Challenging Partisan Gerrymandering

Simon Jackman*

March 27, 2018

*Professor of Political Science and Chief Executive Officer, United States Studies Centre, University of Sydney.

Abstract

Partisan gerrymandering is rife in the United States, where electoral redistricting in many jurisdictions remains in the hands of state legislatures and governors. U.S. courts have been reluctant to intervene, unsure of what partisan gerrymandering is or when it generates constitutional harm. Social scientists have long known that partisan gerrymandering can be reliably detected and remedied. A new measure of partisan gerrymandering — the efficiency gap (Stephanapolous and McGee 2015) — has served as the basis of three recent court cases in the US, each producing verdicts in favour of plaintiffs alleging unconstitutional partisan gerrymandering (notably, *Gill v Whitford*, currently pending adjudication on appeal by the Supreme Court of the United States). I explain the efficiency gap, presenting exhaustive analyses of its properties in state legislative and Congressional elections in the United States, work that figured prominently in the litigation mentioned above. My analysis provides a practical, easily-applied, quantitative standard for assessing the constitutionality of redistricting plans, challenging egregious partisan gerrymanders and restoring integrity to American elections.

Partisan gerrymandering is routinely cited as one of the more pernicious features of contemporary American politics. Partisan control of the drawing of electoral boundaries (redistricting) creates incentives and opportunity for party operatives to "pack and crack" opposition voters, creating an asymmetric distribution of vote shares across districts, conferring lasting electoral advantage for the party controlling redistricting. Incumbents isolated from meaningful general electoral challenges are hypothesized to be more responsive to their party's primary constituencies; parties systematically disadvantaged by partisan gerrymandering face difficulties winning seats, recruiting candidates and sustaining themselves organisationally.

American courts have been reluctant to entertain claims that partisan gerrymandering causes constitutional harm, in violation of the 1st and/or 14th amendments to the US Constitution. The U.S. Supreme Court has ruled that partisan gerrymandering exists and is justiciable, but has hitherto been failed to be swayed by measures of gerrymandering proposed by litigants.

Plaintiffs in three recent cases have utilised a new measure of partisan gerrymandering, the *efficiency gap*. Here I draw on my experience as an expert witness in two recent U.S. court cases in which plaintiffs relied on my research — and testimony — detailing the properties of the efficiency gap, how it compares with other measures of gerrymandering and why courts can rely on the efficiency gap as a valid and reliable measure of gerrymandering and its constitutional harms.

1 Measuring features of electoral systems

The use of quantitative methods to characterise electoral systems has a long history in political science and statistics. Litigation and judicial scrutiny of electoral systems — principally in the United States — has seen considerable cross-fertilisation between law and the social sciences.

Some definition of terms and my scope is required. An *electoral system* is that set of legally binding arrangements that determine how citizens' votes are translated into seats in a legislature or other representative body.

Both votes and seats lend themselves to quantitative analysis. While individual votes are discrete variables — a selection from or a rank ordering over a finite set of candidates — aggregations of individual votes are continuous variables. Similarly for seats in a legislative, which are usefully labelled by the *political party* of the incumbent. The decision rule used in most legislative bodies — majority rule deciding binary questions — means that aggregations of seats by political party are critical quantities, both practically and analytically.

Moreover, similarities in the structure of electoral systems — the rules governing the mapping of votes into seats, the party affiliations of candidates — across elections and/or jurisdictions means that vote shares and seat shares are comparable across time and space, or, in the language of statistics, *exchangeable*. This opens the way to analysis that allows scholars and practitioners to (1) use statistical modeling to make rigorous inferences about the properties of electoral systems; (2) to identify anomalous cases or *outliers*, particular data points that depart from general patterns. In turn, rigorous, evidence-based identification of outlying data points is relevant to lawyers seeking to determine if a particular election outcome or electoral system is so unusual — at least relative to other comparable cases — that it ought to attract further scrutiny or judicial intervention.

Many elements of electoral systems have attracted the attention of lawyers and the courts, too many to be all considered here. The "apportionment revolution" in American electoral law spurred interest in measures of malapportionment, the extent to which electoral districts have unequal numbers of eligible voters. The Voting Rights Act has given rise to considerable interplay between law and social science; for example, measuring racially polarised voting is typically a delicate inferential problem, an instance of "cross-level" or "ecological" inference, or more generally, an "inverse" problem. Legal challenges to voter-id laws, studies of precinct consolidation or ballot order effects, and forensic investigations of electoral fraud also make extensive use of quantitative data and methods of analysis.

2 Partisan symmetry

My focus here is on a particular facet of an electoral system: *partisan symmetry* in the translation of votes into seats in the legislature, a question that has long interested American courts and central to ongoing litigation in various settings in the United States.

2.1 District-based systems

At this point the focus of my analysis does narrow. First, the question of partisan symmetry is only meaningfully operational in electoral systems with *districts*, a geographic partitioning of a jurisdiction, typically into a set of mutually exclusive and exhaustive regions, and — at least in the contemporary United States — of approximately the same population size.

Electoral systems that allocate legislative seats under a deterministic, *proportional representation* rule (a party's seat share is its vote share, perhaps subject to a minimum viability, threshold rule) do not rely on districts and are symmetric with respect to political parties by definition: the seat share for party A is simply its vote share and the same is true for *all* other parties, again subject to any thresholding rule. Accordingly, my focus here is on district based systems, which are the norm at both Federal and state level in the United States.

Second — and again reflecting my focus on the American case — while partisan symmetry can be defined for arbitrarily many parties, almost all political science and legal examinations of partisan symmetry focus on the two-party case. This is perfectly reasonable for the analysis of American electoral systems, but would pose an issue in an examination of British elections. A further simplification is that the two-party district systems are single-member, simple plurality (SMSP) systems; again, this has been the norm in American state and Federal elections for decades.

The focus on two-party, district-based SMSP systems vastly simplifies the analysis of electoral systems in general and the measurement of partisan asymmetry. In this context, partisan asymmetry is simply the property that the mapping of votes into seats is more advantageous to one party than the other. Constitutional questions arise in the American context when this asymmetry violates the Equal Protection provisions of the Fourteenth Amendment and/or elements of the First Amendment of the U.S. Constitution.

2.2 Districting plans, partisan asymmetry and gerrymandering

Partisan asymmetry can arise through malapportionment, but this has been unconstitutional in the United States for decades. Instead, *partisan gerrymandering* — drawing districts that induce asymmetry, typically by creating a set of districts that help one party win an excess of districts relative to its jurisdiction-wide level of support — is the typical way that partisan asymmetry arises in the United States today and is at the heart of recent litigation in the United States.

What might constitute evidence of partisan gerrymandering? One indication might be a series of elections conducted under the same districting plan in which a party's seat share (S) is unusually large (or small) relative to its vote share (V). Gerrymandering generates this outcome via "packing", creating a relatively small number of districts that have unusually large proportions of partisans from party A relative to party B. In other districts in the jurisdiction, party A supporters are "cracked" such that they never (or seldom) constitute a majority (or a plurality), making those districts "safe" for party B. This districting plan helps ensure that party B wins a majority of seats even though party A has a majority of support across the jurisdiction; or, at the very least, the districting plan helps ensure that party B's seat share exceeds its vote share in any given election.

It is conventional in political science to say that such a districting plans with these properties allows party B to "more efficiently" translate its votes into seats, relative to the way the plan translates party A's votes into seats. This nomenclature is telling and will be revisited in the discussion of the *efficiency gap* measure, below.

2.3 Seats-votes curves and a definition of partisan asymmetry

Electoral systems translate parties' vote shares (V) into seat shares (S). Both V and S are proportions. Plotting the two quantities V and S against one another yields the "seats-votes" curve, a staple in the analysis of electoral systems and districting plans.

Two seats-votes curves are shown in Figure 1. The curve labeled "Cube Law" in Figure 1 is a good approximation to the relationship between votes and seats observed in SMSP systems, generated assuming that $S/(1-S) = [V/(1-V)]^3$, an approximation for the lack of proportionality typically observed in single-member district systems, though hardly a "rule" or "law.". The other curve (actually, a line) shows the linear relationship between seats and votes observed under proportional representation systems.

Single-member district systems tend to produce steeper and non-linear seats-votes curves. In single-member, simple plurality (SMSP) systems with an approximately symmetric mix of districts (in terms of partisan leanings) close to the 50-50 point on the horizontal "votes" axis, large changes in seat shares (S) can result from relatively small changes in vote shares (V) at the middle of the distribution of district types. Note however that the steepness of the seats-votes curve — a property known as *responsiveness* — implies nothing about partisan symmetry.

In the two-party case being considered here, partisan asymmetry is the property that for the seats-vote curve $S(V)$, there exists at least one value of V, v^* such that $S(v^*) \neq 1 - S(1 - v^*)$. Note that in a proportional representation (PR) electoral systems, seats-votes curves are 45 degree lines by design or $S(V) = V$ and symmetry always holds.

3 Partisan asymmetry as partisan bias

Political scientists have devoted considerable effort to examining the case $V = .5$, corresponding to an even split of the two-party, jurisdiction-wide vote. A seats-votes curve that does not run through the 50-50 point is not only asymmetric, but is said to exhibit *partisan bias* in favor of the party for whom $S(V = .5) > .5$. Indeed, the quantity $S(V = .5) - .5$ — the vertical offset of the seats-votes curve from the 50-50 point — operationalizes partisan bias.

Figure 2 shows three seats-votes curves, with the graph clipped to the region $V \in [.4, 6.]$ and $S \in [.4, .6]$ so as to emphasize the nature of partisan bias. The blue, positive bias curve "lifts" the seats-votes curve; it crosses $S = .5$ with $V < .5$ and passes through the upper-left quadrant of the graph. That is, with positive bias, a

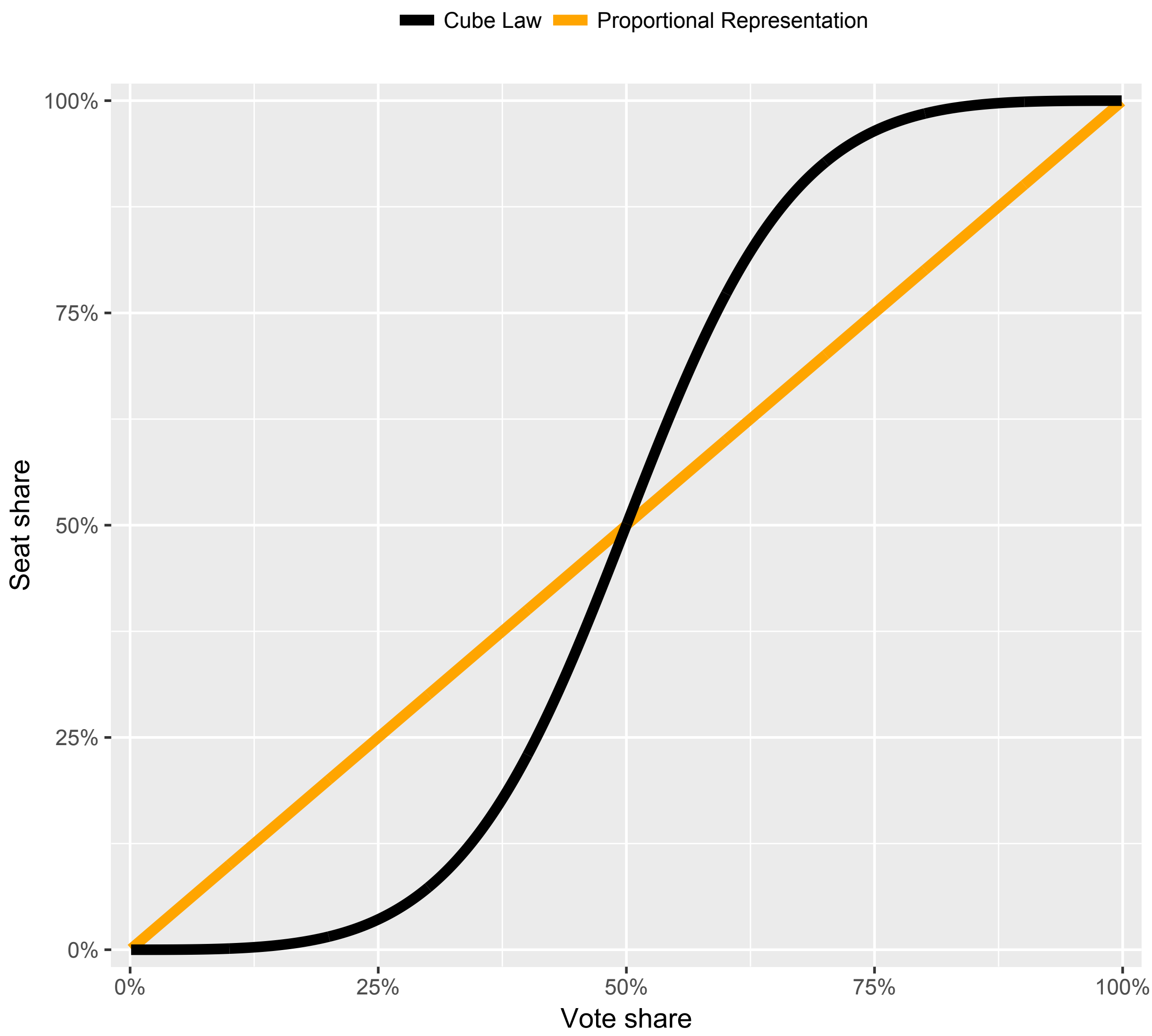

Figure 1: Two Theoretical Seats-Votes Curves

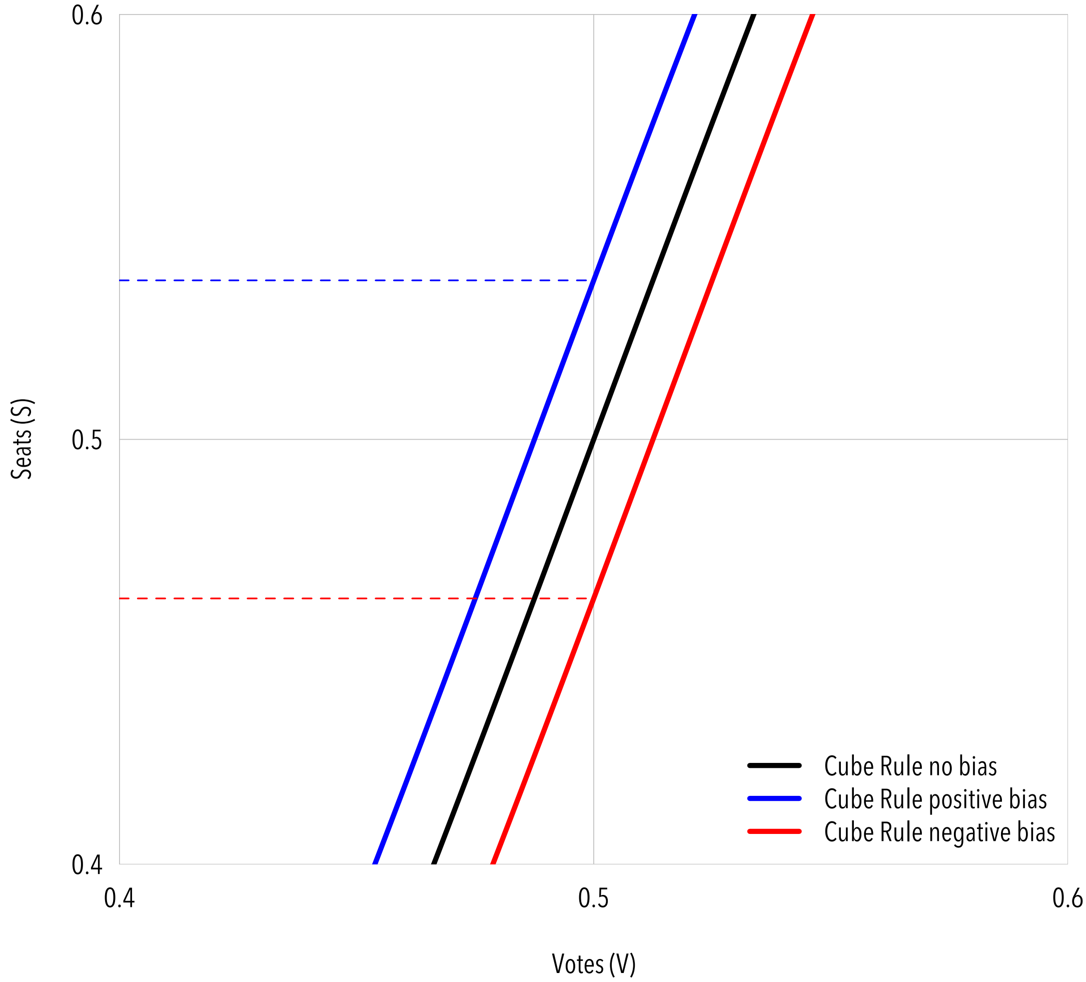

Figure 2: Theoretical seats-votes curves, with different levels of partisan bias. This graph is "zoomed in" on the region $V \in [.4, .6]$ and $S \in [.4, .6]$; the seats-votes "curves" are approximately linear in this region.

