is $\beta(0.5) = -0.1$ — meaning that the Republicans receive 10 percentage points more seats than the Democrats with the same vote proportion.

Yet, deviations from the seats-votes curve can also be *votes-denominated* (McDonald, 2017). Instead of asking whether a party receives an unfair proportion of seats (more seats for the same vote proportion than the other party), we could instead ask whether the party must earn a larger average district vote than the other party in order to win a given seat proportion. A simple example is the horizontal distance in Figure 1) from where the two dashed lines cross (at $S(V) = 0.5, V = 0.5$) to where the red line crosses the ($S(V) = 0.5$) horizontal dashed line (see McGhee, 2017, Fig.2). This horizontal distance is $\text{VDB}(0.5) = 0.045$ — meaning that to obtain 50% of the seats, the Democrats must earn 4.5 percentage points more in votes than the Republicans. (The blue line in the right graph is an example where it happens that the vertical and horizontal distances are the same: $\beta(0.5) = \text{VDB}(0.5)$, in this case 0.08 seats and votes respectively.) Seat- and vote-denominated partisan biases are analogous to the difference between the usual causal quantity, e.g. "how much longer exercise twice a week causes a person to live," and the alternative quantity, e.g. "the number of days of exercise needed to cause a person to live one year longer".

Seats- and votes-denominated biases are different theoretical quantities, but both convey the degree to which an electoral system deviates from partisan symmetry. We formalize this intuition here. Thus, a symmetric electoral system can be represented in the usual seat-denominated way given in Definition 1, $S(V) = 1-S(1-V)$, or equivalently in this alternative vote-denominated way, with votes as a function of seats: $V(S) = 1-V(1-S)$, where $V(S)$ is the average district vote the Democratic party needs in order to receive S proportion of seats in the legislature. We can thus define *vote-denominated partisan bias* (in parallel to Definition 2) as a function of seats: $\text{VDB}(S) = -\{V(S)-[1-V(1-S)]\}/2$, with the leading negative sign because the Democrats are advantaged when $V(S)$ is smaller given any S and $S(V)$ is larger given any V.

3 Other Partisan Fairness Standards

We consider here alternatives to and modifications of the partisan symmetry standard by studying the effects of two variables that characterize every redistricting — the existence of partisan gerrymanderers and the competitiveness of the party system. We first show how the goals of partisan gerrymandering affects electoral systems in terms of bias and responsiveness, and how these can differ, depending on competitiveness, from the often misleading “cracking and packing” stereotype used in the literature (Section 3.1). We then show how a pure partisan gerrymandering perspective suggests alternative, but ultimately unsatisfactory, normative definitions of partisan fairness (Section 3.2). And finally, we consider standards of partisan fairness for noncompetitive party systems (Section 3.3).

3.1 Gerrymandering Goals

Consider an imaginary partisan gerrymanderer focused solely on advantaging their political party.[4] Partisan gerrymanderers use their knowledge of voter preferences and their ability to draw favorable redistricting plans to maximize their party’s seat share. Gerrymanderers do not necessarily care about voter support, the efficiency of the translation of votes into seats, partisan bias, electoral responsiveness, or differential turnout — unless it helps them win more seats.

We show that these goals, when mapped into the concepts of partisan bias and electoral responsiveness, can be either consistent with or the opposite of those commonly described in the literature. To show this, consider three situations, each of which leads to a different optimization function, effect on symmetry, and goal for bias and responsiveness (see Cox and Katz 1999, §3.3, Friedman and Holden 2008, and Puppe and Tasnadi 2009).

First is when the gerrymanderer is *running scared* (Mann, 1978) and so is worried about what the statewide vote U may be in future elections (V is not defined without districts). Here, optimizing means trying to win maximal sets with a safe margin, in order to insulate the party from potentially unfavorable future partisan swings. In this case, optimizing means seeking *high bias and low responsiveness*. Operationally, the gerrymanderer may do this by “packing” overwhelming numbers of opposition party votes into a few otherwise unwinnable districts and “cracking” the remaining opposition voting

[4] This person or entity is imaginary because in practice those actually in control of or involved in redistricting balance numerous other factors in addition to partisan gain. These other factors include optimizing or balancing the protection or pairing of specific incumbents, changing ideological polarization (McCarty, Poole, and Rosenthal, 2009) or the legislature’s median voter (Herron and Wiseman, 2008), maintaining or splitting communities of interest, changing district compactness, not splitting local political subdivisions, keeping an incumbent’s children’s schools or parents houses in or out of their districts, keeping good challengers’ homes out of certain districts, state legislators drawing congressional districts for them to run in, optimizing turnout differentials, swapping populations to hurt or encourage incumbents to retire, and many others (Hardy 1977, Owen and Grofman 1988, Cox and Katz 2002, p.39ff, and Yoshinaka and Murphy 2009).

strength across a large number of districts in order to win each by a small number of votes. High bias helps the party in control of redistricting and low responsiveness protects their incumbents by locking in these gains for future elections.

Second is the opposite situation where the gerrymanderer is *confident of a statewide majority* of votes and so tries to make each district a microcosm of the entire state (i.e., $v_d = V$ for all d), producing a winner-take-all outcome overall Cox and Katz (2002). In other words, the goal is an electoral system with *low bias and high responsiveness*. The "low bias" result is merely the consequence of optimizing primarily for high responsiveness, without preparing for in the situation where $V = 1 - V^O$, since they do not think it will happen. This situation involves neither packing nor cracking: If a Democratic gerrymanderer thinks his or her party can count on a statewide vote of $U = 0.55$, then packing to give the Republicans a few seats is out of the question and cracking, to win any seats by 50% plus a few votes, is irrelevant. Instead, the goal would be to win with $v_d = 0.55$ for all d. (Of course, if the gerrymanderer turns out to be overconfident and wrong about the partisan swing, optimizing in this way may cause their party to lose all the districts.)

Also worth mentioning is where a partisan gerrymanderer must reach agreement with the other party. The result is a *bipartisan gerrymander*, which winds up optimizing for *low bias and low responsiveness*. Bias would be low because it is a zero-sum compromise between the parties, and low responsiveness reduces uncertainty in future elections by locking in the deal and protecting incumbents in both parties.

3.2 Gerrymandering-Based Fairness Standards

We offer here two ways of deriving a normative standard of partisan fairness from a purely partisan gerrymandering perspective. First, consider (as a thought experiment since implementation may be infeasible) letting the same person or group control redistricting but preventing them from using knowledge of where their party's supporters live. This idea, which is equivalent to randomly permuting party labels on voters or on the gerrymanderer's voter forecasts, clearly removes *intent* to do harm. This step alone may be of value, since human psychology and most judicial systems judge intentional harm more severely than accidental harm (Greene, 2009). However, since plans drawn without knowledge of party support are drawn randomly with respect to party, any plan can be selected regardless of the degree of bias, responsiveness, or any other feature. In other words, gerrymandering without knowledge of party removes intent but does not remove harm. In fact, one possible districting plan that can occur is the identical plan that would be drawn by a partisan gerrymanderer with full knowledge of where its party's voters live. At the end of the day, the absence of intentional unfairness is not the same as fairness.

Second, we compare the efficiency of each party's translation of votes into seats. In one observed election, the Democratic party receives $S(V^O)$ seats given V^O votes and the Republican party receives $1 - V^O$ votes and $1 - S(V^O)$ seats. Which of these parties has a better or more efficient translation of seats into votes? Unless it happens that $V^O = 1 - V^O = 0.5$, this is an apples to oranges comparison because of the two different starting points. The only way to make the vote comparison between the two parties in any one election meaningful is by imposing a counterfactual assumption. We consider two possibilities for this assumption.

In one, we could make an assumption that enables us to estimate what would happen if the parties switched their vote proportions, so that the election result was $1 - V^O$ rather than V^O (we describe these assumptions in Section 4). Then, we would be able to estimate the unobserved seat proportion $1 - S(1 - V^O)$ and compare it to $S(V^O)$. This of course leads exactly to partisan symmetry.

In the other, we could try assuming away the differential meaning of all, or some particular type of, votes cast for each party (e.g., "wasted votes," which are those cast for losing party in a district or above 0.5 plus one vote in winning districts; see Section 5.6). However, although all votes are observed, asserting that all or any subset has the same meaning for each party, when the parties have different expected vote proportions, requires an assumption with the same ontological status as assumptions imagining partisan swings that lead to partisan symmetry. For example, suppose the Democrats receive $V^O = 0.6$ and are confident of a statewide majority in subsequent elections under the same redistricting plan. Then, the votes cast for each party in specific districts (and the resulting characteristics of the electoral system like bias and responsiveness) have markedly different meanings for Democrats than for Republicans now in the minority, with $1 - V^O = 0.4$ votes. The Democrats in this scenario would benefit by having votes distributed so that each district is a microcosm of the state, but Republicans would benefit most by packing and cracking (see Section 3.1), and so assuming that these votes have the same meaning would be a stretch at best. This does not seem like a promising direction for developing a new standard for partisan fairness.

