Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies: Online Appendices

Jonathan N. Katz* Gary King† Elizabeth Rosenblatt‡

August 4, 2019

Online Appendices for: Jonathan N. Katz, Gary King, and Elizabeth Rosenblatt, "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies," forthcoming, *American Political Science Review*, GaryKing.org/symmetry.

*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.

Contents

Appendix A Differential Partisan Turnout Effects

The Supreme Court requires equal population, not equal turnout, across districts within states (*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 the section on the Seats-Votes Curve. The curve conditions on V — the unweighted *average district vote*, $V = \text{mean}(v_d)$ — and so 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 Sen, 2016). To construct this counterfactual curve, we switch from the average district vote to the total *statewide vote*, which can be represented as 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,

$$U = \frac{\sum_d n_d v_d}{\sum_d n_d} = \frac{\sum_d \bar{n} v_d + \sum_d t_d v_d}{\sum_d n_d} = V + \frac{\sum_d t_d v_d}{\sum_d n_d}. \tag{1}$$

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) and Tamas (2019). For example, quantities like $\beta(U) - \beta(V)$ may help isolate turnout effects (Grofman, Koetzle, and Brunell, 1997).

Appendix B Noncompetitive Party System Fairness Standards

We address here standards of fairness for electoral systems when one party has an overwhelming majority of votes 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 the section, Gerrymandering Goals). 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 1 (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.

Appendix C 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 4) 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 the section on Partisan Swing Assumptions can be used directly.

Appendix D Models of Individual Voters

Researchers in this literature frequently condition on district vote proportions and treat them as fixed quantities in calculating seat shares. This, however, is unnecessary for all our analyses. To provide a feel for how these assumptions can be relaxed explicitly, first denote an individual vote as v_{id} for voter i ($i = 1, \ldots, P$ where P is the number of voters in the state) and district d, redistricters could make a weaker assumption (which we also do not require) by assuming only that $v_{id} \perp d$, or $v_{id} \perp d | X_d$ where X_d are characteristics of the candidates and the election in district d. This assumption can easily be relaxed further because only the means $\{v_1, \ldots, v_d\}$ are used in subsequent analyses, and so violations that cancel do not matter.

Thus 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, and 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 partisan 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).

Appendix E Uncertainty Estimates

We now discuss uncertainty estimates (such as standard errors, confidence intervals, hypothesis tests, and posterior distributions) for existing measures.

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 (see the section on "Functional Form Assumptions"), stochastic uniform partisan swing (Assumption 4), or stochastic forecasting-based methods (see the section on "Forecasting Assumptions"). 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, such as 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.

Appendix F 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 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: $\beta(0.5) =$ 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: 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.

Appendix G Intuition and Empirical Evidence on the Efficiency Gap

We offer intuition about the (corrected) efficiency gap in this section by first highlighting an incorrect assumption used in constructing the measure and then explaining how the approach fails empirically.

First, Stephanopoulos and McGhee (2015, p.853) claim that "if we assume that all districts are equal in population (which is constitutionally required)..., then the computation reduces through simple algebra to something quite straightforward". The claim, in our notation, is that their assumption implies $C = 0$, thus simplifying Equation 20 (and in turn making it possible to derive the seats-votes curve). McGhee (2014) further claims these are "constraints virtually universal in the research on symmetry and responsiveness"; McGhee (2017) repeats the same claims. Unfortunately, these claims are incorrect. First, the US constitution does not require equal district population in any American legislature (including districts in the House of Representatives which is allowed to vary in population across states). Second, population does not appear in Equation 19 and so assumptions about it do not help. Third, an equality claim that would simplify the expression is that the number of *voters* (not people) is the same in every district, $n_d = n$ (McGhee, 2014, fn6), although this assumption is sufficient but not necessary.

Second, for intuition, we plot the seats-votes curve conveying CEG's standard of fairness as the red line in the right panel of Figure 1. To be more specific, we add to the left panel of Figure 1 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 swing (Assumption 3). We add this (blue) line to the left panel in Figure 1. 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.

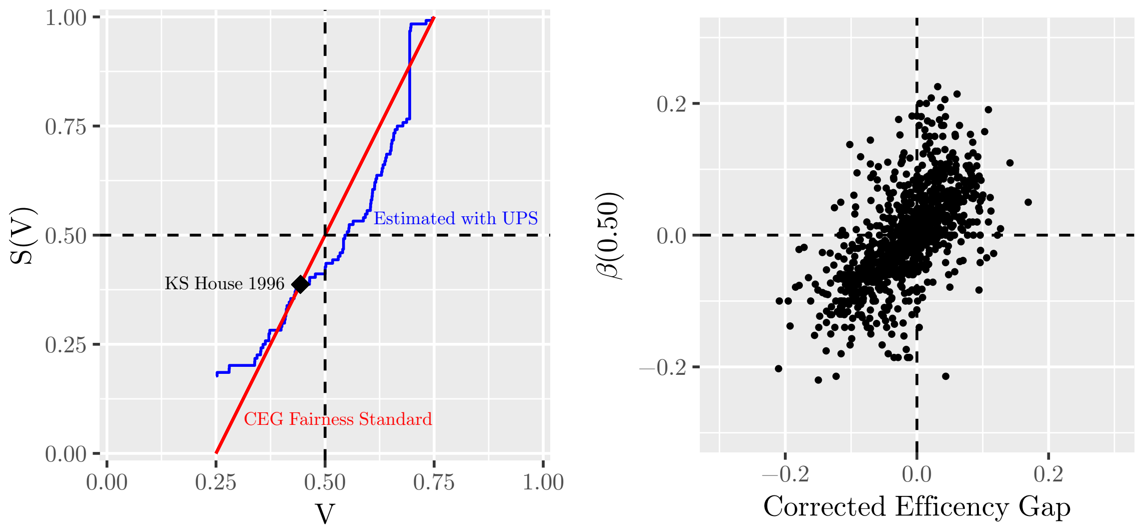

Figure 1: Seats-Votes Curves based on assumptions from the corrected efficiency gap (CEG) and uniform partisan swing. Left panel: CEG's standard of fairness in red, and the Kansas House of Representatives in 1996 estimated with uniform partisan swing. Right panel: scatterplot of bias vertically by CEG horizontally.

We also study whether this measure happens to work empirically. We do this in the right panel of Figure 1, which plots $\mathrm{CEG}(V^O)$ 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 $\mathrm{CEG}(V^O) \approx 0$, $\beta(0.5)$ varies over $[-0.2, 0.2]$.

More generally, the partisan symmetry standard requires estimating $\beta(V) = 0$ for all V.[1] 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 $V \in [0.25, 0.75]$, 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 1, 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

[1]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.

the election may still be treating both parties equally.

Additional 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 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.

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

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

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.

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

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.

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.

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

McGhee, Eric M. (2014): “Measuring Partisan Bias in Single-Member District Electoral Systems”. In: *Legislative Studies Quarterly*, no. 1, vol. 39, pp. 55–85.

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

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

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

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

Tamas, Bernard (2019): “American Disproportionality: A Historical Analysis of Partisan Bias in Elections to the US House of Representatives”. In: *Election Law Journal: Rules, Politics, and Policy*, no. 1, vol. 18, pp. 47–62.