Rebuttal Report

Simon Jackman

December 21, 2015

Introduction

In this rebuttal report, I respond to criticisms made by Sean P. Trende and Professor Nicholas Goedert in their respective expert reports. I also conduct new empirical analyses further confirming the validity of the efficiency gap as a measure of partisan gerrymandering and the reasonableness of the proposed 0.07 threshold. More specifically, my principal contributions are the following:

- *First*, I respond to Goedert's various critiques of the efficiency gap and of the proposed efficiency gap threshold. Among other things, he misunderstands the relevance of efficiency gap data, cherry-picks information from my initial report while ignoring its broader context, and wrongly claims that plaintiffs' test would mandate "hyper-responsiveness" or prevent states from pursuing goals such as competitiveness or proportional representation.

- *Second*, I calculate several widely accepted prognostic measures—all based on the rates of true positives, false positives, true negatives, and false negatives—with respect to the odds of a district plan's efficiency gap changing signs over the plan's lifetime given a certain efficiency gap value in the plan's first election. Based on these measures, I conclude that the proposed 0.07 threshold is highly conservative. In fact, this threshold *sacrifices* some accuracy (which would be maximized at a lower threshold) in order to reduce the proportion of false positives.

- *Third*, I calculate the same prognostic measures with respect to the odds of a district plan's *average* efficiency gap, over its lifetime, having a different sign than that observed in the first election under a plan, given a certain efficiency gap value in this first election. Under this method, the proposed 0.07 threshold appears even more conservative, driving down the share of false positives to below 5%.

- *Fourth*, I compare the values of the efficiency gap in the *first* election under a plan and *on average* over the plan's lifetime. This relationship is impressively tight (r^2=0.73), indicating that a plan's initial bias is a very good predictor of its overall lifetime bias. For Act 43, this analysis allows us to predict that it will *average* a pro-Republican efficiency gap of almost 10% over the 2010 cycle as a whole.

- *Fifth*, I examine to what extent changes in party control over redistricting are responsible for the pro-Republican trend in the efficiency gap since the 1990s. In the current cycle, about *four times* more state house plans were designed by Republicans in full control of state government than in the 1990s. Had the distribution of party control over redistricting remained unchanged, essentially *all* of the pro-Republican movement in the efficiency gap over the last two decades would not have occurred. It is thus changes in party control, and *not* changes in the country's political geography, that primarily account for Republicans' growing redistricting advantage over the last generation.

- *Sixth*, I address recent work by Chen and Rodden (2013), cited by both Trende and Goedert for the proposition that Republicans enjoy a natural geographic advantage over Democrats. Chen and Rodden's simulated maps are not *lawful* because they ignore the Voting Rights Act and state redistricting criteria; they are based on presidential election results rather than more relevant state legislative election results; they do not constitute a representative sample of the entire plan solution space; and they are contradicted by other recent work (Fryer & Holden 2011) finding that randomly drawn plans *reduce* bias and *increase* electoral responsiveness.
- *Lastly*, I comment on Trende's analysis of particular state legislative and congressional plans. This analysis is marked by conceptual and methodological errors severe enough to render it useless. For example, Trende ignores two of the three prongs of plaintiffs' proposed test; he calculates congressional efficiency gaps without converting them from percentage points to House seats and for House delegations too small to generate reliable estimates; and he simply *substitutes* presidential election results for congressional election results whenever the latter are missing due to uncontested races. None of this work meets accepted standards of social science rigor.

1 Responses to Goedert's criticisms

In his report, Goedert offers several critiques of the efficiency gap and of the 0.07 threshold I recommended in my initial report, based primarily on the alleged instability of the efficiency gap. None of these critiques have merit. In this section, I respond to Goedert's points relying only on the analysis of my initial report and on the existing literature. My new empirical analyses appear in subsequent sections.

First, Goedert appears to believe that a plan's efficiency gap is only relevant to the extent that it sheds light on the partisan intent (or lack thereof) underlying the plan. He writes that "such intent cannot be inferred" from a large efficiency gap, that "a durable bias . . . is not even a sign of deliberate partisan intent," and that the "efficiency gap [is] a standard to measure partisan intent" (pp. 11, 13, 19). But this is not at all the legal function of the efficiency gap in plaintiffs' proposed test. Rather, partisan intent is its own independent inquiry, and the efficiency gap then comes into play at the *second* stage of the test, to determine if a plan's electoral *consequences* are sufficiently severe that it should be deemed presumptively unconstitutional. To put it simply, the efficiency gap is plaintiffs' measure of partisan *effect*, not of partisan *intent*. Goedert's misunderstanding of this basic point infects all of his discussion.

Second, Goedert observes that of *all* plans, anytime in the decade, with a *pro-Democratic* efficiency gap of greater than 0.07, a substantial proportion of them switch signs over their lifetimes (p. 11). In making this observation, Goedert cherry-picks a single bit of data from my initial report, and an irrelevant piece of data at that. This fact is irrelevant because it applies to plans no matter when their elections were held, while the appropriate universe for plaintiffs, defendants, and courts is limited to the *first* elections held under plans. It is the first elections that typically will be used in litigation, given Justice Kennedy's admonition in *Vieth* that plans should not be struck down based on a "hypothetical state of affairs," but rather "if and when the feared inequity arose" (*Vieth v. Jubelirer* (2004), p. 420). And the fact is misleading because it applies only to pro-Democratic efficiency gaps above 0.07, and not to the larger set of pro-Republican efficiency gaps above this threshold.

If we consider only plans that exhibit a pro-Democratic efficiency gap above 0.07 in their *first* elections, the probability that they will switch signs over their lifetimes drops by about five percentage points (Jackman Report, p. 61). And if we then turn to plans that exhibit a *pro-Republican* efficiency gap above 0.07 in their first elections—a more sizeable set, for which more accurate estimates are possible—this probability drops all the way to about 15% (Jackman Report, p. 61). In other words, of plans that open with large pro-Republican efficiency gaps, close to 85% of them continue to favor Republicans in every election for the remainder of the cycle. *This* is the most pertinent data point in my report, not the one cherry-picked by Goedert, and it reveals the persistence of many gerrymanders.

Third, Goedert discusses *congressional* district plans throughout his report, even though this case is exclusively about state legislative redistricting (pp. 7-8, 10, 12, 20). In doing so, he makes some of the same errors as does Trende: namely, not converting the efficiency gap from percentage points to House seats, and improperly handling uncontested races (in his case, by not adjusting for the uncontestedness *at all*, and simply treating the races as if all of the vote went to one party and none to the other). I discuss these errors in more detail later in this report.

Fourth, Goedert claims that it is "arbitrary" to focus on the first election after redistricting, and that doing so "biases toward a finding of *EG* durability" by ignoring wave elections (p. 14). As noted above, the first election after redistricting is the critical one for purposes of litigation, since under *Vieth*, it is after this election that a lawsuit will typically commence and have to be decided by the courts. Later elections are largely irrelevant for litigation purposes, since it is unreasonable to expect suits to be brought six or eight or even ten years into a cycle. Moreover, my analysis in no way ignored wave elections; to the contrary, I determined the odds that a plan's efficiency gap would switch signs by examining *all* elections held under the plan, waves and non-waves alike. If anything, the fact that most wave elections over the last forty years have not taken place in the first election after redistricting biases *against* a finding of durability, since these elections may well cause the efficiency gap to flip signs.

Fifth, Goedert is wrong that an efficiency gap of zero represents "'hyper-responsive' representation" (p. 2). In fact, as he has recognized in his own prior work, an efficiency gap of zero corresponds almost exactly to the responsiveness actually displayed by American elections over the course of the twentieth century, under which "a 1% increase in vote share will produce about a 2% increase in seat share" (Goedert 2014, p. 3). Indeed, this correspondence is one of the efficiency gap's most attractive properties, and it explains why Goedert himself calculated a quantity nearly identical to the efficiency gap in his work (Goedert 2014; Goedert 2015).

And sixth, Goedert is wrong as well that plaintiffs' proposed test might discourage states from pursuing worthwhile goals such as competitiveness or proportional representation (pp. 6-10). If a state's aim in redrawing districts was to make them more competitive or to produce more proportional representation, then the partisan intent required by the first prong of plaintiffs' test would not be present. Even if partisan intent were somehow found, the state would likely be able to show that its plan's large efficiency gap was necessitated by its pursuit of competitiveness or proportional representation. And in any event, competitiveness and proportional representation are extremely rare objectives in American redistricting. Only *one* state, Arizona, has a competitiveness requirement, and not a *single* state has a proportional representation criterion. (And needless to say, line-drawers do not tend to seek out either of these goals on their own.)