3.3 Noncompetitive Party System Fairness Standards

We address here standards of fairness for electoral systems when one party has an overwhelming majority and is likely to keep it. In this situation, the partisan symmetry promise to a minority party of eventually receiving a controlling seat proportion, when in a future election the party has more voter support, seems empty. Put in the context of our framework, when the rotation in office assumption (Assumption 2) does not hold, questions about the partisan symmetry standard may be meaningless. When Assumption 2 does hold, but counterfactual estimation is highly uncertain or model dependent, the questions are coherent but efforts to determine the answer may be fruitless.

Fortunately, the political science literature on constitutional design for ethnically or racially divided societies can be used to define standards of fairness composed of the basic concepts introduced in this paper. Thus, to protect minority parties, and to prevent them viewing the electoral system as illegitimate, political scientists advise adding constitutionally mandated power sharing to electoral rules (Lijphart, 2004). Exactly how much protection and in what form can be derived from first principles, but this precision often comes at the price of model dependence (King, Bruce, and Gelman, 1996). Yet, since the direction needed is clear, we describe two specific ways improving the situation.

First, we could require redistricters to follow a strategy opposite to that of a partisan gerrymanderer confident of a statewide majority (see Section 3.1). Thus, instead of creating each district as a microcosm of the state, and giving the majority a winner-take-all victory, we would pack minority party voters into a small number districts and thus ensure them at least some seats. This is indeed what happens with protected racial minorities in US legislatures covered by the Voting Rights Act. The way to do this within our framework is to require *low levels of electoral responsiveness*, which thus makes it more difficult for the majority party to wipe out the minority. This requires, at a minimum, particularly low levels of $\rho(V)$ for V near V^O.

Second, we can adapt an alternative and surprisingly common approach to mandated power sharing in constitutional design — formally reserving legislative seats for the minority party to guarantee that their views will at least be heard in the legislature (Reynolds, 2005). In this case, we can restate the symmetric democracy standard in Definition 4 by replacing the unanimity condition (c) with a minority protection provision:

Definition 5 (Symmetric Democracy with Minority Party Protection)**.** *An electoral system characterized by symmetric democracy with minority party protection satisfies (a) partisan symmetry (Definition 1), (b) nonnegative responsiveness, $\rho(V) \geq 0$ for all V, and (c) minority protection, $S(V) = c > 0$ for $V \leq \tau \ll 0.5$, where τ is the protection vote threshold for a political party and c is the party's guaranteed seat proportion.*

Conditions (b) and (c) ensure that $S(V)$ is monotonically increasing over its entire range.

4 Assumptions for Estimating Seats-Votes Curves

We show here, under different types of assumptions, how to estimate the full seats-votes curve, from which we can easily compute partisan bias, electoral responsiveness, or other electoral system features. We begin with values of the curve that can be ascertained without assumptions and then discuss estimation under functional form assumptions using statewide averages, partisan swing assumptions using district-level data, and forecasting assumptions when no elections under the redistricting plan in question have been held. We conclude with a brief discussion of how models of individual voters.

4.1 No Additional Assumptions

In what is usually the best case, where we have five elections occurring between the decennial censuses and thus which we could consider (close to being) under the same electoral system, we observe five data points $\{\hat{S}(V_t^O), V_t^O : t = 1, \ldots, 5\}$, where the observed statewide seat proportion $\hat{S}(V_t^O)$ is an estimate of the expected value $S(V_t^O)$ in election t.

From these data, two unusual circumstances enable us to compute a summary measure of partisan bias with no modeling assumptions. In the first, if we happen to observe an election with a tied average district vote, $V^O = 0.5$, then one quantity of interest, $\beta(0.5)$ is estimated simply by the observed seat proportion. In the second, which is an even luckier situation (encompassing the first), two elections are observed under the same electoral system and happen to have average district vote proportions symmetric around 0.5. For example, in Wisconsin State House elections run under the same redistricting plan, the average district vote was approximately $V^O = 1 - 0.48$ in 2012 and $V^O = 0.48$ in 2014 and where, as a result, statewide seat proportions were observed in each election. In this particular case, the results indicate severe bias favoring the Republicans because of the dramatic seat proportion differences: $1 - \hat{S}(1 - 0.48) = 0.6$ but $\hat{S}(0.48) = 0.36$ (approximately), and so $\hat{\beta}(0.48) = -0.12$. (This election was the subject of the Supreme Court case, *Gill v Whitford*, 585 U.S. (2018).)

4.2 Functional Form Assumptions

To estimate the entire seats-votes curve without more data requires assumptions. One type of assumption is to specify a class of parametric functional forms for the seats-votes relationship and to estimate the parameters of that form with (usually up to about) five data points. Two examples of this form are *linear* (Tufte, 1973),

$$S(V) = \alpha_0 + \alpha_1 V, \tag{3}$$

and (reusing parameters α_0 and α_1) *bilogit*, (King and Browning, 1987):

$$S(V) = \frac{1}{1 + \exp\left[-\alpha_0 - \alpha_1 \ln\left(\frac{V}{1-V}\right)\right]}, \tag{4}$$

In each equation, α_0 and α_1 are related in different ways to partisan bias and electoral responsiveness, respectively (and since $S(V)$ in each expression is an expected value, real data need not fit either form exactly). For example, we drew the fair seats-votes curves in Figure 1 with $\alpha_0 = 0$ for all four and $\alpha_1 = \{0.5, 1, 3, 10, 000\}$ (10,000 being a sufficiently close approximation, for our figure, to winner-take-all, which is $\alpha_1 \to \infty$).

Once we estimate the seats-votes curve, we can then read off the point estimate of $S(V)$ (along with its uncertainty) given any chosen V. This method works well, and enables one to compute partisan bias or any quantity of interest from the resulting estimated curve, along the appropriate level of uncertainty.

Unfortunately, the few available observations from one redistricting plan means that the result is often quite uncertain and model dependent (and nonparametric approaches are not reasonable options). As such, this strategy tends to be used more often for academic study of broad patterns across many electoral systems than for practical use evaluating individual redistrictings shortly after or before they take effect.

4.3 Partisan Swing Assumptions

An alternative approach is to use as inputs the set of district-level vote proportions in at least one election held under the redistricting plan of interest. From this, we can compute a easily estimate a single point on the seats-votes curve, $S(V^O)$, at the observed statewide vote V^O. To estimate other points, we need an assumption to generate other hypothetical elections from the same electoral system, for different points V.

To develop an assumption we note that patterns in electoral data throughout the US and most parts of the world can be decomposed into (a) the absolute average partisan swing from one election to the next that tends to affect almost all districts and (b) the relative positions of district votes within any one election. The relative district vote positions tend to remain highly stable over time and so are quite predictable, whereas the statewide swings over time are more volatile and harder to predict. Fortunately, the relative positions are more important for evaluating redistricting than the absolute swings.

A simple and remarkably accurate assumption that identifies $S(V)$ for any V is *uniform partisan swing*:

Assumption 3. [Uniform Partisan Swing (Butler, 1951)] *When the average district vote swings between elections under the same electoral system from V to V', every district vote proportion moves uniformly by $\delta \equiv V' - V$, so that $\{v_1, \ldots, v_L\}$ from one election becomes $\{v_1 + \delta, \ldots, v_L + \delta\}$ in the next (with elements truncated to [0,1] if necessary).*

Given Assumption 3, we can use the observed district-level votes from one election, $\{v_1, \ldots, v_L\}$, and a chosen swing δ to estimate the seat proportion in the new election under the same electoral system: $\hat{S}(V + \delta) = \text{mean}_d[s(v_d + \delta)]$, which is single-valued.

We also study the empirical accuracy of estimates of the seats-votes function under uniform partisan swing. Our quantity of interest here is the out-of-sample error rate for the statewide seat proportion using uniform partisan swing for one election that is identical in all respects to the previous one — including candidates, the campaign, spending, weather on election day, patterns of incumbency, etc. — except for the statewide partisan swing and the usual random uncertainties in voter preferences. Finding pairs of observed elections like these is obviously impossible, and so we instead use successive elections within the same redistricting regime. The consequence of this decision is that our estimated out-of-sample error rate is an upper bound on the actual errors of uniform partisan swing-based predictions. To be specific, we begin with all data from all regular elections to the lower house and state sentate in US state legislatures 1968–2016. We narrow these to the 646 elections for legislatures with all single member districts, at least 20 districts, with at least half the seats contested, and where no redistricting has occurred between this election and the one before.[5]

Thus, for each of 646 elections, we use the district-level vote proportions in election 1, the statewide swing to election 2, $\delta = V_2^O - V_1^O$, and the uniform partisan swing assumption to predict the expected statewide seat proportion for election 2, $S_2(V_2^O)$. We do not observe this expected value and so use the observed election 2 seat share $\hat{S}_2(V_2^O)$ (as a model-free estimate of the expected value) for validation. The error metric for the prediction $\hat{S}_1(V_2^O)$ is then simply $\hat{S}_2(V_2^O) - \hat{S}_1(V_2^O)$.

The left panel of Figure 2 gives a histogram of these out-of-sample prediction errors from uniform partisan swing. As expected, results reveal highly accurate predictions, with a median error of 0.0000, a mean error of -0.001 (one tenth of one percentage point), and an interquartile range of only $[-0.025, 0.021]$. And recall that these numbers are *upper* bounds.

