Supplemental Materials

A Common-space CFscore Scaling Methodology

The common-space CFscores are estimated as follows. The first step involves organizing the contribution data into a an n by m contingency matrix $\boldsymbol{R}$ where the rows index contributors, the columns index recipients and recipient committees, and each entry R_{ij} stores the total amount contributor i gives to recipient j. I implement an initial layer of normalization that helps to adjust for variation in contribution limits by converting contribution amounts from dollar amounts to count values. The conversion is based on federal contribution limits. Contributions between \$1 and \$100 are coded as 1, contributions between \$101 and \$200 are coded as 2, and so on. Contributions of \$5,000 or greater are capped at 50.[1] The matrix is then standardized by dividing each cell by $\sum_i \sum_j R_{ij}$. The next step performs singular value decomposition on the matrix $\boldsymbol{K} = \boldsymbol{D_r}^{-\frac{1}{2}}(\boldsymbol{R} - \boldsymbol{r}\boldsymbol{c}^\top)\boldsymbol{D_c}^{-\frac{1}{2}}$, where $\boldsymbol{r}$ and $\boldsymbol{c}$ are vectors of row and column sums, and $\boldsymbol{D_r}$ and $\boldsymbol{D_c}$ are diagonal matrices of $\boldsymbol{r}$ and $\boldsymbol{c}$. Ideal points are then calculated as $\boldsymbol{\theta} = \boldsymbol{U}\boldsymbol{D_c}^{-\frac{1}{2}}$ for contributors and as $\boldsymbol{\delta} = \boldsymbol{V}\boldsymbol{D_r}^{-\frac{1}{2}}$ for recipients, where the left singular vectors $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{K}\boldsymbol{K}^\top$, $\boldsymbol{D}$ is a diagonal matrix of singular values, and the right singular vectors $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{K}^\top\boldsymbol{K}$.

Although using CA to scale federal elections is straightforward, the model must be augmented to bridge across state and federal elections. Difficulties arise in the presence of overlapping pop-

[1] In addition to the normalization introduced by the chi-square distance metric, capping contribution amounts adjusts for organizations such as 527s and ballot campaign committees that can raise funds in unlimited amounts. Nonetheless, I find that the conversion barely affects the results. The correlation between candidate ideal point estimates recovered with and without the conversion to counts is 0.992.

ulations. It is typical for donors to give to both federal campaigns and candidates for state-level office in their home states, but giving to state-level candidates in different states, although not uncommon, is not the norm.[2] As a result, bridges that connect state elections to federal elections are abundant but bridges between states are relatively sparse. The solution utilizes the communality of federal elections to recover measures of liberal-conservative ideology that can be used to anchor the state-level estimates.

Having estimated federal-level ideal points, the next step is to scale each state separately, using contributors that have given to both state and federal campaigns as bridge observations. The bridging technique as applied to each state is implemented as follows. Let $\mathbf{S}$ be the set of contributors who have donated to elections in the state and $\mathbf{F}$ be the set of contributors that have donated to federal elections. Bridge contributors are denoted as $i \in \mathbf{S} \cap \mathbf{F}$. I denote ideal points derived from the federal scaling as $\boldsymbol{\theta_F}$ and denote ideal points the state is represented by $\theta_{\mathbf{s}}$. The χ^2 matrix for state s is represented by $\mathbf{K}_s$. I factor $\mathbf{K}_s$ with an iterative method known as alternating least-squares (Gabriel and Zamir 1979).[3] The algorithm factors $\mathbf{K}_s$ one dimension at a time by iteratively switching between estimating the contributor and recipient ideal points while holding the other set of ideal points fixed, such that the following objective function is optimized:

$$minimize \sum_{i}^{n} \sum_{j}^{m} (K_{s_{ij}} - \theta_i \delta_j) \tag{1}$$

The bridging algorithm is as follows:

---

[2] In contrast, the majority of contributions to federal candidates come from out of state.

[3] It is also possible to implement the identification strategy using reciprocal averaging which is less computationally demanding. Greenacre (1984) demonstrates equivalency between the two techniques in recovering a single dimension. The more general alternating least-squares technique allows for greater flexibility and, if desired, additional dimensions.

1. Recover starting values for state recipients by using $\boldsymbol{\theta}_{F_{i \in S \cap F}}$ and minimizing equation 1 subject to $\boldsymbol{\delta}$.
2. Recover contributor estimates by minimizing equation 1 subject to $\boldsymbol{\theta_S}$.
3. For the set of bridge contributors, regress $\boldsymbol{\theta}_{F_{i \in S \cap F}}$ on $\boldsymbol{\theta}_{S_{i \in S \cap F}}$ using an error-in-variable specification and rescale $\boldsymbol{\theta_S}$ using the estimated coefficients.
4. For the set of bridge contributors, combine information from contributions to state and federal elections using the mean of their state and federal ideal points weighted by the proportion of their contributions going to state and federal elections. Let ρ_i be the total percentage of contributor i's contributions going to state elections: $\theta_{S_i} = \rho_i \theta_{S_i} + (1 - \rho_i)\theta_{F_i}$.
5. Recover recipient estimates by minimizing equation 1 subject to $\boldsymbol{\delta}$.
6. Set $\boldsymbol{\delta}$ to values recovered from the federal scaling for state candidates that also ran for federal office.
7. Return to step 2 and repeat until convergence.
8. Set $\boldsymbol{\theta_F}_{i \in S \cap F} = \boldsymbol{\theta_S}_{i \in S \cap F}$.

After scaling each state separately, I jointly re-estimate the entire set of recipient ideal points in order to reintroduce information from contributors that give to elections in multiple states. Finally, I normalize the scaling such that the weighted mean and weigted standard deviation of contributions by amounts are set to zero and one, respectively.

A.1 Between-Set Identification

The method faces one last identification problem in making distance comparisons between row (contributor) and column (recipient) coordinates (ideal points). The well-known between-set identification problem applies to CA as well as most other scaling methods (Carroll, Green, and Schaffer 1986; Greenacre 2009). In fact, variants of the between-sets identification problem are common in the literature on ideological measurement. In roll call analysis, for example, the cutpoints are identified with respect to the legislator ideal points but the positions of the yea and nay outcomes are not. Issues arising from this problem have been addressed in the context of ideological scaling of political texts (Laver, Benoit, and Garry 2003; Lowe 2008). With respect to scaling data using CA, the axes for row and column coordinates will brought into coincidence so that they share

common dimensionality but not a common scale. This is because the contributor ideal points may be arbitrarily shifted and stretched with respect to the recipient ideal points. To see this, consider the transition formula for contributor ideal points: $\theta_i = \frac{\sum_j \delta_j y_{ij}}{\sum_j y_{ij}}$. The weighted averaging shrinks donor ideal points toward the center of the space. Left unadjusted, the donors will artificially appear more centrist than the candidates.