2 Reliability of a district plan's first efficiency gap

Having rebutted Goedert's criticisms using preexisting data, I now provide further analysis of the reliability of the first efficiency gap (*EG*) observed in the life of a district plan. This played a key role in the determination of the threshold *EG* value in my initial report. In that report, I focused on the probability of a "sign-flip": that is, given the magnitude of the efficiency gap observed in the first election under a district plan, what can we infer about the likelihood that all subsequent efficiency gaps observed under that plan will have the same sign as that from the first election.

Under this approach, just one election that produces an efficiency gap with a different sign from the efficiency gap in the first election will generate a "failure," in the sense we would say that the plan has generated an efficiency gap that conflicts with that from the first election. In short, the "constant sign" analysis in my original report considers the most extreme set of efficiency gap estimates produced under a plan and insists that they have the same sign. In this sense, the "constant sign" analysis I performed is a quite stringent and conservative test of what we can or ought to infer from the efficiency gap observed in the first election under the district plan. Another approach would be to inquire as to the *average* efficiency gap over the life of the district plan. A summary statistic such as the average is—by definition—less sensitive to extreme values. At the same time—and again, by definition—the average measures central tendency or typicality, and is the most widely used summary statistic in existence. I thus consider how well the first *EG* observed under a district plan predicts the average *EG* observed over the life of the plan.

But I first provide some additional analysis of the prognostic properties of the first efficiency gap observed under a district plan. In each instance the test is whether the first *EG* observed under a plan exceeds a given threshold value. The outcome of interest is whether the plan's remaining efficiency gaps have the same sign as the *EG* from the first election. For purposes of this exercise, plans are classified as "positive" (all *EG* scores under the plan have the same sign) or "negative" (*EG* scores differ in sign). With these definitions in place, we can then classify plans according to the accuracy of the prediction implicit in the first *EG* observed under the plan:

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

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

1. sensitivity, or the *true positive rate*: proportion of positives that test positive, TP/(TP + FN)

2. specificity, or the *true negative rate*: proportion of negatives that test negative, TN/(TN + FP)

3. *balanced accuracy*, the average of the sensitivity and the specificity
4. *accuracy*, the proportion of cases that are true positives or true negatives, (TP + TN)/(TP + FP + FN + TN).
5. the *false positive rate*; proportion of negative cases that test positive, 1 minus the specificity or FP/(TN + FP).
6. the *false discovery rate*; proportion of cases testing positive that are actually negative, FP/(TP + FP).
7. the *false omission rate*; proportion of cases that test negative that are actually positive, FN/(FN + TN).

Figure 1 shows how these prognostic performance indicators vary as a function of the absolute *EG* threshold (on the horizontal axis in the figure). That is, as we move to the right in each panel of the graph, the test is becoming increasingly stringent: larger absolute values of the efficiency gap in the first election under a district plan are required to trip the increasingly higher threshold. When the threshold is set to zero, all plans trip the threshold (all first-election *EG*s are greater than zero in magnitude, by definition) and so all cases test positive; in this case the sensitivity is 1, while conversely the specificity is 0 and the false positive rate is 1 (all negatives test positive).

The test has better properties as the threshold grows, with the accuracy measures maximized around absolute values of .03 to .04. Yet accuracy is not all in this context. The rate of false positives is quite high at thresholds where the accuracy is high, as is the false discovery rate. At a threshold of .03, for example, over half of plans that would go on to exhibit sign flips in their *EG*s would test positive and be flagged for inspection; of the plans selected for scrutiny, more than a third would turn out to have *EG* sign flips over the life of the plan. The .07 threshold is thus a conservative standard, the point at which the rate of false positives is becoming reasonably low (25%), without letting the false omission rate go above 50%.

It is worth noting the weight being put on false discoveries or false alarms versus the weight on false omissions in this context, which in turn reflects the conservatism and caution of the thinking underlying the .07 threshold. We propose accepting *twice* the rate of false omissions (plans that should have been scrutinized but were not) than the rate of false discoveries (plans that would be flagged for scrutiny given the *EG* observed in the first election, but would then go on to display sign flips). To reiterate: the proposed standard for judicial scrutiny is cautious and conservative, erring on the side of letting even durably skewed plans stand.

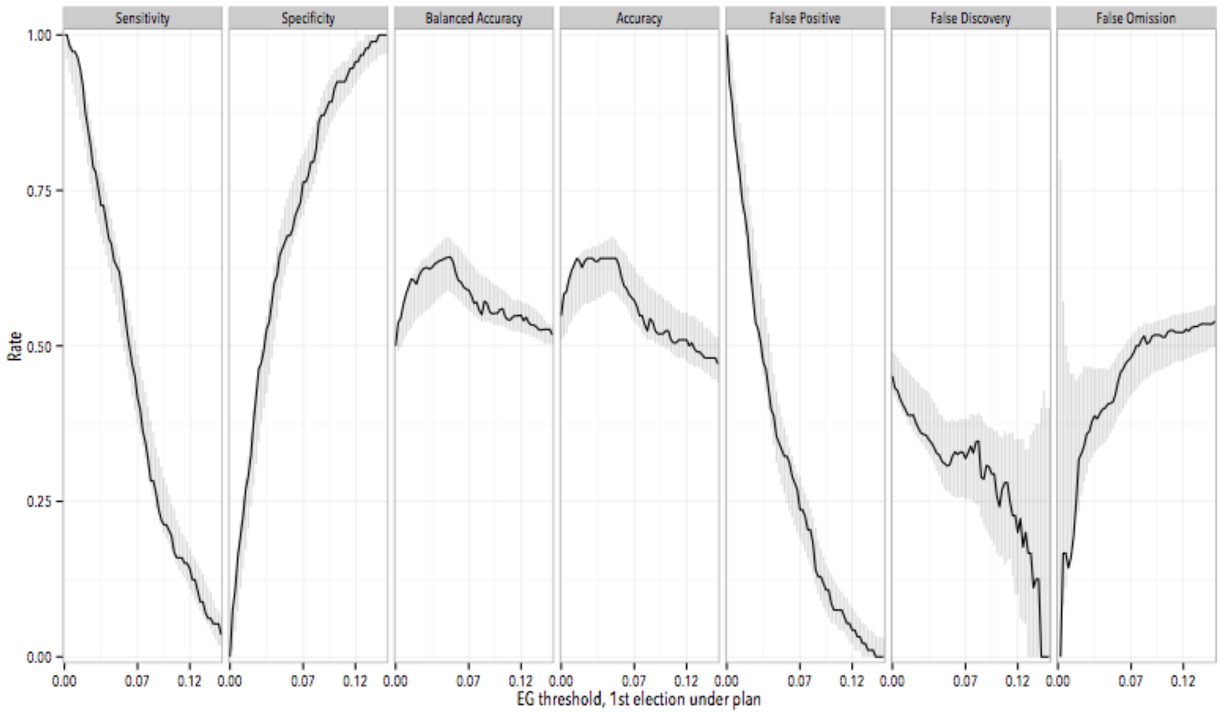

Figure 1: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the subsequent efficiency gaps recorded under the district plan all have the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis spans all state legislative elections and district plans as per my initial report, 1972-2014.

Figure 2 repeats this analysis, but only considering the performance of *negative* values of the first-election efficiency gap threshold, consistent with Republican advantage (and more relevant to the Wisconsin plan at issue). Here the threshold becomes less stringent as we move across the horizontal axis from left to right, from larger negative thresholds to closer to zero at the right hand edge of each panel. With a large negative threshold (left hand edge of each panel), almost all plans test negative and so the sensitivity is close to zero, the specificity is 1, and the false positive rate is zero. The accuracy measures increase as the threshold becomes less stringent, attaining maxima in the range -.05 to -.02. Again—and consistent with the cautious approach we take—we emphasize that accuracy is not the sole criterion we use to evaluate a decision rule. At low values of the threshold, where accuracy is maximized, the false positive and false discovery rates are relatively high. On the other hand, at the proposed threshold value of -.07, the false positive rate is under 10% (fewer than 10% of plans with efficiency gaps changing signs would be scrutinized), and the false omission rate is about 35% (close to

35% of plans would not be flagged despite having *EG*s of the same sign over their lifetimes). The proposed threshold again errs on the side of restraint, tolerating a higher rate of false omissions than false discoveries.