[5] Following Gelman and King (1994b), we impute uncontested districts at 0.75 for Democratic wins and 0.25 for Republican wins, although this has no material impact on our results.

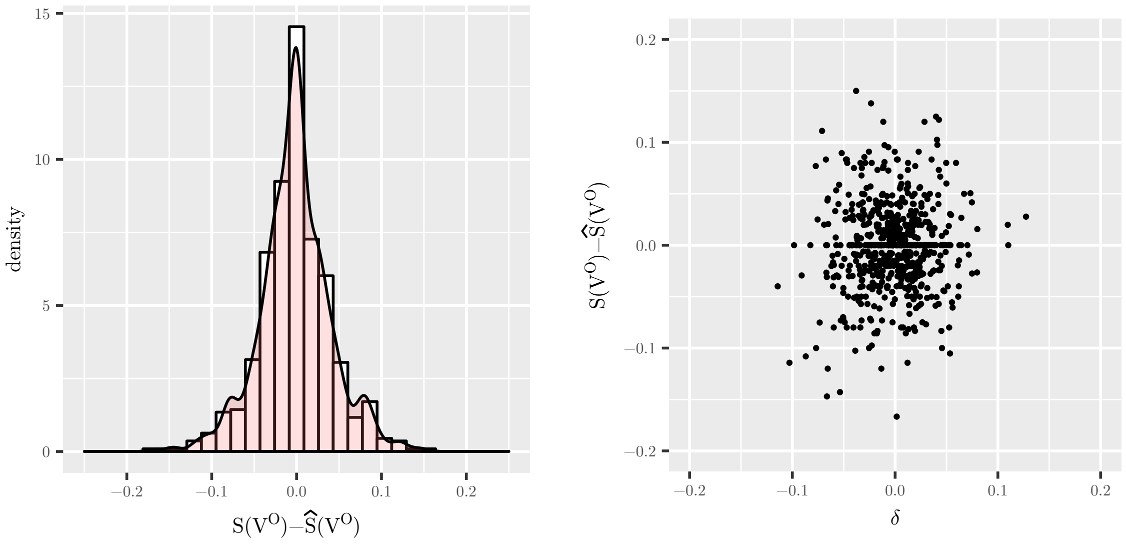

Figure 2: Error rates of out-of-sample statewide seat proportion predictions based on the assumption of uniform partisan swing: histogram of errors (left panel) and errors by statewide partisan swing, δ, from the observed election to the next where we are predicting the seat proportion.

We also use these data to study how fast uniform partisan swing-based predictions degrade as we extrapolate farther from the original data, that is when the statewide vote swing is larger. Thus, the right panel of Figure 2 plots the prediction error by the size of the statewide vote swing, δ. Remarkably, the graph shows that predictions do not seem to degrade at all for larger swings (i.e., as δ deviates from zero). The implication is that uniform partisan swing is a relatively fixed feature of elections, with the more difficult-to-predict component mostly relegated to statewide voter swings, which happen not to be important for studying partisan symmetry.

In our data, and in many elections all over the world, the uniform partisan swing assumption is a reasonable first approximation, especially for theoretical purposes like ours. What Assumption 3 ignores is that the world is stochastic and so is less useful for some empirical purposes. The simplicity can also generate inefficiency due in part to discreteness (Nagle, 2015, p.351). We can thus generalize the deterministic uniform partisan swing assumption either directly via stochastic modeling (King, 1989) or statistical modeling:

Assumption 4. [Stochastic Uniform Partisan Swing (Gelman and King, 1994a)] *Hypo-*

thetical (denoted "(hyp)") district-level vote proportions, under the same electoral system, are generated as

$$v_d^{(hyp)} = X_d\theta + \gamma_d + \delta^{(hyp)} + \epsilon_d^{(hyp)}, \tag{5}$$

where X_d is a vector of covariates describing the districts, candidates, voters, and lagged vote; θ is a vector of effect parameters; γ_d is an independent random normal district effect that is constant over hypothetical elections but varying over districts; $\delta^{(hyp)}$ is the researcher-chosen uniform swing; and $\epsilon_d^{(hyp)}$ is a stochastic normal error term, independent of γ and over d.

Assumption 3 is a special case of Assumption 4 and is thus less restrictive, more realistic, and more statistically efficient, and so should be used whenever it makes a difference, such as in most substantive or applied work. For our theoretical and methodological purposes, we will usually use the simpler Assumption 3 in this paper to ease exposition, in part because we analyze so many districts that inefficiency is a minor issue and because relevant empirical patterns of voter behavior are extremely regular across most elections in most countries (King, Rosen, Tanner, and Wagner, 2008, p.952).

4.4 Forecasting Assumptions

Political scientists, redistricters, legislators, and those involved in redistricting litigation are often in the position of having to evaluate one or more redistricting plans before any elections have been held under the plan. To do this, the underlying data are forecast at the precinct level, the lowest level at which electoral data are observed, and aggregated into the new districts. Fortunately, the relative positions of the districts are the most important and also the most predictable, and so these are what we focus on.

The creation of these forecasts typically involves two steps. First, influential district-level variables measuring candidate characteristics, such as incumbency advantage and uncontestedness, are corrected for. This is typically done by estimating the effects of these variables in a simple district-level analysis (such as by estimating θ in Equation 5) and then subtracting them out from the raw precinct-level variables. And second, several years of these corrected precinct-level variables are forecast, typically using simple autoregressive models, which are quite accurate. After aggregating, the methods in Section 4.3 can be used directly.

4.5 Models of Individual Voters

Throughout most of the literature, researchers condition on the district vote proportions and treat them as fixed quantities. Understanding the motivations of voters that give rise to these vote proportions, and building the models useful for understanding them, turn out to be unnecessary to the definition, standard, or measures of partisan symmetry. However, the aggregate patterns, such as (stochastic) uniform partisan swing, are so stable and predictable over time and across jurisdictions that they ought to be of use for building models of individual voters and their motivations and, at the same time, verified models of individual voters may well turn out to further inform the study of fairness in district-level democracies. Further research in these areas is surely warranted. See Ansolabehere and Leblanc (2008), Ansolabehere, Leblanc, and Snyder (2012), Coate and Knight (2007), and Cox and Katz (1999).

5 Evaluating Fairness Measures

We now evaluate several measures of partisan fairness in district-based electoral systems. For each, we identify the corresponding estimand and implied notion of fairness.

5.1 Estimation from Seats-Votes Curves

A straightforward way to estimate a feature of the seats-votes curve is to begin with an estimate of the entire curve, using one of the assumptions in Section 4. With the full curve, partisan bias $\beta(V)$, electoral responsiveness $\rho(V)$, and other quantities are easy to compute appropriately for any relevant $V \in [0,1]$, ensuring that Assumptions 1 and 2 hold. A few of the important articles computing bias and responsiveness in this way include Brunell (1999) and Jackman (1994), which use Assumption 4; Erikson (1972), using the functional form assumption in Equation 3; Gilligan and Matsusaka (1999) and Niemi and Jackman (1991), using the functional form assumption in Equation 4; and

Brady and Grofman (1991) and Garand and Parent (1991), which uses a combination of the functional form assumption in Equation 4 along with Assumption 3.

As an illustration, we estimate partisan bias and electoral responsiveness using data from 963 legislatures (those 1968–2016 with all single-member districts, at least 20 seats, and at least half of the seats are contested) via uniform partisan swing (Assumption 3). Figure 3 plots bias vertically by responsiveness horizontally, both for $V \in [0.45, 0.55]$. The scatterplot shows that bulk of bias results is in $[-0.1, 0.1]$ and responsiveness is in $[1, 3]$. The two quantities are uncorrelated in these data, but not independent in that as ρ increases, $|\beta|$ declines. This pattern is consistent with the scenario from Section 3.1 where the redistricter is confident of a statewide majority and so seeks high responsiveness and is left with low bias.

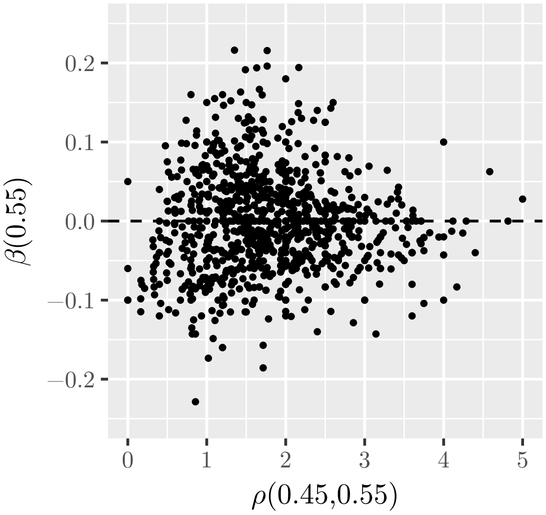

Figure 3: Estimates of Partisan Bias and Electoral Responsiveness, both evaluated in the range $V \in [0.45, 0.55]$ in 963 Legislatures

5.2 Proportional Representation

We now discuss several individual measures that do not first estimate the entire seats-votes curve. For expository reasons, we begin with the simple *deviation from proportional representation* measure, $\text{PRD}(V^O) = S(V^O) - V^O$, which is easy to understand and turns out to fail as a measure of partisan fairness. We explain why and then show how it is useful for other purposes.

