Base Rate Neglect and the Diagnosis of Partisan Gerrymanders: Against Efficiency

Sanford C. Gordon [*]and Sidak Yntiso [†‡]

January 2023

Abstract

A popular measure of unfairness in partisan gerrymandering, the "Efficiency Gap" (Stephanopoulos and McGhee, 2015), suffers from base rate neglect because it does not account for the overall partisan composition of a state. We discuss some conceptual and practical implications of this issue. A measure of vote dilution asymmetry that adjusts for the base rate avoids these problems. We show that this adjustment alters not only the magnitude of estimates of partisan disparity but also, frequently, the *sign*. The issue persists in a generalized version of the measure that differentially weights the harms associated with partisan gerrymandering: "packing" and "cracking." Finally, whereas a longitudinal analysis based on the efficiency gap suggests that the pro-Republican consequences of extreme partisan gerrymandering is a recent phenomenon, an analysis using our preferred measure shows a shift from pro-Democratic to pro-Republican bias in the mid-1990s.

Keywords: partisan gerrymandering; redistricting; efficiency gap; vote dilution

Word Count: 7,500

[*]Department of Politics, New York University. 19 West 4th Street, 2nd Floor. New York, New York, 10012. Email: sanford.gordon@nyu.edu

[†]Corresponding author. Harris School of Public Policy, University of Chicago. 1307 East 60th Street. Chicago, Illinois, 60637. Email: syntiso@uchicago.edu

[‡]For helpful suggestions, we gratefully acknowledge Hans Hassell, Doug Spencer, and seminar participants at the 16th Annual Conference on Empirical Legal Studies, the annual meeting of the Southern Political Science Association, New York University and the University of Chicago.

1 Introduction

We consider conceptual and practical implications of base rate neglect in empirical assessments of vote dilution stemming from partisan gerrymandering. A common metric of partisan (un)fairness, the Efficiency Gap (McGhee, 2014; Stephanopoulos and McGhee, 2015), captures inter-party disparities in the empirical compound probability that a voter's vote is "wasted" ("cracked" into persistent minorities or "packed" inefficiently into a small number of districts) *and* that they belong to a specific political party. Drawing on an analogy to empirical studies of discrimination, we argue in favor of disparity metrics that employ the probability a voter's vote is wasted *given that* the voter belongs to that party (cf., McGann et al., 2015; Nagle, 2017). Measures of unfairness based on the former quantity fail to adjust for overall partisan imbalances in a state, whereas those based on the latter do incorporate this base rate. A focus on differences in compound and conditional probabilities as estimands of interest also affords an opportunity to identify and correct a problem in previous proposals to differentially weight cracked and packed voters: weighted probability measures whose values can exceed one.

Using simulated maps and actual and imputed vote returns, we show that correcting for base rate alters not only the magnitude of estimates of partisan disparity but also, frequently, the sign. This issue is not alleviated using a corrected weighting scheme. Finally, we conduct a longitudinal analysis to examine the evolution of the effects of partisan gerrymandering nationwide. An analysis using the Efficiency Gap would suggest that the practice evolved from one in which Republican and Democratic gerrymanders tended to offset one another to one in which the preponderance of gerrymanders benefited Republicans. An analysis using our preferred dilution asymmetry metric, by contrast, shows an overall shift from pro-Democratic to pro-Republican bias in the mid-1990s.

The Supreme Court's 2019 decision in *Rucho vs. Common Cause* would seem to foreclose federal judicial remedies for even the most egregious instances of partisan gerrymandering. But any conclusion that this obviates the need for continued refinement and analysis of measures of the extent and impact of partisan gerrymandering is misplaced: as we discuss below, the issue is very much alive in the states. Moreover, insofar as the majority in *Rucho* rejects the justiciability of partisan gerrymandering cases, it does so primarily by rejecting a constitutional right to fairness toward *parties*, and only secondarily toward *voters*. To sway a sympathetic state court, or a future, more favorable federal court, e.g., on equal protection grounds, it will likely be essential to place voters rather than parties at the center of the argument. A measure of asymmetric vote dilution that properly adjusts for the base rate does precisely this.

2 Background

2.1 The Legal Status of Partisan Disparity Claims

Writing for a 5-4 conservative majority in *Rucho vs. Common Cause* (2019), Chief Justice Roberts wrote that claims of partisan gerrymandering were non-justiciable under the U.S. Constitution, holding that "The Constitution supplies no objective measure for assessing whether a districting map treats a political party fairly" (p. 20). While the court had heard cases pertaining to partisan gerrymandering since the early 1970s, *Rucho* represented the culmination of a process that began in 2004 with *Vieth v. Jubelirer*, when Justice Scalia, in a plurality opinion, held that partisan gerrymanders were not justiciable owing to the absence of a discernible and manageable standard for adjudicating partisan gerrymandering claims. In a concurring opinion, Justice Kennedy did not rule out the possibility that such a standard might ultimately be articulated. Then, in *Gill v. Whitford* (2018), the Court took up Wisconsin's 2011 redistricting plan, following a ruling by a 2-1 ruling by a three-judge panel in the Western District of Wisconsin that the plan violated the 14th amendment's equal protection plan. The Supreme Court remanded the case following a unanimous determination that the plaintiff had failed to demonstrate standing, but previewed its determination in *Rucho*, holding, "[T]his Court is not responsible for vindicating generalized partisan preferences" (p. 21).

For our purposes, the Court's determination in *Rucho* that there was no discernible and manageable standard for adjudicating constitutional claims concerning partisan gerrymandering is an important factor in considering the utility of diagnostic measures thereof. Stephanopoulos and McGhee (2015), for example, in answering the implicit challenge in Kennedy's *Vieth* concurrence, suggest a threshold value of the efficiency gap that would constitute a sufficient statistic for constitutional impermissibility. An additional point to which we return below is Roberts' approving reference to Justice O'Connor's argument in *Bandemer v. Davis* (1986), explicitly rejecting "a conviction that, the greater the departure from proportionality, the more suspect an apportionment plan becomes." The Court has repeatedly ruled that the Constitution does not oblige to allocate seats with an eye toward proportionality.