A less than satisfactory, yet typical, approach to this problem is to assume that both sets of ideal points have weighted means of zero and weighted standard deviations of one. This approach, however, is highly problematic given that many potential applications of the common-space CFscore measures involve making distance comparisons between the ideal points of contributors and candidates. Fortunately, the rich structure of campaign finance data provides a solution to the problem. A sizable number of committees and candidates are included in the data both as contributors and as recipients. This has the practical effect of giving contribution data the unusual property of having an substantial percentage of column observations also appear in the database as row observations. This, in turn, makes it possible to rescale the estimates by regressing donor ideal points on to corresponding candidate ideal points using an error-in-variable specification to adjust for attenuation bias. The estimated coefficients are then used to project contributors onto the same space as candidates. The projection is performed using a subsample of 29,913 candidates (roughly half the total sample) that have been matched up with their contribution records. The regression coefficients are 0.017 for the intercept and 0.975 for the slope, indicating that only a slight adjustment is needed.

A.2 Adjusting Ideal Points for Presidential Nominees

The special status of those nominated by their party for president as temporary standard bearers of their party presents a distinctive set of issues. The narrowed choice set and campaign finance

provisions that permit presidential nominees to jointly fundraise with their party typically cause sudden shifts in their estimated ideal points subsequent to being named the presumptive nominee. In order to account for these changes, I estimate separate pre- and post- nomination ideal points for presidential nominees. The pre-nomination ideal point are based on all contributions received prior to the month that the candidate secured his position as presumptive nominee. The post nomination ideal points are based on all contributions received from that point on until the candidates end their bids for president. For presidential nominees, such as Senators Kerry and McCain, who return to their prior offices after losing the election, all contributions raised for their senate re-election campaigns count toward their pre-nomination ideal points. Given the concerns noted above, the pre nomination estimates are likely more accurate measures of the candidates true ideology.