Under this approach's standard of fairness, $\text{PRD}(V) = 0$, we first solve for expected seats as a function of votes, $S(V^O) = V^O$, and then swap V^O with V, yielding, $S(V) = V$. This is a coherent seats-votes curve because each value of V produces one value of $S(V)$, if we also add Assumptions 1 and 2. This curve appears as the green line in Figure 1, a symmetric electoral system with $\rho(V) = 1$, for all V. By writing partisan bias as $\beta(V) = \{S(V) - [1 - S(1 - V)]\}/2 = \{\text{PRD}(V) + \text{PRD}(1 - V)\}/2$, we can see that the proportional representation standard is a special case of partisan symmetry (because $\text{PRD}(V) = \text{PRD}(1 - V) = 0$ implies $\beta(V) = 0$ for all V), but partisan symmetry is not a special case of proportional representation (because $\beta(V) = 0$ whenever $\text{PRD}(V) = \text{PRD}(1 - V)$, even if $\text{PRD}(V) \neq 0$; see the other lines in Figure 1).

Although the proportional representation standard of $\text{PRD}(V^O) = 0$ is theoretically coherent, $\text{PRD}(V^O)$ is inadequate as a measure of partisan symmetry. The problem is that V^O and $S(V^O)$ produce only a single model-free estimate of a point on the seats-votes curve, which is insufficient for estimating the entire curve or partisan symmetry, because the second term in $\beta(V^O) = [\text{PRD}(V^O) + \text{PRD}(1 - V^O)]/2$ is unobserved without further assumptions. For example, the election outcome $S(0.6) = 0.6$ is consistent with the proportional representation standard because it falls on the line $S(V) = V$, which is proportional and symmetric. However, the same observed point is also consistent with the flat line $S(V) = 0.6$ (for all V) or with $1 - S(1 - 0.6) = 0$, neither of which are proportional, symmetric, or fair.

Although the deviation from proportional representation in one election is thus not a general measure of partisan symmetry, it can be informative about its more specific standard ($\beta(V) = 0$ such that $\rho(V) = 1$): Although $\text{PRD}(V^O) = 0$ offers no information one way or the other, $\text{PRD}(V^O) \neq 0$ implies that this specific standard should be rejected. This may be useful on its own, but to go further requires estimating relevant points on the seats-votes curve via the assumptions from Section 4.

However, even when completely uninformative about partisan symmetry, $\text{PRD}(V^O)$ is still an interesting and politically relevant summary of the outcome of an election. Certainly, small minority parties will want to know whether they receive at least some seats and so may compare it to their vote proportion as at least a benchmark. A forecast of $\mathrm{PRD}(V^O)$ would likely influence whether a small minority party would even compete in many districts or be likely to attract significant campaign contributions.[6]

5.3 Mean-Median

The *mean-median* measure summarizes skewness in the distribution of district votes via an easy-to-calculate difference: $\textsc{mm} = V^O - M$, where V^O is the average district vote and M is the median district vote, implicitly defined as $\frac{1}{L}\sum_{v_d>M} 1 = \frac{1}{2}$. Fairness according to this measure is when $\textsc{mm} = 0$. The measure is claimed "to reliably assess [partisan] asymmetry in state-level districting schemes" (Wang, 2016a, p.367). Essentially the same claim appears in Wang (2016b), Krasno, Magleby, McDonald, Donahue, and Best (2018), and McDonald and Best (2015), among others. Although no proof of this claim has appeared in the literature, we show that it is correct in different ways for two distinct theoretical quantities, vote- and seat-denominated partisan bias. In the first, we prove that $\textsc{mm}$ provides important but limited information about $\beta(V)$; in the second, we show that $\textsc{mm}$ is a more generally useful measure of the alternative theoretical quantity $\mathrm{VDB}(S)$ from Section 2.2.6.

A Limited Measure of Seat-Denominated Partisan Bias. We begin by showing, under Assumptions 1, 2, and 3, that $\textsc{mm} = 0$ if and only if $\hat{\beta}(0.5) = 0$. Formally,

$$\begin{aligned}
\hat{\beta}(0.5) &= \hat{S}(0.5) - 0.5 \quad \text{(by definition)} && (6)\\
&= \frac{\sum_{v_d>V^O} 1}{L} - 0.5 \quad \text{(Assumption 3)} \\
&= \frac{\sum_{v_d>M} 1}{L} - 0.5 \quad \text{(Assuming MM=0)} && (7)\\
&= 0.5 - 0.5 = 0 && \square
\end{aligned}$$

As an estimate of $\beta(0.5)$, the mean-median measure has two limitations (in addition

[6] In practice, single member district, first-past-the-post electoral systems rarely turn out to be approximately proportional (see Figure 3). Paradoxically, even many electoral systems that impose proportional representation at the statewide level wind up with considerable asymmetry, given how they are applied in complicated multiparty contexts (see Grofman and King, 2007, fn.37). In the US, the Supreme Court has been explicit: "the Constitution provides no right to proportional representation" as a standard for American elections (*Vieth v. Jubilerer*, 541 U.S. 267 (2004)).

to the effects of discreteness; Nagle 2015). First, although our proof shows that MM is a useful indicator for whether $\beta(0.5)$ is zero, and so could be used for a hypothesis test, it is not a general measure of $\beta(0.5)$, as we have no proof that it is an unbiased or consistent estimator since the magnitude is not known to be correct when other than zero.

Second, if the electoral system is biased at a point other than $V = 0.5$, the mean-median measure will not necessarily reflect overall partisan symmetry (see right panel, Figure 1). Consider an election with 10 districts and the following vote proportions:

$$\{.48, .49, .49, .49, .59, .61, .65, .65, .65, .90\}.$$

From these data, and the assumptions above, MM $= \hat{\beta}(0.5) = 0$, which would enables us to conclude that an aspect of the electoral system is fair. However, without any additional assumptions, we can show that in fact other aspects of the electoral system can be biased. For example, if the Democratic party receives an average district vote of $V^O = 0.6$ (the observed value of these district proportions), they would win a $\hat{S}(V^O) = 0.6$ seat proportion but, when the Republicans receive the same $1 - V^O = 0.6$, they would win a remarkable $1 - \hat{S}(1 - V^O) = 0.9$ of the seats, which is a 30 percentage point difference. This means that $\hat{\beta}(0.6) = -0.15$. In this example, the mean-median measure indicates that the electoral system represented is fair, but it is instead quite unfair. How often $\beta(0.5)$ differs from other values of $\beta(V)$ is an empirical question that is worth further study and not necessarily a problem for the mean-median measure, which is still a valid measure of one important quantity of interest, $\beta(0.5)$.

A Better Measure of Vote-Denominated Partisan Bias. More interestingly, we can prove that MM is a valid estimator of the quantity VDB(0.5). With Assumptions 1, 2, and 3, we have:

$$\begin{aligned} \text{VDB}(0.5) &= -\{V(S) - [1 - V(1-S)]\}/2 \quad \text{(by definition)} \\ &= 0.5 - V(0.5) \\ &= V - M \quad \text{(Assumption 3)} \\ &\equiv \text{MM} \qquad \square \end{aligned}$$

If we use the more realistic Assumption 4 in place of Assumption 3, it is easy to show that MM is a statistically consistent estimator of VDB(0.5). Either way, this proof is important because, although the *magnitude* of MM has no clear relationship to $\beta(0.5)$, it is correct for the alternative quantity VDB(0.5), making the mean-median measure an easy-to-calculate and accurate estimator of this unusual but still coherent theoretical quantity.

5.4 Lopsided Outcomes

Wang introduces the *lopsided outcomes test* and claims it can be used "to reliably assess [partisan] asymmetry in state-level districting schemes," (Wang, 2016b, p.1263), or "to detect..." or "identify partisan asymmetry" (Wang, 2016a, p.368). We show here that this claim is false, and along the way describe other more productive uses of the measure.

Begin by denoting the average Democratic vote in Democratic-won districts as $D = \sum_d s(v_d)v_d / \sum_d s(v_d)$ and the average Democratic vote in Republican-won districts as $R = \sum_d [1 - s(v_d)]v_d / \sum_d [1 - s(v_d)]$ which implies $R < 0.5 < D$. Then we write an accounting identity with the average district vote as a weighted average of Democratic and Republican seat shares, $V = S(V^O)D + [1 - S(V^O)]R$, and solve for the generic seats-votes relationship, all without assumptions:

$$S(V^O) = \frac{V^O - R}{D - R}. \tag{8}$$

Equation 8 is true by definition, but swapping V^O for V is not sufficient to define a coherent seats-votes curve because $S(V)$ is not single-valued. The presence of R and D on the right side, which are functions of V, means that we need more constraints to ensure $S(V)$ has only one value. Only at that point can we add Assumption 2 and evaluate the claim that the resulting seats-votes curve meets the partisan symmetry standard. So the question for any measure is whether it imposes these sufficient constraints.