party can win a majority of the seats with *less* then a majority of the jurisdiction-wide or average vote; equivalently, if the party wins $V = .5$, it can expect to win *more* than 50% of the seats. Conversely, with negative bias, the opposite phenomenon occurs: the party can't expect to win a majority of the seats until it wins more than a majority of the jurisdiction-wide or average vote.

3.1 Multi-year method

With data from multiple elections under the same district plan, partisan bias can be estimated by fitting a seats-votes curve to the observed seat and vote shares, typically via a simple statistical technique such as linear regression; this approach has a long and distinguished lineage in both political science and statistics (e.g., Edgeworth, 1898; Kendall and Stuart, 1950; Tufte, 1973).

For instance, given a data set consisting of n data points, each a (S_i, V_i) ordered pairs, $i = 1, \ldots, n$, the regression

$$\ln\left(\frac{S_i}{1-S_i}\right) = \beta_0 + \beta_1 \ln\left(\frac{V_i}{1-V_i}\right) + \epsilon_i \tag{1}$$

could be estimated with β_0 interpretable as measuring partisan bias; i.e, after some re-arranging of terms,

$$E(S_i|V_i = .5) = \frac{\exp(\beta_0)}{1+\exp(\beta_0)} \tag{2}$$

with partisan bias defined as this quantity minus .5. Niemi and Fett (1986) referred to this method of estimating the partisan bias of an electoral system as the "multi-year" method, reflecting the fact that the underlying data comes from a sequence of elections.

This approach is of limited utility when assessing a new or proposed districting plan. More generally, it is of no great help to insist that a sequence of elections must be conducted under a districting plan before the plan can be assessed. Few plans stay intact long enough to permit reliable analysis in this way: districting plans in the United States might generate as many as five elections between decennial censuses.

One solution is to combine multiple plans and/or jurisdictions in a single analysis, so as to estimate average levels of partisan bias. For instance, Niemi and Jackman (1991) estimated average levels of partisan bias in state legislative districting plans, collecting data spanning multiple decades and multiple states, and grouping districting plans by the partisanship of the plan's authors (e.g., plans drawn under Republican control, Democratic control, mixed, or independent).

Assessing the properties of a districting plan after a tiny number of elections — or *no* elections — requires some assumptions and/or modeling. A single election yields just a single (V, S) data point, through which no unique seats-vote curve can be fitted. In this case partisan bias can not be estimated without further assumptions. Absent *any* actual elections under the plan, one might examine votes from a previous election, say, with precinct level results re-aggregated to the new districts.

3.2 Uniform swing

One approach—dating back to Sir David Butler's (1974) pioneering work on British elections—is the uniform partisan swing approach. I introduce some notation as to explain the method.

Let $\mathbf{v} = (v_1, \ldots, v_n)'$ be the set of vote shares for party A observed in an election with n districts. Party A wins seat i if $v_i > .5$, assuming just two parties (or defining v as the share of two-party vote). s_i is a binary indicator denoting who wins seat i; i.e., $s = 1$ if $v_i > .5$ and otherwise $s_i = 0$. Party A's seat share is $S = \frac{1}{n}\sum_{i=1}^{n} s_i$. V is the jurisdiction-wide vote share for party A, and if each district had the same number of voters $V = \bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$, the average of the district-level v_i. Districts are never *exactly* equal sized, in which case V is defined as follows: let t_i be the number of voters in district i, and $V = \sum_{i=1}^{n} t_i v_i / \sum_{i=1}^{n} t_i$.

The uniform swing approach perturbs the observed district-level results $\mathbf{v}$ by a constant factor δ, corresponding to a hypothetical amount of *uniform swing* across all districts. For a given δ, let $v_i^* = v_i + \delta$ which in turn generates $V^* = V + \delta$ and an implied seat share S^*. Now let δ vary over a grid of values ranging from $-V$ to $1 - V$; then V^* varies from 0 to 1 and a corresponding value of S^* can also be computed at every grid point. The resulting set of (V^*, S^*) points are then plotted to form a seats-vote curve (actually, a step function). Partisan bias is simply "read off" this set of results, computed as $S^*|(V^* = .5) - .5$.

There is an elegant simplicity to this approach, taking an observed set of district-level vote shares $\mathbf{v}$ and shifting them by the constant δ. The observed distribution of district level vote shares observed in a given election is presumed to hold under *any* election that might be observed under the redistricting plan, save for the shift given by the uniform swing term δ.

3.3 Critiques of partisan bias

Among political scientists, the uniform swing approach was criticized for its determinism. Swings are never exactly uniform across districts. There are many permutations of observed vote shares that generate a statewide vote share of 50% other than simply shifting observed district-level results by a constant factor. A less deterministic approach to assessing partisan bias was developed over a series of papers by Gary King and Andrew Gelman in the early 1990s (e.g., Gelman and King, 1990). This approach fits a statistical model to district-level vote shares — and, optionally, utilizing available predictors of district-level vote shares — to model the way particular districts might exhibit bigger or smaller swings than a given level of state-wide swing. Perhaps one way to think about the approach is that it is "approximate" uniform swing, with statistical models fit to historical election results to predict and bound variation around a state-wide average swing. The result is a seats-vote curve and an estimate of partisan bias that comes equipped with uncertainty measures, reflecting uncertainty in the way that individual districts might plausibly deviate from the state-wide average swing yet still produce a state-wide average vote of 50%.

3.4 Partisan bias after LULAC

The King and Gelman model-based simulation approaches remain the most sophisticated methods of generating seats-votes curves, extrapolating from as little as one election to estimate a seats-votes curve and hence an estimate of partisan bias. Despite the technical sophistication available in estimating partisan bias, legal debate has centered on a more fundamental issue, the *hypothetical* character of partisan bias itself. Recall that partisan bias is defined as "seats in excess of 50% *had the jurisdiction-wide vote split 50-50.*" The premise that $V = .5$ is the problem, since this will almost always be a counter-factual or hypothetical scenario. The further V is away from .5 in a given election, the counter-factual being contemplated (when assessing the partisan bias of a districting plan) becomes all the more speculative.

The last time the U.S. Supreme Court considered the question of partisan gerrymandering — the *LULAC* case, with judgement rendered in 2006[1] — a majority of the Court declined to endorse measures of partisan bias as evidence of partisan gerrymandering. While *Veith*[2] opened the door on the justicability of partisan gerrymandering claims, *LULAC* sent political scientists and lawyers back to the drawing board. Subsequently, Bernie Grofman and Gary King — at the time, two of the most visible political scientists in redistricting litigation — authored a review article that at time read almost as a lament for partisan bias as a measure of partisan asymmetry in districting plans (Grofman and King, 2008). If a concept as well-studied and as rigorously founded as partisan bias couldn't satisfy the Court, then what could?

4 The Efficiency Gap

In two articles in 2014 and 2015 — one in a political science journal (McGhee, 2014), one in a law review (Stephanopoulos and McGhee, 2015) — a new partisan asymmetry measure entered the fray, the *efficiency gap* (*EG*). The efficiency gap addresses several of the criticisms of partisan bias raised by SCOTUS justices open to the justiciability of partisan gerrrymandering claims. Critically, the calculation *and* interpretation of the *EG* is not tied to any counter-factual election outcome.

4.1 Wasted votes

Stephanopoulos and McGhee (2015) derive the efficiency gap measure beginning with the concept of wasted votes. A party only needs v_i = 50% + 1 of the votes cast for two-party candidates to win district i. Anything more are votes that could have been deployed in other districts. Conversely, votes in districts where the party doesn't

[1] *League of United Latin American Citizens v. Perry*, 548 U.S. 399 (2006).
[2] *Vieth v. Jubelirer*, 541 U.S. 267 (2004).

win are "wasted," from the perspective of generating seats: districts in which $v_i < .5$ yield no seats.

4.2 Partisan asymmetry in wasted votes: a hallmark of gerrymandering

Wasted votes get at the core of what partisan gerrymandering is, and how it operates. A gerrymander against party A creates a relatively small number of districts that "lock up" a lot of its votes ("packing", creating districts with $v_i > 50\%$) and a larger number of districts that disperse votes through districts won by party B ("cracking", creating districts with $v_i < 50\%$). To be sure, both parties are wasting votes. But partisan advantage ensues when *one party is wasting fewer votes than the other*, or, equivalently, more efficiently translating votes into seats. Note also how the efficiency gap measure is closely tied to asymmetry in the distribution of v_i.

Some notation will help make the point more clearly. If $v_i > 50\%$ then party A wins the district and $s_i = 1$; otherwise $s_i = 0$. The efficiency gap is defined by McGhee (2014, 68) as "relative wasted votes" or

$$EG = \frac{W_B}{n} - \frac{W_A}{n}$$

where

$$W_A = \sum_{i=1}^{n} s_i(v_i - .5) + (1 - s_i)v_i$$

is the sum of wasted vote proportions for party A and

$$W_B = \sum_{i=1}^{n} (1 - s_i)(.5 - v_i) + s_i(1 - v_i)$$

is the sum of wasted vote proportions for party B and n is the number of districts in the jurisdiction. If $EG > 0$ then party B is wasting more votes than A, or A is translating votes into seats more efficiently than B; if $EG < 0$ then the converse, party A is wasting more votes than B and B is translating votes into seats more efficiently than A.

Again note that under these definitions, both parties waste votes as they win and lose districts. The efficiency gap measures the difference in wasted vote rates between the parties. If $EG = 0$ then both parties are wasting votes at the same rate and the electoral system is not demonstrating any partisan asymmetry.

4.3 Example: North Carolina's 2016 Congressional elections

To demonstrate the relative simplicity of the efficiency gap and its calculation, consider the following table of election returns from North Carolina's 2016 Congressional elections. Democrats won 3 of North Carolina's 13 Congressional districts: votes cast for Democratic candidates in the 10 districts won by Republicans constitute "wasted votes" with respect to the *EG* calculation. Wasted votes in seats where a party wins are simply those votes beyond the 50% (plus one) needed to win.

Table 1: North Carolina 2016 Congressional elections.

District	Dem Votes	Rep Votes	Winner	Dem Wasted	Rep Wasted
1	240,661	101,567	D	69,546	101,567
2	169,082	221,485	R	169,082	26,201
3	106,170	217,531	R	106,170	55,680
4	279,380	130,161	D	74,609	130,161
5	147,887	207,625	R	147,887	29,868
6	143,167	207,983	R	143,167	32,407
7	135,905	211,801	R	135,905	37,947
8	133,182	189,863	R	133,182	28,340
9	139,041	193,452	R	139,041	27,205
10	128,919	220,825	R	128,919	45,952
11	129,103	230,405	R	129,103	50,650
12	234,115	115,185	D	59,464	115,185
13	156,049	199,443	R	156,049	21,696
Total	**2,142,661**	**2,447,326**	**3D/10R**	**1,592,124**	**702,859**

In this case we observe that Democrats wasted many more votes than Republicans, more than twice as much. Expressed a proportion of total votes cast, the difference in the parties' wasted votes is

$$\text{EG} = \frac{702,859 - 1,592,124}{2,142,661 + 2,447,326} = -0.194$$

The negative value of the efficiency gap in this example reflects the fact that Democrats have more wasted votes than Republicans; without loss of generality we could have just as easily defined the efficiency gap the other way around, with Republican wasted votes minus Democratic wasted votes.

Note that no statistical modeling is required to compute *EG*; nothing more than a hand calculator or a spreadsheet program is needed. The measure takes actual election results as its input; unlike partisan bias there is no recourse to a hypothetical 50-50 election, nor the invocation of an abstract seats-vote curve. The efficiency gap measure has a political realism to it and a close connection to the processes at work in gerrymandering: packing and cracking.

As the North Carolina data makes clear, the three seats that the Democrats won were won with large margins: 67-33 in District 12, 68-32 in District 4 and 70-30 in District 1. Only one of the 10 Republican wins is as lop-sided, the 67-33 Republican win in District 3. This is a vivid demonstration of packing and cracking at work. The end results is that Republicans were able to convert their votes into seats much more efficiently than Democrats, wasting far fewer votes. With 53.3% of the statewide, two-party vote for Congress, Republicans won 10/13 = 76.9% of the seats. The underlying partisan asymmetry that generates that result is reflected in the large, negative value of the efficiency gap in North Carolina in 2016, -0.194.

5 Towards a judicially administerable standard

Having ventured that efficiency gap has validity as a measure of partisan asymmetry, our practical challenge is to specify for the courts benchmarks or thresholds and to give courts — and those directly impacted by their decisions — confidence in those benchmarks. Specifically, careful quantitative analysis helps answer the following questions for the courts.

1. How large is large? When is the efficiency gap so large that a court can be confident that breaches of constitutional protections are occuring, so as warrant judicial intervention?

2. Districting plans typically stay in place for multiple elections in the United States and are typically subject to court challenge early in their life, or even before the plan is operational. Can a court be confident that the EG score computed from one election under a districting plan — or even with *no elections* under the plan — is a valid indicator of the partisan asymmetry of the districting plan and not of chance factors?