B Candidate Versus Contributor CFscores

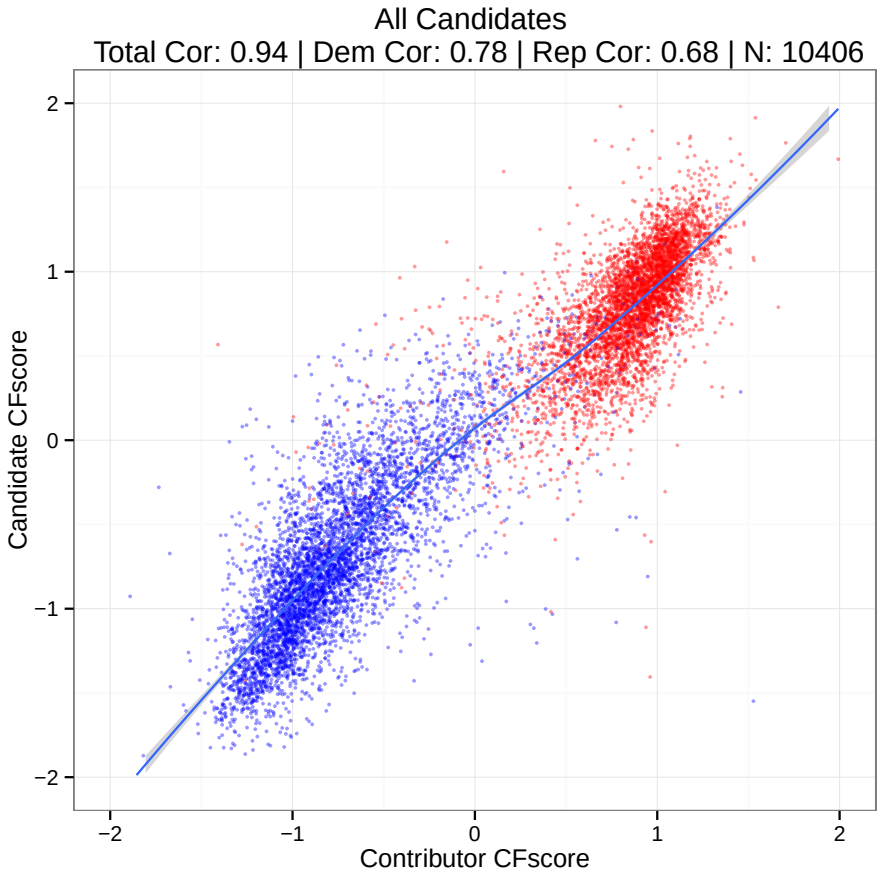

Figure 1. Candidate vs. Contributor CFscores

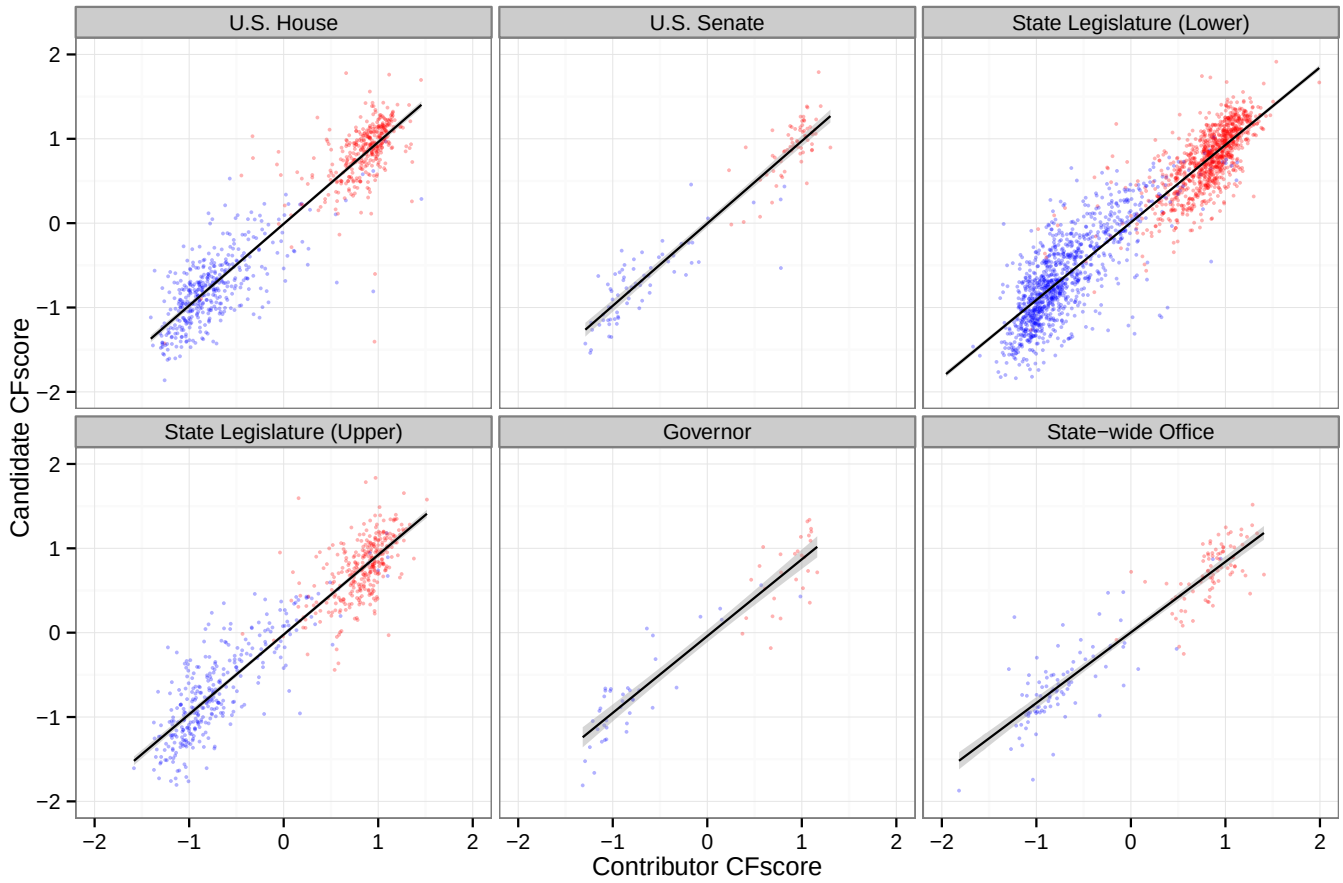

Figure 2. Candidate vs. Contributor CFscores

C Record Linkage and Validation

The identity resolution software links records using information on names, addresses, occupations, and employers disclosed to the FEC and state reporting agencies. The software loads a reference record associated with each individual, queries the database for all records with key similarities, and applies a carefully refined set of decision rules to determine which of these contribution records were made by the same individual. This task is complicated by donors who fail to disclose the requested personal information in entirety or do so inconsistently across records, while others relocate and change addresses once or more during the past three decades. The disclosure rates for all categories are above 95 percent for recent election cycles but are significantly lower for address and occupation in the 1980's and early 1990's. Another issue is that donors often report slight variatiants of their personal information. The algorithm accounts for these inconsistencies with a combination of fuzzy matching and supervised machine learning methods. The identity resolution software is written in R and MySQL. Both the software and access to the database will be made available by the author upon request.[4]

I selected corporate board members as the training set because they are often among the most difficult cases to code. They are typically affiliated with more than one institution and have multiple residences. In 57 percent of the cases, the algorithm correctly assigned a single ID to all contributions made by the individual. In 42 percent of the cases, the algorithm split contributions made by the individual into two or more groups (type 1 errors). However, in only 13 percent of

[4]The Center for Responsive Politics also assigns IDs for individual donors reported by the FEC, but the IDs only apply for a single election cycle and cannot be used to track donor activity across election cycles. In addition, the CRP's coding scheme is not made transparent and appears to be far less reliable than my linkage algorithm, with both a lower linkage success rate and a higher number of erroneously linked records.

these cases did the algorithm fail to assign at least 90 percent of the individual's contributions to a single ID. That is, in most of these cases, the algorithm correctly grouped most of the individual's records but left off a few hard to code stragglers. There were only two instances where the algorithm erroneously grouped contributions made by separate individuals (type 2 errors). Both cases involved family members associated with the same employer or organization. The first is William Gates, Jr. and his father William Gates, Sr., both of whom serve as directors for the Gates foundation. The second case erroneously grouped members of the Walton family, heirs to the WalMart fortune. It is important to note that type 1 errors, which result in a loss of information, are far less problematic than type 2 errors, which have the potential to introduce bias. It is reassuring that even among the most difficult to code individuals type 2 errors are exceedingly rare.

D Ideological Consistency in Contribution Patterns

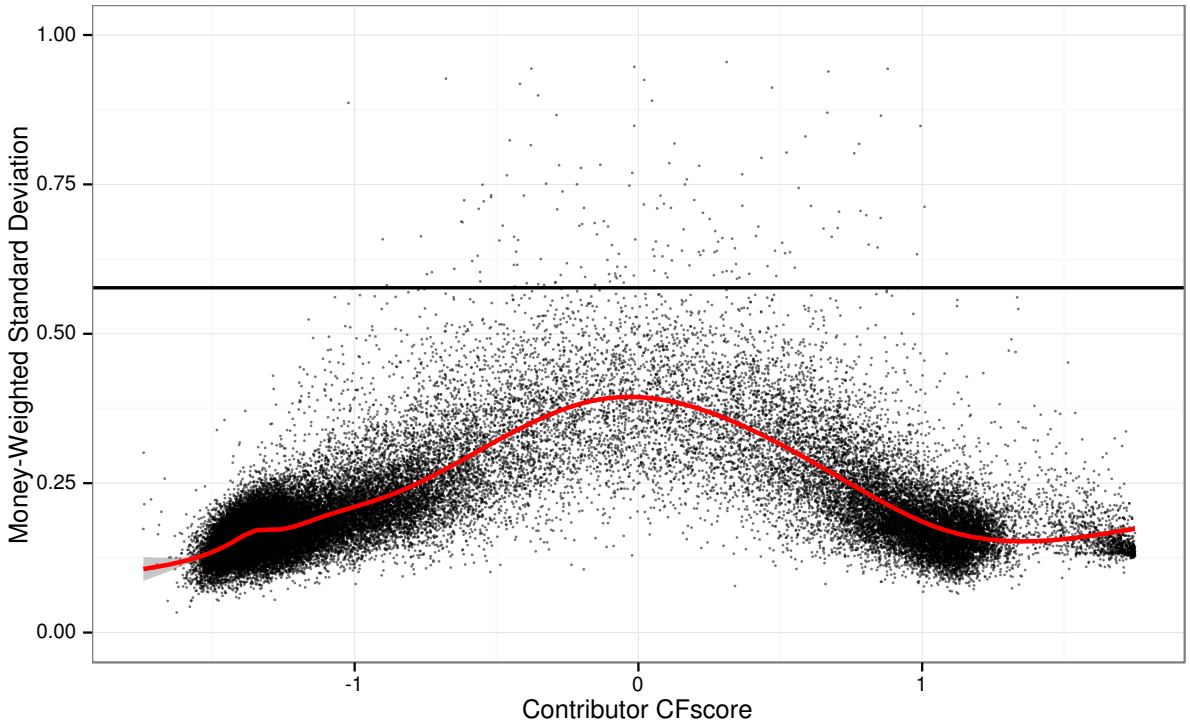

Figure 3. Money-Weighted Ideological Standard Deviation versus CFscores

Note: Each point represents an individual donor that made at least 25 contributions to federal candidates and committees during the 2004 through 2012 election cycles (N = 49,418). The smoothing line is a LOESS curve that weights each donor equally. The horizontal line indicates the theoretical limit at .577 for non-ideological giving. See McCarty, Poole, and Rosenthal (2006) for details on the summary statistic.