While the prospects of the federal courts rejecting explicitly partisan gerrymanders on constitutional grounds are slim, this does not rule out such considerations in the state courts. In *League of Women Voters v. Commonwealth of Pennsylvania* (2018), the Pennsylvania Supreme Court ruled that the state's redistricting plan violated Article I, Section 5 of the Pennsylvania Constitution, which requires that "Elections shall be free and equal." In 2022, the high courts in New York (in *Harkenrider v. Hochul*), Kansas (in *Alonzo v. Schwab*), and North Carolina (in *Harper v. Hall*) overturned districting maps as inconsistent with their state constitutions.[1]

An additional area of development is state law and state constitutional amendments. To reach its decision in the *Harkenrider* case, the New York Court of Appeals relied on a 2015 amendment to the New York Constitution, which states that "Districts shall not be drawn to discourage competition or for the purpose of favoring or disfavoring incumbents or other particular candidates or political parties." In 2018, voters in Colorado, Michigan, Missouri, Ohio, and Utah approved ballot initiatives addressing the partisan composition of legislative districts in response to the *Gill* decision.[2]

2.2 The Efficiency Gap

The Efficiency Gap has featured prominently in recent scholarship(Stephanopoulos and McGhee, 2015; Caughey, Tausanovitch, and Warshaw, 2017; Cover, 2018; Jiang et al., 2020), litigation (including the cases cited above), and press coverage. The measure depends upon the distribution

[1]As of writing, the North Carolina case, now called *Moore v. Harper*, is currently pending before the U.S. Supreme Court. The critical issue in that case is whether the elections clause in Article I, section 4 of the U.S. Constitution effectively prohibits state courts from reviewing any rules set by state legislatures regulating federal elections, including maps of congressional districts. In oral arguments a majority of the justices appeared generally unsympathetic to the strongest version of the claim of prohibition (the so-called "Independent State Legislature" theory); in any case, these arguments would not apply to the drawing of district boundaries for state legislatures.

[2]For a comprehensive overview, see Professor Doug Spencer's "All About Redistricting" website at `https://redistricting.lls.edu/`.

of a party's wasted votes, i.e., all votes cast in favor of the losing candidate or votes in excess of 50%+1 for the winning candidate. Formally, let $W^p_{s,t}$ be the number of wasted votes for party p in state s at time t, and $V^p_{s,t}$ the total number of votes for that party, with $p \in \{D, R\}$. Then the Efficiency Gap in state s, ($EG_{s,t}$) is defined as the difference in the number of individuals wasted by each major party divided by the total number of votes cast for the two major parties:

$$EG_{s,t} \equiv \frac{W^D_{s,t} - W^R_{s,t}}{V^D_{s,t} + V^R_{s,t}}. \tag{1}$$

At first blush, EG has several features that recommend it. Relative to previous partisan symmetry approaches, which depend upon counterfactual electoral shifts, EG can be calculated from real-world election results. The metric can be plainly interpreted as the party's surplus seat share over the seat share that would arise if both parties wasted an equal number of voters.

McGhee (2014) argues that another favorable feature of the efficiency gap is that it adheres to an *efficiency principle*: that any measure of efficiency must increase monotonically with seat share holding vote share constant.

In the years since it was first proposed, the efficiency gap has become a standard diagnostic measure of partisan gerrymandering, extensively relied on by expert witnesses in gerrymandering cases and cited approvingly by the state supreme court majorities in Kansas and North Carolina, as well as the federal district court in *Gill.* A 2018 ballot initiative amended Michigan's constitution to require that the state's nonpartisan redistricting commission minimize disproportionate partisan advantage "using accepted measures of partisan fairness" such as the efficiency gap. Utah's Proposition 4 (2018) directed the legislature and redistricting commission to employ "best available data and scientific and statistical methods, including measures of partisan symmetry" to assess whether a proposed map conforms to a set of standards that includes partisan fairness. And Missouri's Amendment 1 (2018) explicitly added calculation and consideration of the efficiency gap into the state's constitution.

3 Conceptual Issues with Efficiency as a Criterion

3.1 Extant Critiques of the Efficiency Gap Metric

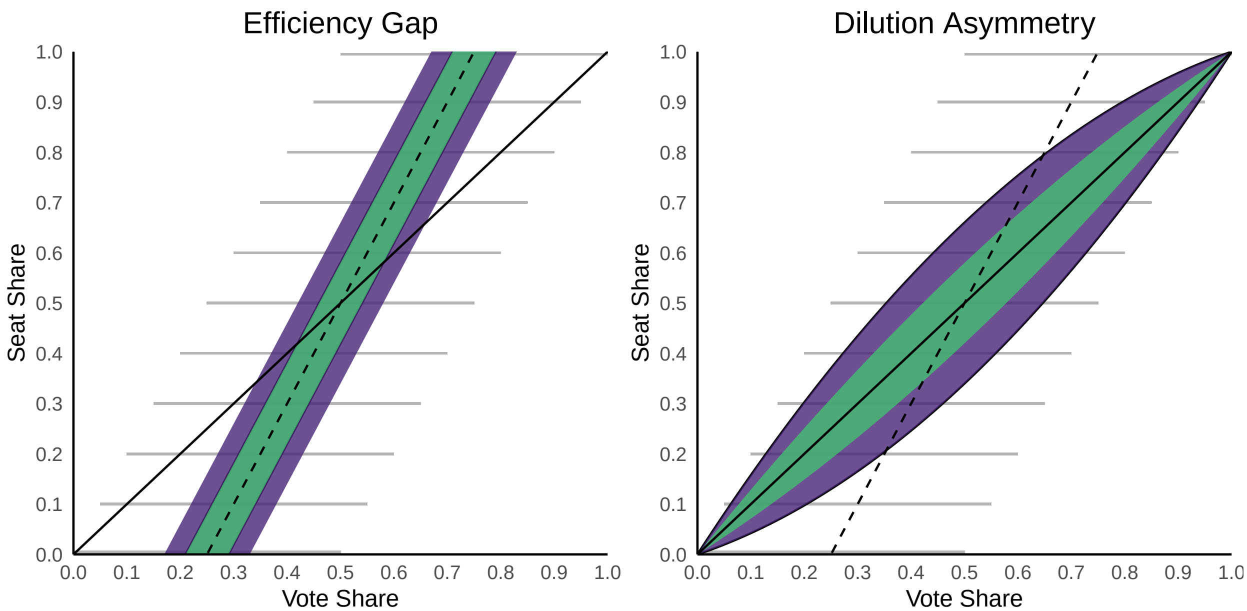
The Efficiency Gap has been the subject of several criticisms. First, for a given a vote split, EG only takes as many values as the number of districts. One consequence of this coarseness is that a permissible EG range may never occur in small states (Tam Cho, 2017). Second, EG fails to distinguish between types of wasted votes (Bernstein and Duchin, 2017), an issue that we return to below. Third, the measure fails to differentiate between partisan inefficiencies that arise from the enacted plan from those inefficiencies that would be expected given the spatial distribution of voters.

A fourth concern emerges from the fact that EG does not just tolerate, but may actually *require* departures from a 1:1 proportionality between voteshare and seatshare in a state (Duchin, 2018). The left panel of Figure 1 illustrates this issue by plotting the seat share of one party against its vote share in a hypothetical allocation of a 10-seat congressional delegation and assumed equal turnout across districts. The solid 45° line denotes perfect proportionality. The gray line segments denote the logical bounds on vote share associated with a particular seat share: so, for example, a party with two seats out of ten could theoretically have received anywhere between 10% and 60% of the statewide vote share.