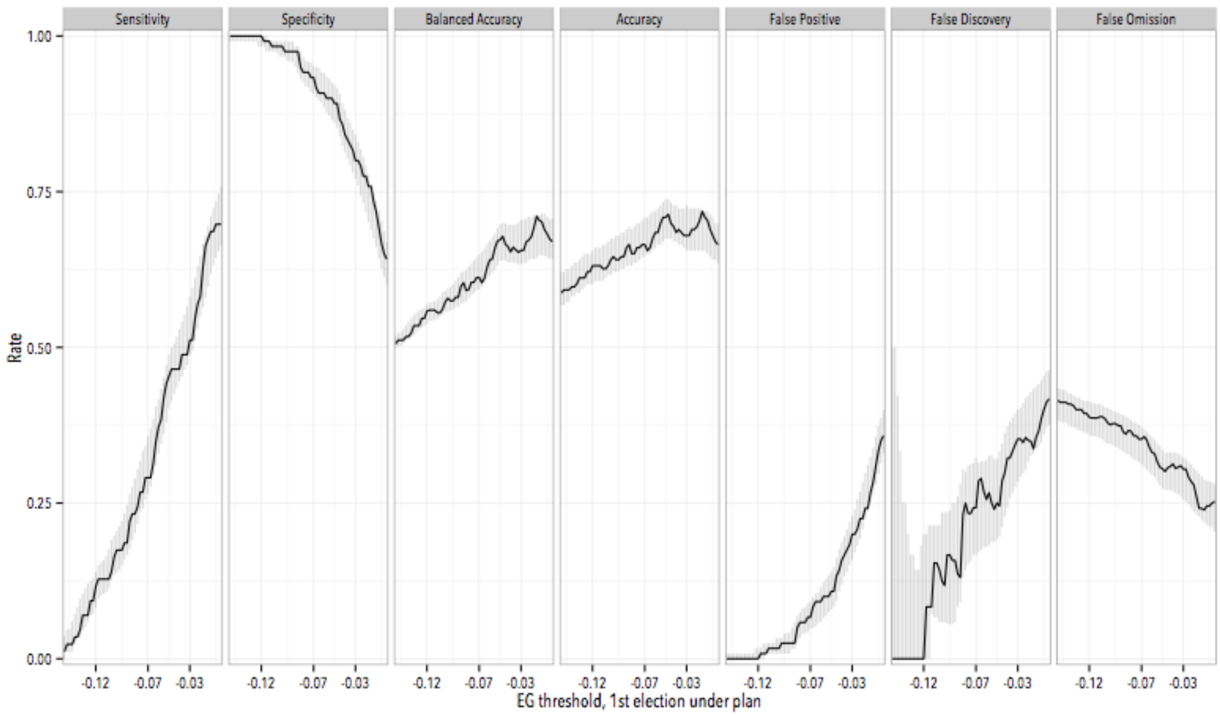

Figure 2: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the subsequent efficiency gaps recorded under the district plan all have the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis examines negative, first-election threshold values of the efficiency gap, consistent with Republican advantage.

Figure 3 presents the corresponding analysis of *positive* values of the first-election *EG* threshold, consistent with Democratic advantage. Here the proposed threshold becomes more stringent as we move to the right of each panel, in the sense that fewer plans trip the threshold. At high values of the threshold (the right hand edge of each panel), no plans trip the threshold and all are classified as "negatives," leading to a specificity of 1, and false positive and false discovery rates of zero. Once again, accuracy is maximized at a less stringent threshold than the proposed .07 standard, around .03. The false positive rate is much lower at the proposed threshold of .07 than at the accuracy-maximizing threshold of .03. Note that the false discovery rates are moderately large but unstable and estimated with considerable imprecision; this is because there are so few plans exhibiting high (pro-Democratic) levels of *EG* in their first election. Moreover, of the few plans that do trip a given pro-Democratic threshold in their first election, it is reasonably likely that they will record efficiency gaps that will change sign over the life of the plan; this sign-flip or "false discovery" probability is about 35% at the proposed threshold of .07.

Comparing the analyses in Figures 2 and 3, we see an asymmetry in the results. The .07 threshold is more permissive with respect to plans that begin life exhibiting Democratic advantage than it is for plans that initially exhibit Republican advantage. At a +/- .07 threshold, the false discovery rate for plans initially exhibiting Republican advantage is under 10%, but around 35% for plans initially exhibiting Democratic advantage. As Figure 3 shows, it is difficult to find a threshold for apparently pro-Democratic plans that drives the false discovery rate to reliably low levels, if only because the historical record has relatively few instances of these types. We also note that the .07 threshold generates false omission rates of about 30% for both sets of plans.

Because the preceding discussion is somewhat technical, it is worth restating its principal conclusion: It is that an efficiency gap threshold of 0.07 is quite conservative, in that it sacrifices some accuracy (which would be maximized at a threshold of around 0.03) in order to drive down the false positive and false discovery rates. At a threshold of 0.07, in fact, the false positive and false discovery rates are about *half* of the false omission rate, indicating that there are about twice as many plans that are *not* being flagged even though their *EG* signs would remain one-sided throughout the cycle, than there are plans that *are* being flagged even though their *EG* signs would flip. This is further powerful confirmation of the reasonableness of the 0.07 efficiency gap threshold.

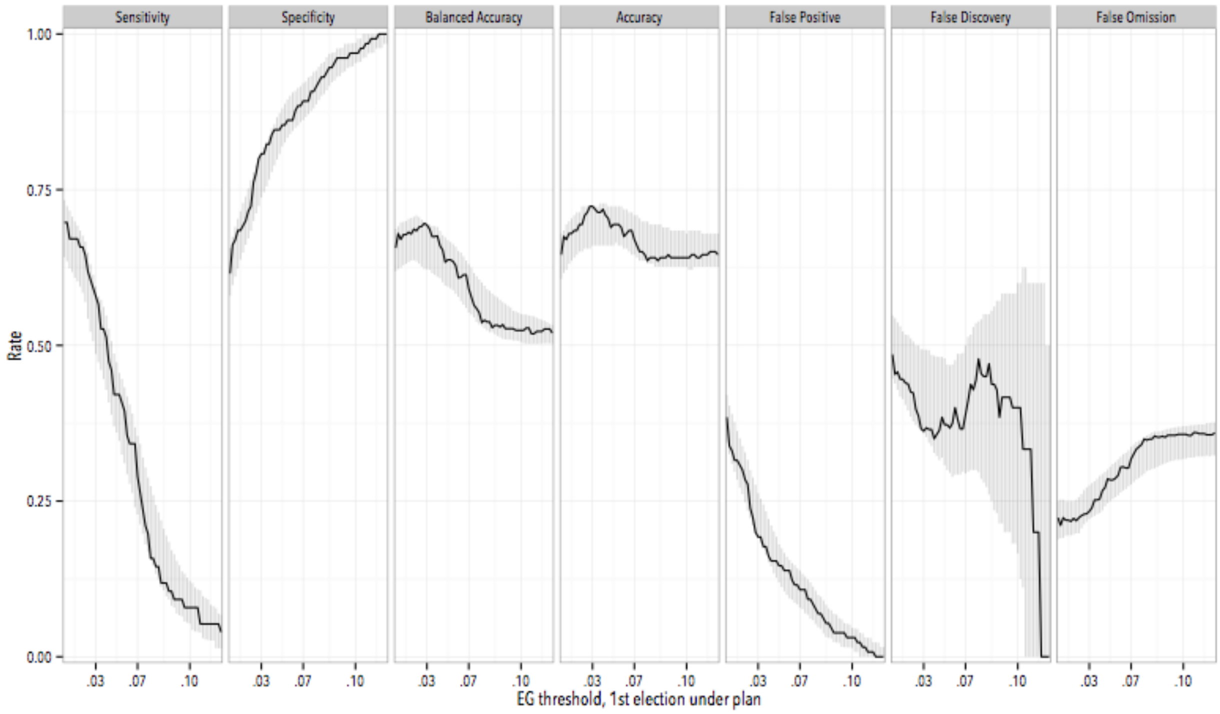

Figure 3: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the subsequent efficiency gaps recorded under the district plan all have the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis examines positive, first-election threshold values of the efficiency gap, consistent with Democratic advantage.