The lopsided outcomes measure is defined as the simple party difference in the average win size:

$$\text{LO} = D - (1 - R). \tag{9}$$

This measure seems intuitive because packed districts is sometimes a characteristic of successful partisan gerrymandering but, as Section 3.1 shows, the intuition is often wrong because packing (and cracking) can be counterproductive. The measure deserves credit as a fine measure of the skewness of the vote proportion distribution, since nonskewness implies that the center of mass on either side of $v_d = 0.5$ will be equidistant from this midpoint. A forecast of this measure may indeed be useful to partisans or others trying to understand the competitive playing field, and what it takes on average to win a district for their party.

Unfortunately, lopsided outcomes is not necessarily related to partisan symmetry. To show this, we now study how Equation 9 is constrained by the measure's notion of fairness, LO $= 0$. Thus, by substituting $D = 1 - R$ into Equation 8, we have:

$$S(V^O) = \frac{V^O - R}{1 - 2R}. \tag{10}$$

Unfortunately, after substituting V^O with V, we are still left with multiple values of $S(V)$ for any one V, because of the presence of R on the right side which indicates the lack of a coherent seats-votes curve. This means that the lopsided outcomes test, and the implied *set* of multiple seats-votes curves it considers "fair", can be consistent with either symmetry or asymmetry. As a result, LO does not imply particular values of $\beta(V)$ and is not a measure of (the deviation from) partisan symmetry.

Thus, to construct an example, we add information the framework omits in the form of Assumption 3. We then construct examples of votes from three hypothetical legislatures:

$$\{.15, .15, .15, .65, .65, .65, .65, .65, .65, .65, .50, .70\} \tag{11}$$

$$\{.18, .18, .28, .43, .43, .53, .53, .78, .78, .88., .50, .50\} \tag{12}$$

$$\{.40, .40, .40, .40, .40, .60, .60, .60, .60, .60., .50, .50\} \tag{13}$$

and then compute partisan bias and LO for each. The inconsistency is apparent: Legislature (11) is judged fair by the lopsided outcomes test but is in fact asymmetric (LO $= 0$, $\hat{\beta}(V^O) = \hat{\beta}(0.5) = 0.2$). Legislature (12) is judged unfair by lopsided outcomes but is in fact symmetric (LO $= 0.08$, $\hat{\beta}(V^O) = 0$). And Legislature (13) is also judged fair by lopsided outcomes and is in fact symmetric (LO $= 0$, $\hat{\beta}(V^O) = 0$).

5.5 Declination

Warrington (2018a, p.2) introduced a measure called *declination* and claims it "is a measure of partisan symmetry" (or "a new measure of partisan asymmetry"; Warrington 2018b). We prove that this claim is incorrect, but along the way convey the measure's intuition and potential descriptive uses.

Warrington found a clever geometric interpretation of his measure, intuitive from the perspective of his field of mathematics, by defining it as $\text{DECLINATION} = 2(\theta_D - \theta_R)/\pi$, where $\theta_D = \arctan[(2D-1)/S(V^O)]$ and $\theta_R = \arctan\{(1-2R)/[1-S(V^O)]\}$. For our intended audiences (mostly in social science and statistics), the measure is easier to understand without the arctan transformation or constant normalizations, which only adjust the scale. We thus define the un-normalized declination,

$$\text{DEC} = \frac{D - 0.5}{S(V^O)} - \frac{0.5 - R}{1 - S(V^O)}. \tag{14}$$

Equation 14 (which can be thought of as a normalized version of lopsided outcomes; cf. Equation 9) is similar to the difference in the magnitude of electoral responsiveness on each side of $V = 0.5$. This is important because under partisan symmetry the difference in (actual) responsiveness is zero. The problem is that responsiveness is a change in votes divided by a change in seats (see Definition 3), whereas each of the two terms in DEC is a change in votes divided by an absolute seat proportion. That means that DEC is not an unbiased measure of partisan symmetry, but is closely related and also serves as another measure of the skewness of the distribution of district vote proportions.

To formally prove the connection between declination and partisan symmetry, we consider how its notion of fairness, $\text{DEC} = 0$, constrains the generic Equation 8. The result is:

$$S(V^O) = \frac{D - 0.5}{2D - 0.5 - V^O}. \tag{15}$$

As with lopsided outcomes, even after swapping V^O for V, $S(V)$ is not a single-valued function of V and so, even under its notion of "fairness", is not a coherent seats-votes curve. That is, we can try to adjust V to see how $S(V)$ changes, but how D changes with V is left unspecified which leaves many possible values of $S(V)$. The proposed standard is thus consistent with both symmetric and asymmetric seats-votes curves and, as such, declination not a measure of partisan symmetry.

In parallel to Section 5.4, we now offer examples of these inconsistencies with three hypothetical 10-district legislatures:

$$\{.30, .30, .40, .40, .40, .45, .65, .65, .65, .80, .50, .40\} \tag{16}$$

$$\{.33, .33, .48, .48, .48, .53, .53, .56, .63, .63, .50, .50\} \tag{17}$$

$$\{.33, .33, .48, .48, .48, .53, .53, .56, .63, .63, .50, .50\} \tag{18}$$

As in Section 5.4, we add missing information in the form of Assumption 3 and compute partisan bias and DEC. We find that Legislature (16) is judged fair by declination but is in fact asymmetric (DEC $= 0$, $\hat{\beta}(0.5) = -0.1$, $\hat{\beta}(V^O) = -0.05$). Legislature (17) is judged unfair by declination but is in fact symmetric (DEC $= -0.36$, $\hat{\beta}(0.5) = \hat{\beta}(V^O) = 0$). And Legislature (18) is also judged fair by declination and is symmetric (DEC $= 0$, $\hat{\beta}(0.5) = \hat{\beta}(V^O) = 0$).

5.6 Efficiency Gap

Stephanopoulos and McGhee (2015) introduce the *efficiency gap* and claim it is "a new measure of partisan [a]symmetry" (quote repeated on pages 831, 834, 838, 849, and 899). We prove that this claim is false, and also convey the intuition and productive uses of the measure.

The efficiency gap redefines the classic definition of "wasted votes" from all votes cast for losing candidates (Campbell, 1996) to votes for losing candidates plus those for winning candidates above the 50%-plus-one-vote threshold. The article then claims that partisan symmetry is satisfied when these wasted votes are equally divided between the parties. We show that this claim is incorrect. Although the efficiency gap is controversial (Chambers, Miller, and Sobel, 2017; Cho, 2017; Tapp, 2018), it comes with important intuition worthy of further study and the authors deserve substantial credit for bringing many, including the U.S. Supreme Court, back to this venerable field (see *Gill v. Whitford*, 585 US (2018); see Stephanopoulos and McGhee 2018).

The intuition for the efficiency gap works best in highly competitive situations, when one party is in control of redistricting and running scared (Section 3.1). Here, redistricters will often try to pack and crack and thus often reduce wasted votes. In other situations, such as when confident of a statewide vote majority, packing is against the redistricter's interests. At this point, the efficiency gap becomes confused; for example, if a party receives 80% of the votes and *all* the seats, the measure indicates that the electoral system treats it unfairly (see also Veomett, 2018).

To formalize, denote the proportion of wasted votes in district d for Democrats as $w_d = v_d - s(v_d)/2 \in [0, 0.5]$ and Republicans as $(0.5 - w_d) = (1 - v_d) - [1 - s(v_d)]/2$. Then define the efficiency gap as:

$$\text{EG}(V^O) = \frac{\sum_d n_d(0.5 - w_d) - \sum_d n_d w_d}{\sum_d n_d} \tag{19}$$

$$= S(V^O) - 2V^O + 0.5 - C, \qquad \text{where } C = 2\frac{\sum_d t_d w_d}{\sum_d n_d} \tag{20}$$

where $t_d = n_d - \text{mean}_d(n_d)$ (see McGhee 2017).

We can solve this expression as $S(V^O) = 2V^O - 0.5 + C$. However, because C is a function of V, a single-valued seats-votes function does not result, which violates Assumption 1. Stephanopoulos and McGhee (2015, p.853) tried to remove the problem by assuming turnout is constant, which implies $C = 0$ but, because this claim is always observable, making it an "assumption" does not make sense. A minimally necessary condition for which $C = 0$ is Cov$(t_d, w_d) = 0$, but this too does not solve the problem since this covariance is rarely zero.

This result means that the claims for the efficiency gap are mistaken: it is not a measure of partisan symmetry. The slope of the implied seats-votes curve is not 2 since it does not imply a coherent seats-votes curve. The claim that the efficiency gap and partisan bias "are mathematically identical in the special case in which both parties receive exactly 50 percent of the vote" (p.856) is incorrect. The claim that "a party can win more than half the seats with half the votes only by exacerbating the efficiency gap in its favor" (p.856) is also untrue.