The green region in the left panel denotes the set of (vote share, seat share) combinations consistent with a standard of $|EG| < 0.08$, a standard proposed by Stephanopoulos and McGhee (2015) (for state legislatures). The purple region (inclusive of the green one) doubles the set of permissible values for EG. As is evident from the figure, outside of a range of vote shares in the neighborhood of 50%, a standard based on EG would be presumptively invalid. To understand why this is, suppose turnout is equal in each district. Let $\sigma_{s,t}$ be the share of seats held by the Democratic Party in state s at time t, and $\nu_{s,t}$ the Democratic share of the two-party vote. Then, as shown by McGhee (2014), the efficiency gap given equal turnout across districts may be expressed as

$$EG_{s,t} = 2\nu_{s,t} - \sigma_{s,t} - \frac{1}{2}. \tag{2}$$

Figure 1: Features of Efficiency Gap and Dilution Asymmetry Standards (10-Seat Delegation)



In each panel, the gray line segments denote the range of vote shares that could theoretically give rise to a particular seat share. The solid $45°$ *line corresponds to perfect proportionality of votes to seats; the dashed line corresponds to double proportionality. Green areas denote (seat share, vote share) combinations consistent with* $|EG| < 0.08$ *and* $|DA^P| < 0.16$. *Purple areas (inclusive of green areas) denote combinations consistent with* $|EG| < 0.16$ *and* $|DA^P| < 0.32$.

For any fixed value of $EG_{s,t}$, the implied seat-vote relationship has a slope of 2. This "double-proportionality" is depicted by the dashed line in the figure. 1:1 proportionality minimizes the absolute value of the Efficiency Gap if and only if $\sigma_{s,t} = \nu_{s,t} = \frac{1}{2}$, i.e., if the state's voting population is perfectly evenly divided between Republicans and Democrats (and the solid and dashed lines in the figure intersect).

It has long been observed (e.g., Tufte, 1973) that *empirically*, the slope obtained from a regression of seats on votes is closer to 2 than 1 (and may even exceed 2). As Warrington (2019) notes, the realized slope will tend to depend on the geographic distribution of partisanship within a state. That being said, it would seem peculiar that a diagnostic approach would rule out perfect proportionality as inconsistent with a norm of fairness except in knife's edge cases. Put another way, the repeated holding by the courts that the constitution does not require this proportionality is surely not equivalent to a *prohibition* thereof.

A final concern with the efficiency gap was raised by the majority in the *Gill* decision: as with other measures of partisan asymmetry, the measure captures the effect of a gerrymander on the fortunes of political parties rather than the welfare of individual citizens.

3.2 Consequences of Base Rate Neglect and a Simple Correction

Unless a state is close to evenly split between Democrats and Republicans, meeting the constitutional standard envisioned by Stephanopoulos and McGhee (2015) may necessitate departing from perfect proportionality. We trace this issue to the fact that EG as a measure of disparity suffers from *base rate neglect*: a failure to account for the overall partisan composition of voters in a state. To illustrate the problem, consider the following example from the study of discrimination. Suppose there are one hundred motorists on the highway: 80 white, and 20 black. The police pull over three white and three black motorists. Clearly, the black motorists experience a disproportionately higher likelihood of being stopped: the rate at which a motorist is stopped given that they are Black is $3/20 = 15\%$, compared to the analogous quantity for white motorists of $3/80 = 3.75\%$. One summary measure of the disparity is the difference between these two quantities: $15 - 3.75 = 11.25$

percentage points.[3] But a measure of disparity equivalent to EG, $\frac{3-3}{20+80}$, would yield a difference of zero percentage points. What gives?

Consider the expression for the efficiency gap in equation (1). Suppressing the s and t subscripts for clarity, let $V = V^D + V^R$. Then (1) is equivalent to

$$\frac{W^D}{V} - \frac{W^R}{V}.$$

The fraction $\frac{W^p}{V}$ is the empirical probability that a vote is wasted *and* that it was in favor of party p.[4] The efficiency gap is the difference between these estimated compound probabilities for the two parties. It is immediate that equalizing W^D and W^R minimizes $|EG|$. However, as is clear from our police stop example, equalizing the number of black and white motorists who are stopped does not equalize the expected cost imposed on the motorists. Similarly, equalizing the number of wasted Democratic and Republican (or, for that matter, black and white or urban and rural) voters does not equalize the expected risk of vote dilution between those populations when their numbers differ.[5]

Rather, we are interested in the difference in the empirical probability that a vote is diluted/wasted *given that* it was for one party and the same quantity for the other. This yields an appropriate measure of partisan *Dilution Asymmetry* at time t in state s, ($DA^P_{s,t}$):

$$DA^P_{s,t} \equiv \frac{W^D}{V^D} - \frac{W^R}{V^R}. \tag{3}$$

The superscript P is used to denote the fact that partisanship is just one dimension on which disparities may exist in the dilution of votes.

While our rationale is novel, we wish to be clear that we are not the first to propose this metric: McGann et al. (2015) argue that comparing party's shares of wasted votes is more appropriate than comparing absolute numbers. Nagle (2017) develops a measure of the difference in shares, labeling

[3]Another would be the ratio: $15/3.75 =$ a $4\times$ greater risk of being stopped.

[4]By definition, $\Pr(A \wedge B) = \Pr(A|B)\Pr(B)$. $\Pr(\text{vote for } p \text{ and wasted}) = \Pr(\text{vote wasted}|\text{vote for } p)\Pr(\text{vote for } p)$, which is estimated as $\left(\frac{W^p}{V^p}\right)\left(\frac{V^p}{V}\right) = \frac{W^p}{V}$.

[5]See also footnote 14 in Judge Griesbach's dissent in *Whitford vs. Gill*, 218 F. Supp. 3d 837 (W.D. Wis. 2016).

it a "voter-centric" approach in contrast to the "party-centric" approach of Stephanopoulos and McGhee (2015). MchGee (2017) terms the measure EG^{VC} (for "Efficiency Gap, Voter-Centric"), while Tapp (2019) calls the measure the "relative efficiency gap" (REG). Nagle, perhaps anticipating the court's argument in *Gill*, justifies the measure with reference to the normative goal of equalizing the average effectiveness of like-minded voters (p. 201), with Cover (2018) citing a similar justification.[6]

Our preference for the term Dilution Asymmetry over terminology employing "efficiency" stems from the reason hinted at by Nagle: efficiency refers to an optimization problem for *parties*, rather than a harm imposed on *voters*. For similar reasons, we prefer "dilution" to "waste" – waste is something that a (partisan) gerrymanderer seeks to minimize by efficiently allocating voters to districts, whereas dilution is something a voter (or group of voters) seeks to minimize.

While the normative justification for the waste disparity measure is attractive, equally compelling is the fact that it reestablishes proportionality as permissible for any base rate. To see why, again let $\sigma_{s,t}$ be the share of seats held by the Democratic Party in state s at time t, and $\nu_{s,t}$ the Democratic share of the two-party vote. It is straightforward to demonstrate that given equal turnout in districts, DA^P may be expressed as

$$DA^P_{s,t} = \frac{\nu_{s,t} - \sigma_{s,t}}{2\nu_{s,t}(1 - \nu_{s,t})}, \tag{4}$$

so $|DA^P_{s,t}|$ is minimized when the seat share equals the vote share.