3 First-election efficiency gap reliability with respect to the plan-average efficiency gap sign

Next we consider a slightly different kind of test; given that the first election under a district plan produces a value of the efficiency gap above or below a given threshold, how likely is it that the *average* value of the efficiency gap produced over the life of the plan lies on the same side of zero as that of the first election? Recall that the sign of the efficiency gap speaks to the corresponding direction of partisan advantage ($EG < 0$ is consistent with Republican advantage; conversely for $EG > 0$). We expect that this will be a less strenuous test than asking if *any* EG has an opposite sign to the first EG observed under a district plan.

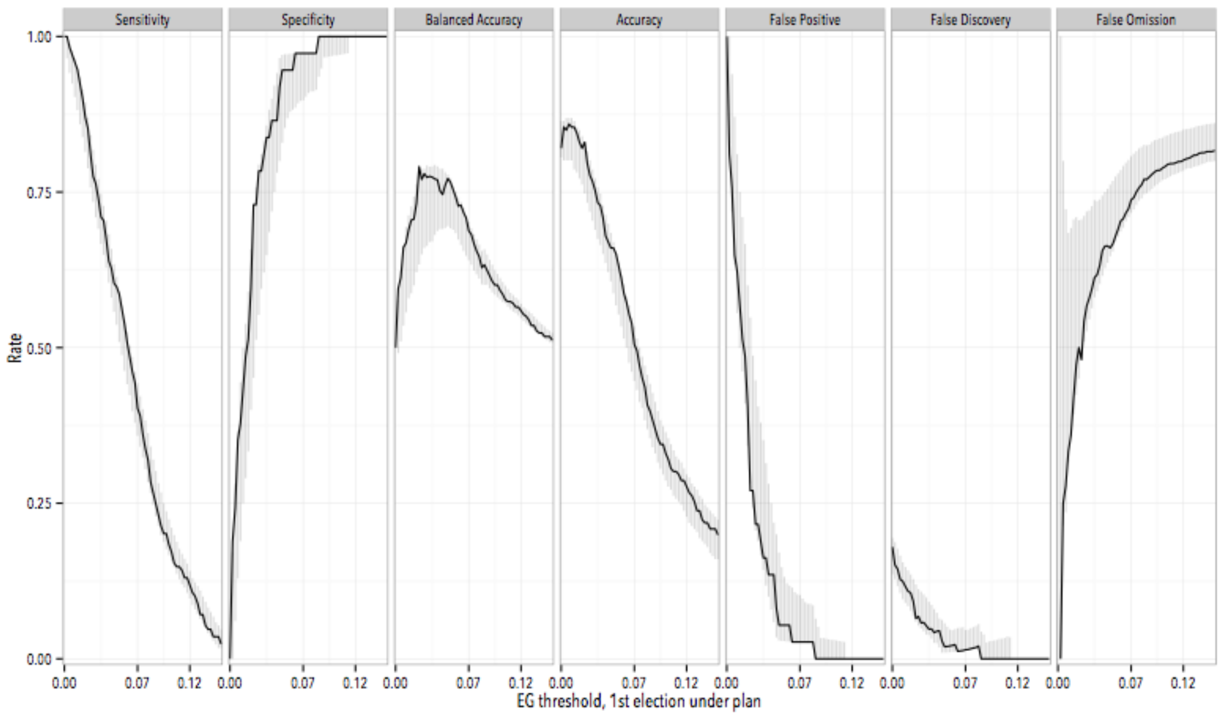

Figure 4: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the average efficiency gap recorded under the district plan has the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis spans all state legislative elections and district plans as per my initial report, 1972-2014.

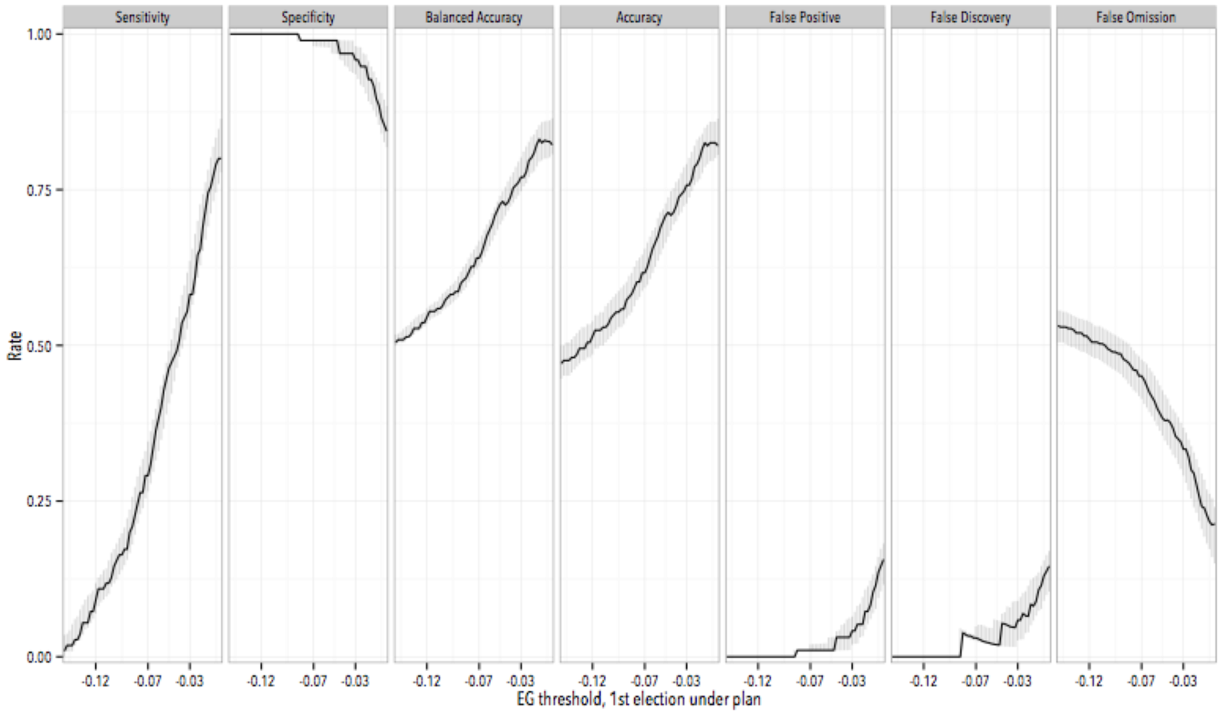

Figure 5: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the average efficiency gap recorded under the district plan has the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis examines negative, first-election threshold values of the efficiency gap, consistent with Republican advantage.