5.7 Corrected Efficiency Gap

We give the efficiency gap idea the benefit of the doubt here with the same simplicity sought in Stephanopoulos and McGhee (2015) by computing a *corrected efficiency gap* (CEG). This measure involves moving C to the left side of the Equation 20, and defining:

$$\text{CEG}(V^O) \equiv \text{EG}(V^O) + C = S(V^O) - 2V^O + 0.5 \tag{21}$$

(cf. McGhee 2017, p.427ff). We study this measure's standard of fairness $\text{CEG}(V^O) = 0$ by solving Equation 21 for $S(V)$, adding Assumptions 1 and 2, and writing what turns out to be a coherent (single-valued) seats-votes curve:

$$S(V) = 2V - 0.5. \tag{22}$$

The *assumed fair* seats-votes curve in Equation 22 meets the partisan symmetry standard in Definition 1 because $S(V) = 2V - 0.5 = 1 - [2(1 - V) - 0.5]$, but it is a special case because of the additional constraints of a slope of $\rho(V) = 2$ for $V \in [0.25, 0.75]$ and $\rho(V) = 0$ for $V \notin [0.25, 0.75]$. For intuition, we plot this seats-votes curve as the red line in Figure 4; note that all four symmetric electoral systems in Figure 1 would be judged unfair according to this standard. Equation 22 is a restrictive and unpopular normative standard (e.g., Chambers, Miller, and Sobel 2017, p.16 and McGann, Smith, Latner, and Keena 2015, fn.1), but it is coherent and so meets Assumption 1.

We move now from the fair seats-votes curve assumed under the efficiency gap framework to estimation. Unfortunately, an estimated CEG in one election is insufficient to determine whether the electoral system is symmetric. In particular, $\beta(V^O) = [\text{CEG}(V^O) + \text{CEG}(1 - V^O)]/2$ equals zero only when $\text{CEG}(V^O) = -\text{CEG}(1 - V^O)$. However, an election with $1 - V^O$ is unobserved and so $\text{CEG}(1 - V^O)$ is not identified, nor is $\beta(0.5)$ or $\beta(V)$.

To be more specific, we add to the left panel of Figure 4 a real election outcome, the 1996 state house in Kansas (a black diamond). In this election $V^O = 0.44$ and $S(V^O) = 0.39$. Because the data indicate that $\text{CEG}(V^O) \approx 0$, it falls on the red line. Yet, this does *not* indicate that the electoral system in Kansas treated the two parties equally. To see this, compare it to the full seats-votes curve estimated via the highly accurate uniform partisan

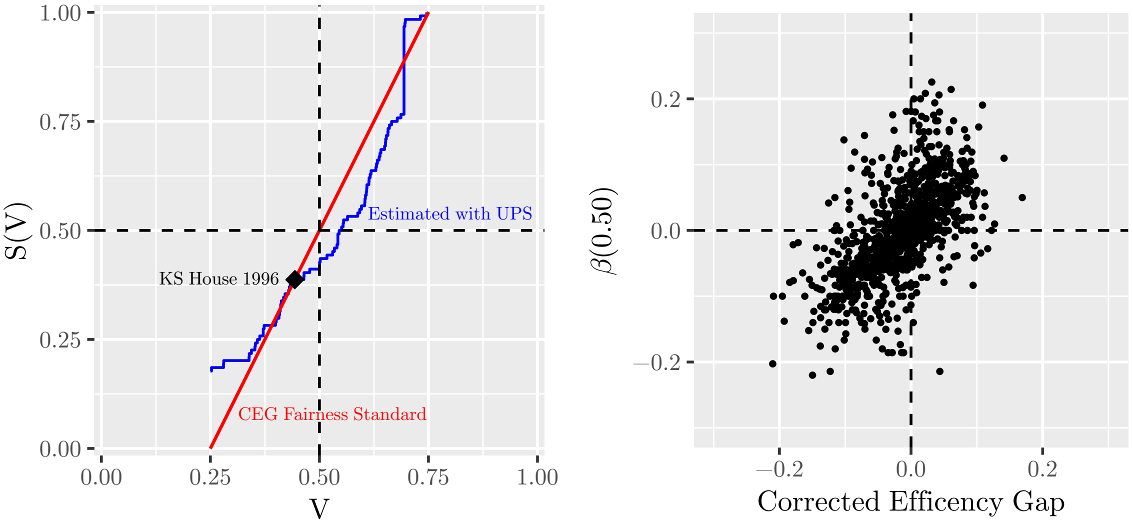

Figure 4: Seats-Votes Curves based on assumptions from the efficiency gap and uniform partisan swing

swing (Assumption 3). We add this (blue) line to the left panel in Figure 4. The results demonstrate that the 1996 Kansas election with $\text{CEG}(V^O) \approx 0$ was in fact substantially biased in favor of the Republicans: $\beta(0.5) = -0.08$. Thus, the Kansas electoral system in 1996 was highly unfair, even though the (corrected) efficiency gap measure indicated that it was fair.

We also study whether this measure happens to work empirically. We do this in the right panel of Figure 4, which plots $\text{CEG}(V^O)$ horizontally by $\beta(0.5)$ (computed by assuming uniform partisan swing) vertically for 963 legislatures. This panel does indicate a positive correlation between the two measures, as we might expect, but with remarkable error in CEG around the prediction for any observed vote. For example, when $\text{CEG}(V^O) \approx 0$, $\beta(0.5)$ varies over $[-0.2, 0.2]$.

More generally, the partisan symmetry standard requires estimating $\beta(V) = 0$ for all V.[7] In contrast, the standard of fairness according to the corrected efficiency gap is a more demanding inferential task requiring one to verify that $\beta(V) = 0$ for all V, $\rho(V) = 2$ for

[7] Although Stephanopoulos and McGhee (2015, p.854) claim that an advantage of the efficiency gap is that it “does not require any counterfactual analysis” (p.854), but on the same page and elsewhere they require counterfactuals even before they encourage counterfactual sensitivity testing. Moreover, almost any evaluation of fairness, including even the basic plurality voting rule in a single member district, requires a counterfactual analysis.

$V \in [0.25, 0.75]$, and $\rho(V) = 0$ otherwise. Thus, this framework typically requires data from more elections, or more assumptions, than other approaches. With one data point falling on the red line from the left panel of Figure 4, one cannot determine whether the election is fair or not; if the point falls off the red line, then this standard is not met but the election may still be treating both parties equally.

6 Inference: Uncertainty Estimates and Simulation

We now discuss uncertainty estimates (such as standard errors, confidence intervals, hypothesis tests, and posterior distributions) for existing measures, and then turn to approaches to uncertainty based on simulating distributions of possible redistricting plans.

Since the seats-votes curve as we have conceptualized it is a conditional expected value, classical uncertainty estimates can be easily computed for measures based on functional form assumptions (Section 4.2), stochastic uniform partisan swing (Assumption 4), or stochastic forecasting-based methods (Section 4.4). Classical uncertainty estimates can be computed for the mean-median as an estimate of vote-denominated bias; for seats-denominated bias, it can only be used as a hypothesis test with a null constructed via bootstrapping. Other proposed measures either are deterministic (Assumption 3) or are not defined separately from the quantity of interest and so implicitly have no uncertainty, but their actual uncertainty could be computed by switching to a similar method that respects uncertainty (such as from uniform to stochastic uniform partisan swing), by bootstrapping, or by identifying some quantity of interest that they estimate.

An alternative approach, called “simulation”, “outlier”, or “ensemble” analysis, attempts to compare an actual redistricting plan to an enumerated list of all possible redistricting plans constrained to fit the geography of the state and sometimes other criteria such as contiguity, compactness, and not splitting local political subdivisions (Chen and Rodden, 2013; Chikina, Frieze, and Pegden, 2017; Duchin, 2018; Magleby and Mosesson, 2018). Because the number of possible plans is unmanageably large, the goal of the literature has been to approximate the full list by random sampling (in contrast say to distributions of actual plans from other states that do not retain the constraints Wang 2016b).

Unfortunately, drawing maps purely randomly is also an unsolved problem for realistic sized legislatures (Fifield, Higgins, Imai, and Tarr, 2018). If this sampling problem is eventually solved, or if we restrict applications to very small legislatures such as a small city council, there remains the question of how to interpret the results. We discuss four options.

First, sampling can be used for "producing a large set of legally viable maps with respect to multiple criteria" (Cain, Tam Cho, Liu, and Zhang, 2017, p.1538), which can be useful for conveying what is *possible*, such as examples of plans with de minimis levels of partisan bias while also meeting other criteria. The approach would also be useful to demonstrate that plans with certain characteristics are *impossible* to draw given the state's geography, which can be compelling in applications (e.g., Chen and Rodden 2013, §5 and Duchin, Gladkova, Henninger-Voss, Klingensmith, Newman, and Wheelen n.d.).

Second, some seek to use this approach to compute how extreme a proposed plan is to all possible plans, without any probabilistic structure. One issue here is that the measure is not defined separately from the fairness standard, which is thus invulnerable being to being proven wrong or improved. Another issue is that extremity, as a purely relative indicator, is of dubious value: Should we judge a plan to be fair if it is at the 50th percentile of all possible plans (i.e., not extreme) but, when the parties split the vote equally, the Republicans receive 85% of the seats? Should we judge a plan to be unfair if it was more extreme on partisan bias than 99.99% of other plans but, when the parties split the votes equally, they split the seats equally?

Third, some uses of this approach implicitly put a probabilistic structure over randomly drawn plans, most often by imposing a uniform probability distribution. Although this imputed distribution has been compared to estimating the "cone of uncertainty" in hurricane predictions (Brief for Amicus Curiae Eric S. Lander in support of Appellees, *Gill v Whitford*, 585 US (2018)), the analogy does not hold. The cone of uncertainty is a posterior distribution of probable outcomes based on informative data, whereas the uniform distribution in this case is a prior without justification or evidence regarding what is likely, desirable, or fair (based on something like Laplace's discredited "Principle of Insufficient Reason"). A uniform distribution of locations for hurricane predictions would encompass the entire globe. In real redistricting cases, naive judges or special masters sometimes propose drawing plans randomly, or arbitrarily like a checkerboard, at which point experts on all sides object strenuously because of the likelihood of unintended consequences.