The right panel of Figure 1 reproduces the left panel using Dilution Asymmetry instead of Efficiency Gap. The lens-shaped regions surrounding the 45° line indicate absolute values of Dilution Asymmetry consistent with a hypothetical standard. For example, any combination of voteshare and seatshare that falls in the green region has an absolute dilution asymmetry less than 0.16, while any combination that falls in the purple lens (inclusive of the green lens) is consistent with an absolute vote dilution less than 0.32. (Efficiency Gap ranges from -0.5 to 0.5, whereas dilution

[6] As Stephanopoulos and McGhee (2018, p. 1520) demonstrate via simulation, unless the proportion of votes going to a third party candidate is quite large, this measure also satisfies their "efficiency principle" – that is, the notion that holding a party's statewide voteshare constant, an increase in that party's seatshare should be reflected in greater advantage for that party as captured by the measure.

asymmetry ranges from -1 to 1; hence, for direct comparability, it is helpful to compare DA^P thresholds equal to twice those of the EG thresholds used in the left panel of the figure.) As is immediate from the figure, and in contrast to the corresponding figure above for EG, proportionality is preserved as permissible for any base rate vote share and threshold level of permissible absolute dilution asymmetry.

To get a sense of the potentially perverse consequences of base rate neglect in the Efficiency Gap measure that Dilution Asymmetry remedies, consider a hypothetical state consisting of 44,000 voters split across four districts. 33,000 of the voters are Republicans and 11,000 are Democrats. One way to achieve perfect proportionality of seats to votes would be to create three districts with 11,000 Republicans each, and one with 11,000 Democrats. This would yield an Efficiency Gap of $\frac{5{,}499-3\times 5{,}499}{44{,}000} \approx -0.25$, indicative of substantial unfairness *against Republicans*. Now move around some voters so that all four districts can be expected to elect a Republican: three of the districts split 9,000-2,000 Republican-to-Democrat, and the fourth splits 6,000-5,000 Republican-to-Democrat. The new value of the Efficiency Gap is $\frac{11{,}000-(3\times 3{,}499+499)}{44{,}000} \approx 0$, i.e., perfectly fair from an efficiency gap perspective (despite Republicans now holding all four districts).

What is driving large negative values of the Efficiency Gap measure in the first example isn't actually disparate harm to Republicans, but rather that there are simply more Republicans available to waste than Democrats. Adjusting for the base rate corrects for this issue: in the map with three Republican seats, the Dilution Asymmetry is zero (no bias), whereas in the map with four it is 0.67, indicating a substantially higher likelihood that Democratic votes are diluted.

3.3 Issues with Weighting "Cracked" and "Packed" Voters

From a normative perspective it is not altogether clear whether the harm to voters associated with being cracked – that is, being placed in a district where one is a minority with no hope of electing a favored candidate – is equally severe to that of being packed – that is, being placed in a district where one is part of a partisan supermajority (e.g., Bernstein and Duchin, 2017). For this reason, Nagle (2017) proposes differentially weighting minority votes and votes in excess of a majority. He argues that because packing is the "the most obvious way" to create partisan bias, we might wish to weight packed votes more than cracked ones. The normative justification he provides for a weighting scheme that gives higher weight to cracked votes is that "some voters feel happier when their candidate wins big because it enhances their confidence in being on the right side" (p. 200).

Letting $\lambda \in \mathbb{R}^+$ denote the weight on packed votes relative to cracked ones, Tapp (2019) refers to $\lambda = 1, 2$ as "the only natural cases." $\lambda = 1$ weights packed and cracked votes equally (as above), whereas $\lambda = 2$ assigns double the weight to packed votes. The latter is motivated by Judge Griesbach's dissenting opinion in *Gill*, who argues that

> ... in reality, all you need to win an election in a two-candidate race is one more vote than the other candidate, not 50%-plus-one of the total votes. For example, if the Indians defeat the Cubs 8 to 2, any fan might say that the Indians "wasted" 5 runs, because they only needed 3 to win yet scored 8. Under the Plaintiff's theory, however, the Indians needed 5 runs to beat the Cubs that day: 4 runs to reach 50% of the total runs, plus one to win. That, of course, is absurd.[7]

Insofar as the margin in the two-party vote is twice the excess of the majority over 50%+1, $\lambda = 2$ would seem to account for this concern.

As applied to baseball, Judge Griesbach's reasoning is sound; as applied to gerrymandering, however, it is severely flawed because of the constraint that districts must be as near as practicable of equal population size. If we assume for clarity that all voters vote for one of the two major parties, changing the partisan balance of a district while maintaining population size entails swapping Democratic voters for Republican ones or vice versa, not removing voters on the winning side until its size is reduced to the size of the losing one (Cover, 2018; Stephanopoulos and McGhee, 2018). Hence, we can reject this particular justification for $\lambda = 2$ out of hand.

From the perspective of the individual voter, a stronger basis for assessing the normative foundation for weighting packing and cracking unevenly than those offered by Nagle is agency-theoretic. Consider the principal-agent relationship between the representative and the voter. On the one hand, to the extent that a voter believes her interests will be better represented by a like-minded legislator than an unlike-minded one, being part of a large majority will result in less agency loss

[7] 218 F. Supp. 3d 837 (W.D. Wis. 2016), 958.

between the representative and that voter than being part of a minority. In other words, *packing may mitigate the adverse selection problem.* Note also that given increasing partisan polarization, the magnitude of loss from adverse selection is also increasing. This mitigates in favor of weighting cracked votes more heavily than packed ones.

On the other hand, favoring the minimization of cracking over packing will tend to reduce the average competitiveness of districts (Tapp, 2019, p. 603). Ceteris paribus, we might anticipate that noncompetitive districts are more prone to moral hazard problems than competitive ones, as greater electoral security could encourage incumbent shirking or corruption.[8] Accordingly, this would mitigate in favor of weighting packing as more damaging than cracking.

With this said, the weighting schemes proposed in Nagle (2017) and Tapp (2019) suffer from a more basic mathematical issue. Letting L^p denote the total number of votes for party p in districts won by the opposing party, and E^p the total number of votes for party p in excess of 50% plus 1 in districts won by party p, the authors propose a disparity metric based on a weighted measure of dilution risk:

$$\frac{L^p + \lambda E^p}{V^p}, \tag{5}$$

where λ is permitted to take on nonnegative values on the real line. The difficulty with this approach is that it is not a generalization of the conditional probability interpretation of vote dilution. Specifically, the dilution risk for a voter from party p can be decomposed into the sum $\Pr(packed|p) + \Pr(cracked|p)$. Examining (5), it is clear that for sufficiently high values of λ, the measure will exceed one and thus can no longer be interpreted as a (weighted) conditional probability. A simple correction is to constrain $\lambda \in [0, 1]$ and define the weighted dilution risk as

$$\frac{2(1-\lambda)L^p + 2\lambda E^p}{M^p + 2(1-\lambda)L^p + 2\lambda E^p} \in [0, 1],$$

where M^p denotes the number of voters from party p constituting bear majorities in districts won by that party. (Note that $V^p = M^p + L^p + E^p$.) When $\lambda = \frac{1}{2}$, this collapses to the conditional dilution probability that serves as the basis of Partisan Dilution Asymmetry expressed in equation

[8]This effect might be mitigated in the presence of competitive primaries.