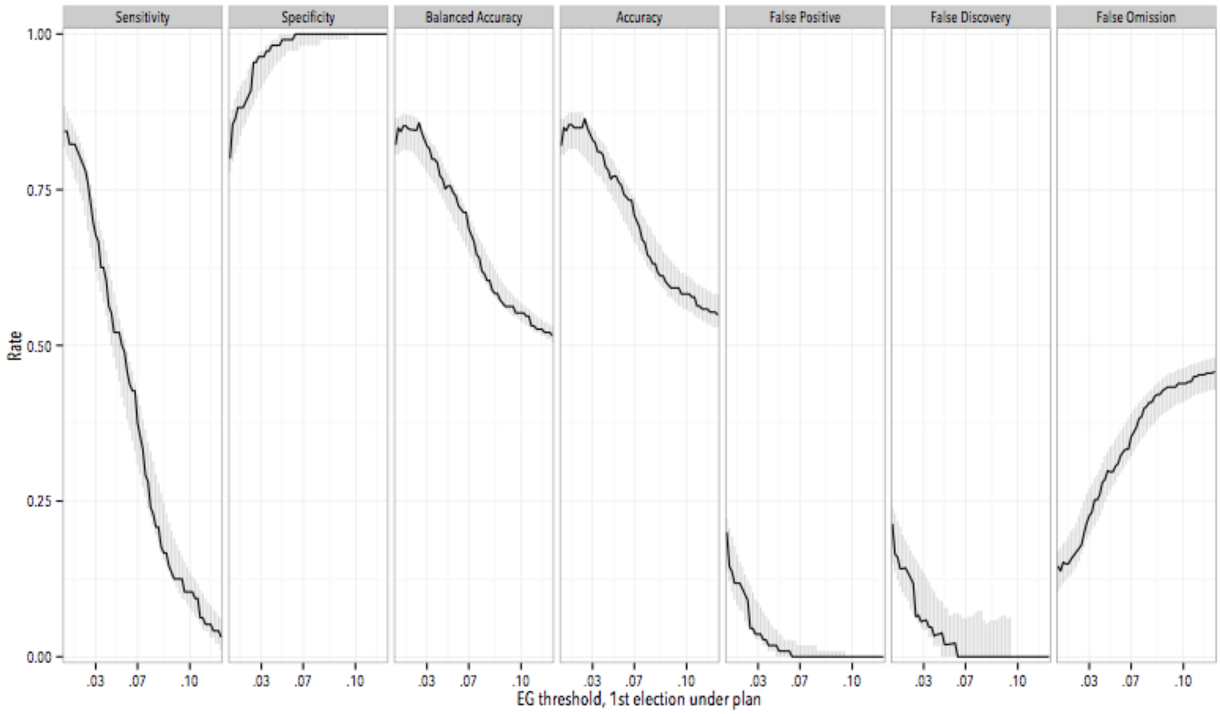

Figure 6: Prognostic performance measures, first efficiency gap under a district plan more extreme than threshold (horizontal axis) as a predictor of whether the average efficiency gap recorded under the district plan has the same sign as the first efficiency gap. Vertical lines indicate 95% confidence intervals. Analysis examines positive, first-election threshold values of the efficiency gap, consistent with Democratic advantage.

Figures 4, 5 and 6 show the prognostic performance of the first-election *EG* with respect to the sign of the corresponding plan's average *EG*, looking at the absolute value of the first-election *EG* (Figure 4), negative first-election efficiency gaps (Figure 5) and positive first-election efficiency gaps (Figure 6). The first thing to observe is the generally superior prognostic performance when it comes to forecasting the sign of the *plan-average* efficiency gap, relative to the prognostic performance with respect to *all* of the plan's efficiency gaps having the same sign. As anticipated, the former is better predicted by the plan's first-election efficiency gap than the latter. Second, the accuracy-versus-caution tradeoff noted earlier is also apparent. The proposed threshold of +/- 0.07 trades away accuracy for very low false positive and false discovery rates, below 5%, at the cost of higher false omission rates, a pattern we observed earlier. Finally, note that at the proposed threshold of +/- 0.07, almost one-half of all plans with a negative (pro-Republican) average *EG* would *not* be candidates for scrutiny (right-hand panel of Figure 5); about one-third of plans with a positive (pro-Democratic) average *EG* also would not trigger the threshold for scrutiny.

4 Relationship between the first-election efficiency gap and the plan-average efficiency gap

I next present analysis on a related issue, the relationship between the magnitudes of the *first* efficiency gap observed under a plan and the *average* efficiency gap we observe over the life of the plan. Does a larger or smaller first-election efficiency gap portend anything for the average value of the efficiency gap generated over the life of a district plan?

Clearly the first value of the efficiency gap and the plan-average efficiency gap are related; the former contributes to the calculation of the latter, and after the first election under a district plan we observe at most four more elections under the plan (given elections every two years in most states and redistricting once a decade). Accordingly we expect a positive correlation between the two quantities. The interesting empirical question—and one with considerable substantive implications for the issue at hand—is *how strong* the relationship is between the first-election efficiency gap and the corresponding plan-average efficiency gap. This speaks to the reliability of the first-election *EG* measure as a predictor of *EG* over the life of the plan.

Figure 7 shows the relationship between the first-election *EG* and the average *EG* observed over the entire plan. Note that we restrict this analysis to plans with at least three elections, so that the first election does not unduly contribute to the calculation of the average; this restriction has the consequence of omitting elections from the most recent round of redistricting after the 2010 Census, which have contributed at most two elections. The black diagonal line on the graph is a 45-degree line: if the relationship between first-election *EG* and plan-average *EG* were perfect, the data would all lie on this line. Instead we see a classic "regression-to-the-mean" pattern, with a positive regression slope of less than one (as indeed we should, given that the first-election *EG* on the horizontal axis contributes to the average plotted on the vertical axis). But the relationship here is especially strong. The variation in plan-average efficiency gaps explained by this regression is quite large, about 73%; after taking into account the uncertainty in the *EG* scores (stemming from the imputation procedures used for uncontested districts; see my initial report) a 95% confidence interval on the variance explained measure ranges from 67% to 74% (the uncertainty has the consequence of tending to make the regression fit slightly less well). That is, even given the uncertainty that accompanies *EG* measures due to uncontestedness, the relationship between first-election *EG* and plan-average *EG* is quite strong.

In particular, at the threshold values of +/- 0.07 there is very little doubt as to the plan-average value of the efficiency gap. The historical relationship between first-election *EG* and plan-average *EG* shown in Figure 7 indicates that a first-election *EG* of -.07 is typically associated with a plan-average *EG* of about -0.053 (95% CI -0.111 to 0.004); the probability that the resulting, expected plan-average *EG* is negative is 96.5%. Conditional on a first-election *EG* of .07 we typically see a plan-average *EG* of about 0.037 (95% CI -0.021 to 0.093); the probability that the resulting, expected plan-average *EG* is positive is 89.8%. This constitutes additional, powerful evidence that (a) first-election *EG* estimates are predictive with respect to the *EG* estimates that will be observed over the life of the plan; and (b) the threshold values of +/- 0.07 are conservative, generating high-confidence predictions as to the behavior of the district plan in successive elections.

In the particular case of Wisconsin in 2012—the first election under the plan in question—I estimated the efficiency gap to be -0.133 (95% CI -0.146 to -0.121). The analysis of historical data discussed above—and graphed in Figure 7—indicates that the plan-average *EG* for this plan will be -0.095 (95% CI -0.152 to -0.032)[1], a quite large value by historical standards, placing the current Wisconsin district plan among the five to ten most disadvantageous district plans for Democrats in the data available for analysis. The probability that the Wisconsin plan—if left undisturbed—will turn out to have a positive, pro-Democratic, average efficiency gap is for all practical purposes zero (less than 0.1%).

1 It is also worth stressing that the confidence interval is computed so as to take into account uncertainty from all known sources: in the underlying efficiency gap scores themselves, the fact that the 2012 *EG* scores for Wisconsin are large by historical standards, and in the regression relationship between first-election *EG* and plan-average *EG*.