E Robustness to Changes to Non-Spatial Covariates

I extend the analysis on temporal stability to assess the sensitivity of the estimates to changes in candidate characteristics and electoral contexts associated with strategic behavior. If investment behavior is systematically biasing the candidate estimates, we should observe sizable shifts in the estimated ideal points following changes in incumbency status, seniority, majority-party control, and other characteristics associated with the ability to more efficiently provide legislative services. Likewise, if electoral considerations are biasing the results we should expect a candidate's ideal point to vary with election-specific characteristics such as electoral competitiveness or committee assignment. I assess the sensitivity of the measures to these factors using with a panel of period-specific candidate CFscores and time-varying electoral and candidate characteristics linked to strategic giving.

Table 1 reports results from two model specifications. The first model regresses the period-specific CFscores on the static CFscore estimates. The second model adds covariates for several time-varying electoral and candidate characteristics unrelated to ideology. The included variables cover incumbency status, office sought, competitiveness of seat, committee assignment, rank within party, status as a party leader or committee chair, and membership of the majority party. The table also reports results from models restricted by party that should detect centripetal and centrifugal effects that will be asymmetric with respect to party.

The model results show that controlling for the set of non-spatial covariates barely increases explanatory power over a model that only includes the static CFscores, which by itself explains nearly all of the variance. This is strong evidence that the CFscores are robust to large shifts associated with incentive to give strategically. At the same time, the results do reveal statistically significant effects for several of the covariates. Incumbency status and majority party status shifts

DV: Period-specific Candidate CFscores (House and Senate Candidates)

	All MCs		Democratic MCs		Republican MCs	
	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	0.001	0.029	−0.014*	−0.060	−0.013*	0.077
	(0.001)	(0.028)	(0.004)	(0.035)	(0.005)	(0.045)
Static CFscore	0.974*	0.972*	0.954*	0.937*	0.992*	0.959*
	(0.002)	(0.002)	(0.005)	(0.005)	(0.006)	(0.006)
Incumbent		0.007		0.060*		−0.042*
		(0.004)		(0.006)		(0.005)
Senate Seat		0.006		0.031*		−0.014*
		(0.004)		(0.006)		(0.005)
Competitive Seat		−0.007*		−0.041*		0.023*
		(0.003)		(0.005)		(0.004)
High-ranking		−0.014		−0.022		0.000
		(0.016)		(0.024)		(0.020)
Mid-ranking		0.002		−0.010		0.015
		(0.006)		(0.008)		(0.009)
Committee Chair		−0.007		−0.005		0.000
		(0.004)		(0.007)		(0.006)
Tenure in Office		−0.002*		−0.002*		−0.002*
		(0.000)		(0.000)		(0.000)
Majortiy Party		−0.010*		0.017*		−0.024*
		(0.003)		(0.004)		(0.003)
Party Leader		−0.018		−0.008		−0.017
		(0.028)		(0.034)		(0.044)
Committee FEs	N	Y	N	Y	N	Y
R^2	0.974	0.975	0.887	0.895	0.864	0.874
Num. obs.	10564	10564	5477	5477	5075	5075

$*p < 0.05$

Table 1. Sensitivity of Period-Specific CFscores to Changes in Candidate Characteristics
Note: Some models include fixed effects for assignments to thirteen committees.

members of both parties slightly toward the center, whereas competing in a closely contested race shifts members toward the extremes. However, these effects only seem to matter at the margins. To put the size of these effects in context, the largest are associated with a shift of between 0.03 and 0.05, or about a twentieth of a standard deviation of candidate CFscores. This corresponds to a shift of a dozen or so ranks.

As a point of comparison, replicating the analysis using DW-NOMINATE scores suggests that roll call based measures are slightly *more* sensitive to changes in candidate characteristics. Since the Nokken-Poole period specific estimates are not estimated in a common-space framework, I separate the analysis by legislative chamber. Controlling for a similar set of candidate characteristics increases the within party R^2 from 0.860 to 0.864 for House Democrats and from 0.698 to 0.751 for House Republicans. Including the covariates increases the within party R^2 from 0.715 to 0.745 for Senate Democrats and from 0.843 to 0.845 for Senate Republicans.

By these measures, the CFscores exhibits greater inter-temporal stability than DW-NOMINATE scores. It is also the case that CFscores are, on the whole, less sensitive than DW-NOMINATE scores to changes in non-spatial covariates. This is despite the fact that incumbency status, which is included as a covariate in the regression for CFscores, is excluded due to lack of availability the regressions for DW-NOMINATE. In combination with the finding in Table 2 that covariates linked to strategic models of giving have very little explanatory power as compared to a simple spatial model, the analysis performed here should do much to address concerns about strategic giving biasing the ideal point estimates.

To be clear, this is not to claim that ideological proximity is the sole determinant of contribution patterns. Strategic giving can matter at the margins, as shown by the small but statistically significant estimated coefficients for several of the covariates included in Table 1. Rather, the claim is simply that the omitted non-spatial covariates (1) explain a relatively minuscule proportion of

DV: Period-specific DW-NOMINATE scores (House)

	All MCs		Democratic MCs		Republican MCs	
	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	−0.002	−0.113*	−0.117*	−0.108*	0.034*	−0.143
	(0.002)	(0.036)	(0.005)	(0.030)	(0.011)	(0.074)
Static DW-NOMINATE	0.962*	0.966*	0.827*	0.817*	0.934*	0.930*
	(0.002)	(0.002)	(0.006)	(0.006)	(0.012)	(0.011)
Competitive Seat		−0.010		0.015*		−0.013
		(0.006)		(0.006)		(0.011)
High-ranking		0.041*		−0.026		0.106*
		(0.020)		(0.019)		(0.032)
Mid-ranking		0.030*		0.005		0.053*
		(0.008)		(0.007)		(0.014)
Committee Chair		−0.011		0.001		−0.050*
		(0.007)		(0.006)		(0.011)
Tenure in Office		0.002*		−0.002*		0.008*
		(0.000)		(0.000)		(0.001)
Majority Party		0.041*		0.004		0.099*
		(0.004)		(0.004)		(0.008)
Party Leader		0.029		−0.002		0.004
		(0.035)		(0.029)		(0.073)
Committee FEs	N	Y	N	Y	N	Y
R^2	0.962	0.964	0.860	0.864	0.698	0.751
Num. obs.	6007	6007	3302	3302	2705	2705

$*p < 0.05$

Table 2. Sensitivity of Period-Specific DW-NOMINATE scores to Changes in Candidate Characteristics (House)

Note: This table replicates the analysis found in Table **??** using DW-NOMINATE scores rather than CFscores. The sample is restricted to members of the House. The scores have been rescaled to have a standard deviation of 1.