(3). We adopt this approach below.

4 Analysis

4.1 Severity of Disparities in Simulated Redistricting Plans

Given the conceptual issues detailed above, how does EG fare in practice? Clearly, adjusting for the base rate will lead to magnitudes of EG and DA^P that will differ from each other. This has consequences for proposed constitutional standards. But the problem is potentially more severe: as shown in the toy example above, base rate neglect might actually lead the analyst to *misidentify the disadvantaged party.*

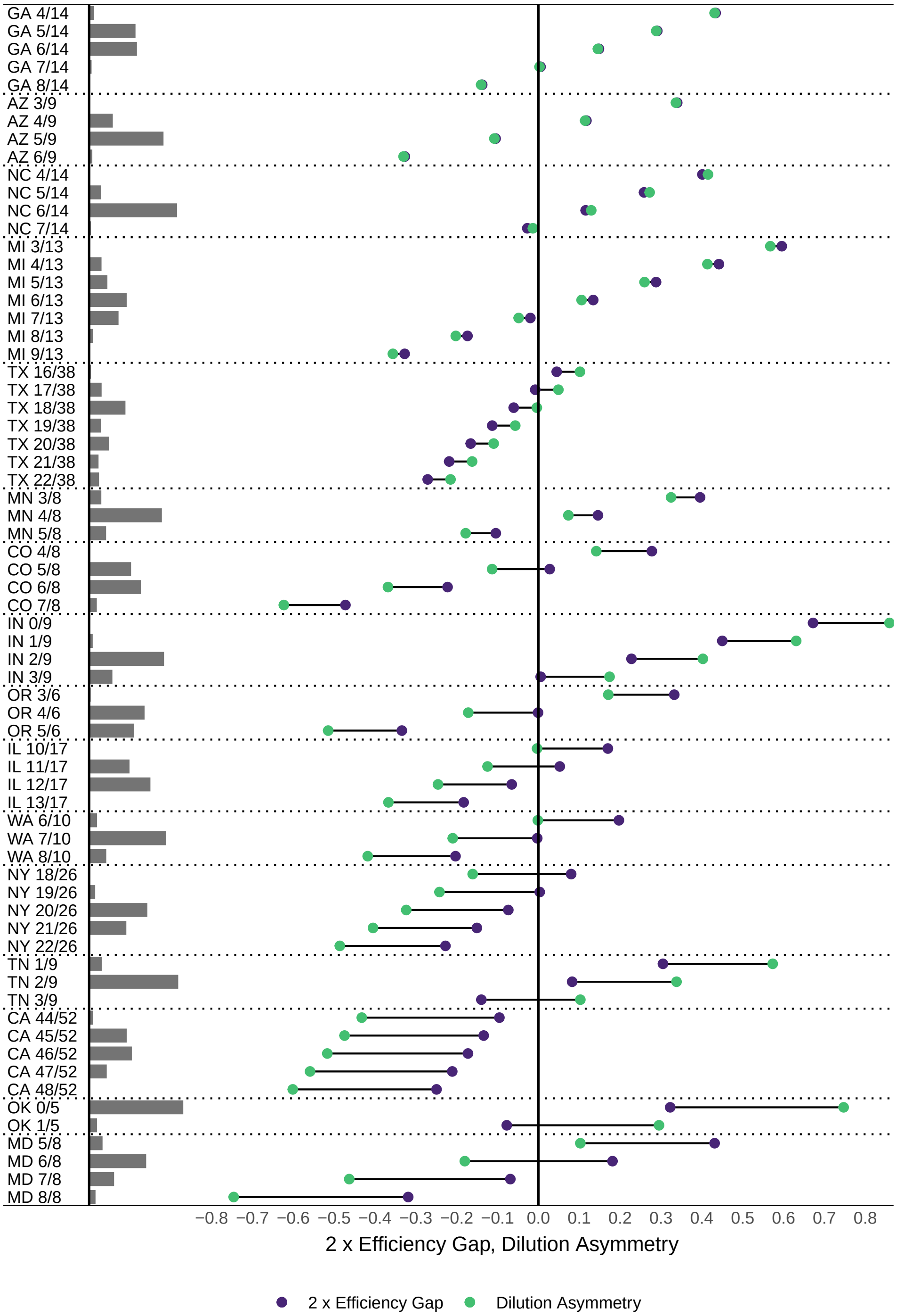
At the same time, it could be the case that maps that would yield substantively meaningful differences of this sort do not arise in practice. In this section, we consider differences in the magnitudes and signs of EG and DA^P across the range of feasible maps (as determined by simulation) in a sample of sixteen states with five or more congressional seats that vary in statewide partisan competitiveness.

For each state, we collected 2020 Presidential elections data and 5,000 simulated redistricting plans from the Algorithm-Assisted Redistricting Methodology (ALARM) Project (McCartan et al., 2022). ALARM's simulated plans are designed to satisfy an equal population requirement and additional redistricting criteria that may vary from state to state. The histogram running down the left side of Figure 2 depicts, for each state, the frequency of simulations that gave rise to a particular seat share for Democrats. (States are sorted by the absolute statewide vote margin.) So, for example, a large majority of simulations for Arizona (73%) anticipated a map with 5 of 9 districts with Democratic majorities; only a minuscule fraction generated a map with 3 of 9 Democratic majority districts. (Seat allocations never generated by simulation are assumed infeasible and not shown.)

The green circles in Figure 2 depict partisan Dilution Asymmetry calculated using the equal turnout formula in Equation (4), while purple circles depict the 2 $\times$ the Efficiency Gap using the formula in Equation (2).[9] (For reasons discussed above, EG is multiplied by two to put it on the

[9] We employ the equal turnout formulas for clarity of exposition: different simulations give rise to subtly different

Figure 2: Efficiency Gap and Partisan Dilution Asymmetry in a Sample of Simulated Districting Plans



same scale as DA^P.) First, consider the estimates for Georgia and Arizona. The 2020 presidential winner in these states was determined by less than half a percentage point, so the base rate issue should be negligible no matter what the seat allocation. This is precisely what we see: $2 \times EG$ and DA^P are nearly identical for those states. By comparison, the gap between the two measures for lopsided states like Maryland (33% margin for Biden), California (29% margin) and Oklahoma (33% margin for Trump) is quite substantial, averaging -0.37, -0.34, and 0.39, respectively.