Fourth, a better use of this approach would be to try to follow the logic of hypothesis testing, for which we require a null probability distribution, such as the distribution of $\hat{\beta}(0.5)$ given $\beta(0.5) = 0$. If we could draw plans in this way, we would have a proper hypothesis test and a measure of extremity with substantive meaning. However, this distribution is not identified, and so cannot be estimated, from random uniform draws. Rejection sampling to ensure that all draws have $\hat{\beta}(0.5)$ would be useful to show what is possible, but it would not be a null distribution given $\beta(0.5) = 0$.

Finally, we construct a coherent probabilistic interpretation that could be used in this setting for a specific purpose. Suppose a redistricter claims that the *only* criteria used in selecting a plan was (say) contiguity, compactness, and equal population. The choice among plans that meet these criteria, then, is by definition random (formally because all plans that meet the three criteria are exchangeable). That then imposes a known probability distribution on the set of all possible plans that meet the criteria, giving equal probability to each (and 0 to any other). We then have a coherent hypothesis test: Choose a criterion that was not one of the original three, such as partisan bias. Then under the null — which is that no information was used in drawing the plan other than the three criteria — the probability of observing partisan bias as or more extreme is where estimated partisan bias in the proposed redistricting plan falls on the percentile of this distribution. This is then a coherent hypothesis test that gives meaning the claim that a plan is "extreme". Such an observation might be of use where redistricters were required to only use a specific set of delineated criteria, or they were defending themselves by claiming they did not use political variables. In most contexts, however, redistricting works in the opposite way, with the legislature retaining full discretion to draw any plan it chooses, using whatever criteria it chooses, so long as it does not violate the state constitution, state and federal law, or judicial rulings. The choice among plans that satisfy these rules is explicitly left to the legislature. The hypothesis test might be useful in a court case if a redistricter makes an (ill-advised) comment that he or she did not use any criteria other than some specific set, but other than that it is hard to see the use of this.

7 Concluding Remarks

The literature on partisan fairness in district-based electoral systems dates back more than a century, which spans the invention of most of modern statistics. This time period even includes the invention of what is now a fundamental principle of statistical inference — having separate notation and conceptualization for the estimator and the quantity of interest being estimated — and all the ways of using this principle to evaluate and improve statistical estimators. We update the venerable partisan fairness literature by applying this statistical principle. We reveal essential assumptions not discussed or formalized and shore them up with extensive empirical evaluations when observable implications are available. We also prove that some ideas claimed to be measures of partisan symmetry are in fact measures of this concept or are biased or otherwise limited.

Our main goal has been to build a more solid foundation for this literature so that the progress that has been made will continue. We hope that as future researchers develop new measures of partisan fairness they are better able to evaluate them given our framework. This may include simple measures that are easy to compute, more sophisticated statistical models that push forward the frontier of statistical estimation, and other types of concepts such as those which are process-oriented (e.g., nonpartisan redistricting commissions) or those which adjudicate trade offs between partisan fairness and other goals such as racial fairness, representing communities of interest, compactness, and others.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen (Aug. 2016): “Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects”. In: *American Political Science Review*, vol. 110, iss. 3, pp. 1–18.

Ansolabehere, Stephen, David W Brady, and Morris P Fiorina (1988): *Turnout and the Calculation of Swing Ratios*. Graduate School of Business, Stanford University.

Ansolabehere, Stephen and Gary King (May 1990): "Measuring the Consequences of Delegate Selection Rules in Presidential Nominations". In: *Journal of Politics*, no. 2, vol. 52, pp. 609–621. URL: bit.ly/DelSeln.

Ansolabehere, Stephen and William Leblanc (2008): "A spatial model of the relationship between seats and votes". In: *Mathematical and Computer Modelling*, no. 9-10, vol. 48, pp. 1409–1420.

Ansolabehere, Stephen, William Leblanc, and James M Snyder (2012): "When parties are not teams: party positions in single-member district and proportional representation systems". In: *Economic Theory*, no. 3, vol. 49, pp. 521–547.

Best, Robin E, Shawn J Donahue, Jonathan Krasno, Daniel B Magleby, and Michael D McDonald (2018): "Considering the prospects for establishing a packing gerrymandering standard". In: *Election Law Journal*, no. 1, vol. 17, pp. 1–20.

Blackburn, Simon (2003): *Ethics: A very short introduction*. Vol. 80. Oxford University Press.

Brady, David W and Bernard Grofman (1991): "Sectional differences in partisan bias and electoral responsiveness in US House elections, 1850–1980". In: *British Journal of Political Science*, no. 2, vol. 21, pp. 247–256.

Brunell, Thomas (1999): "Partisan Bias in US Congressional Elections, 1952-1996: Why the Senate is Usually More Republican than the House of Representatives". In: *American Politics Quarterly*, no. 3, vol. 27, pp. 316–337.

— (2010): *Redistricting and representation: Why competitive elections are bad for America*. Routledge.

Butler, David E. (1951): "Appendix". In: *The British General Election of 1950*. Ed. by H.G. Nicholas. London: Macmillan.

Cain, Bruce E, Wendy K Tam Cho, Yan Y Liu, and Emily R Zhang (2017): "A Reasonable Bias Approach to Gerrymandering: Using Automated Plan Generation to Evaluate Redistricting Proposals". In: *Wm. & Mary L. Rev.* Vol. 59, p. 1521.

Campbell, James E (1996): *Cheap seats: the Democratic Party's advantage in US House elections*. The Ohio State University Press.

Chambers, Christopher P, Alan D Miller, and Joel Sobel (2017): "Flaws in the efficiency gap". In: *Journal of Law & Politics*, vol. 33, p. 1.

Chen, Jowei and Jonathan Rodden (2013): "Unintentional gerrymandering: Political geography and electoral bias in legislatures". In: *Quarterly Journal of Political Science*, no. 3, vol. 8, pp. 239–269.

Chikina, Maria, Alan Frieze, and Wesley Pegden (2017): "Assessing significance in a Markov chain without mixing". In: *Proceedings of the National Academy of Sciences*, no. 11, vol. 114, pp. 2860–2864.

Cho, Wendy K Tam (2017): "Measuring partisan fairness: How well does the efficiency gap guard against sophisticated as well as simple-minded modes of partisan discrimination". In: *U. Pa. L. Rev. Online*, vol. 166, p. 17.

Coate, Stephen and Brian Knight (2007): "Socially optimal districting: a theoretical and empirical exploration". In: *The Quarterly Journal of Economics*, no. 4, vol. 122, pp. 1409–1471.

Cox, Gary W. (1997): *Making votes count: strategic coordination in the world's electoral systems*. Cambridge University Press.

Cox, Gary W. and Jonathan N. Katz (1999): "The reapportionment revolution and bias in US congressional elections". In: *American Journal of Political Science*, pp. 812–841.

— (2002): *Elbridge Gerry's salamander: The electoral consequences of the reapportionment revolution*. Cambridge University Press.

Duchin, Moon (2018): "Gerrymandering metrics: How to measure? What's the baseline?" In: *arXiv:1801.02064*.

Duchin, Moon, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen (n.d.): "Obstructions to Proportional Representation: Republicans in Massachusetts". Tufts University.

Erikson, Robert S (1972): "Malapportionment, gerrymandering, and party fortunes in congressional elections". In: *American Political Science Review*, no. 4, vol. 66, pp. 1234–1245.

Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr (2018): "A new automated redistricting simulator using markov chain monte carlo". In: *Work. Pap., Princeton Univ., Princeton, NJ*.

Friedman, John N and Richard T Holden (2008): "Optimal gerrymandering: sometimes pack, but never crack". In: *American Economic Review*, no. 1, vol. 98, pp. 113–44.

Garand, James C and T Wayne Parent (1991): "Representation, swing, and bias in US presidential elections, 1872-1988". In: *American Journal of Political Science*, pp. 1011–1031.

Gelman, Andrew and Gary King (June 1990): "Estimating the Electoral Consequences of Legislative Redistricting". In: *Journal of the American Statistical Association*, no. 410, vol. 85, pp. 274–282. URL: http://j.mp/GerryDem.

— (May 1994a): "A Unified Method of Evaluating Electoral Systems and Redistricting Plans". In: *American Journal of Political Science*, no. 2, vol. 38, pp. 514–554. URL: j.mp/unifiedEc.

— (Sept. 1994b): "Enhancing Democracy Through Legislative Redistricting". In: *American Political Science Review*, no. 3, vol. 88, pp. 541–559. URL: j.mp/redenh.

Gerber, Elisabeth R and Jeffrey B Lewis (2004): "Beyond the median: Voter preferences, district heterogeneity, and political representation". In: *Journal of Political Economy*, no. 6, vol. 112, pp. 1364–1383.

Gilligan, Thomas W and John G Matsusaka (1999): "Structural constraints on partisan bias under the efficient gerrymander". In: *Public Choice*, no. 1-2, vol. 100, pp. 65–84.