6 Historcal analysis

To address these question, I computed *EG* scores for a wide array of Congressional elections in the United States, 1972-2016, spanning a wide array of districting plans and political circumstances. I begin the analysis in 1972 since districting plans and sequences of elections from 1972 onwards can be reasonably considered to be from the post-malapportionment era. I further restrict my analysis to states with at least seven Congressional districts, so as to form a relatively homogenous group of states:

redistricting in jurisdictions with a small number of is qualitatively different problem than redistricting in jurisdictions with moderate to large states.[3]

- a large, canonical data set on candidacies and results in Congressional elections, 1898-1992, collected and published by Professor Gary King (Harvard University);
- more recent Congressional elections data from a data file maintained and collected by *Congressional Quarterly*; and
- presidential election returns aggregated to Congressional districts, a data collection maintained by Professor Gary Jacobson (University of California, Los Angeles) that has canonical status in the political science profession.

These data collections are augmented with data collected from the Associated Press, the Cook Political Report and Daily Kos for the 2016 Congressional elections.

Figure 3 provides a graphical depiction of the elections that satisfy the selection criteria described above.

- Alaska, Arkansas, Connecticut, Delaware, Hawaii, Idaho, Iowa, Kansas, Louisiana, Maine, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Oklahoma, Oregon, Rhode Island, South Dakota, Utah, Vermont, West Virginia and Wyoming all drop out of the analysis entirely, because they have six or fewer congressional districts throughout the period under analysis or because of the use of a run-off system (Louisiana).
- South Carolina, Kentucky, Colorado and Arizona only supply data for part of the timespan because they have too few districts for part of the period.

In summary, the data available for analysis span 7,949 district-level Congressional elections, from 512 state-level Congressional elections across 25 states and 23 election years.

6.1 Grouping elections into redistricting plans

Districting plans remain in place for sequences of elections. An important component of my analysis involves tracking the efficiency gap across a series of elections held under the same districting plan. A key question is how much variation in the *EG* is observed *within* districting plans, versus variation in the *EG between* districting plans.

[3] I also exclude Lousiana from this analysis because that state uses a unique run-off election system when no candidate receives a majority of the vote in the general election.

If the efficiency gap is a feature of a districting plan *per se*, there should be a small amount of within-plan variation in *EG* scores relative to between-plan variation. To perform this analysis I first group sequences of elections within states by the districting plan in place at the time.

Figure 3 displays how the elections available for analysis group by districting plan. Districts are typically redrawn after each decennial census; the first election conducted under new district boundaries is often the "2" election (1982, 1992, etc). Occasionally there is just one election under a plan: examples include Virginia, Florida and North Carolina in 2016 and Texas in 2002, 2004 and 2006.

6.2 Uncontested races

Uncontested races are not uncommon: for 14.0% of the districts in this analysis, it isn't possible to directly compute a two-party vote share, either because the seat was uncontested or not contested by both a Democratic and Republican candidate. This percentage is too large to be ignored.

A graphical summary of the prevalence of uncontested districts appears in Figure 4, showing the percentage of districts without Democratic and Republican vote counts, by election and by state. While most Congressional races are contested in every state, Massachusetts, Alabama and Florida have rates of uncontested seats above 30%.

I employ an imputation strategy for uncontested districts with *two* distinct statistical models, predicting Democratic, two-party vote share in Congressional districts (y_i) and the total turnout one would expect if the district had actually been contested. Uncertainty stemming from the fact that some district level results have been imputed propagates forward through my analyses, such that characterisations about *EG* thresholds or the durability of the *EG* over the life of a districting plan are accompanied with uncertainty bounds.

7 The efficiency gap, by individual state elections

I compute 512 efficiency gap measures, spanning 25 states and 23 election years. In each election I compute the efficiency gap for a given state-level, Congressional election as

$$\frac{\sum_{i=1}^{n} W_i^R - W_i^D}{\sum_{i=1}^{n} v_i} \tag{3}$$

where i indexes n districts, W_i^R is the number of wasted votes for the Republican candidates in district i, W_i^D are wasted votes for the Democratic candidate and v_i is the number of votes for Democratic and Republican candidates.

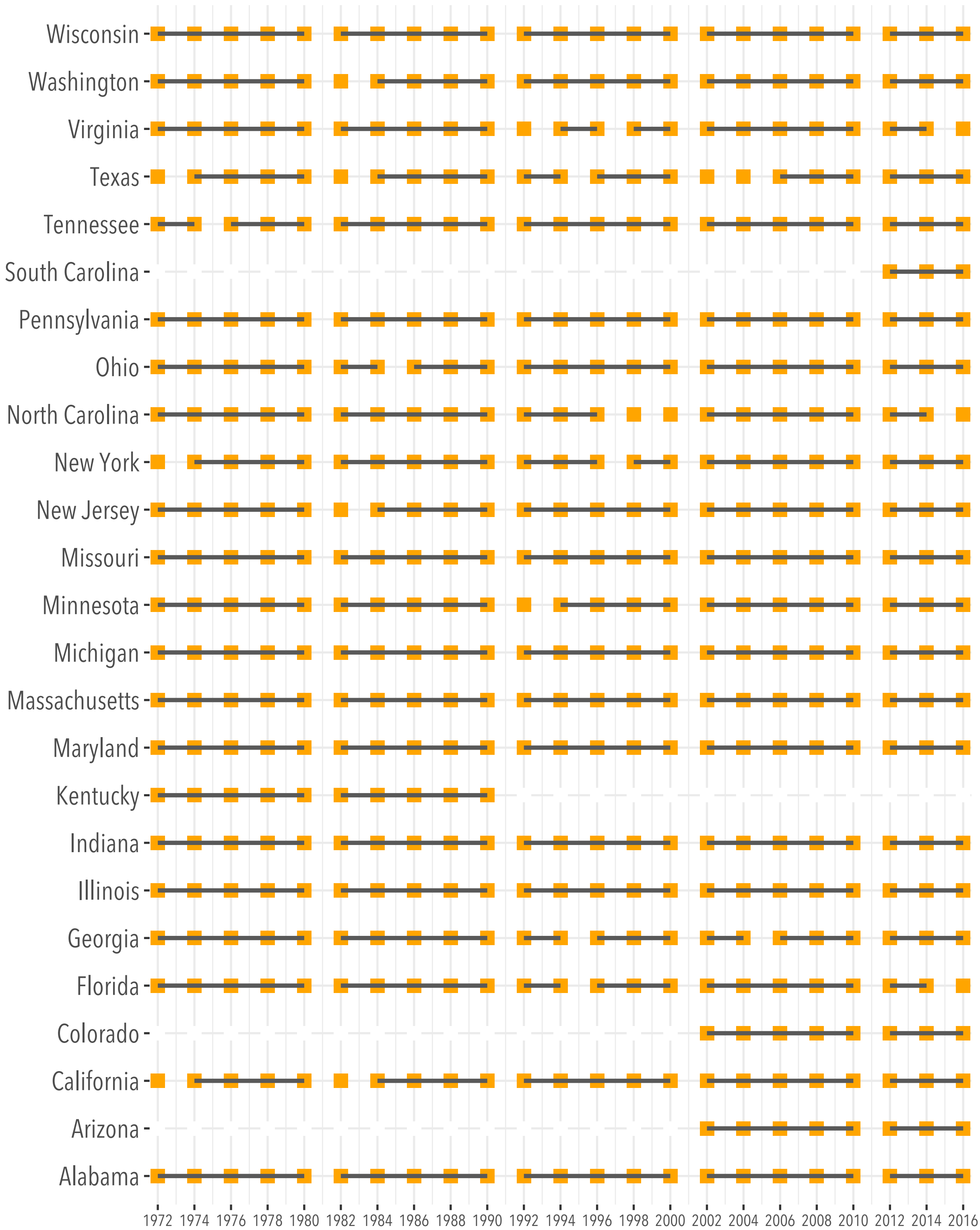

Figure 3: 512 Congressional elections available for analysis, 1972-2016, by state, grouped by districting plan (horizontal line).

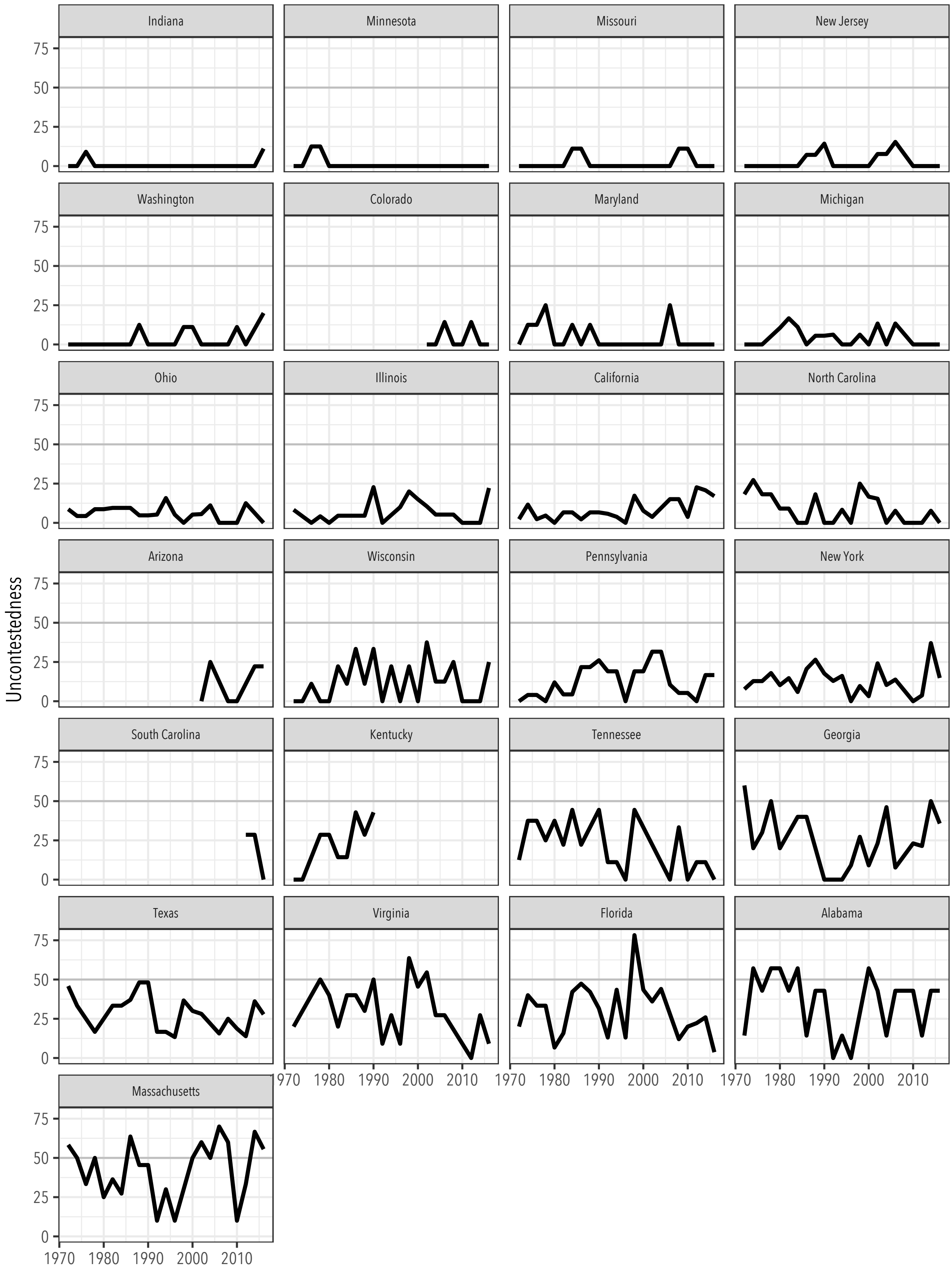

Figure 4: Percentage of districts missing two-party vote shares, by election, in 512 state-level, Congressional elections, 1972-2016. Missing data is always due to districts being uncontested by both major parties.

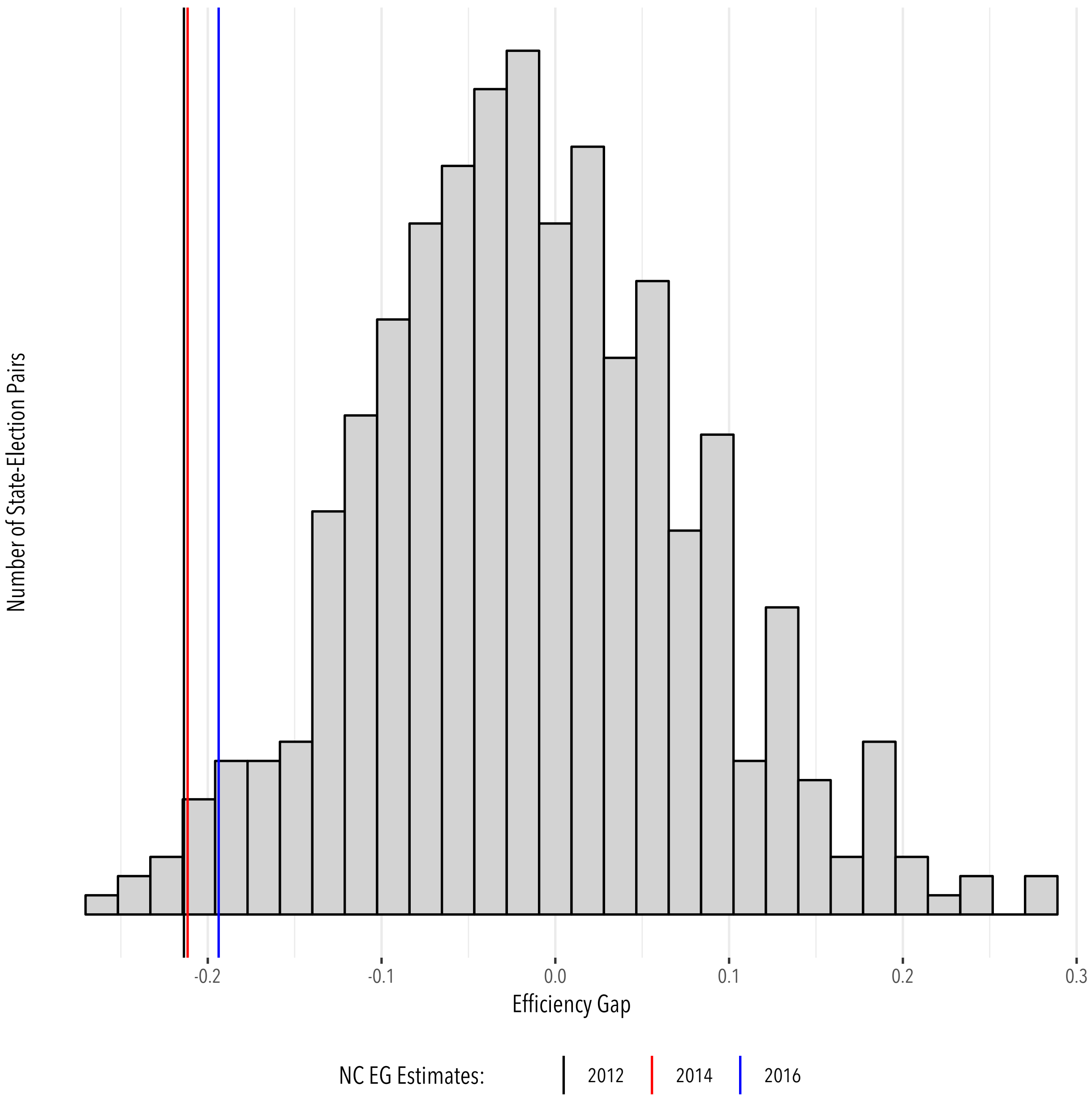

Figure 5: Histogram of efficiency gap estimates in 512 elections, 1972-2016. The three vertical lines indicate where North Carolina's three most recent elections lie in the distribution of efficiency gap scores.