Second, while we can sign the difference between the two measures, with Democratic-leaning states producing $EG - DA^P < 0$ and Republican ones $EG - DA^P > 0$, the substantive import of any difference may depend critically on the seat allocation. This is perhaps best illustrated using the state of Washington, which split approximately 60-40 for Biden in 2020. A strictly proportional delegation would anticipate 6 Democratic seats out of 10, and indeed DA^P falls extremely close to zero for maps generating that anticipated seat allocation. The estimate of $2 \times EG$, by contrast, is approximately 0.2, suggesting a slightly anti-Democratic gerrymander and a bias of EG away from zero. By contrast, a map with 7 Democratic seats minimizes the Efficiency Gap at close to zero, but the pro-Democratic shift in seats suggests a negative Dilution Asymmetry – in other words, the bias is toward zero.

Third, conditions do arise in the simulations where adjusting for the base rate flips the sign of the estimate of unfairness. This can be seen most dramatically for a subset of feasible maps in Colorado, Illinois, Maryland, New York, Oklahoma, and Tennessee. For example, nearly 40% of simulated redistricting plans for Illinois suggest a Republican advantage using EG, and a Democratic advantage using DA^P. For Maryland, 56% of simulated plans appear to favor the Republicans using EG but Democrats using DA^P.

4.2 Base Rate Neglect and Enacted Plans

In the next part of our analysis, we compare the Efficiency Gap and Partisan Dilution Asymmetry calculated using congressional elections returns data from 1982-2020 (MIT Election Data & Science Lab, 2017). We exclude states with fewer than five congressional districts; Louisiana (owing to its

turnouts across districts.

runoff system); and California since 2010, and Washington since 2004 (owing to their top-two system).

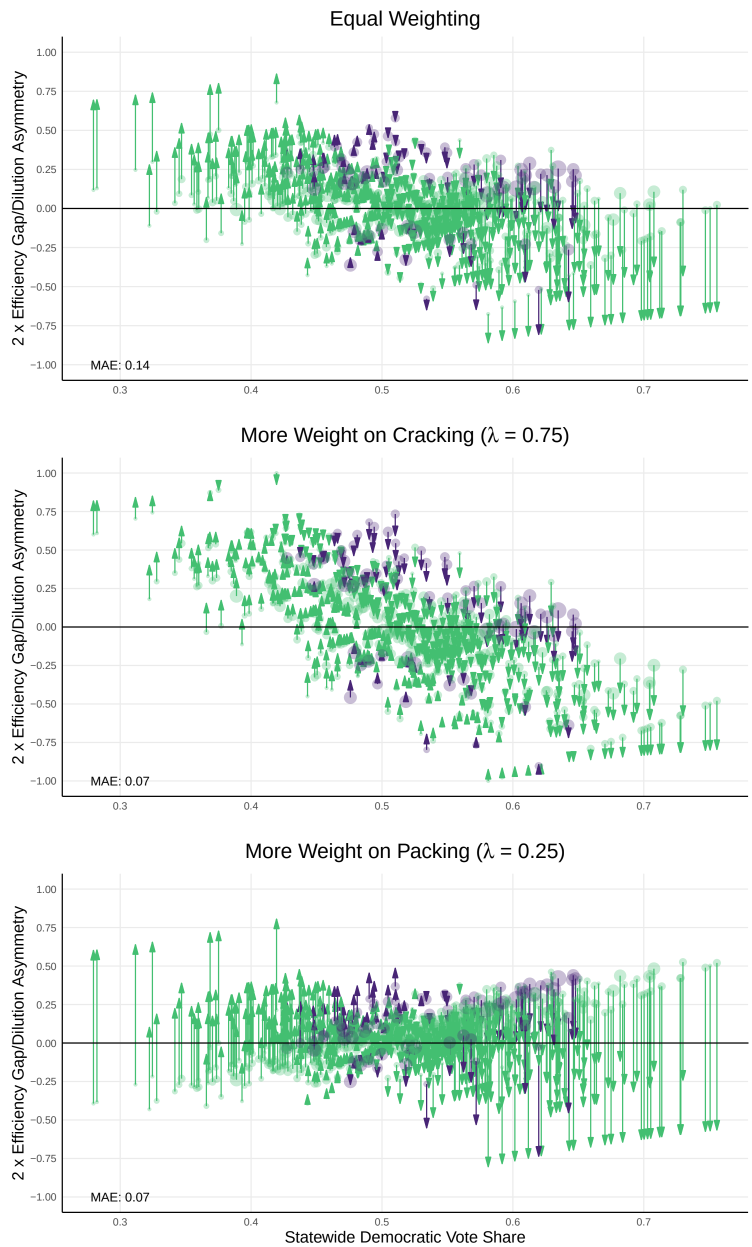
This leaves 662 state-year observations. Uncontested elections pose a potential challenge to estimation. Namely, the proportion of wasted votes cannot be determined when only one candidate appears on the ballot. To incorporate uncontested elections into our analysis of wasted votes in congressional elections, we estimate a Bayesian imputation model based on previous elections, as well as district- and election-cycle- factors (e.g., Stephanopoulos and Warshaw, 2020).[10]

The first panel of Figure 3 plots unweighted $2 \times EG$ and DA^P measures for the 547 observations against the (real or imputed) statewide Democratic vote share. Arrows point from EG to DA^P; the size of the circular token depicting $2 \times EG$ is proportional to the number of seats in the state's congressional delegation. As with the simulation analysis above, we multiply EG by 2 to facilitate comparability between the two measures. We also distinguish between outcomes that meet the proposed constitutional threshold given in Stephanopoulos and McGhee (2015) for district maps (displayed in green) and those that don't (in purple).

There are two immediate takeaways from the figure. First, unsurprisingly, EG and DA^P are highly, but imperfectly, correlated at $\rho = 0.64$. As indicated at the bottom left of the figure, the mean absolute error between the two measures is approximately 14 percentage points. Second, and more importantly, there are numerous instances in which EG and DA^P have different *signs*: in other words, EG may indicate mild-to-severe partisan bias in favor of Republicans (Democrats) when the DA^P measure, which adjusts for the base rate, indicates bias in favor of Democrats (Republicans). The occurrence of sign errors persists even when weighting packed and cracked