Greene, Joshua D (2009): "The cognitive neuroscience of moral judgment". In: *The cognitive neurosciences*, vol. 4, pp. 1–48.

Grofman, Bernard and Gary King (Jan. 2007): "The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after LULAC v. Perry". In: *Election Law Journal*, no. 1, vol. 6. http://gking.harvard.edu/files/abs/jp-abs.shtml, pp. 2–35.

Grofman, Bernard, William Koetzle, and Thomas Brunell (1997): "An integrated perspective on the three potential sources of partisan bias: Malapportionment, turnout differences, and the geographic distribution of party vote shares". In: *Electoral studies*, no. 4, vol. 16, pp. 457–470.

Gudgin, Graham and Peter J Taylor (2012): *Seats, votes, and the spatial organisation of elections*. ECPR Press.

Hardy, Leroy C (1977): “Considering the gerrymander”. In: *Pepperdine Law Review*, no. 2, vol. 4, p. 3.

Herron, Michael C and Alan E Wiseman (2008): “Gerrymanders and theories of law making: A study of legislative redistricting in Illinois”. In: *The Journal of Politics*, no. 1, vol. 70, pp. 151–167.

Iacus, Stefano M., Gary King, and Giuseppe Porro (2018): “A Theory of Statistical Inference for Matching Methods in Causal Research”. In: *Political Analysis*, pp. 1–23. URL: j.mp/Nt9TkZ.

Jackman, Simon (1994): “Measuring electoral bias: Australia, 1949–93”. In: *British Journal of Political Science*, no. 3, vol. 24, pp. 319–357.

Katz, Jonathan N. and Gary King (Mar. 1999): “A Statistical Model for Multiparty Electoral Data”. In: *American Political Science Review*, no. 1, vol. 93, pp. 15–32. URL: bit.ly/mtypty.

King, Gary (Nov. 1989): “Representation Through Legislative Redistricting: A Stochastic Model”. In: *American Journal of Political Science*, no. 4, vol. 33, pp. 787–824. URL: http://j.mp/2o46Gkk.

— (May 1990): “Electoral Responsiveness and Partisan Bias in Multiparty Democracies”. In: *Legislative Studies Quarterly*, no. 2, vol. XV, pp. 159–181. URL: bit.ly/ErespMP.

King, Gary and Robert X Browning (Dec. 1987): “Democratic Representation and Partisan Bias in Congressional Elections”. In: *American Political Science Review*, no. 4, vol. 81, pp. 1252–1273. URL: j.mp/parSym.

King, Gary, John Bruce, and Andrew Gelman (1996): “Racial Fairness in Legislative Redistricting”. In: ed. by ed. Paul E. Peterson. Princeton University Press. URL: j.mp/Fairrace.

King, Gary, Ori Rosen, Martin Tanner, and Alexander F Wagner (2008): “Ordinary economic voting behavior in the extraordinary election of Adolf Hitler”. In: *The Journal of Economic History*, no. 4, vol. 68, pp. 951–996. URL: bit.ly/nazivote.

King, Gary and Langche Zeng (2006): “The Dangers of Extreme Counterfactuals”. In: *Political Analysis*, no. 2, vol. 14, pp. 131–159. URL: j.mp/dangerEC.

Klarner, Carl (2018): *State Legislative Election Returns, 1967-2016: Restructured For Use*. DOI: 10.7910/DVN/DRSACA. URL: https://doi.org/10.7910/DVN/DRSACA.

Krasno, Jonathan, Daniel B Magleby, Michael D McDonald, Shawn J Donahue, and Robin E Best (2018): “Can Gerrymanders Be Detected? An Examination of Wisconsin’s State Assembly”. In: *American Politics Research*, pp. 1–40.

Leighley, Jan E and Jonathan Nagler (2013): *Who votes now?: Demographics, issues, inequality, and turnout in the United States*. Princeton University Press.

Lijphart, Arend (2004): “Constitutional design for divided societies”. In: *Journal of democracy*, no. 2, vol. 15, pp. 96–109.

Magleby, Daniel B and Daniel B Mosesson (2018): “A New Approach for Developing Neutral Redistricting Plans”. In: *Political Analysis*, no. 2, vol. 26, pp. 147–167.

Mann, Thomas E (1978): *Unsafe at any margin: Interpreting congressional elections*. Vol. 220. Aei Press.

May, Kenneth O (1952): “A set of independent necessary and sufficient conditions for simple majority decision”. In: *Econometrica: Journal of the Econometric Society*, pp. 680–684.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal (2009): “Does gerrymandering cause polarization?” In: *American Journal of Political Science*, no. 3, vol. 53, pp. 666–680.

McDonald, Michael D (2017): “The Arithmetic of Electoral Bias, with Applications to US House Elections”. Revised 2009 APSA meeting paper, Toronto, CA.

McDonald, Michael D and Robin E Best (2015): “Unfair partisan gerrymanders in politics and law: A diagnostic applied to six cases”. In: *Election Law Journal*, no. 4, vol. 14, pp. 312–330.

McGann, Anthony J, Charles Anthony Smith, Michael Latner, and J Alex Keena (2015): “A discernable and manageable standard for partisan gerrymandering”. In: *Election Law Journal*, no. 4, vol. 14, pp. 295–311.

McGhee, Eric M. (2017): “Measuring efficiency in redistricting”. In: *Election Law Journal: Rules, Politics, and Policy*, no. 4, vol. 16, pp. 417–442.

Nagle, John F (2015): “Measures of partisan bias for legislating fair elections”. In: *Election Law Journal*, no. 4, vol. 14, pp. 346–360.

Niemi, Richard G and Simon Jackman (1991): “Bias and responsiveness in state legislative districting”. In: *Legislative Studies Quarterly*, pp. 183–202.

Owen, Guillermo and Bernard Grofman (1988): “Optimal partisan gerrymandering”. In: *Political Geography Quarterly*, no. 1, vol. 7, pp. 5–22.

Petracca, Mark P (1996): “A history of rotation in office”. In: *Legislative Term Limits: Public Choice Perspectives*. Springer, pp. 247–277.

Plener Cover, Benjamin (2018): “Quantifying Partisan Gerrymandering: An Evaluation of the Efficiency Gap Proposal”. In: *Stan. L. Rev.* Vol. 70, p. 1131.

Puppe, Clemens and Attila Tasnadi (2009): “Optimal redistricting under geographical constraints: Why ‘pack and crack’ does not work”. In: *Economics Letters*, no. 1, vol. 105, pp. 93–96.

Reynolds, Andrew (2005): “Reserved seats in national legislatures: A research note”. In: *Legislative Studies Quarterly*, no. 2, vol. 30, pp. 301–310.

Rubin, Donald B. (1991): “Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism”. In: *Biometrics*, vol. 47, pp. 1213–1234.

Sen, Amartya (1976): “Liberty, unanimity and rights”. In: *Economica*, no. 171, vol. 43, pp. 217–245.

Stephanopoulos, Nicholas O and Eric M. McGhee (2015): “Partisan gerrymandering and the efficiency gap”. In: *The University of Chicago Law Review*, pp. 831–900.

— (2018): “The measure of a metric: The debate over quantifying partisan gerrymandering”. In: *Stan. L. Rev.* Vol. 70, p. 1503.

Tapp, Kristopher (2018): “Measuring Political Gerrymandering”. In: *arXiv:1801.02541*.

Tufte, Edward R (1973): “The relationship between seats and votes in two-party systems”. In: *American Political Science Review*, no. 2, vol. 67, pp. 540–554.

VanderWeele, Tyler J. and Miguel A Hernan (2012): “Causal Inference Under Multiple Versions of Treatment”. In: *Journal of Causal Inference*, vol. 1, pp. 1–20.

Veomett, Ellen (2018): “Efficiency Gap, Voter Turnout, and the Efficiency Principle”. In: *Election Law Journal: Rules, Politics, and Policy*.

Wang, Samuel (2016a): “Three Practical Tests for Gerrymandering: Application to Maryland and Wisconsin”. In: *Election Law Journal*, no. 4, vol. 15, pp. 367–384.

— (2016b): “Three tests for practical evaluation of partisan gerrymandering”. In: *Stanford Law Review*, vol. 68, p. 1263.

Wang, Samuel and Brian Remlinger (2018): “An Antidote for Gobbledygook: Organizing the Judge’s Partisan Gerrymandering Toolkit into a Two-Part Test”. In: URL: https://ssrn.com/abstract=3158123.

Warrington, Gregory S (2018a): “Introduction to the declination function for gerrymanders”. In: *arXiv preprint arXiv:1803.04799*.

— (2018b): “Quantifying gerrymandering using the vote distribution”. In: *Election Law Journal*, no. 1, vol. 17, pp. 39–57.

Wimmer, Karl (2010): “Agnostically learning under permutation invariant distributions”. In: *51st Annual IEEE Symposium on Foundations of Computer Science*. IEEE, pp. 113–122.

Yoshinaka, Antoine and Chad Murphy (2009): “Partisan gerrymandering and population instability: Completing the redistricting puzzle”. In: *Political Geography*, no. 8, vol. 28, pp. 451–462.