DV: Period-specific DW-NOMINATE scores (Senate)

	All Senators		Democratic Senators		Republican Senators	
	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	0.002	0.030*	−0.040*	0.151	−0.094*	−0.070*
	(0.004)	(0.012)	(0.015)	(0.122)	(0.015)	(0.025)
Static DW-NOMINATE	0.963*	0.961*	0.889*	0.868*	1.082*	1.071*
	(0.005)	(0.005)	(0.022)	(0.025)	(0.019)	(0.020)
Competitive Seat		−0.009		0.048*		−0.064*
		(0.012)		(0.015)		(0.017)
Committee Chair		−0.009		0.010		−0.019
		(0.010)		(0.012)		(0.015)
Tenure in Office		0.000		−0.002*		0.002
		(0.001)		(0.001)		(0.001)
Majority Party		−0.017*		0.005		−0.029*
		(0.008)		(0.009)		(0.012)
Committee FEs	N	Y	N	Y	N	Y
R^2	0.963	0.965	0.715	0.745	0.843	0.855
Num. obs.	1274	1274	637	637	637	637

$*p < 0.05$

Table 3. Sensitivity of Period-Specific DW-NOMINATE scores to Changes in Candidate Characteristics (Senate)
Note: This table replicates the analysis found in Table 1 using DW-NOMINATE scores rather than CFscores. The sample is restricted to members of the House. The scores have been rescaled to have a standard deviation of 1.

variance in contribution decisions compared to spatial proximity and (2) appear to be largely orthogonal to ideological considerations. As is the case with all ideal point measures, researchers should be mindful of the potential ways that even small amounts of bias present in the "off-the-shelf" estimates can impact their results and make an effort to make appropriate adjustments. It is my hope that by making all the data and code used to estimate the CFscore accessible as publically available database, researchers will be able to make adjustments directly to the scaling model.

F Roll Call Classification, cont.

Another point to consider when interpreting these classification results is the increased rates of partisan overlap apparent in CFscores as compared with DW-NOMINATE scores. Figure 4 plots the gain in correct classification associated with DW-NOMINATE scores over CFscores against legislator CFscores. It reveals CFscores tend to be relatively poor predictors of the voting behavior of moderate legislators who overlap members of the opposing party. What is more, when a two-cutpoint model devised by McCarty, Poole, and Rosenthal (2001) test for party effect is used, the classification gap between the two measures narrows substantially for close votes but not for lopsided ones, suggesting that much of classification boost over the CFscores associated with DW-NOMINATE scores is due to party effects (Bonica 2013).

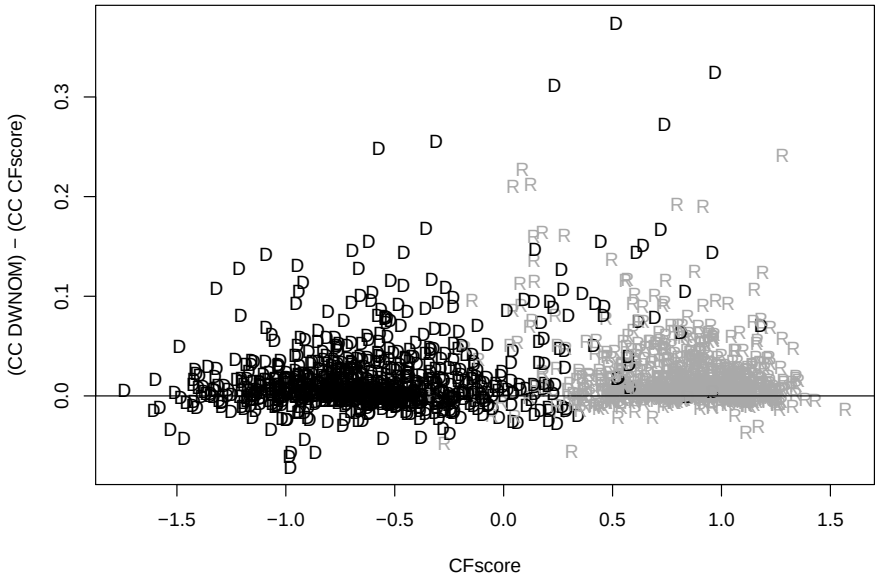

Figure 4. Increase in Correct Classification of DW-NOMINATE over CFscores
Note: The y-axis plots for each legislator the percentage of votes correctly classified by DW-NOMINATE scores less the percentage of votes correctly classified by CFscores. Values above 0 indicate the DW-NOMINATE better classifies the voting record of the legislator.