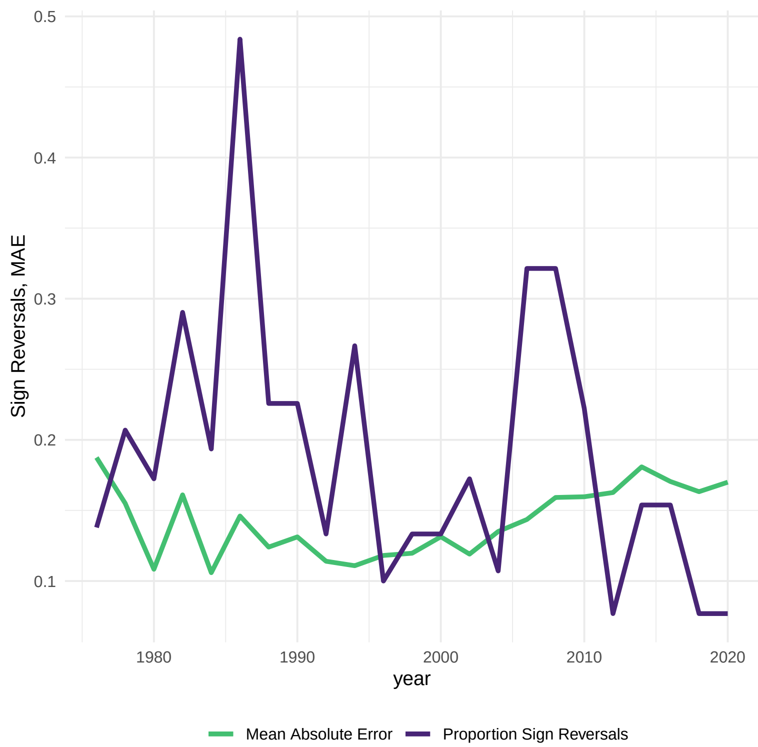
[10]We estimate the two-party vote share as follows: $V_{it} \sim Binomial(N_{it}, \pi_{it})$ where V_{it} is the vote tally for Democratic candidates in district i at time t and N_{it} is the turnout. V_{it} and N_{it} are set to zero for uncontested elections. The target parameter, $\pi_i \in (0, 1)$, is modeled based on linear predictors: $\pi_{it} = g^{-1}(x_{it}\beta)$ where g^{-1} is the inverse logit, and x_i includes lagged congressional vote share; district-level presidential vote share collected from Jacobson (2015) (for the period 1976-2014) and from Daily Kos (for 2016, 2018, and 2020)); incumbency status (whether an open-seat, Democratic incumbent or Republican incumbent), sourced from Hodges (2016) and online resources; region indicators based on the economic regions defined by the Bureau of Economic Advisors; and election-cycle indicators. We employ weak priors on the intercepts and slopes for each covariate: $\beta \sim N(0, 100)$. We modeled each decennial redistricting period from 1970-2020 separately in *rstan* using four chains, each running 1,000 iterations. The average imputed vote share, π_{it}, in uncontested races won by Democrats (72%) and Republicans (31%) closely resembles estimates from Stephanopoulos and Warshaw (2020)(71% and 31%) and Stephanopoulos and McGhee (2015) (70% and 32%). Our analysis utilizes the imputed vote share for uncontested elections and the observed two-party vote share in contested races.

Figure 3: Comparison of Efficiency Gap and Dilution Asymmetry using Real and Imputed Congressional Election Returns, 1976-2020



Data include all state-year observations with five or more seats, excluding (1) Louisiana, (2) Washington after 2002; and (3) California after 2008. Larger tokens indicate larger state delegations. Green denotes observations that would withstand the Stephanopoulos and McGhee (2015) constitutional standard, while purple denotes observations that would not meet the standard.

Figure 4: Differences in Efficiency Gap and Partisan Asymmetry Over Time, 1976-2000



voters asymmetrically (as is done in the second and third panels of the figure).

How severe is the sign reversal problem? In the full sample, we find that in 19% of cases, a failure to incorporate the base rate leads EG to take on a different sign than that of DA^P. As indicated by the purple line in Figure 4, the fraction of observations with sign reversals fluctuates over time, reaching a peak of almost 48% in 1986. By contrast, the issue was less pronounced in 2020, with fewer than 10% of observations characterized by sign reversals. The average absolute difference between the two measures, depicted in green, has risen slightly since its low in the early 1980s.

Of course, the import of sign reversals may be minimal if the magnitude of both measures is relatively small. In this respect, a more relevant piece of information is the fraction of cases that would, using EG as a diagnostic, warrant judicial intervention according to the standard set out by Stephanopoulos and McGhee, where the bias measured using the DA^P measure would be in the opposite direction. Of the 662 state-year combinations in our sample, 70 would fail to meet the

Table 1: Permissibility of Enacted Plans Per Efficiency Gap and Dilution Asymmetry based on a ±1.5 Standard Deviation Benchmark

		DA^P Standard		
		Impermissibly Pro-Democratic	Permissible	Impermissibly Pro-Republican
EG	Impermissibly Pro-Democratic	3.93	3.17	0.00
Standard	Permissible	3.02	80.06	3.78
	Imperssibly pro-Republican	0.00	2.11	3.93

standard; of these, 13 exhibit sign reversals between the measures. In other words, more than 19% of the states exhibiting *presumptively invalid* levels of partisan bias in favor of one party actually exhibit greater vote dilution of voters from the *other* party.

Suppose instead that we adopt a standard of labeling a plan impermissible if it is an outlier on an unfairness measure – for example, ±1.5 standard deviations above the mean observed level of that measure. If we do this for both EG and DA^P, we can represent the relative frequencies of impermissibly pro-Democratic, impermissibly pro-Republican, and permissible plans in a contingency table. Table 1 does this.

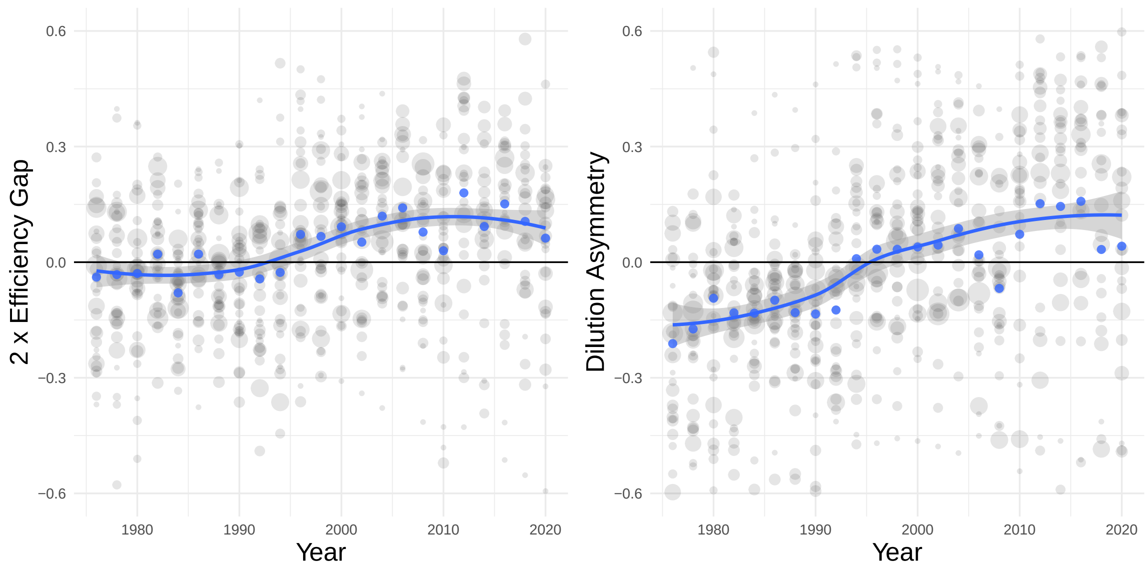
The Table suggests that based on this standard, the lion's share (around 80%) of cases are permissible using both standards. Suppose, arguendo, that the DA^P standard represents the *true* level of partisan unfairness. The figures represented in the table would imply a false positive rate of 6% for the Efficiency Gap, and a false negative rate of 48%.