Figure 5 summarizes the distribution of the 512 efficiency gap scores. There is considerable variation in the efficiency gap estimates across states and elections, but North Carolina stands out for having its three most recent elections generate the 12th, 14th, and 21th most pro-Republican efficiency gaps of the 512 efficiency gap measures. No other state makes three appearances in the list of the fifteen most pro-Republican efficiency gaps.

The overall distribution of efficiency gap scores is roughly symmetric and centered on zero (mean of -0.0138163 and median of -0.0174644). This indicates that when averaged over many elections and many states, the districting plans used in Congressional elections in the United States display no systematic efficiency gap advantage towards one party or the other. Particular districting plans *do* have large efficiency gap scores, including the North Carolina elections and plans in the left tail of the distribution of efficiency gap scores in Figure 5.

7.1 Within-plan variation in the efficiency gap

The efficiency gap is measured at each election, with a given districting plan typically generating up to five elections and hence five efficiency gap measures. Efficiency gap measures will change from election to election as the distribution of district-level vote shares varies over elections. Some of this variation is to be expected. Even with the same districting plan in place, districts will display "demographic drift," gradually changing the political complexion of those districts. Incumbents lose, retire or die in office. Sometimes incumbents face major opposition, sometimes they don't. Variation in turnout — most prominently, from on-year to off-year — will also cause the distribution of vote shares to vary from election to election, even with the districting plan unchanged. All of these election-specific factors will contribute to election-to-election variation in the efficiency gap.

Precisely because I expect a reasonable degree of election-to-election variation in the efficiency gap, I assess the magnitude of "within-plan" variability in the measure. If a plan is a partisan gerrymander — with a systematic advantage for one party over the other — then the "between-plan" variation in *EG* should be relatively large relative to the "within-plan" variation in *EG*.

59% of the variation in the *EG* estimates is between-plan variation. The *EG* measure varies election-to-election, but there is a moderate to strong "plan-specific" component to variation in the *EG* scores. I conclude that efficiency gaps *are* measuring an enduring feature of a districting plan.

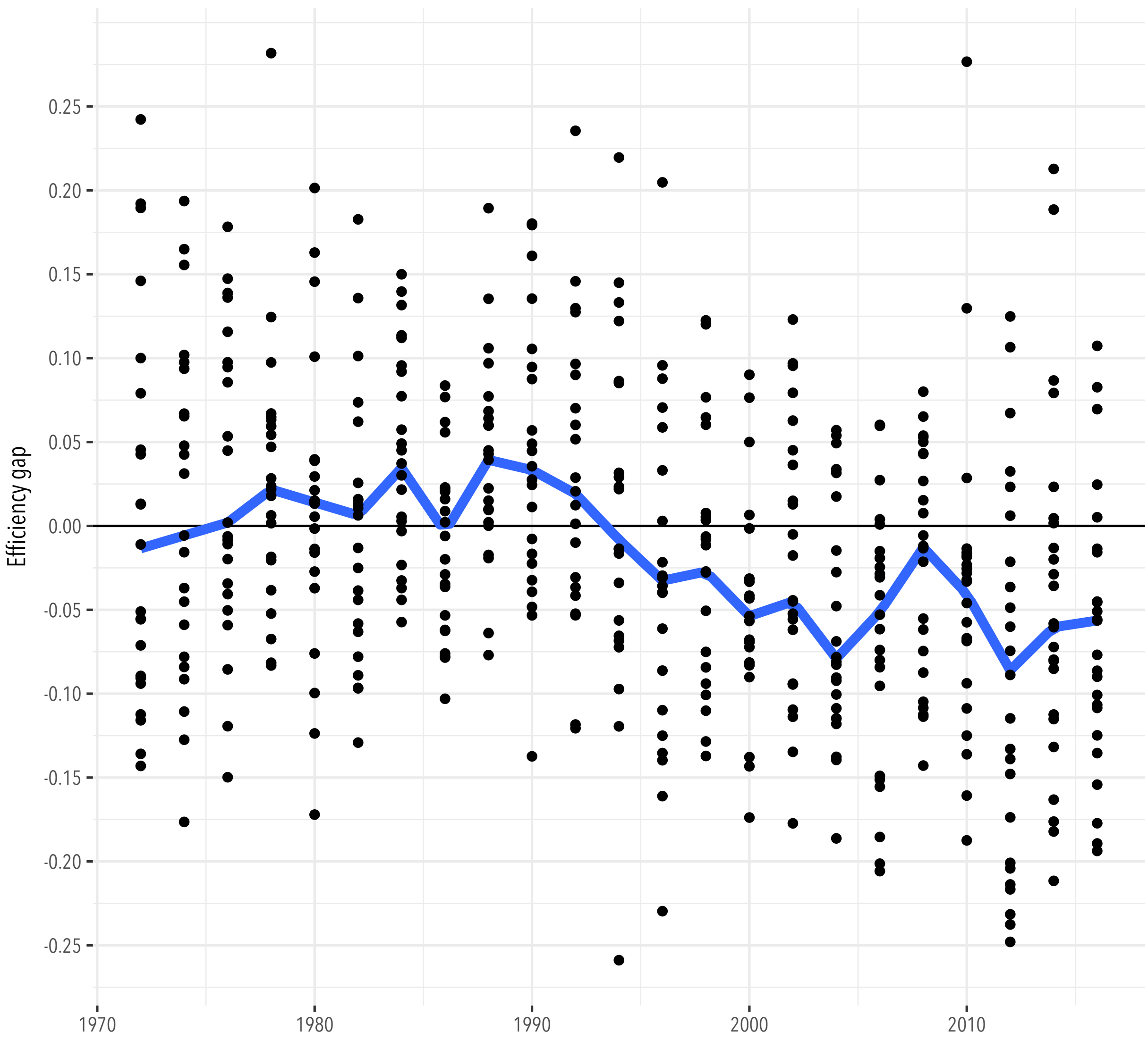

Figure 6: Efficiency gap estimates, over time. The line is a smoothed estimate of the median efficiency gap.

7.2 Over-time change in the efficiency gap

Figure 6 plots *EG* estimates over time, with the median measure indicated by the heavy blue line. The distribution of *EG* measures in the 1970s and 1980s appeared to slightly favor Democrats; 58% of all *EG* measures in this period were positive. The distribution of *EG* measures trends in a pro-Republican direction through the 1990s, such that by the 2000s, *EG* measures were more likely to be negative, indicative of pro-Republican advantage (see Figure 6). In the 2010s, 78% of efficiency gap scores were negative, indicative of pro-Republican advantage in their underlying districting plans. Figure 7 plots the *magnitude* of *EG* estimates over time, with the median absolute value indicated by the heavy blue line. The magnitude of the median efficiency gap was roughly constant until the current cycle, when it spiked to the highest value on record.

8 Party control of redistricting drives change in the efficiency gap

Districting plans for Congressional elections have tended to produce pro-Republican average efficiency gaps in recent years, but over the entire 1972-2016 period, the average efficiency gap is very close to zero (see Figure 6). There is thus no sign of a pro-Republican advantage in the dataset as a whole. Efficiency gap scores indicative of Republican advantage are much more likely to be found in recent decades.

A leading cause of this is the fact that Republican control of Congressional redistricting has increased markedly in recent decades. As Figure 8 displays, no Congressional maps were designed by Republicans in full control of the redistricting process in the 1990s, compared to about 30% by Democrats in full control and about 70% by another institution (divided government, a commission, or a court). But in the 2000s, Republicans were fully responsible for about the same proportion of plans as Democrats (about 30%). And in the 2010s, the distribution changed again such that about 50% of plans were designed entirely by Republicans, versus about 10% designed entirely by Democrats.

To determine the impact of this change in party control on the change in the efficiency gap over the last generation, I created five separate boxplots, one for each decadal redistricting cycle in the dataset. In each case, the distribution of efficiency gaps for Congressional districting plans is shown for each type of plan-making body: Democrat-controlled bodies, Republican-controlled bodies and other institutions (divided government, courts and commissions). Figure 9 displays the average efficiency gap and the distribution of the efficiency gaps.

I employ regression analysis to assess the effects of partisan control on the effi-

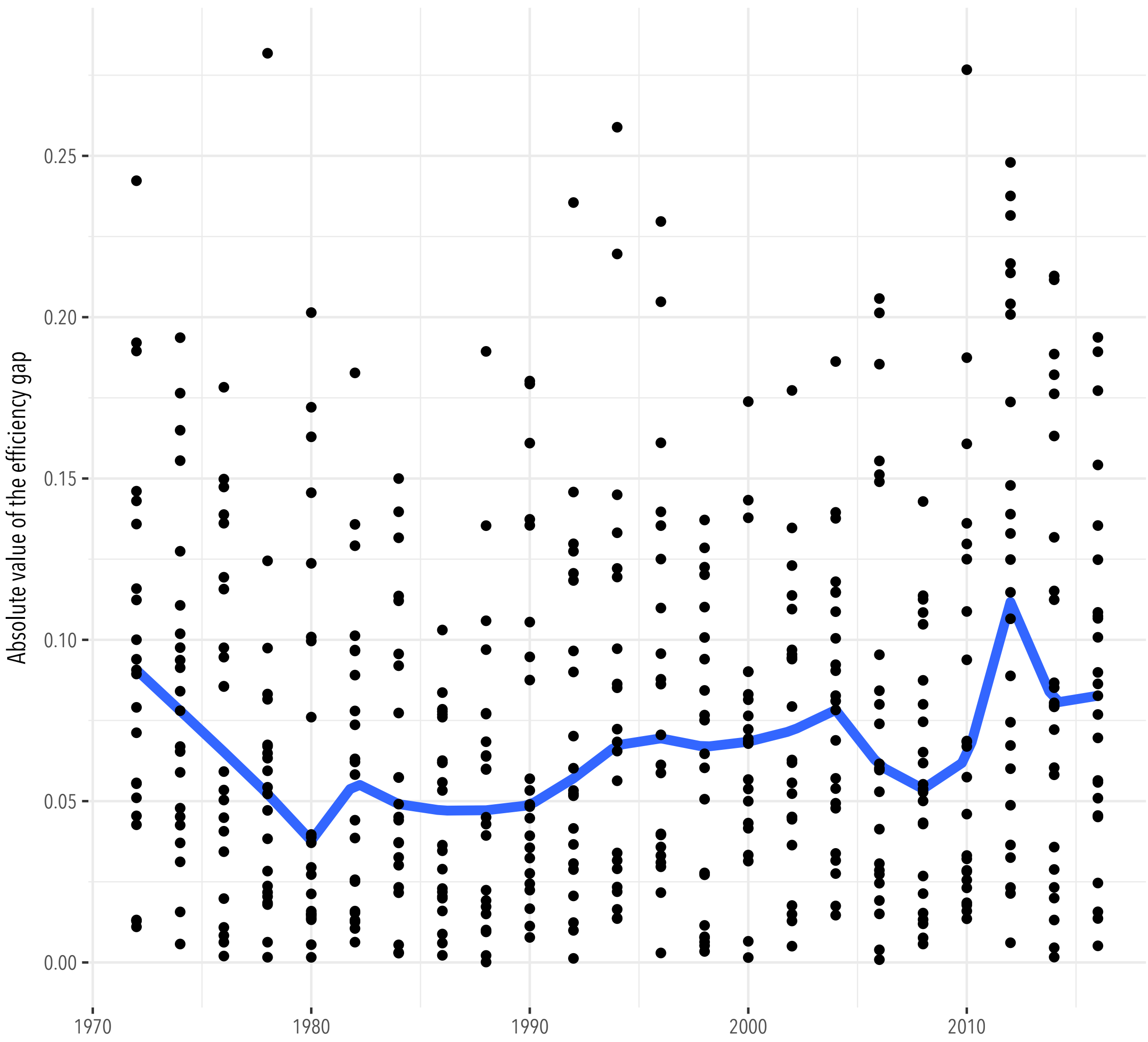

Figure 7: Absolute value of efficiency gap measures, over time. The blue line is a smoothed estimate of the median absolute value of the efficiency gap measure.

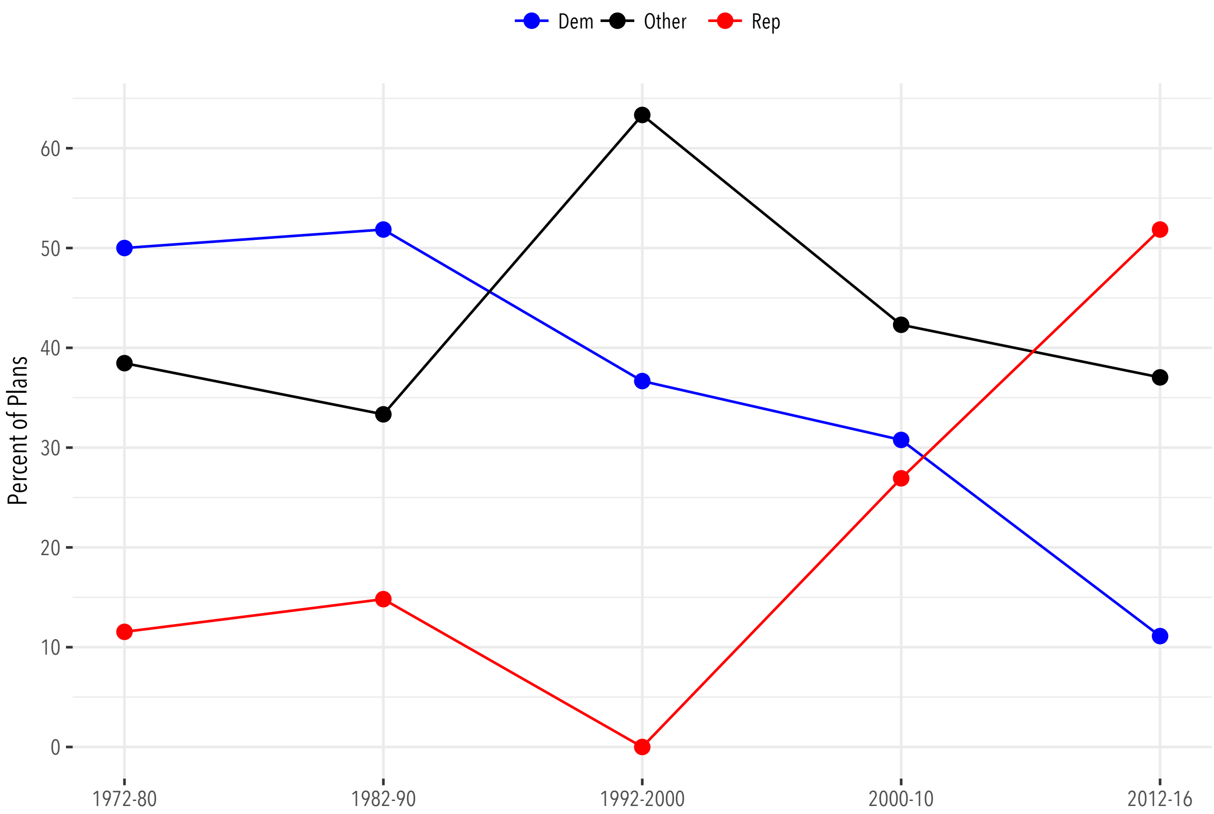

Figure 8: Share of 136 Congressional districting plans used in the efficiency gap analysis, grouped by decade, designed by Democrats in unified control of state government, by Republicans in unified control of state government, or by another institution (divided state government, commission, or court).