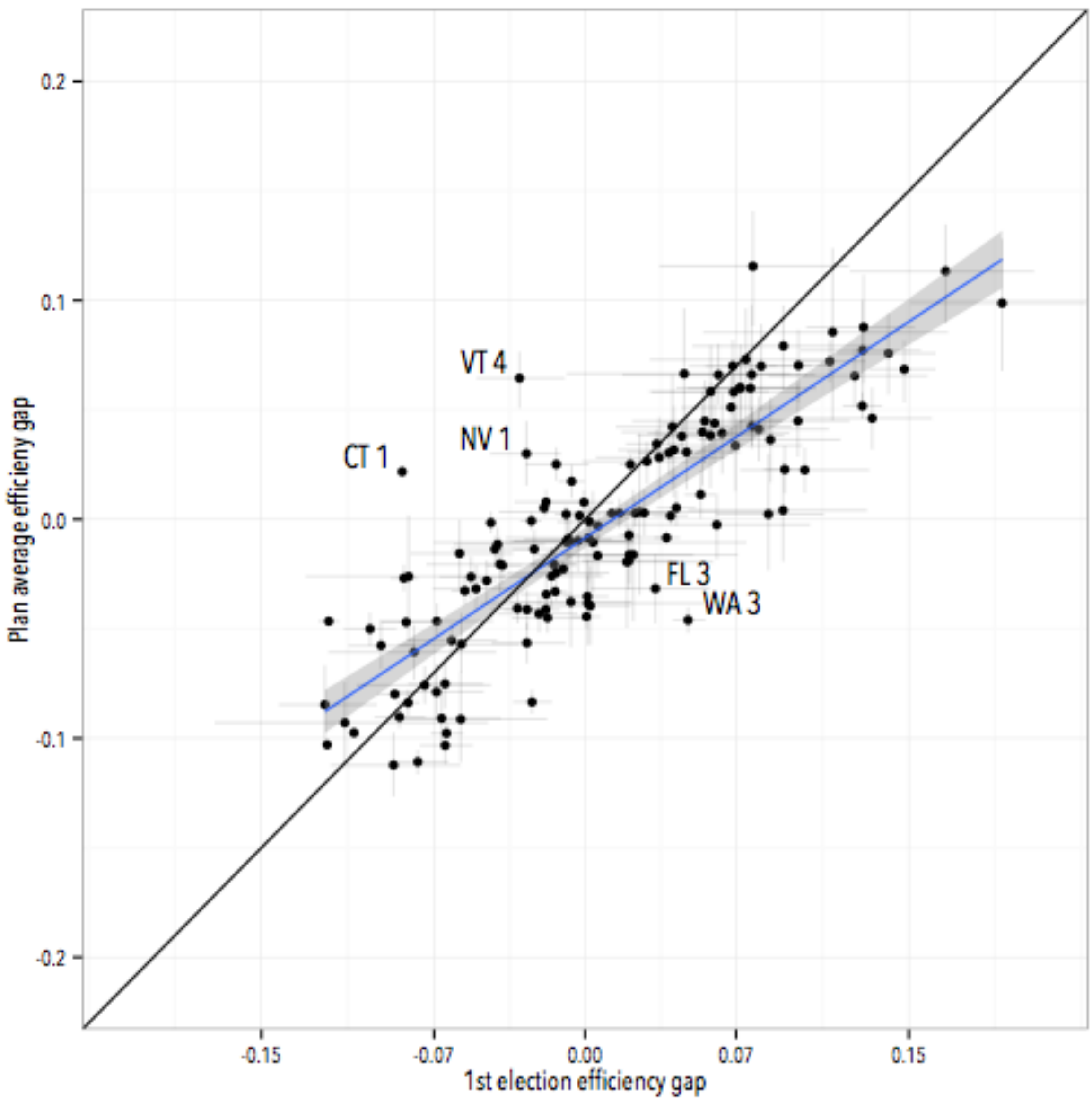

Figure 7: Scatterplot of first-election efficiency gap scores (horizontal axis) and plan-average efficiency gap scores (vertical axis). The diagonal black line is a 45-degree line; the data would lie on this line if first-election efficiency gaps coincided with plan-average efficiency gaps. The solid blue line is a linear regression with slope .64 (95% CI 0.57 to 0.72); the shaded region around the blue line is a 95% confidence interval for the regression line. 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 plan-average *EG*, stemming from imputations for uncontested districts. Outliers are labeled (state, plan). Analysis restricted to plans with at least three elections (1972-2010), omitting plans adopted after the 2010 Census. The first-election *EG* for the current Wisconsin plan is -0.133 (95% CI -0.146 to -0.121).

5 Party control as an explanation for change in the efficiency gap

Both Trende and Goedert point out that, on average, state house plans have exhibited pro-Republican efficiency gaps in recent years (Trende, paragraphs 129-30; Goedert p. 19). They then argue that this pro-Republican mean is attributable to a natural pro-Republican political geography in many states. However, as I found in my initial report, the *overall* efficiency gap average, over the entire 1972-2014 period, is very close to zero (Jackman Report, p. 35, 45, 57). There is thus no sign of a natural pro-Republican advantage in the dataset as a whole, nor any evidence (despite Trende and Goedert's unsupported assertions to the contrary) that states' political geography is changing in ways that favor Republicans.

In fact, the one historical change that *is* undeniable is the trend toward unified Republican control over redistricting. As Figure 8 displays, only about 10% of all state house plans were designed by Republicans in full control of the state government in the 1990s, compared to about 30% by Democrats in full control and about 60% by another institution (divided government, a commission, or a court). But in the 2000s, Republicans were fully responsible for slightly *more* plans than were Democrats (about 20% versus about 15%). And in the 2010s, the partisan gap jumped again, to about *40%* of plans designed entirely by Republicans, versus less than 20% designed entirely by Democrats.

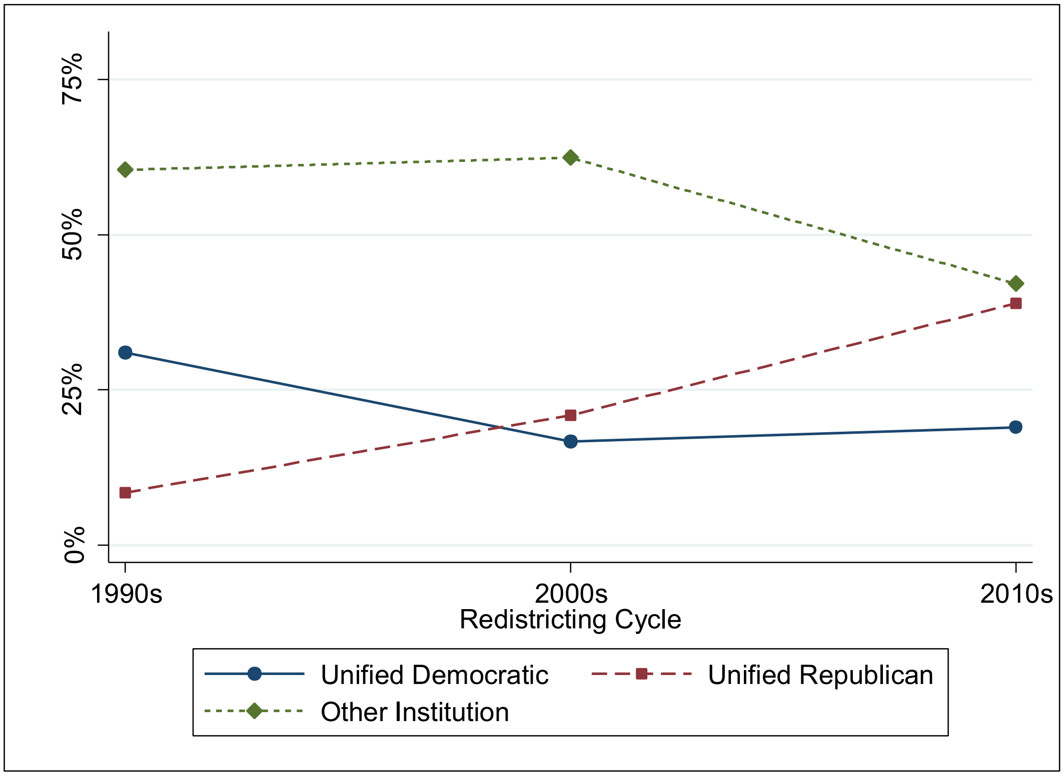

Figure 8: Share of all state house plans, by cycle, designed by Democrats in unified control of state government, by Republicans in unified control of state government, or by another institution (divided state government, commission, or court).

To determine the impact of this change in party control on the change in the efficiency gap over the last generation, I carry out three regressions, one for the 1990 redistricting cycle, one for the 2000 cycle, and one for the 2010 cycle. In each case, state house plans' efficiency gaps are the dependent variable, and unified Democratic control over redistricting and unified Republican control over redistricting are the independent variables. (The omitted category is any other institution responsible for redistricting, such as divided government, a court, or a commission.) Figure 9 then displays the *actual* average efficiency gap for each cycle, as well as the *predicted* average efficiency gap if the distribution of party control over redistricting had remained unchanged since the 1990s.

As is evident from the chart, state house plans' average efficiency gap in the 2000 cycle would have been substantially less pro-Republican (by about 0.5 percentage points) had Republicans not gained control of more state governments in this cycle relative to the 1990s. And in the current cycle, *all* of the efficiency gap's movement in a Republican direction would have been erased had the distribution of party control over redistricting not changed since the 1990s. That is, if the same distribution of party control had existed in this cycle as in the 1990s, state house plans' average efficiency gap would have been very close to zero, not over 3% in a Republican direction. Accordingly, it is the change in party control that appears to account for essentially all of the pro-Republican trend in the efficiency gap over the past two decades—and not, as claimed by Trende and Goedert, a dramatic alteration of the country's political geography.