G Additional Ideal Point Comparisons

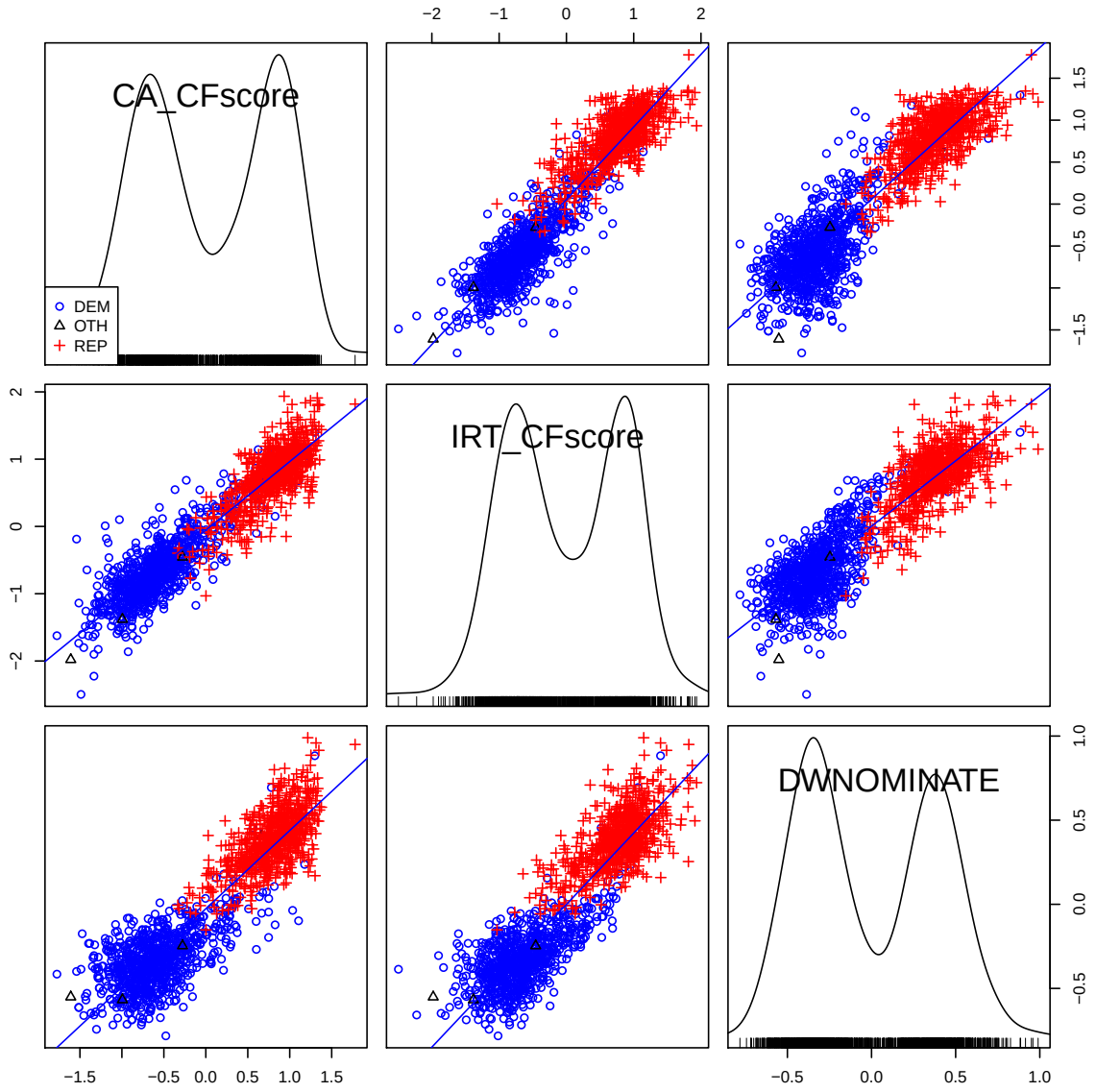

Figure 5. Comparison of common-space CFscores, scores recovered from an IRT negative binomial count model applied to PAC contributions, and DW-NOMINATE scores

Note: This figure compares the common-space CFscores for candidates CFscores with the ideal point estimates recovered from the IRT negative binomial model applied to PAC contributions. The two measures correlate at $r = 0.94$.