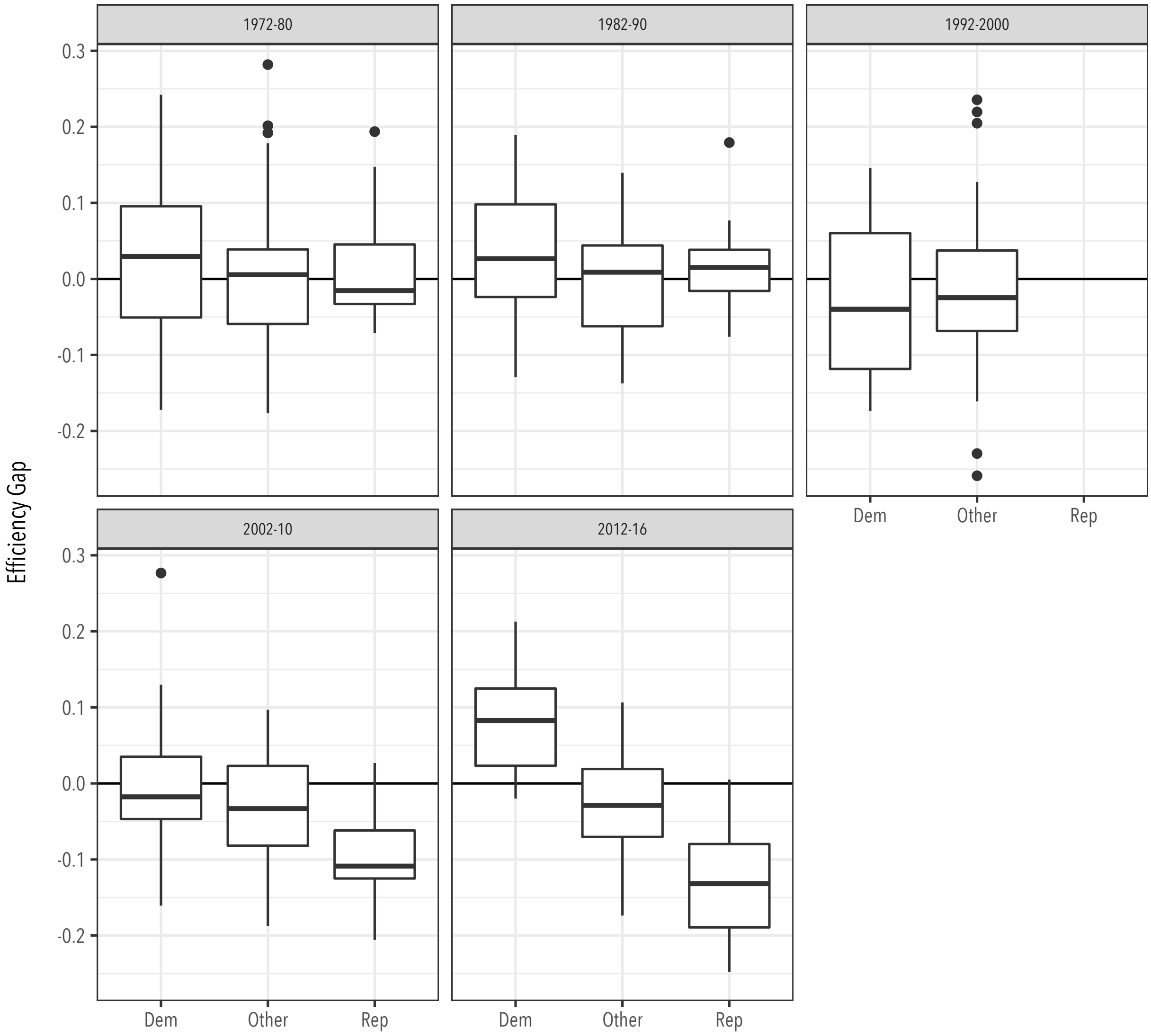

Figure 9: Distribution of efficiency gap scores, by decade, by control of corresponding redistricting process: Democrats in unified control of state government, Republicans in unified control of state government, or control by other institutions (divided state governments, commissions or courts).

		Democratic	Republican
2002-2016	Estimate	.119	-.070
	t	(3.91)	(-2.53)
1992-2010	Estimate	-.025	-.087
	t	(-1.04)	(-2.41)
1992-2016	Estimate	.023	-.139
	t	(0.75)	(-3.57)
1972-2016	Estimate	.027	-.054
	t	(0.92)	(-1.97)

Table 2: Estimates of average effect of change in efficiency gap resulting from change from non-partisan control of redistricting to Democratic or Republican control of redistricting. Regression models include unreported fixed effects for states and years. *t*-statistics are based on estimates of standard errors computed via clustering on state/districting plans.

ciency gap. Table 2 reports estimates of the average effects on the efficiency gap of a change from non-partisan control of Congressional redistricting to partisan control, for various time slices spanning the years encompassed by my analysis. Each regression includes fixed effects for election years and states and weights taking into account the uncertainty associated with some of the efficiency gap scores. Standard errors are estimated by clustering on state/districting-plans, acknowledging that efficiency gap scores from a given state under a given districting plan are unlikely to be conditionally independent given the predictors in the regression model. The fixed effects for states means that the party control estimates are within-state estimates and identified by states that change party control of redistricting over the period spanned by a particular regression model.

Efficiency gaps in the 2002-2010 cycle would have been substantially less pro-Republican had Republicans not gained control of more state governments in this cycle relative to the 1990s. The shift from non-partisan control of Congressional redistricting in the 1990s to Republican control in the 2000s produces a -.087 average shift in the resulting efficiency gaps (t = -2.41). Conversely, shifting from non-partisan control of Congressional redistricting in the 1990s to Democratic control in the 2000s produces a -.025 average shift in the resulting efficiency gaps (t = -1.04) which is not distinguishable from "no change" at conventional levels of statistical significance).

Between the 2000s and the 2010s (1st row of Table 2), changing from non-partisan control of Congressional redistricting to Republican control produces a shift in the efficiency gap of -.070 (t = -2.53), while a shift from non-partisan control to Democratic control produces a shift in the efficiency gap of .119 (t = 3.91).

Over the last two decades, the regression analysis reported in the 3rd row of

Table 2 finds large effects associated with switching to Republican control: an average change of the efficiency gap in a negative/pro-Republican direction of -.139 (t = -3.57) but no statistically significant movement associated with change from non-partisan control to Democratic control.

Finally, the last row of Table 2 confirms that the effects of partisan control of redistricting are largely confined to recent decades. Averaging over the entire 44 year period spanned by my analysis, there are much smaller effects of changes in the partisan control of redistricting than in recent decades. Averaged over the 1972-2016 period, there is no statistically significant average effect of switching from independent/non-partisan control of Congressional redistricting to Democratic control (t = .92) and the average effect of a switch to Republican control is a modest -.054 but distinguishable from zero at conventional levels of statistical significance (t = -1.97). Again, this pattern of results confirms that the association between partisan control of Congressional redistricting and large and consequential changes in the efficiency gap is a relatively recent development.

Much of the observed change in the efficiency gap in recent decades is due to the effects of change in control of the redistricting process. If the composition of partisan control of redistricting had stayed as it was in the 1990s — but the effects of partisan control of redistricting on efficiency gaps were as observed in the 2000s and 2010s — then the average efficiency gap in the 2002-2010 period would be -.030 versus the -.041 actually observed; see Figure 10. In the 2012-2016 period, the corresponding estimate is .002 versus the observed average efficiency gap of -.067; given that these are multiple-election, multiple-state *average* efficiency gap scores, these differences are large and of substantial consequence. In summary, changes in the distribution of efficiency gap scores in recent decades can be confidently attributed to changes in partisan control of redistricting, and, in particular, to the increased prevalence of Republican control of Congressional redistricting.

9 When is a large efficiency gap score politically consequential?

How large must the efficiency gap be before it can be said to have triggered a politically meaningful outcome such as a seat changing hands?

To answer this question I look at the historic relationship between seats and votes, by state, in the Congressional elections spanned by this analysis. For each election, in each state, I compute the Democratic share of the statewide, two-party vote (after any imputations for uncontested seats) and the percentage of the states' CDs won by Democratic candidates. Note that again I discard states with fewer than seven CDs.

Figure 10: Average efficiency gap by decade and predicted efficiency gap had partisan control of redistricting stayed as it was in the 1990s. Vertical lines cover 95% credible intervals for predictions of the average efficiency gap.

The orange line in Figure 11 is a linear regression, not constrained to run through the 50-50 point. Instead, the fact that the regression line *does* run through the 50-50 point reflects a regularity in the data: i.e, averaging over many districting plans and many elections over some 50 years of American political history, if a party wins 50% of the statewide, two-party vote in a Congressional election, it generally wins half of the CDs in that state. The slope of the regression line is 2.36, meaning that for every additional percentage point in state-wide vote share won (or lost) by Democrats, Democrats win (or lose) an extra 2.36 percentage points of seat share.[4]

Each deviation (in the vertical direction) from the orange regression line in Figure 11 is that portion of seat share in a given election that lies above or below that which is expected given the party's statewide share of the vote. This exercise provides guidance as to when the efficiency gap gets sufficiently large to be causing at least a one seat disparity relative to the long-run, historical mapping of state-wide votes into seats won. I consider the point at which the efficiency gap is associated with a *half* seat surplus/deficit, which is the point at which the surplus/deficit is closer to one seat than no seats.

In so doing, it is worth remembering that the analysis here pools over states with differing numbers of Congressional districts. I omit states with fewer than 7 CDs, but even so, the analysis spans Alabama, Colorado (2002-2016) and South Carolina (2012-2016) and other states with just seven CDs, through to California with 53 seats (2002-2016). Accordingly, I disaggregate the data into six bins: states with 7 or 8 CDs, 9 or 10 CDs, 11 to 15 CDS, 16 to 20 CDs, 21 to 30 CDs and 31 to 53 CDs.

In each of these bins, I examine what values of the efficiency gap are associated with discrepancies from the historical votes-seats relationship. In general, I expect a positive correlation between the efficiency gap and discrepancies relative to the long-run historical relationship between votes and seats, and indeed, this is confirmed in Figure 12.

Table 3 presents the results of this analysis. For states with small numbers of CDs, an efficiency gap of about -.07 or -.08 is associated with a Democratic seat deficit of approximately 1/2 a seat in the corresponding districting cycle; an efficiency gap larger than this 1/2 seat threshold is more likely to generate a one seat deficit (surplus) than no deficit (or surplus). This threshold is roughly the same irrespective of whether one is considering a Democratic half-seat deficit or surplus. Larger values of the efficiency gap are associated with a one seat deficit or surplus: about -.16 for a one-seat Democratic deficit, about .14 for a one seat Democratic surplus.

4I acknowledge that the linear form of the regression means that it will predict seat shares greater than 100% or lower than 0% conditional on high (or low) values of vote shares. But the linear regression model generates a good fit to the data in the region of observed vote shares, with little indication that predictions of seat shares will lie outside of the feasible zero to 100% interval.

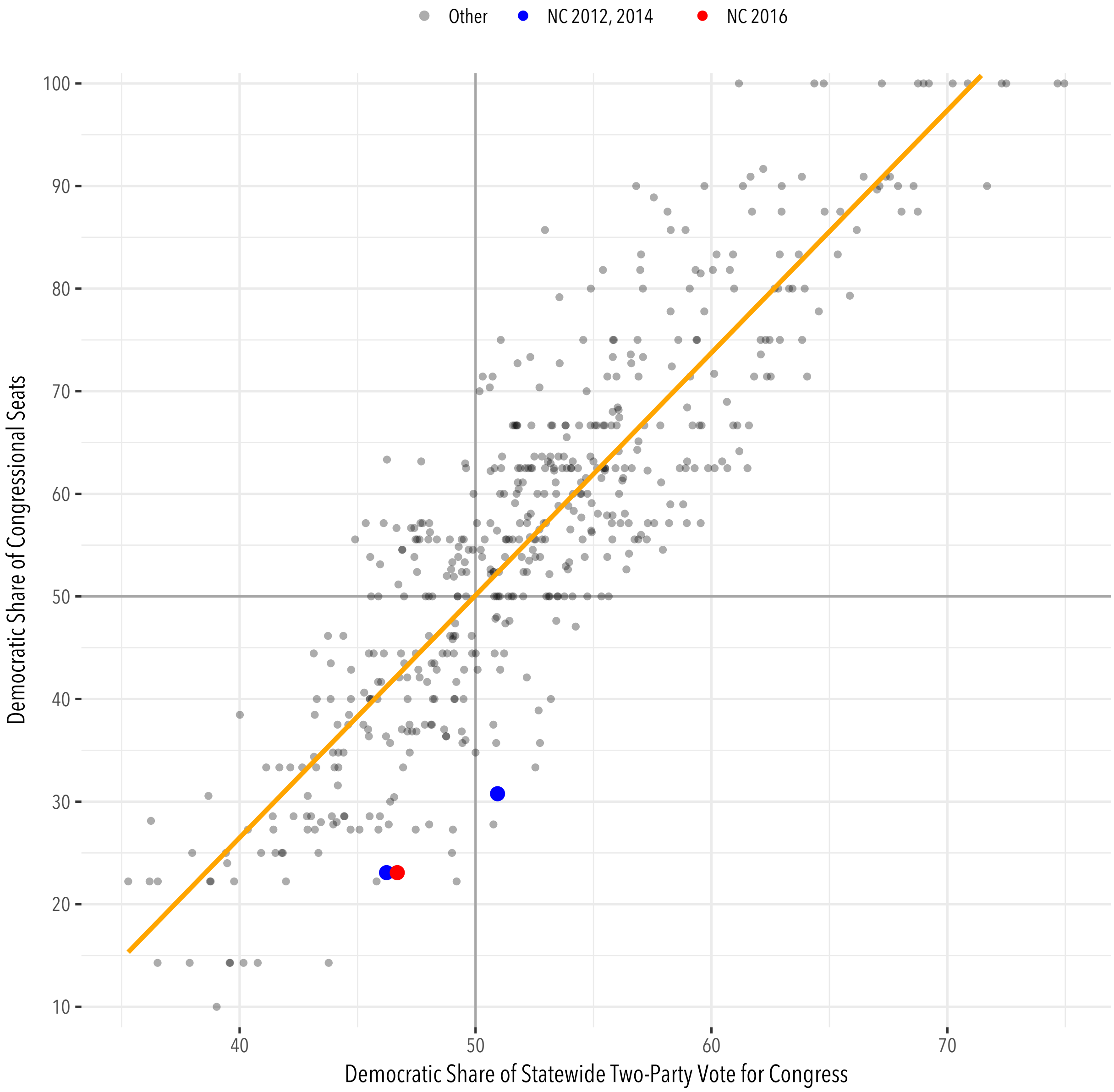

Figure 11: Seats and Votes in Congressional elections, 1972-2016. Each plotted point corresponds to a particular state in a particular election. The Democratic share of the statewide, two-party vote (after any imputations for uncontested seats) is plotted on the horizontal axis and the percentage of the states' CDs won by Democratic candidates is plotted on the vertical axis. The analysis covers states with seven or more CDs. The orange line is a linear regression that is not constrained to run through the 50-50 point; instead, the fact that the regression line *does* run through 50-50 point reflects the fact that averaged over many districting plans and many elections over some 50 years of American political history, if a party wins 50% of the statewide, two-party vote in a Congressional election, it generally wins half of the seats in that state.