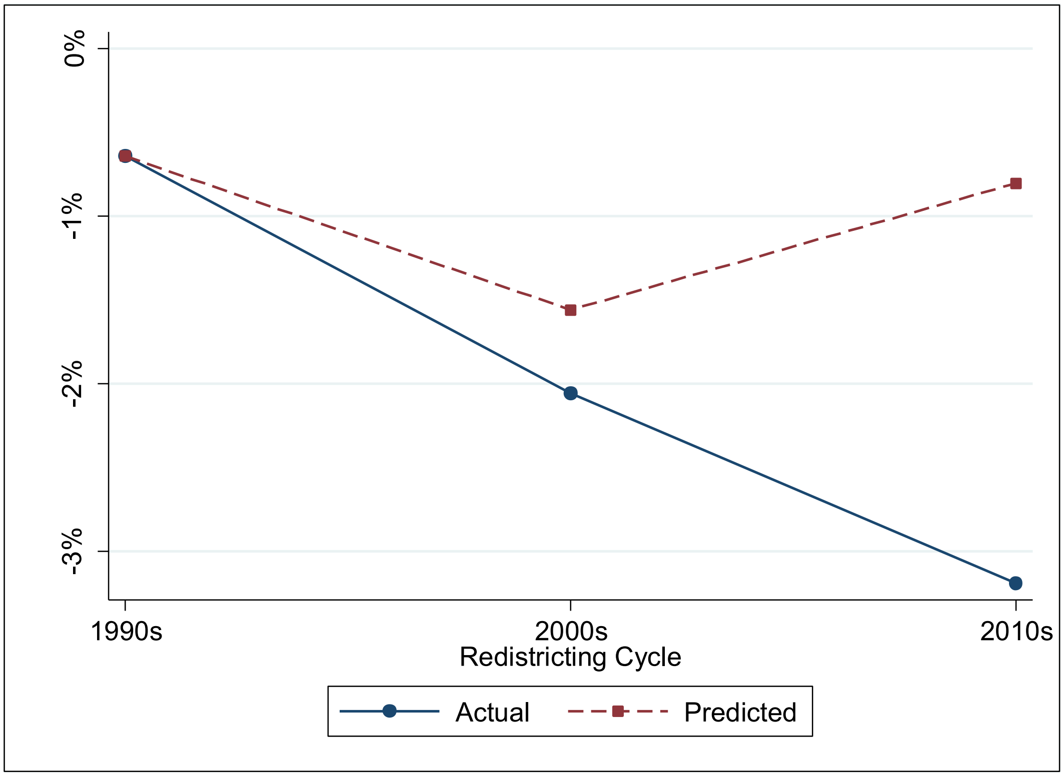

Figure 9: Actual and predicted values of state house plans' average efficiency gaps by cycle. Predicted values calculated assuming that the 1990s distribution of party control over redistricting remained constant in subsequent cycles.

6 Response to the Chen and Rodden map simulations

Both Trende and Goedert cite a recent article by Chen and Rodden (2013) that purports to find, based on simulations of hypothetical district maps, that random redistricting would benefit Republicans because of their more efficient spatial allocation (Trende, paragraphs 89, 126; Goedert, pp. 13, 18, 21). While I respect Chen and Rodden's contribution, there are several issues with their work that make it inapplicable here.

First, Chen and Rodden do not even attempt to simulate *lawful* plans. Rather, they simulate plans "using only the traditional districting criteria of equal apportionment and geographic contiguity and compactness" (Chen and Rodden, 248). They do not take into account Section 2 of the Voting Rights Act, which often requires majority-minority districts to be constructed. They also do not take into account Section 5 of the VRA, which until 2013 meant that existing majority-minority districts could not be eliminated in certain states. And they do not take into account state-level criteria such as respect for political subdivisions and respect for communities of interest, which are in effect in a majority of states (NCSL 2010, pp. 125-27).

Second, Chen and Rodden only use *presidential* election results in their analysis, but these outcomes may diverge from *state legislative* election results due to voter roll-off as well as voter preferences that vary by election level. As Stephanopoulos and McGhee have noted, "If certain voters consistently support Republicans at the presidential level and Democrats at the legislative level, then presidential data may produce more pro-Republican estimates than legislative data" (Stephanopoulos & McGhee, 870). In fact, this is exactly what seems to be occurring; at the congressional level, efficiency gaps are about 6% more Republican when they are calculating using presidential data than when they are computed on the basis of congressional election results.

Third, Chen and Rodden's simulated maps do not constitute a representative sample of the entire plan solution space. Their simulation algorithm has "no theoretical justification," is "best described as ad-hoc," and is not "designed to yield a representative sample of redistricting plans" (Fifield et al. 2015, pp. 2-3; Altman & McDonald 2010, p. 108). The explanation for this lack of representativeness is highly technical and involves the details of the particular simulation approach adopted by Chen and Rodden. But its implication is clear: that no conclusions can yet be drawn about the partisan consequences of randomly drawn maps.

Lastly, Chen and Rodden's results are directly contradicted by Fryer and Holden, who also simulated contiguous, compact, and equipopulous districts for multiple states. Unlike Chen and Rodden, Fryer and Holden found that, "[u]nder maximally compact districting, measures of Bias are slightly *smaller* in all states except [one]" (Fryer & Holden 2011, p. 514). Fryer and Holden also found that "[i]n terms of responsiveness . . . there are large and statistically significant" *increases* in all states, sometimes on the order of a fivefold rise (p. 514). Their analysis thus leads to the opposite inference from Chen and Rodden's: that randomly drawn contiguous and compact districts favor *neither* party and substantially boost electoral responsiveness.

7 Trende's analysis of particular plans

Trende devotes a large portion of his report (paragraphs 106-31) to analyzing the efficiency gaps of particular state legislative and congressional plans. He first examines a set of seventeen state legislative plans that had efficiency gaps favoring the same party over their entire lifespans, arguing that not all of these plans were gerrymanders (paragraphs 106-14). He then cites a series of congressional plans, some of which he claims had large efficiency gaps despite not being gerrymanders, and others of which allegedly had small efficiency gaps despite being gerrymanders (paragraphs 115-24). All of this analysis is riddled with conceptual and methodological errors that, in my judgment, renders it unreliable and unhelpful to the court.

Beginning with the set of seventeen state legislative plans that had efficiency gaps of the same sign throughout their lifespans, Trende asserts that they "would be included in the definition of a gerrymander," and are a "list of gerrymandered states" (paragraphs 109-10). But neither plaintiffs nor I argue that these plans should have been held unconstitutional. That is, neither plaintiffs nor I argue that these plans were designed with partisan intent (the first element of plaintiffs' proposed test), that their initial efficiency gaps exceeded a reasonable threshold (the second element), or that their efficiency gaps could have been avoided (the third element). To the contrary, I simply included these plans in my report to illuminate historical cases in which the efficiency gap's direction did not change over the course of a decade. I never stated or implied that these plans should have been deemed unlawful.

However, if we focus on the plans among the seventeen that likely *would* have failed plaintiffs' proposed test (at least the first two elements), we see that both the test and the efficiency gap perform exceptionally well. Five of the seventeen plans featured unified control by a single party over redistricting (from which, like Goedert (2014) and Goedert (2015), we can infer partisan intent) as well as an initial efficiency gap above 7% (the threshold I recommended in my initial report): Florida in the 1970s, Florida in the 2000s, Michigan in the 2000s, New York in the 1970s, and Ohio in the 2000s. Assuming that these plans' large efficiency gaps were avoidable (a granular inquiry that cannot be carried out here), it would have been quite reasonable for all of these maps to attract heightened judicial scrutiny. In particular:

- Florida's plan in the 1970s was designed exclusively by Democrats, opened with a 9.9% pro-Democratic efficiency gap, averaged a 7.0% pro-Democratic efficiency gap over its lifespan, and never once favored Republicans.

- Florida's plan in the 2000s was designed exclusively by Republicans, opened with a 8.9% pro-Republican efficiency gap, averaged a 11.2% pro-Republican efficiency gap over its lifespan, and never once favored Democrats.
- Michigan's plan in the 2000s was designed exclusively by Republicans, opened with a 12.0% pro-Republican efficiency gap, averaged a 10.3% pro-Republican efficiency gap over its lifespan, and never once favored Democrats.
- New York's plan in the 1970s was designed exclusively by Republicans, opened with a 10.7% pro-Republican efficiency gap, averaged a 9.7% pro-Republican efficiency gap over its lifespan, and never once favored Democrats.
- Ohio's plan in the 2000s was designed exclusively by Republicans, opened with a 8.6% pro-Republican efficiency gap, averaged a 9.0% pro-Republican efficiency gap over its lifespan, and never once favored Democrats.