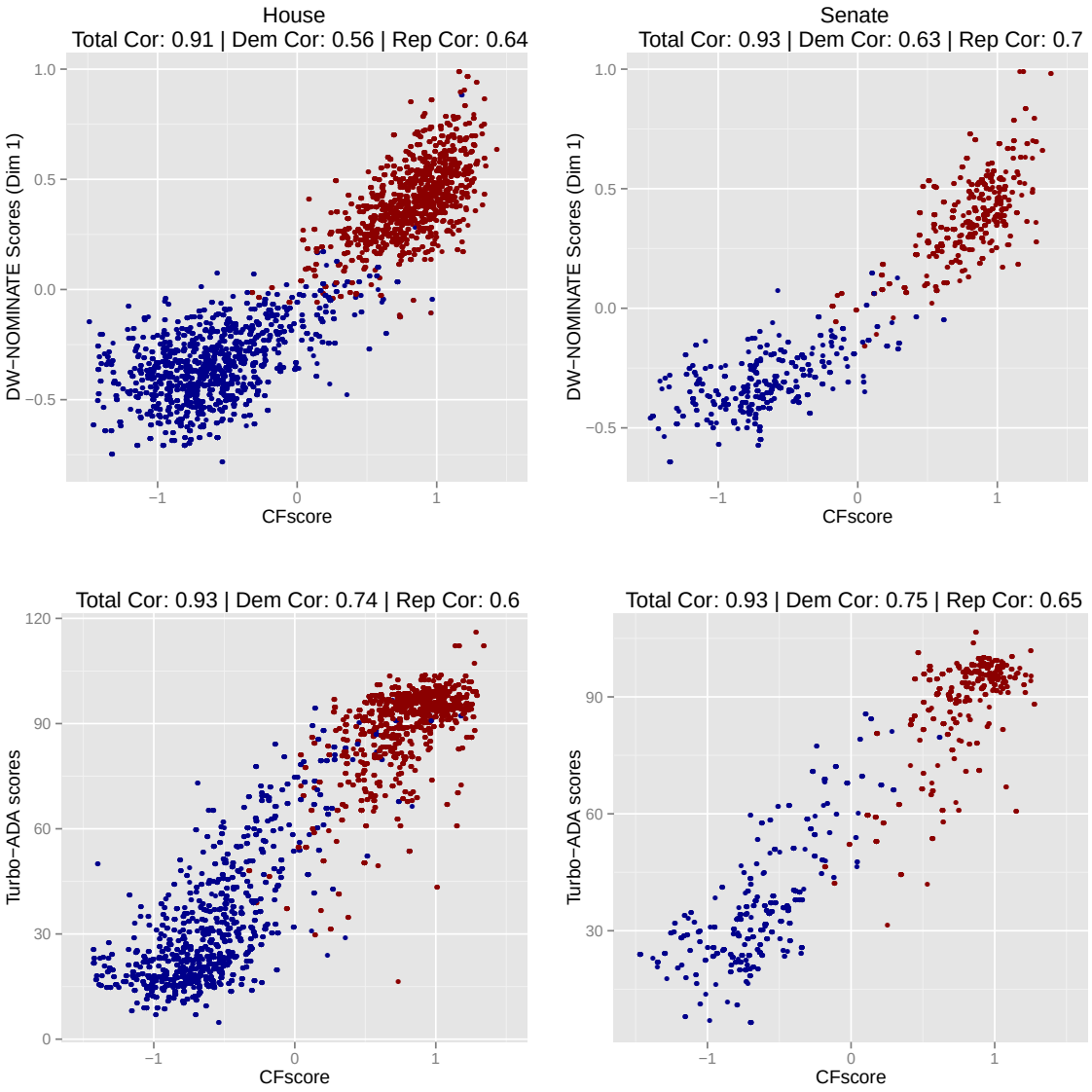

Figure 6. Comparison of common-space CFscores, Turbo-ADA, and DW-NOMINATE scores

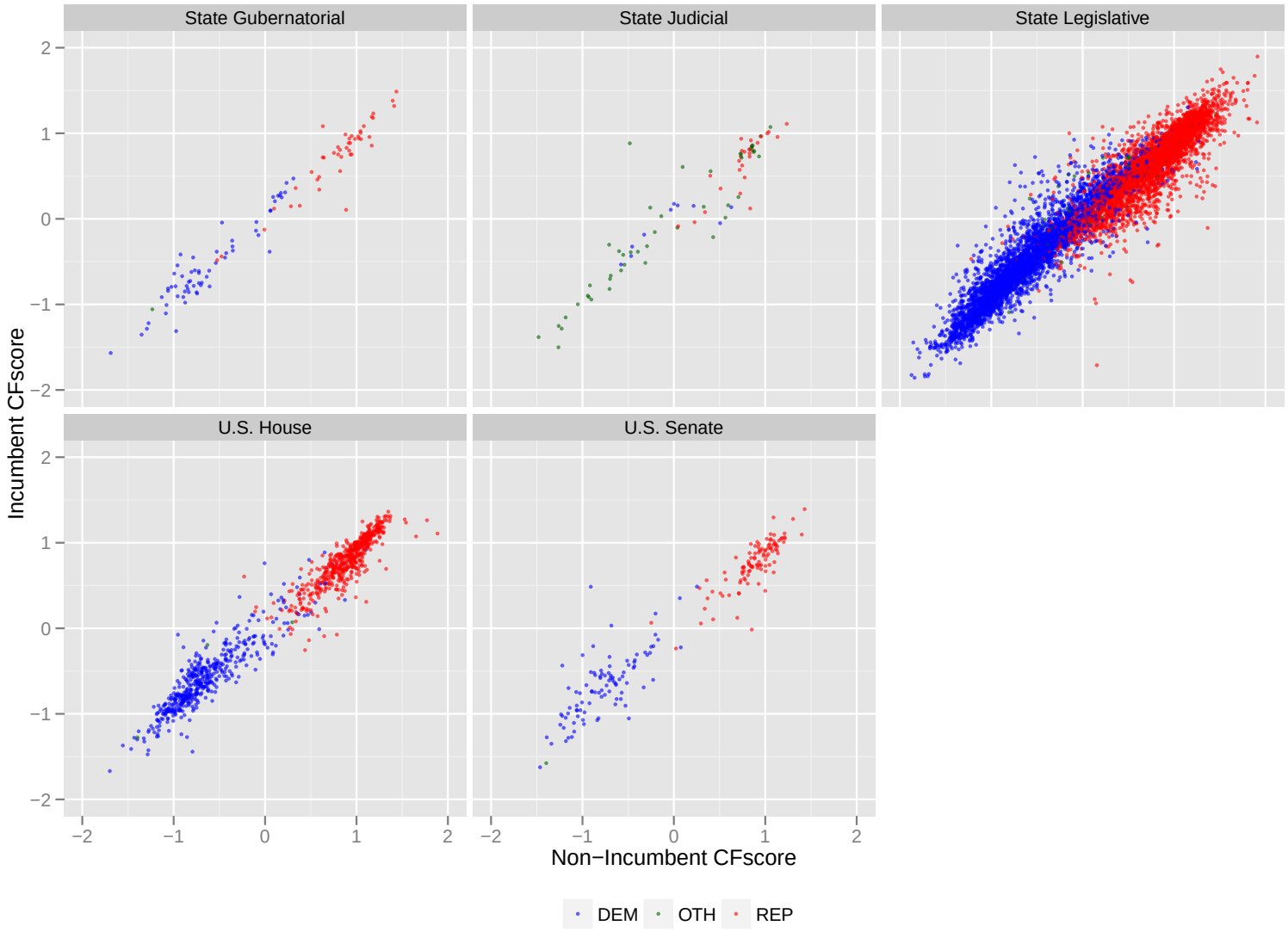

Figure 7. Candidate CFscores recovered as Incumbent as Non-incumbents by office

H Party Mean Trends

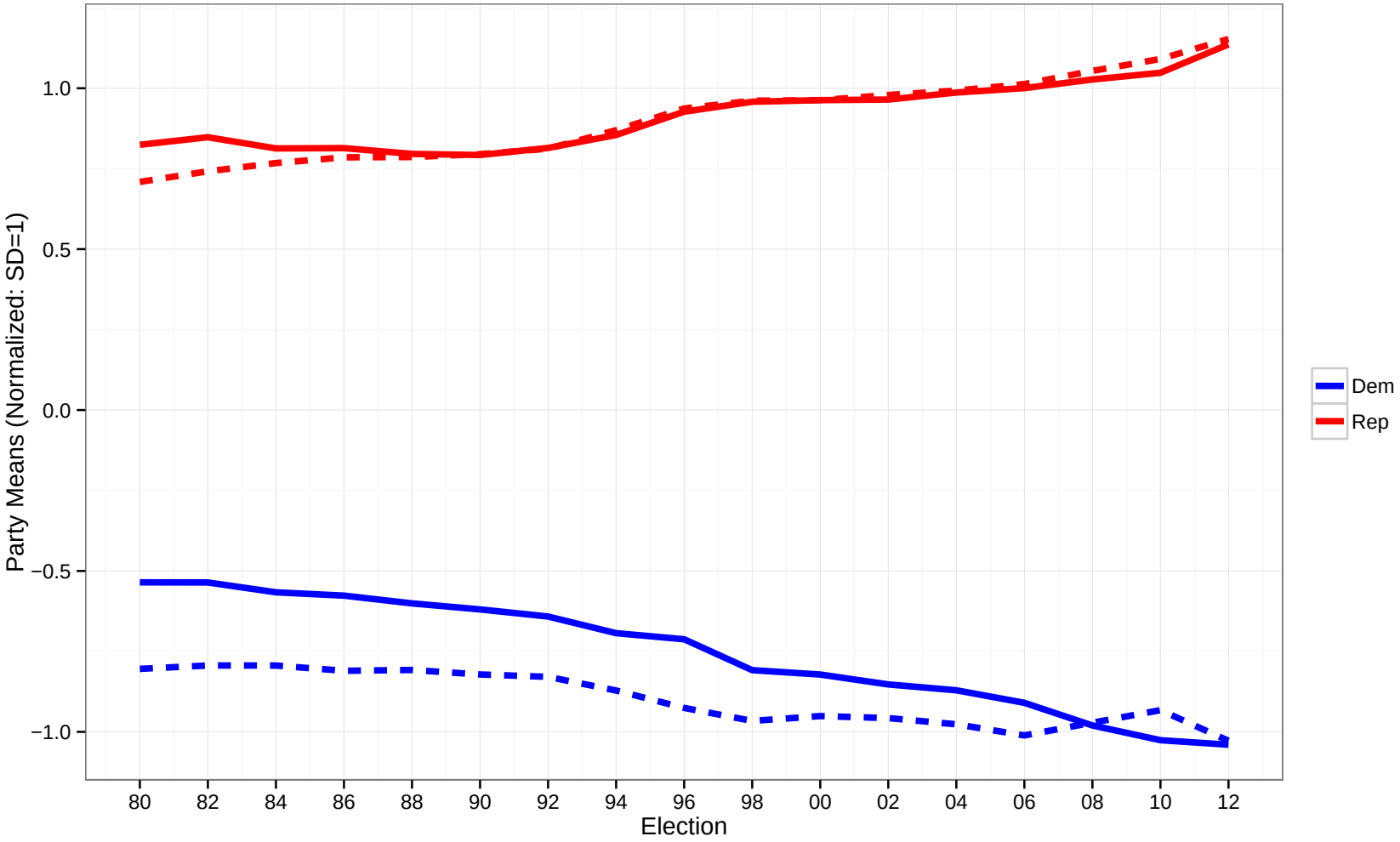

Figure 8. Comparison of Partisan Trends from CFscores and DW-NOMINATE

Note: The solid lines are the CFscore party mean and the dotted lines are the DW-NOMINATE party means. Both measures have been commonly rescaled to facilitate meaningful comparisons.

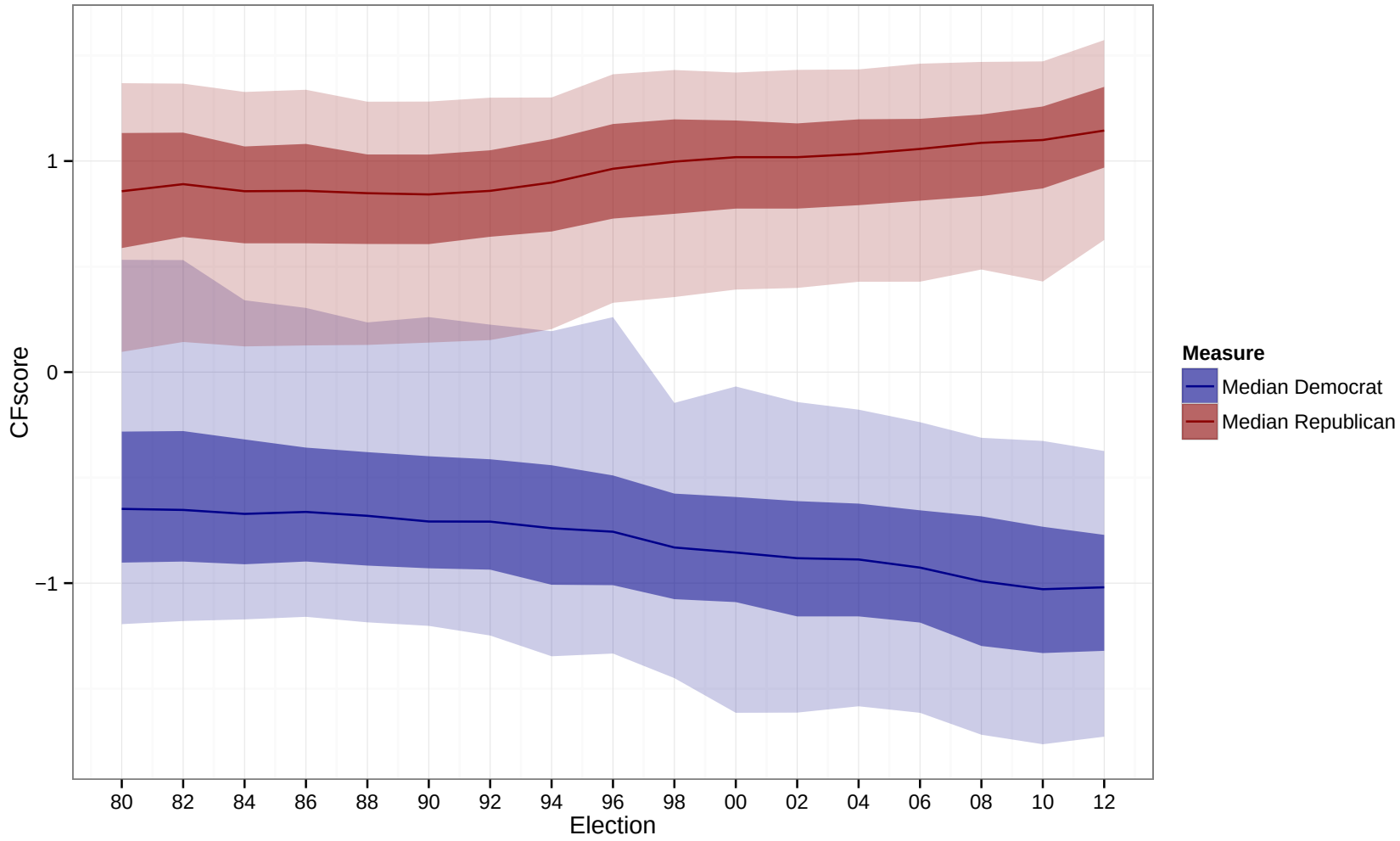

Figure 9. Party Medians and Distributions for Members of Congress

Note: Ribbon bars are plotted for the .25th to .75th and .05 to .95 percentiles for each party.

References

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57 (2): 294–311.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1986. "Interpoint Distance Comparisons in Correspondence Analysis." *Journal of Marketing Research* 23 (3): 271–80.

Gabriel, K. Ruben, and S. Zamir. 1979. "Lower Rank Approximation of Matrices by Least Squares with Any Choice of Weights." *Technometrics* 21: 489–98.

Greenacre, Michael J. 1984. *Theory and Applications of Correspondence Analysis*. New York: Academic Press.

Greenacre, Michael J. 2009. "Power Transformations in Correspondence Analysis." *Computational Statistics and Data Analysis* (53): 3107–16.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 92 (2): 311–32.

Lowe, Will. 2008. "Understanding Wordscores." *Political Analysis* 16 (4): 356–71.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2001. "The Hunt for Party Discipline in Congress." *The American Political Science Review* 95 (3): 673–87.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, Mass.: MIT Press.