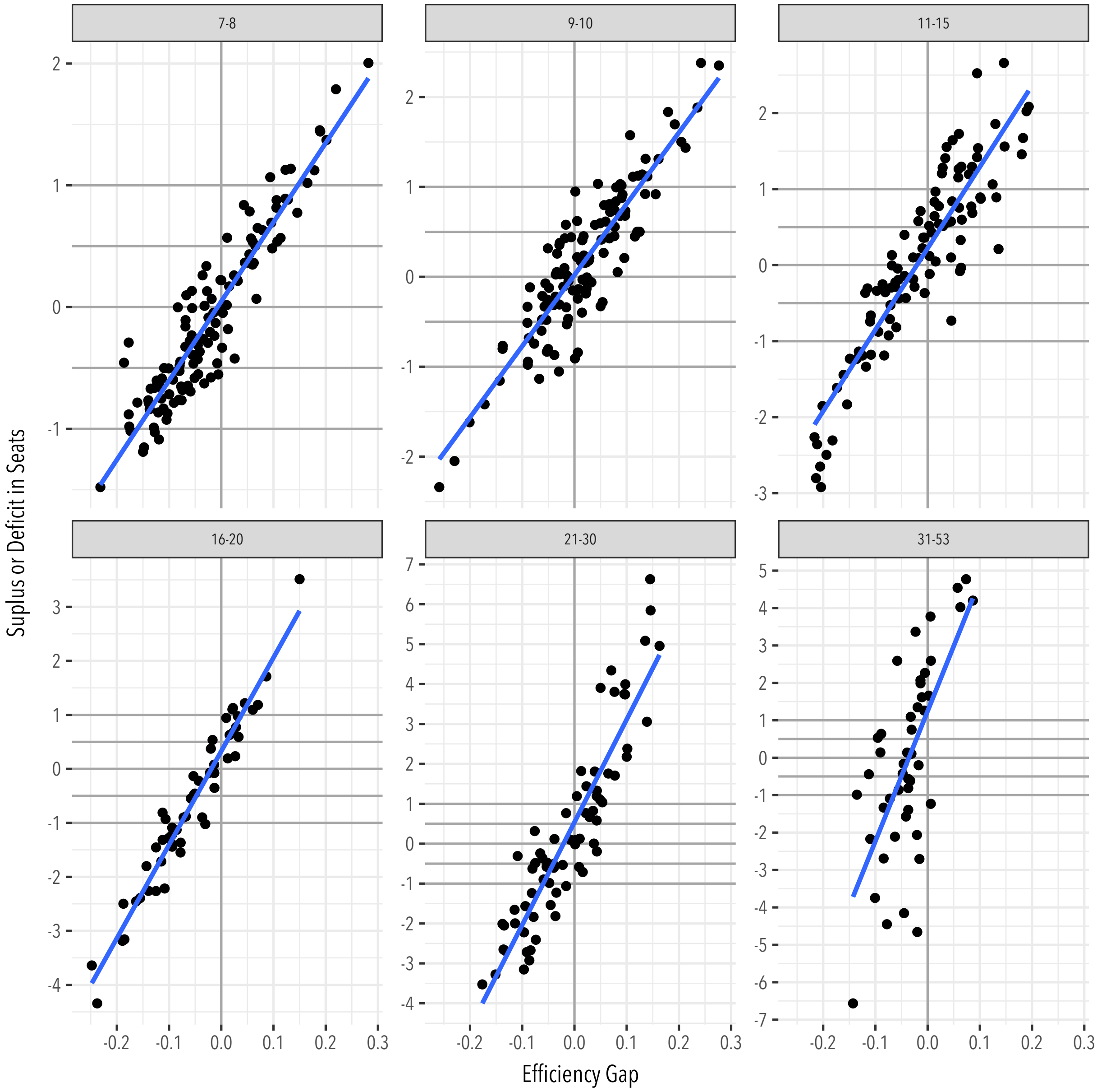

Figure 12: Surplus seats and the efficiency gap, 1972-2016, by number of seats in each state. The surplus seats measure is defined as the difference between the percentage of seats won by the Democrats in a given election and the seats predicted given the historical, regression relationship shown in Figure 11. The blue line in each panel is a smoothing spline fit as to minimize cross-validated prediction error (Wood, 2006).

	Seat Deficit/Surplus			
# CDs	-1	-0.5	0.5	1
7-8	-.16	-.08	.07	.14
	[-.18, -.15]	[-.09, -.07]	[.06, .08]	[.13, .15]
9-10	-.12	-.06	.06	.12
	[-.14, -.11]	[-.07, -.05]	[.05, .07]	[.11, .13]
11-15	-.11	-.07	.02	.07
	[-.12, -.10]	[-.09, -.06]	[.00, .03]	[.06, .08]
16-20	-.08	-.05	.01	.04
	[-.08, -.07]	[-.06, -.04]	[.00, .02]	[.03, .05]
21-30	-.05	-.03	.01	.03
	[-.06, -.04]	[-.04, -.02]	[.00, .02]	[.02, .04]
31-53	-.06	-.05	-.02	-.01
	[-.08, -.05]	[-.06, -.04]	[-.03, -.01]	[-.02, .00]

Table 3: Efficiency gap scores associated with indicated Congressional seat deficit/suplus. Columns refer to Democratic seat deficit/surplus; rows correspond to the number of CDs in a state. A surplus or a deficit in seats is defined as deviation from the historical relationship between seat shares and vote shares shown in Figure 11.

As states' Congressional delegations get larger, the efficiency gap values associated with a 1/2 seat or a one seat deficit/surplus start to get smaller. For large delegations (above 20 seats), a one seat departure from the historical mapping between vote shares and seat shares (Figure 11) is associated with a -.06 efficiency gap (Democratic deficit) or a .02 efficiency gap (Democratic surplus).

A simple summary of the estimates presented in Table 3 is that for states with delegations with 15 or fewer members, an efficiency gap of $\pm.08$ means it is more likely that a seat changes hands than not. The corresponding number is $\pm.05$ for states with more than 15 CDs. These values are conservative in the sense that it is often the case that smaller values of the efficiency gap are associated with a seat changing hands. Erring on the side of larger values of the efficiency gap has the effect of reducing the number of plans that would be recommended for judicial scrutiny.

10 Predictive performance of 1st efficiency gap observed under a plan

I now investigate the prognostic properties of the first *EG* observed under a districting plan. In each instance the test is whether the magnitude (or absolute value) of the first *EG* observed under a plan exceeds a given threshold value EG^*. The outcome of interest is whether the average of the plan's remaining *EG* scores trips the thresholds, based on the analysis reported in Table 3: i.e., $\pm .05$ for states with large Congressional delegations, $\pm .08$ for states with smaller Congressional delegations. That is, does a positive (or negative) initial *EG* score accurately predict that the remainder-of-plan average *EG* in excess of the Table 3 thresholds for being politically subtantive (more likely to see at least one seat change hands than not).

The classification or prediction task here can be summarized as follows:

	Actual	
Test	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

Again, note that in the discussion that follows the "test" is whether the 1st election *EG* score exceeds a particular magnitude or threshold. The "actual" outcome is whether the remainder-of-plan average *EG* under that districting plan trips the thresholds derived from the analysis reported in Table 3: $\pm .05$ for states with large Congressional delegations, $\pm .08$ for states with smaller Congressional delegations.

I compute and plot a number of quantities from this exercise. First, I note the "detection prevalence rate" or the "test positive" rate, the proportion of districting plans for which the value of the efficiency gap in the first election under the plan exceeds a given threshold. In the left-hand panel of the top row of Figures 13 and 14 I refer to this quantity as the "proportion flagged" or `Pr(flagged)`.

The prognostic measures I rely on are conventional measures of predictive or classification accuracy used throughout the quantitative sciences:

1. *sensitivity*, or true positive rate (TPR): the proportion of positives that test positive ("flagged"), $TP/(TP+FN)$, labelled as `Pr(flagged|positive)` in 2nd panel, top row of Figures 13 and 14.

2. *specificity*, or true negative rate (TNR): the proportion of negatives that test negative, $TN/(TN+FP)$, labelled as `Pr(¬flagged|negative)` in the 3rd panel, top row of Figures 13 and 14, "¬" being a conventional shorthand for negation.

3. *precision*, the proportion of cases testing positive ("flagged") that are actually positive, $TP/(TP+FP)$, labelled as `Pr(positive|flagged)` in the right hand panel of the top row of Figures 13 and 14.

4. *accuracy*, the proportion of cases that are true positives or true negatives, $(TP+TN)/(TP+FP+FN+TN)$, labelled as `Pr(correct diagnosis)` in the left hand panel of the bottom row of Figures 13 and 14.

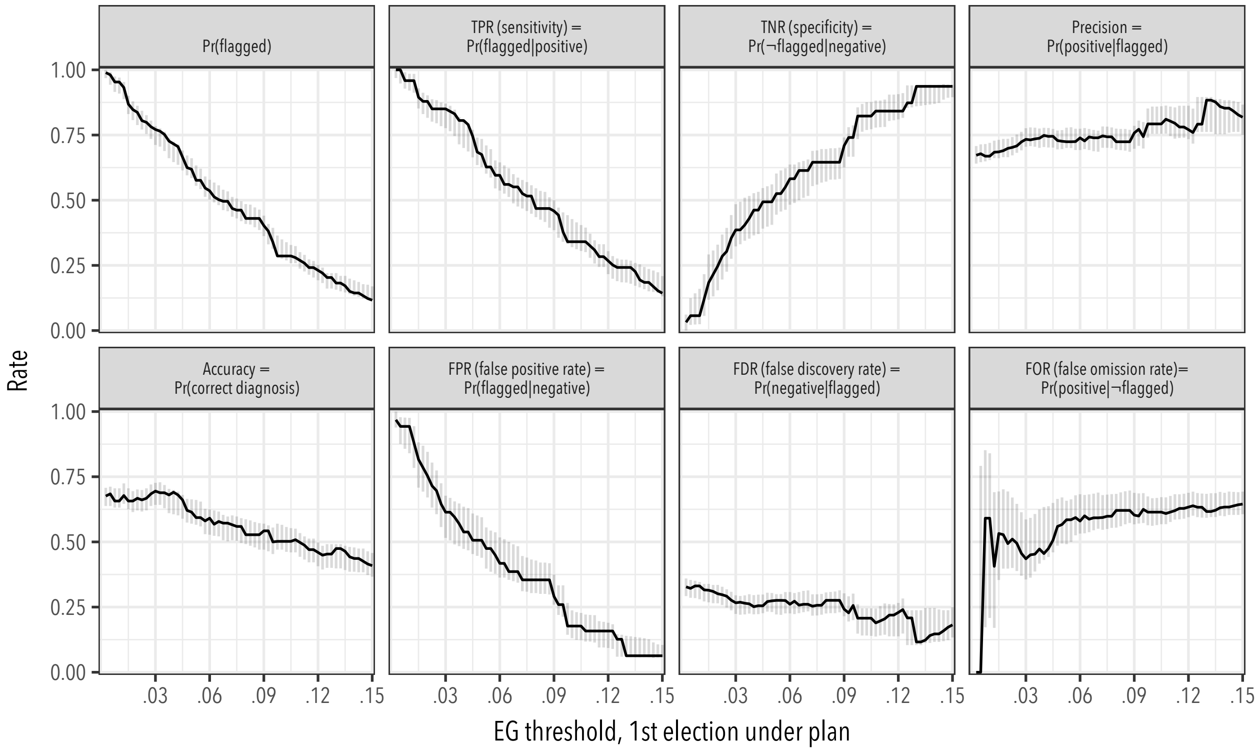

Figure 13: Prognostic performance measures, first efficiency gap under a districting plan more extreme than threshold (horizontal axis) as a predictor of whether the remainder-of-plan average efficiency gap recorded under the districting plan exceeds the thresholds derived from the analysis summarized in Table 3: $\pm .05$ for states with large Congressional delegations, $\pm .08$ for states with smaller Congressional delegations. Vertical lines indicate 95% confidence intervals. Analysis restricted to plans with three or more elections.

5. the *false positive rate* (FPR); proportion of negative cases that test positive, 1 minus the specificity or $FP/(TN + FP)$, labelled as `Pr(flagged | negative)` in the 2nd panel, bottom row of Figures 13 and 14.

6. the *false discovery rate* (FDR); the proportion of cases testing positive that are actually negative, $FP/(TP + FP)$, labelled as `Pr(negative|flagged)` in the 3rd panel, bottom row of Figures 13 and 14. Note that this quantity is one minus the precision.

7. the *false omission rate*; the proportion of cases that test negative that are actually positive, $FN/(FN + TN)$, labelled as `Pr(positive|¬flagged)` in Figures 13 and 14.

Figure 13 displays these various rates based on all plans with three or more elections over the entire 1972-2016 period. Figure 14 displays results based on plans with three or more elections enacted since 2000.