4.3 The Historical Trajectory of Partisan Disparity

A common perception is that partisan polarization, coupled with sophisticated software for drawing district boundaries, has led to a situation in which partisan gerrymandering has reached levels heretofore unheard of, and that these gerrymanders have tended to favor Republicans. While the motive to gerrymander is indisputable, it is valuable to examine whether the historical data evinces strong trends in the prevalence and/or partisan direction of gerrymandering, and whether different measures cause us to draw erroneous conclusions.

With this in mind, the left panel of Figure 5 plots 2× the Efficiency Gap over time. The gray tokens correspond to individual states, with a locally smoothed regression line indicating the

Figure 5: Evaluating the Historical Incidence of Partisan Gerrymanders



The gray tokens indicate state-year observations, with larger tokens corresponding to larger House delegations. Loess curves in blue.

overall trend. Using the Efficiency Gap as our diagnostic, we would conclude that prior to the mid-1990s, the average absolute level of gerrymandering was low – implying, perhaps, that partisan gerrymanders favoring Democrats tended to cancel out those favoring Republicans. More recently, by contrast, there is a clear pro-Republican trend.[11]

Turning to the second panel, which displays dilution asymmetry over time, we also see a clear pro-Republican trend. But the figure also suggests that prior to the mid-1990s, gerrymandering evinced a decidedly pro-Democratic bias, with Republican gerrymanders failing to "cancel out" that bias.

5 Conclusion

We trace several issues with a popular diagnostic for partisan gerrymandering to base rate neglect: a failure to account for the underlying partisan distribution in a given state, and show that a

[11]Figure 7 of Jackman (2017), which conducts the same analysis using data from 1972-2016, displays a very similar pattern for the Efficiency Gap.

measure of vote dilution asymmetry that adjusts for the base rate avoids these problems. In practice, a failure to account for the base rate leads to erroneous conclusions about the magnitude *and direction* of partisan bias in districting plans. Also, whereas a longitudinal analysis based on actual vote returns in congressional races suggests that extreme partisan gerrymandering is a recent phenomenon, an analysis using our preferred measure shows a shift from pro-Democratic to pro-Republican bias in the mid-1990s.

While the recent *Rucho* decision suggests that federal courts consider partisan bias in gerrymandering non-justiciable, state courts that take this kind of bias into account when approving or rejecting districting proposals must still rely on diagnostic tools. Our analysis suggests that Efficiency Gap should be rejected as a measure of partisan disparity in favor of a more coherent dilution asymmetry measure that correctly adjusts for base rates.

Finally, while we explicate a partisan measure of dilution asymmetry in this paper, there is no reason that the methodology can and should not be applied along other dimensions, be they racial or geographic (e.g., urban/rural). A next step in this research agenda is to use the insights developed here to develop more robust measures of vote dilution in these areas.

References

Bernstein, Mira, and Moon Duchin. 2017. "A formula goes to court: Partisan gerrymandering and the efficiency gap." *Notices of the AMS* 64 (9): 1020–1024.

Caughey, Devin, Chris Tausanovitch, and Christopher Warshaw. 2017. "Partisan Gerrymandering and the Political Process: Effects on Roll-Call Voting and State Policies." *Election Law Journal: Rules, Politics, and Policy* 16 (4): 453–469.

Cover, Benjamin Plener. 2018. "Quantifying Partisan Gerrymandering: An Evaluation of the Efficiency Gap Proposal." *Stanford Law Review* 70: 1131.

Duchin, Moon. 2018. "Gerrymandering metrics: How to measure? What's the baseline?" *arXiv e-prints* p. arXiv:1801.02064.

Hodges, Zachary. 2016. "2016 Congressional Candidates.".
URL: *https://doi.org/10.7910/DVN/IARAAF*

Jackman, Simon. 2017. "Assessing the Current North Carolina Congressional Districting Plan: Amended Report.".

Jacobson, Gary C. 2015. "It's nothing personal: The decline of the incumbency advantage in US House elections." *The Journal of Politics* 77 (3): 861–873.

Jiang, Wenxin, Gary King, Allen Schmaltz, and Martin A Tanner. 2020. "Ecological Regression with Partial Identification." *Political Analysis* 28 (1): 65–86.

McCartan, Cory, Christopher Kenny, Tyler Simko, Shiro Kuriwaki, George Garcia III, Kevin Wang, Melissa Wu, and Kosuke Imai. 2022. "ALARM Project: 50-State Redistricting Simulations.".
URL: *https://doi.org/10.7910/DVN/SLCD3E*

McGann, Anthony J., Charles Anthony Smith, Michael Latner, and J. Alex Keena. 2015. "A Discernable and Manageable Standard for Partisan Gerrymandering." *Election Law Journal* 14 (4): 295–311.

McGhee, Eric M. 2014. "Measuring Partisan Bias in Single-Member District Electoral Systems." *Legislative Studies Quarterly* 39 (1): 55–85.

MchGee, Eric. 2017. "Measuring Efficiency in Redistricting." *Election Law Journal: Rules, Politics, and Policy* 16 (4): 417–442.

MIT Election Data & Science Lab. 2017. "U.S. House 1976–2020.".
URL: *https://doi.org/10.7910/DVN/IG0UN2*

Nagle, John F. 2017. "How Competitive Should a Fair Single Member Districting Plan Be?" *Election Law Journal* 16 (1): 196–209.

Stephanopoulos, Nicholas O., and Christopher Warshaw. 2020. "The Impact of Partisan Gerrymandering on Political Parties." *Legislative Studies Quarterly* 45 (4): 609–643.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *University of Chicago Law Review* 82: 831–900.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2018. "The Measure of a Metric: The Debate over Quantifying Partisan Gerrymandering." *Stanford Law Review* 70 (5): 1503–1568.

Tam Cho, Wendy K. 2017. "Measuring Partisan Fairness: How Well Does the Efficiency Gap Guard Against Sophisticated as well as Simple-Minded Modes of Partisan Discrimination?" *University of Pennsylvania Law Review Online* 166 (1): 2.

Tapp, Kristopher. 2019. "Measuring Political Gerrymandering." *The American Mathematical Monthly* 126 (7): 593–609.

Tufte, Edward R. 1973. "The relationship between seats and votes in two-party systems." *The American Political Science Review* 67 (2): 540–554.

Warrington, Gregory S. 2019. "A comparison of partisan-gerrymandering measures." *Election Law Journal: Rules, Politics, and Policy* 18 (3): 262–281.