Accordingly, we see that if my report's set of seventeen plans is analyzed properly, the opposite conclusion emerges from the one advocated by Trende. Only a subset of the seventeen plans likely would have failed plaintiffs' proposed test. But *every member* of this subset turns out to have been an exceptionally severe and durable gerrymander, featuring a very large and consistent efficiency gap over its lifespan. These are *precisely* the historical cases in which judicial intervention may have been advisable.

After commenting on these seventeen state legislative plans, Trende discusses a series of *congressional* plans, all from the 2000 and 2010 redistricting cycles. These congressional plans are entirely irrelevant to this case, which deals only with state legislative redistricting. Neither in their complaint nor in their subsequent filings do plaintiffs ever argue that their approach should be applied to congressional plans. And neither Mayer nor I provide any empirical analysis of congressional plans. In my initial report, in particular, I examined state legislative plans from 1972 to the present, but no congressional plans at all.

This state legislative focus has two explanations. First, and more importantly, each congressional delegation is *not* a legislative chamber in its own right, but rather a portion (often a very small portion) of the U.S. House of Representatives. Methods applicable to entire chambers cannot simply be transferred wholesale to delegations that make up only fractions of Congress. Second, most congressional delegations have many fewer seats than most state houses. The efficiency gap becomes lumpier when there are fewer seats, because each seat accounts for a larger proportion of the seat total, and the efficiency gap thus shifts more as each seat changes hands. This lumpiness is entirely avoided when state legislative plans, which typically have dozens or even hundreds of districts, are at issue.

For these reasons, Stephanopoulos and McGhee make two adjustments when analyzing congressional plans in their work on the efficiency gap. First, they convert the efficiency gap from percentage points to *seats* by multiplying the raw efficiency gap by each state's number of congressional districts. As they explain their method, "What matters in congressional plans is their impact on the total number of *seats* held by each party at the national level. Conversely, state houses are self-contained bodies of varying sizes, for which *seat shares* reveal the scale of parties' advantages and enable temporal and spatial comparability" (Stephanopoulos & McGhee, 869). Second, they only calculate efficiency gaps for states with at least eight congressional districts. Efficiency gaps are lumpier for states with fewer than eight districts, and additionally, congressional "redistricting in smaller states has only a minor influence on the national balance of power" (Stephanopoulos & McGhee, 868).

In his report, Trende fails to make either of these necessary adjustments when examining congressional plans. That is, he does not convert the efficiency gap from percentage points to seats, and he calculates the efficiency gap for small congressional delegations with fewer than eight seats. There is no authority in the literature for his methodological choices, and he is unable to cite any. And his flawed methods have serious substantive consequences that render his results entirely untrustworthy.

Take Trende's failure to convert the efficiency gap from percentage points to House seats. He claims that Alabama's congressional plan had an efficiency gap of -12.5% in 2002, that Arizona's congressional plan had an efficiency gap of 16% in 2012, that Colorado's congressional plan had an efficiency gap of -9% in 2002 and -10% in 2012, that Illinois's congressional plan had an efficiency gap of -9% in 2002, and that Iowa's congressional plan had an efficiency gap of -20% in 2002—all above my suggested 7% threshold for state legislative plans (paragraphs 115-16, 118-19, 121-22). But when converted to seats, *all* of these efficiency gaps become quite small, lower in all cases than the two-seat threshold proposed in the literature for congressional plans (Stephanopoulos & McGhee, 887-88). Specifically, using Trende's own calculations—which, as I discuss below, are incorrect in any event—Alabama had an efficiency gap of -0.9 seats in 2002, Arizona had an efficiency gap of 1.4 seats in 2012, Colorado had an efficiency gap of -0.6 seats in 2002 and -0.7 seats in 2012, Illinois had an efficiency gap of -1.7 seats in 2002, and Iowa had an efficiency gap of -1.0 seats in 2002. *None* of these scores are high enough to rise to presumptive unlawfulness under the literature's suggested two-seat threshold, meaning that we come to exactly the *opposite* conclusion as Trende after making the necessary adjustment.

Next take Trende's consideration of Alabama's congressional plan in 2002 (which had seven districts), Iowa's congressional plan in 2002 (five districts), and Colorado's congressional plans in 2002 and 2012 (seven districts each) (paragraphs 115-16, 119, 122). All four of these plans have fewer than eight districts, and so, based on the literature, should not be included in any efficiency gap analysis because of the measure's lumpiness when applied to so few seats. Trende nowhere acknowledges this limitation, and indeed appears unaware of its existence.

Moreover, Trende's study of congressional plans is marred by two further flaws, one conceptual and the other methodological. The conceptual defect is that, as in his earlier discussion of state legislative plans, he assumes that a large efficiency gap is all that is necessary to render a plan unconstitutional. He writes that efficiency gaps of -12.5%, -9%, -9%, -20%, and 16% "would invite court scrutiny as a Republican gerrymander" or "would invite court scrutiny as a Democratic gerrymander" (paragraphs 115, 116, 118, 119, 121, 122). But again, this is not plaintiffs' proposed test. A large efficiency gap is only a single prong of the test, and does not result in a verdict of unconstitutionality unless it is paired with a finding of partisan intent *and* a finding that it could have been avoided. Trende entirely overlooks these other elements.

The methodological defect is that whenever there were uncontested congressional races, Trende simply *substituted* presidential election results for the missing congressional results. As he put it in his deposition, he "used presidential results" and "imputed those results to the congressional races" whenever the races were uncontested (Trende deposition, p. 83). This is an exceptionally crude method that is guaranteed to produce errors, both because there is voter roll-off from the presidential to the congressional level and because voters may have different presidential and congressional preferences. Of course, presidential results can be used as the *inputs* to a regression model that *predicts* the outcomes of uncontested congressional races. Indeed, this is the preferred approach in the literature, and the approach I employed in my initial report. But presidential results cannot simply be plugged in without any adjustment, and no competent social scientist would have done so.

Accordingly, in my judgment, Trende's examination of particular state legislative and congressional plans is unreliable and entitled to no weight by the court. The state legislative analysis ignores the actual elements of plaintiffs' proposed test, and would have led to the opposite conclusion if these elements had been taken into account. Likewise, the congressional analysis ignores the test's prongs, fails to convert the efficiency gap from percentage points to seats, improperly considers states with small House delegations, improperly substitutes presidential election results whenever congressional results are missing—and deals with federal elections that simply are not part of this case.

Dated December 21, 2015

/s/ Simon Jackman

Simon Jackman, PhD

Department of Political Science

Stanford University

References

Altman, Micah and McDonald, Michael 2010, "The Promise and Perils of Computers in Redistricting." 2010. *Duke Journal of Constitutional Law & Public Policy* 5:69-111.

Chen, Jowei and Jonathan Rodden. 2013. "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures". *Quarterly Journal of Political Science* 8: 239-69.

Fifield, Benjamin, Higgins, Michael, Imai, Kosuke. Tarr, Alexander. 2015. "A New Automated Redistricting Simulator Using Markov Chain Monte Carlo." *Working Paper*, available at http://imai.princeton.edu/research/files/redist.pdf.

Goedert, Nicholas. 2014. "Gerrymandering or Geography?: How Democrats Won the Popular Vote but Lost the Congress in 2012." *Research & Politics* 1(1): 2053168014528683.

Goedert, Nicholas. 2015. "The Case of the Disappearing Bias: A 2014 Update to the "Gerrymandering or Geography Debate." Forthcoming in *Research and Politics*, November 2015.

National Conference of State Legislators. September 29, 2009. *Redistricting Law 2010*.

Nicholas Stephanopoulos & Eric McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap" 82 *University of Chicago Law Review* 831-900.

Expert Report of Professor Nicholas Goedert in *Whitford v. Nichol*. December 2, 2015. "Use of Efficiency Gap in Analyzing Partisan Gerrymandering, Report for State of Wisconsin, *Whitford v. Nichol*.

Expert Report of Professor Simon Jackman in *Whitford v. Nichol*. July 7, 2015. "Assessing the Current Wisconsin State Legislative Districting Plan."

Expert Report of Professor Ken Mayer in *Whitford v. Nichol*. July 3, 2015. "Analysis of the Efficiency Gaps of Wisconsin's Current Legislative District Plan and Plaintiffs' Demonstration Plan"

Declaration of Sean P. Trende in *Whitford v. Nichol*. December 2, 2015.

Deposition of Sean P. Trende in *Whitford v Nichol*. December 14, 2015.

Case

Vieth v. Jubilerer, 541 U.S. 267 (2004).