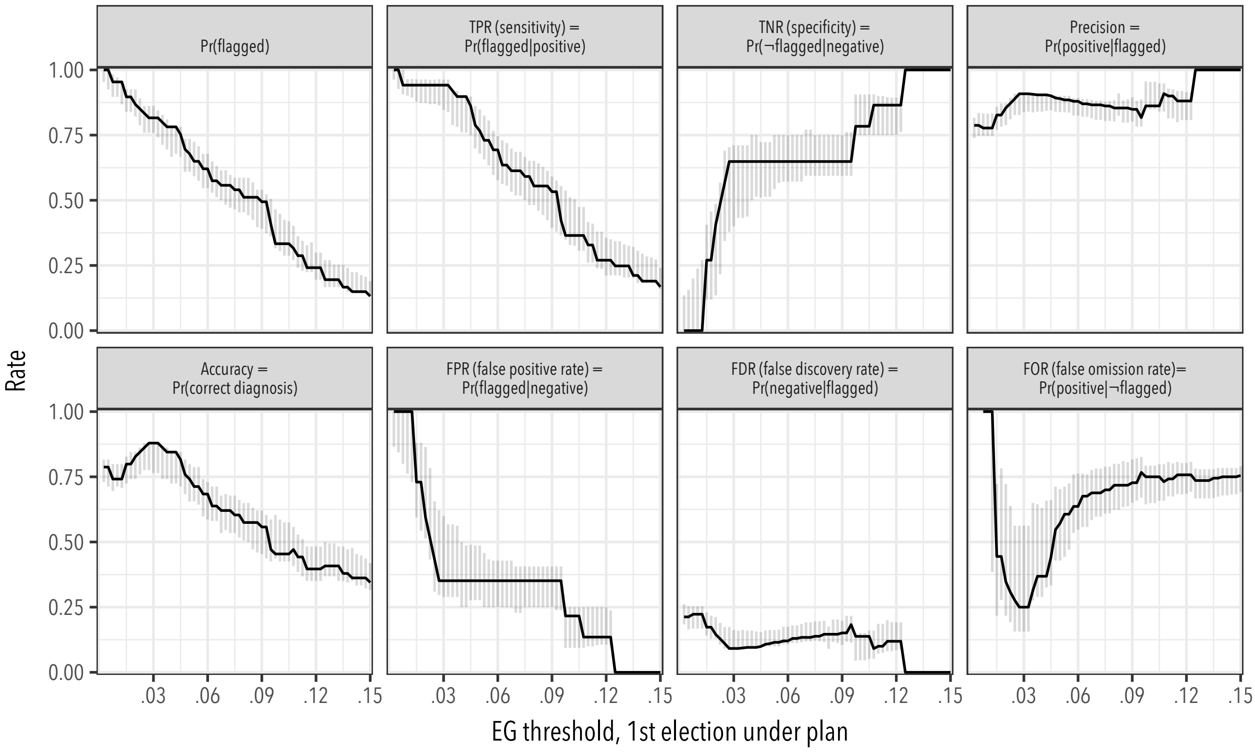

Figure 14: Prognostic performance measures, first efficiency gap under a districting plan more extreme than threshold (horizontal axis) as a predictor of whether the remainder-of-plan average efficiency gap recorded under the districting plan exceeds the thresholds derived from the analysis summarized in Table 3: ±.05 for states with large Congressional delegations, ±.08 for states with smaller Congressional delegations. Vertical lines indicate 95% confidence intervals. Analysis restricted to plans with three or more elections. The rates in this figure are from plans enacted since 2000.

A relatively large proportion of plans have large 1st election *EG* scores. As many as 23% of plans begin life with a 1st election *EG* of .12 in magnitude or greater. Of these 23% of plans, more than three-quarters go on to have a remainder-of-plan average efficiency gap in excess of the proposed thresholds in Table 3, an indication of the *reliability* of the 1st election *EG*. Generally, the precision of a prognostic test based on the 1st election *EG* is high, approaching 75% once the 1st election *EG* is greater than .03 in magnitude.

A more stringent threshold — a 1st election *EG* of at least .15 in magnitude — is tripped by fewer plans, about 12.5% of plans. At this threshold the precision of the test criterion remains high (or conversely, the false discovery rate stays low), but the false omission rate has climbed to around 75%; i.e., the threshold is so stringent that only one out of every four plans with a remaining plan-average *EG* in excess of the Table 3 thresholds is being flagged by the test. The overall accuracy of the test falls to around 40% if one were to adopt a very stringent threshold such as $|1\text{st}\,EG| > .15$.

Note that the false positive rate (FPR) continues to fall as the test threshold is made more stringent: at a threshold of .12, only 12.5% of plans that do *not* have a remainder-of-plan average *EG* in in excess of the Table 3 thresholds. If the test criterion was $|1\text{st}\,EG| > .12$, then (a) 23% of plans would trip that threshold, (b) with a false positive rate of about 10% and (c) a false discovery rate of around 15%.

Districting plans enacted since 2000 see larger and more durable efficiency gap scores, such that the "signal" in the 1st election *EG* is more reliable, relative to plans from the 1970s-1990s. An almost identical proportion of post-2000 plans (Figure 14) trip the 1st *EG* threshold of +/- .12 (25%) as in the analysis of 1972-2016 plans (Figure 13).

But a compelling, distinctive feature of plans enacted since 2000 is how few of them "reverse course" after the 1st election. The false discovery rates in Figure 14 are around 12% for almost any 1st election *EG* threshold, falling to zero once the 1st election *EG* is .13 or greater in magnitude. Figure 14 also makes clear that setting the threshold at higher levels doesn't risk "false alarms", but rather, false omissions.

For the post-2000 era, there is little risk of flagging a plan that will go on to have a remainder-of-plan average efficiency gap that fails to contradicts the "signal" about the direction and magnitude of partisan advantage in the 1st election's efficiency gap. This means that the operative questions would seem to be (a) the volume of plans being flagged for scrutiny; and (b) the magnitude of the remainder-of-plan average efficiency gaps likely to follow given the 1st election *EG*.

Here I draw on the analysis provided in section 9, which supplies guidance in setting an efficiency gap threshold, helping us understand how values of the efficiency gap translate into a tangible outcome such as a change in the seats won for a given level of the statewide vote.

In utilizing the results of Table 3 I (a) focus on the 1/2 seat threshold; and (b) treat the parties symmetrically, taking the larger in magnitude of the efficiency gap scores associated with a 1/2 seat surplus or deficit. That is, the efficiency gap scores I would use as thresholds are: $\pm.08$ for states with delegations with 15 or fewer members, and $\pm.05$ for states with larger delegations.

With this analysis in hand, I now examine which values of the efficiency gap in the 1st election under a plan are predictive of plan-average efficiency gap scores at these threshold values.

10.1 Regression relationship between 1st efficiency gap and remainder-of-plan average efficiency gap

Figure 15 shows the relationship between first-election efficiency gaps and average efficiency gaps observed over the remainder of the plan, for the districting plans analyzed here. I restrict this analysis to plans with at least three elections, so that the average *EG* computed net of the 1st election spans at least two elections, leaving 108 plans in this analysis. Figure 16 repeats this analysis for plans enacted since 2000, with 44 plans with three more elections.

The black diagonal line on each graph is a 45-degree line: if the relationship between first-election *EG* and remainder-of-plan average *EG* were perfect, the data would all lie on this line. Even given the uncertainty that accompanies *EG* measures due to uncontestedness, the relationship between first-election *EG* and remainder-of-plan average *EG* is quite strong. The correlation between 1st election *EG* scores and the remainder-of-plan average *EG* is .54 (95% CI .50 to .57). In the post-2000 era, this correlation is .74 (95% CI .70 to .78).

In the particular case of North Carolina, in 2016 the efficiency gap is -0.1937389. The analysis of historical data discussed above — and graphed in Figure 15 — forecasts that the remainder-of-plan average *EG* for this plan will be -0.08 (95% CI -0.21 to 0.04). Although there is uncertainty as to the eventual, remainder-of-plan average *EG*, I am highly confident that the plan used for North Carolina's 2016 Congressional elections — if left undisturbed — will produce a negative, pro-Republican, average efficiency gap. Figure 17 summarizes this uncertainty revealing that the probability of a positive (pro-Democratic) remainder-of-plan average *EG* is 9.5%.

If these calculations are based on the relationship between 1st election efficiency gaps and remainder-of-plan average efficiency gaps in the post-2000 era (Figure 16), then given the *EG* in 2016 in North Carolina, the forecast remainder-of-plan average *EG* is -.12 (95% CI -.23 to -.02). Using the post-2000 data, the probability that the plan used for North Carolina's 2016 Congressional elections — if left undisturbed — will ultimately turn out to have a positive, pro-Democratic remainder-of-plan average

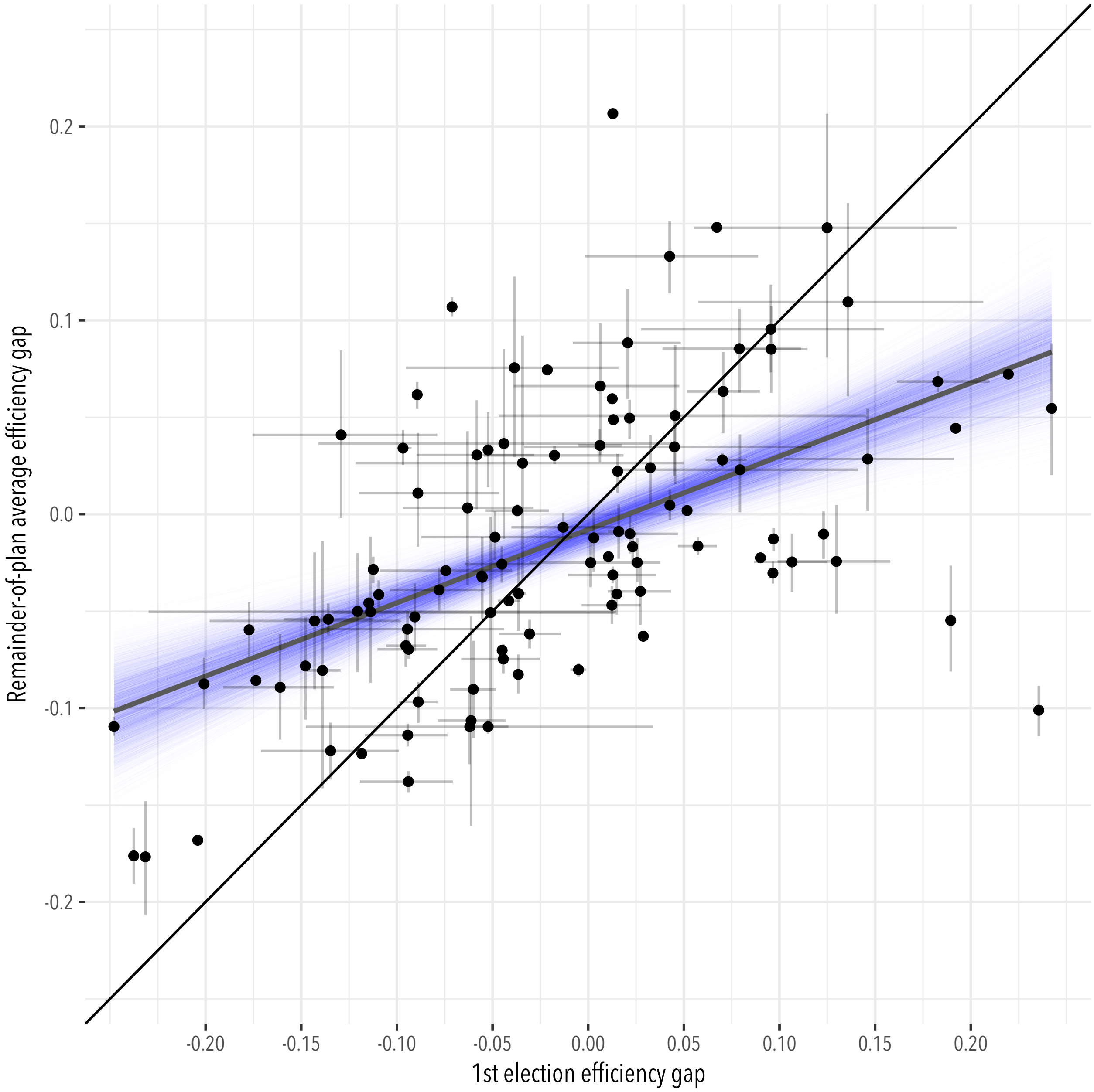

Figure 15: Scatterplot of first-election efficiency gap scores (horizontal axis) and remainder-of-plan average efficiency gap (vertical axis). The diagonal black line is a 45-degree line; the data would lie on this line if first-election efficiency gaps coincided with remainder-of-plan average efficiency gaps. The blue lines are linear regressions, which vary because the underlying data are subject to uncertainty stemming from imputations for uncontested districts. Vertical and horizontal lines extending from each data point cover 95% confidence intervals in either direction, summarizing the uncertainty in both first-election *EG* and remainder-of-plan average *EG* given the imputations for uncontested districts. Analysis restricted to plans with at least three elections. The *EG* in North Carolina in 2016 is -0.1937389.

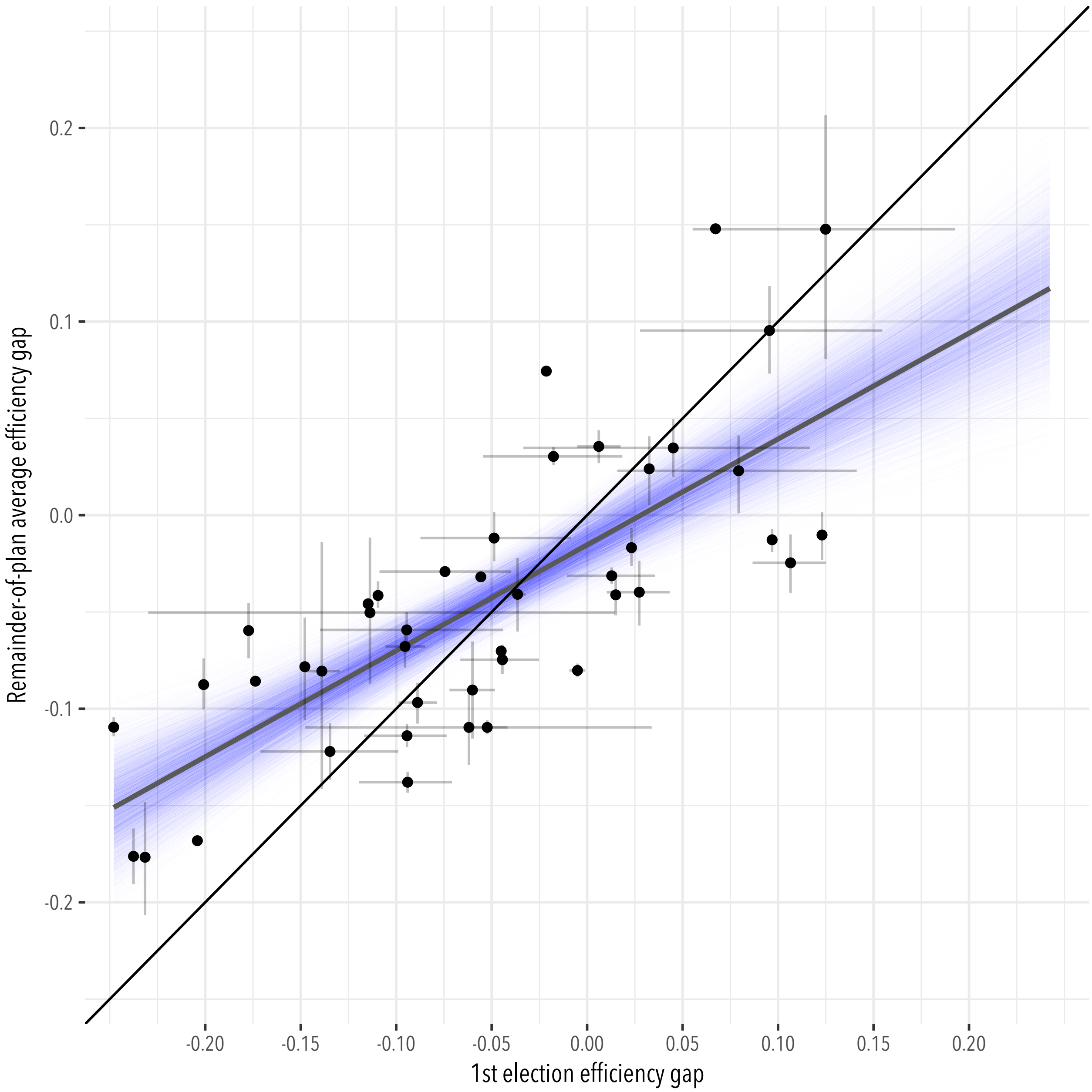

Figure 16: Scatterplot of first-election efficiency gap scores (horizontal axis) and remainder-of-plan average efficiency gap (vertical axis). The diagonal black line is a 45-degree line; the data would lie on this line if first-election efficiency gaps coincided with remainder-of-plan average efficiency gaps. The blue lines are linear regressions, which vary because the underlying data are subject to uncertainty stemming from imputations for uncontested districts. Vertical and horizontal lines extending from each data point cover 95% confidence intervals in either direction, summarizing the uncertainty in both first-election *EG* and remainder-of-plan average *EG* given the imputations for uncontested districts. Analysis restricted to plans with at least three elections, enacted after 2000. The *EG* in North Carolina in 2016 is -0.1937389.

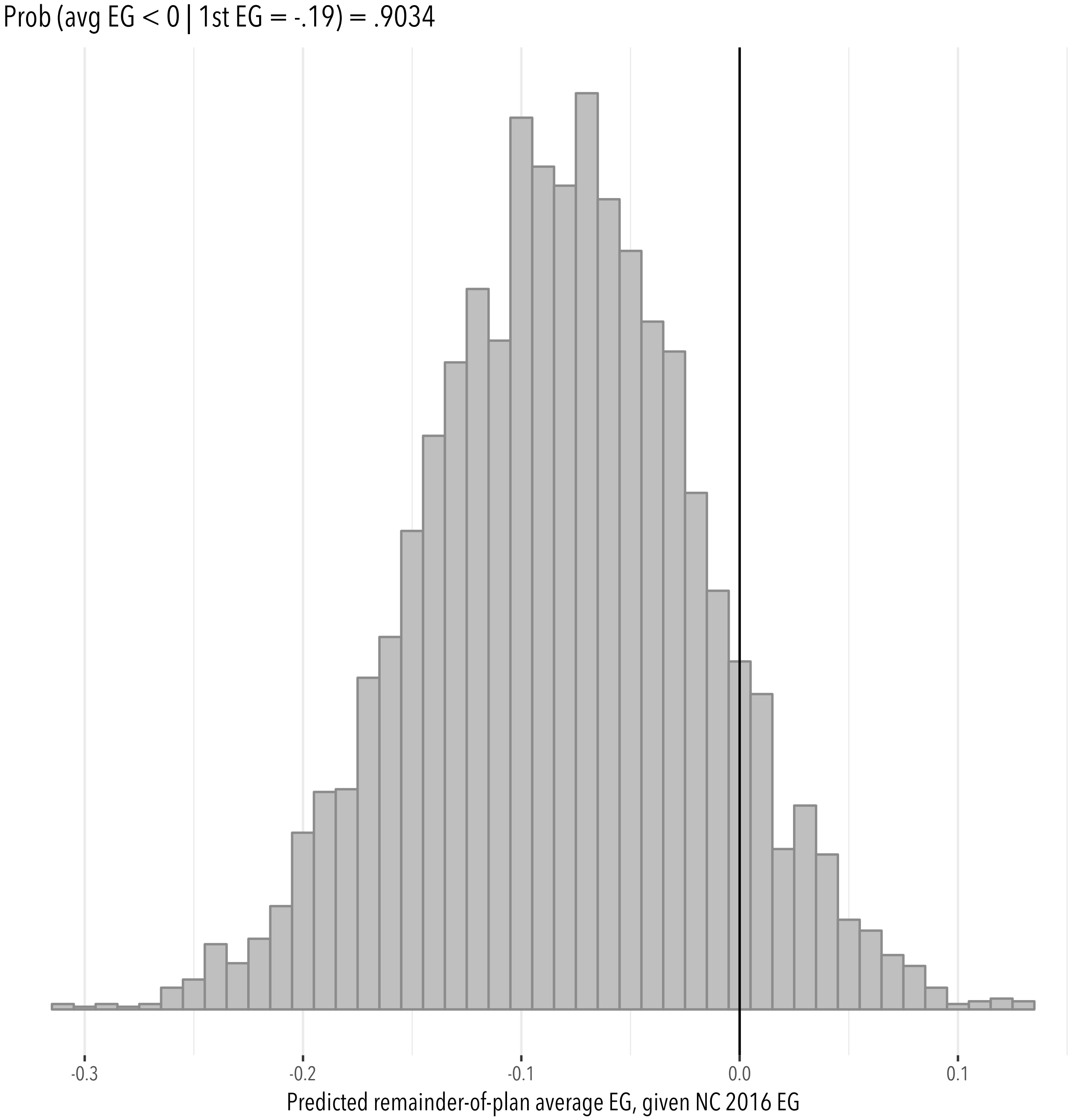

Figure 17: Histogram summarizing predictions as to the remainder-of-plan average efficiency gap expected from the current North Carolina districting plan, based on the historical, regression relationship between 1st election *EG* and remainder-of-plan average *EG* displayed in Figure 15.

efficiency gap is less than 2% (see Figure 18).

10.2 1st election efficiency gap and thresholds

Earlier I suggested using an efficiency gap of .08 as a threshold value in states with relatively small numbers of Congressional seats (15 or fewer members). Now I ask what 1st election efficiency gap is associated with a remainder-of-plan average efficiency gap of .08 or greater in magnitude.

The regression analysis shown in Figure 16 suggests that in the post-2000 era, a districting plan will tend to have a remainder-of-plan average efficiency gap below -.08 if its 1st *EG* is -.10 or lower (indicative of advantage for Republicans). A remainder-of-plan average efficiency gap above .08 in the post-2000 era is associated with a 1st election *EG* of .14 (indicative of advantage for Democrats). I suggest a compromise between the -.10 threshold for plans appearing to favor Republicans and the .14 threshold for plans appearing to favor Democrats, setting the 1st election, efficiency gap threshold for states with relatively small delegations at $\pm$.12.

For states with larger Congressional delegations, I seek the 1st election efficiency gap score that is associated with a remainder-of-plan average efficiency gap of $\pm$.05. In the post-2000 era, the regression analysis in Figure 16 shows that a 1st election efficiency gap of -.06 is associated with a remainder-of-plan average efficiency gap of -.05. For plans exhibiting Democratic advantage, a 1st election efficiency gap of .09 is associated with a remainder-of-plan average efficiency gap of .05. This suggests using $\pm$.075 as the 1st election efficiency gap threshold for states with more than 15 CDs; it would seem only fair that any threshold that would flag an apparently pro-Republican plan for scrutiny should also flag an apparently pro-Democratic plan for scrutiny.

These proposed thresholds are set using only the post-2000 data, and would be slightly different if the entire dataset were used. Employing only the post-2000 data is appropriate here given the advances in redistricting technology and the increases in the durability of the efficiency gap apparent in the post-2000 data.

10.3 Summary

A summary of the efficiency gap thresholds I recommend and their properties is presented in Table 4. To restate the key steps in the argument and supporting findings from the data:

1. In section 9 I determined the values of the efficiency gap that are associated with politically meaningful departures from the long-run relationship between vote

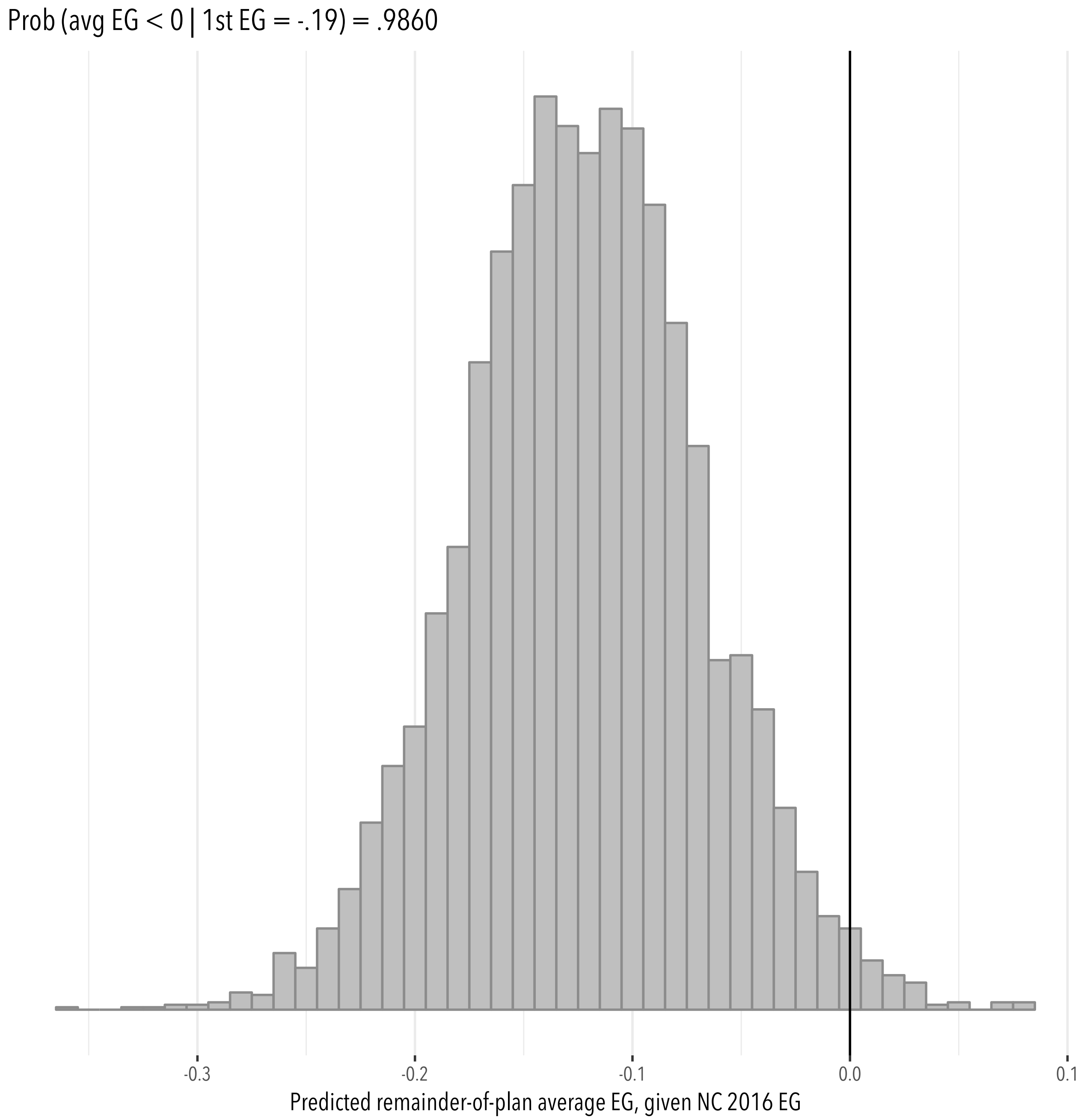

Figure 18: Histogram summarizing predictions as to the remainder-of-plan average efficiency gap for the current North Carolina districting plan, based on the historical, regression relationship between 1st election *EG* and remainder-of-plan average *EG*, subset to plans enacted since 2000, displayed in Figure 16.

shares and seat shares in state-level Congressional elections (see Figure 11 and 12).

2. One-half of a seat is the point at which a change in the allocation of seats is more likely than no change, given a particular split of the two-party vote.

3. In states with a relatively small number of Congressional districts (7 to 14), the value of the efficiency gap associated with at least a half-seat departure from the long-run relationship between vote shares and seat shares is $\pm.08$.

4. In states with a relatively large number of Congressional districts (15 or more), the value of the efficiency gap associated with at least a half-seat departure from the long-run relationship between vote shares and seat shares is $\pm.05$.

5. I then determined the value of the efficiency gap observed in the 1st election under a districting plan that is associated with a remainder-of-plan average efficiency gap that is at least as large as the thresholds defined above (section 10.2). These 1st election efficiency gap values differ between the parties. Hence I propose compromise values of $\pm.12$ in states with 7-14 Congressional districts and $\pm.075$ in states with 15 or more Congressional districts.

6. In states with 7-14 Congressional districts, 9 out of 30 plans enacted in the post-2000 era trip the compromise 1st election *EG* threshold of $\pm.12$. Just three of the nine plans go on to have a remainder-of-plan average efficiency gap smaller than the $\pm.08$ threshold.

7. In states with 15 or more Congressional districts, 8 out of 14 plans enacted in the post-2000 era trip the compromise 1st election threshold of $\pm.075$. All of these eight plans go on to have a remainder-of-plan average efficiency gap beyond the $\pm.05$ threshold.

References

Butler, David and Donald Stokes. 1974. *Political change in Britain: the evolution of electoral choice*. London: Macmillian.

Edgeworth, FY. 1898. "Applications of the Theory of Probability." *JR Statist. Soc* 51:534.

Gelman, Andrew and Gary King. 1990. "Estimating the Consequences of Electoral Redistricting." *Journal of the American Statistical Association* 85:274–82.

	Number of CDs	
	7-14 CDs	$\geq$ 15 CDs
EG $\implies \geq .5$ seat deviation from historical norm	$\pm .08$	$\pm .05$
1st election *EG* such that remainder-of-plan average *EG* exceeds threshold:		
Democratic advantage	.14	.09
Republican advantage	-.10	-.06
Compromise	$\pm .12$	$\pm .075$
Test positive rate, post-2000	9/30 (30%)	8/14 (57%)
False discovery rate, post-2000	3/9 (33%)	0/8 (0%)

Table 4: Rationale and summary of efficiency gap thresholds and their properties. Analysis restricted to plans in place for at least three Congressional elections.

Grofman, Bernard and Gary King. 2008. "The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after LULAC v. Perry." *Election Law Journal* 6(1):2–35.

Kendall, M. G. and A. Stuart. 1950. "The Law of Cubic Proportion in Election Results." *The British Journal of Sociology* 1(3):183–196.

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

Niemi, Richard G. and Patrick Fett. 1986. "The Swing Ratio: An Explanation and an Assessment." *Legislative Studies Quarterly* 11:75–90.

Niemi, Richard G. and Simon Jackman. 1991. "Bias and Responsiveness in State Legislative Districting." *Legislative Studies Quarterly* 16:183–202.

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

Tufte, Edward R. 1973. "The Relationship Between Seats and Votes in Two-Party Systems." *American Political Science Review* 67:540–554.

Wood, Simon N. 2006. *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC.