Supplementary Materials for

“Measuring Executive Ideology and its Influence”

**Supplementary Materials – Part I  
Sampling Bias in the Availability of CFscores**

Two hundred thirty-six individuals served as governor of a U.S. state in the years between 1991 and 2015. (This count excludes governors who retired after the 1990 election cycle but served for a few days in 1991 before their successor was inaugurated.)

Only 150 of these governors are included in Bonica’s (2014) Database on Ideology, Money, and Elections. To create his database, Bonica relied on the availability of state-level campaign finance data from the National Institute for Money in State Politics (NIMSP). States only began requiring disclosures in 1990, and the rollout of these requirements was uneven through that decade. Only by 2001 were state-level candidates in every state required to record and report their campaign contributions.

As a result, CFscores cover more of the total sample of U.S. governors in later decades. Only 53 of the 112 (47%) governors who served at any point from 1991 to 1999 had CFscores. However, this number improves to 88% (110 out of 125) in the period from 2000 to 2009, and 98% (81 out of 83) in the period from 2010 to 2013.

Aside from this limitation of the data, they are otherwise quite representative. This is important because, were governors with and without CFscores different based on some characteristic, scholars that use CFscores to explain gubernatorial behavior may introduce bias along those lines.

Table A1 presents the results of a series of t-tests meant to identify any statistical differences between the populations of governors with and without CFscores. The leftmost column contains characteristics that may differentiate governors from one another. The second (and third) columns present the percentage of governors with (or without) CFscores that claim each characteristic. The final two columns present the t-statistic and corresponding p-value of a Welch two-sample t-test comparing the two populations.

In the first three rows, we see that roughly equal percentages of governors with and without CFscores are Democrats, Republicans, and Independents respectively. The fourth row shows that slightly more governors with CFscores are women than exist in the full population, but this difference is not significant. The fifth row shows that the percentage of non-white governors is equal between populations. Finally, the rows under the “Region” subheading show that a proportionate number of governors from the Northeast and Midwest have CFscores, but that governors from the South are slightly underrepresented, and governors from the West are slightly overrepresented.

These results are important because, to the extent scholars wish to use CFscores in their models of gubernatorial behavior, they necessarily limit their sample size. If this occurs in an unrepresentative way, the scholar may introduce bias into their results. From these results, however, it appears that governors with CFscores are generally representative of the entire population of governors. Scholars need only exercise extra caution if they believe their data-generating process is different in the 1990s than in later decades, or if it varies by region of the country.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table A1. Characteristics of governors with and without CFscores (1991-2013)** | | | | |
|  | Pct of governors with CFscores | Pct of governors w/o CFscores | t | p-value |
| *Partisanship* |  |  |  |  |
| Democrat | 45.4% | 46.5% | -0.15 | 0.88 |
| Republican | 52.0% | 52.1% | -0.02 | 0.98 |
| Independent | 2.6% | 1.4% | 0.64 | 0.52 |
|  |  |  |  |  |
| *Demographics* |  |  |  |  |
| Female | 13.2% | 7.0% | 1.49 | 0.14 |
| Non-white | 5.3% | 4.2% | 0.34 | 0.73 |
|  |  |  |  |  |
| *Region* |  |  |  |  |
| Northeast | 18.4% | 21.1% | -0.47 | 0.64 |
| Midwest | 21.7% | 23.9% | -0.37 | 0.72 |
| South | 29.6% | 40.8% | -1.62 | 0.11 |
| West | 30.3% | 14.1% | 2.89 | 0.00 |
|  |  |  |  |  |
| N | 152 | 71 |  |  |

**Supplementary Materials – Part II  
Combining Measures of Executive Ideology**

In the main text of this article, I compared CFscores to three measures of gubernatorial preferences in order to demonstrate convergent and predictive validity. The results overall affirmed the quality of CFscores as a measure of ideology, although within parties, the correlations between CFscores and other measures were more modest. But weak correlations can be a function of poor data quality in *either* variable; alternatively, and perhaps more aptly here, it can be the result of two meaningful measures capturing different elements of ideology.

This begs the question, “Can multiple measures be combined to a single create a metric that contains additional information about gubernatorial ideology?” Particularly given the discussion in SM 3, which highlights how the noise inherent in CFscores makes them less suitable for fixed-effects models, a combined metric may be desirable for scholars who seek greater statistical power in their studies of gubernatorial ideology.

The largest challenge to this approach pertains to the relative sparsity of data on gubernatorial ideology. Table A2 presents the coverage of each of the measures that I present in this article, as well as every combination thereof.

Among the available options, CFscores provide by far the greatest coverage of US governors from 1991 to 2013, with data available for 152 of the 223 individuals (68%) who served. Coverage varies when we combine CFscores with other metrics. There is considerable overlap between governors with CFscores and CES-based voter placements, as 103 governors have data in the form of both. But fewer than half of governors with CFscores also have legislative roll-call scores or have had their State of the State speeches analyzed to calculate their preferences. Data become sparser still when more than two measures are combined.[[1]](#footnote-1)

|  |  |
| --- | --- |
| **Table A2. Availability of additional measures of ideology  for governors with CFscores (1991 to 2013)** | |
| *Combination of measures* | *Number of governors* |
| *CFscores alone* | *152* |
| CFscores and… | |
| Voter placements | 103 |
| SOTS scores | 66 |
| Legislative roll-call scores | 58 |
| Voter placements and roll-call scores | 44 |
| Voter placements and SOTS scores | 33 |
| SOTS scores and roll-call scores | 19 |
| Voter placements and roll-call scores and SOTS scores | 10 |
| **Note:** A total of 223 individuals served as US governors during this period. Data on gubernatorial ideology for governors without CFscores is limited. | |

For it to be worthwhile for scholars to combine multiple measures into one, the information gained by doing so must overcome the information lost by reducing the available sample size.

As such, I focus here on two combinations of measures. Of the options presented in Table A2, the only combination of measures that covers close to half of governors between 1991 and 2013 are CFscores and CES-based voter placements. Then, based on Table 1 in the main text, SOTS scores appear to provide the most distinctive information of the measurements. Even though this combination reduces our sample size to roughly 30 percent of the population of interest, it is worth exploring how this additional information can add to the statistical power of CFscores, in hopes that additional work on the speeches may expand data availability further.

Using these combined metrics, I assess their ability to predict behavioral and policy outcomes relative to the use of CFscores alone. Before laying out the results, however, it is important to consider the implications of all possible outcomes.

If we find that a combined metric leads to stronger relationships between gubernatorial ideology and outcomes, then it suggests two possibilities, of which either or both may be true. One is that CFscores alone are underpowered. A second is that the additional measure may add a useful dimension of information. Whatever the underlying process, the clear implication to scholars would be use to multiple measures where possible.

A second outcome may be that we find no real differences when using CFscores versus a combined metric. This would suggest that CFscores do contain ample statistical power, at least for the analyses being conducted here. It may also mean that the noise in CFscores and the other measure comes from similar sources of bias, such that the metrics add little to one another. This latter possibility seems more likely for voter placements than State of the State scores, since the former are known to incorporate perceptive biases that likely affect CFscores as well. In either case (or both), the implication would be that using CFscores alone may be a justifiable choice, even as additional measures can add rigor to a study.

A third outcome would see weaker relationships when a combined metric is used. This could be bad news for CFscores, as it could suggest that the noise implicit in them associates with outcomes of interest, and that adding new sources of information addresses some of that endogeneity. Alternatively, it could also indicate that the other measure is of poor data quality. For scholars, this outcome would be the most challenging of the three.

|  |  |
| --- | --- |
| **Table A3. Possible outcomes and interpretations** | |
| *Combined measure predicts outcomes…* | *Potential interpretations* |
| Better than CFscores alone | - CFscores are underpowered at the sample size used  - Alternative measure contributes new information that helps explain the outcome |
| As well as CFscores alone | - CFscores contain enough statistical power at the sample size used to capture “true” relationship  - Metrics contain similar sources of bias |
| Worse than CFscores alone | - Noise in CFscores correlates with outcome, alternative measure provides new information to cancel it out  - Alternative measure is noisy |

To create the combined metrics, I first standardize all three measures (CFscores, voter placements, and SOTS scores) to have a mean of zero. For governors with more than one score—because their constituents placed them in multiple waves of a survey, or because both Coffey (2006) and Kousser and Phillips (2010) analyzed their speeches—I then average across years to create a final, standardized voter placement and SOTS score. Then, I average each of these scores respectively with CFscores.

I apply both CFscores alone and the combined metrics to indicators of gubernatorial behavior. Two derive from Sellers’ (2016) data on gubernatorial action to protect citizens from employment discrimination on the basis of sexual orientation and gender identity. Then, I also associate the combined CFscore-and-voter placement metric with State of the State scores.

Next, I also apply both measures, CFscores and the combined metric, to state economic policy as calculated by Caughey and Warshaw (2016). I focus here on economic policy because from the analysis in the main text, this appears to be where gubernatorial influence is the strongest.[[2]](#footnote-2) I estimate a univariate first-differences model to see how changes in CFscores and in the combined metrics relate to the liberalism of economic policy in a state.

To ensure an even comparison, my analysis is limited to governors for whom both CFscores and the additional data are available. The first set of comparisons is presented in Table A4, which shows the Pearson’s correlations between both ideological measures—CFscores alone and the combined metrics—and each of the three forms of observed behavior discussed above. When we compare between correlations for each behavior, we see that using the combined metrics yields almost the same relationship between ideology and behavior as does using CFscores alone.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table A4. Correlations between metrics of gubernatorial ideology and observed behavior** | | | | | | |
|  | Protections for sexual orientation | | Protections for gender identity | | State of the State proposals | |
|  | Pearson’s *r* [conf. int.] | N | Pearson’s *r* [conf. int.] | N | Pearson’s *r* [conf. int.] | N |
| CFscores alone | -0.75\* | 46 | -0.59\* | 82 | 0.50\* | 33 |
| (govs w/ voter placements) | [-0.86, -0.59] |  | [-0.71, -0.42] |  | [0.19, 0.72] |  |
| CFscores + placements | -0.74\* | 46 | -0.62\* | 82 | 0.53\* | 33 |
|  | [-0.85, -0.57] |  | [-0.74, -0.46] |  | [0.23, 0.74] |  |
| CFscores alone | -0.70\* | 39 | -0.37\* | 63 |  |  |
| (govs w/ SOTS scores) | [-0.83, -0.49] |  | [-0.57, -0.14] |  |  |  |
| CFscores + SOTS | -0.67\* | 39 | -0.38\* | 63 |  |  |
|  | [-0.81, -0.44] |  | [-0.58, -0.15] |  |  |  |
| **Notes:** \* = p < .05.Combined metrics are created by standardizing both CFscores and alternative measures to a mean of zero, and averaging both measures together. Higher scores indicate greater conservatism. | | | | | | |

A second comparison looks at policy outcomes, particularly in the economic arena, where the analysis in the main text suggests governors may have most influence. Here, I am interested in the expected effect of a one-standard deviation change in gubernatorial ideology on the liberalism of state economic policy. To estimate that effect, I set up a univariate first-differences model, which takes as its dependent variable the change in policy liberalism from *t-1* to *t* for unit (state) *i*.

The results of this exercise are presented in Table A5.

Models 1 and 2 estimate the effect of ideology change on policy change among governors with CES-based voter placements. The results do not change substantially whether the analyst uses CFscores alone (Model 1) or combines them with voter placements (Model 2).

Models 3 and 4 estimate the same effect among governors with State of the State-based policy scores. Here again, the results do not change when CFscores are used in combination with other scores, as opposed to alone.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table A5. Relationship between change in gubernatorial ideology  and in state economic policy** | | | | |
|  | CFscores +  voter placements | | CFscores + State of the State scores | |
|  | Model 1 | Model 2 | Model 3 | Model 4 |
| Δ CFscores alone | 0.11\* | -- | -0.01 | -- |
|  | [0.05, 0.17] |  | [-0.08, 0.06] |  |
| Δ Combined metric | -- | 0.10\* | -- | -0.05 |
|  |  | [0.04, 0.16] |  | [-0.13, 0.03] |
| Intercept | 0.00 | 0.00 | 0.01 | 0.01 |
|  | [-0.03, 0.03] | [-0.03, 0.05] | [-0.01, 0.03] | [-0.01, 0.03] |
|  |  |  |  |  |
| R-squared | 0.09 | 0.05 | 0.00 | 0.00 |
|  |  |  |  |  |
| N (n) | 192 (49) | | 218 (46) | |
| Obs. used | 143 | | 172 | |
|  | | |  |  |
| **Notes:** \* p < .05. Dependent variable is change in state economic policy liberalism (Caughey and Warshaw 2016), panel covers even years from 1996 to 2011. | | | | |

What can we take away from these exercises? There are a couple of plausible explanations for why combined scores do not outperform CFscores alone, and none are exclusive from one another. For one, CFscores may be an efficient enough measure as to confidently estimate the “true” univariate relationship between ideology and behavior with only 50 to 100 observations. This seems plausible given the results in Table A4. The correlations shown there include only observations with CFscores and the other measure being studied, and the fact that no difference is observed between tests suggests that CFscores perform well at small values for N. Another possibility is that other measures are subject to the same sources of noise and bias as CFscores, and so add relatively little information. This is less likely for State of the State scores, given their modest correlation with other measures (see Table 1 in main text), but may certainly be the case with voter placements.

Importantly, however, both explanations point to the same takeaway for scholars: that it is not apparent that the trade-off of information for sample size is worthwhile when it comes to combining CFscores with voter placements. Given that CFscores provide superior data coverage for governors, and perform as well or better than other key measures at predicting outcomes, they stand to be considered as the “default” option for measuring gubernatorial ideology.

**Supplementary Materials – Part III  
Modeling using CFscores for Governors**

The use of panel data, which studies units of observation at multiple timepoints, is common in state politics and policy research. Many research designs, including the one in this article, examine how policy outcomes change in the 50 states over time.

Here, I briefly survey the options available to researchers who use panel data, before discussing how the limitations of CFscores pose a challenge to scholars wishing to incorporate gubernatorial ideology into fixed-effects models. I conclude by presenting some partial solutions, as well as a replication of the findings in the main text using fixed effects.

*Options for analyzing panel data*

Analysts who work with panel data face modeling choices that relate to the trade-off between variance and bias. The option that retains most variance is the pooled OLS, which considers all timepoints and units together. In state politics research, this tends to leave bias uncorrected, as observations for a given state across years are not independent from one another.

A middle ground in the trade-off between variance and bias is random effects models, which are what I use in the main text. These models incorporate a parameter λ (sometimes seen as θ) that specifies how much the coefficient estimates should weigh within-unit variation relative to across-unit variation. In the analysis in the main text, for instance, the way that ideology affects policy in a single state over time will count a lot more than the relationships we see when looking at all 50 states together.

This is an attractive option when we theoretically expect an outcome to emerge from some combination of unit-specific and universal processes. For example, it may be fair to assume that one-half of variation in state policy can be explained by features that are comparable across states, such as the preferences of the public and elites. Then, we may attribute another half to the state-specific features (e.g., institutions, culture) that would either mediate the effect of the former or have some direct impact on policy. In this case, a high value for λ would assure us that both are taken into account.

In the past, however, scholars have been hesitant to rely on the results of random effects modeling because of its assumption that unit-level intercepts do not associate with any of the independent variables X. This is rarely the case in political science research when the unit of analysis is a geographic area like a city, state, or country.

But a flurry of recent scholarship has warned against throwing the variance baby out with the bias bath water (Bell and Jones 2015, Esarey and Jaffe 2017, Plümper and Troeger 2019). Clark and Linzer (2015) show that bias can be minimized—and “true” coefficient estimates achieved with smaller sample sizes—when there is a high ratio of within-unit variance in X relative to Y. Their simulations find that when the dependent variable is “sticky,” as is policy, while the independent variables vary more, then incorporating a degree of across-unit variation can improve model performance without taking on much bias.[[3]](#footnote-3) As to the models in the main text, depending on the variables used, there is between 5 and 16 times more within-unit variance in values for X than for Y.

The final, more traditional option is the fixed-effects model, which is attractive in that it represents the most stringent test of significance. By specifying a separate intercept for each unit, only within-unit variation is considered in estimating the coefficients. In the trade-off between variance and bias, then, it trades all across-unit variance in exchange for eliminating unit-level bias.

Of course, removing variance runs the risk of not picking up on a relationship between X and Y unless a large enough *N* or *T* is used. It can also introduce bias that is specific to the variables used, insofar as the information loss (when shifting from between- to within-unit analysis) occurs unevenly across covariates.

This is a particular problem when using CFscores to measure gubernatorial ideology.

*CFscores and fixed effects*

Due to their construction, CFscores are prone to the “dynamic misspecification” problem. This is an issue that emerges in time-series data when a concept is modeled as being static, when in fact it is dynamic (Balestra 1982). When dynamic misspecification exists, fixed-effects models can contain more bias than pooled OLSs (Plümper and Troeger 2019). In this way, eliminating unit-level bias can introduce new biases that may be equally as problematic.

CFscores are obviously not entirely static, as they change with each governor, but they are only constructed once for an executive based on donations in the election cycle *when they are first elected to office*. This means that for most states in a 20-year study, there will typically be only three unique values of gubernatorial ideology in each state. By contrast, other variables that are commonly used in state politics research (e.g., legislative ideology, public opinion, citizen demographics, unemployment, voter turnout) tend to update every year or two, which means that they will contain more variation by which to explain the dependent variable. Without using some degree of across-unit variance, then, the effect of gubernatorial ideology is likely to be underestimated.

Another problem with using CFscores in fixed effects models pertains to their data quality. As measures of ideology derived from campaign contributions, CFscores necessarily contain noise from any nonideological reason that donors give to candidates. This noise makes them less efficient measures of ideology relative to those (e.g., legislative roll-call scores) that directly measure elite preferences. Less efficient measures need a greater sample size in order to converge on “true” means or estimate “true” relationships with other metrics. As such, when analysis is limited to only a couple observations within each unit, CFscores may become disproportionately low-powered relative to other covariates.

For reasons articulated here and in the main text, I recommend using random effects in time-series cross-sectional studies that apply CFscores to state politics outcomes. But for scholars committed to using fixed-effects models, a few considers may help to improve the quality of the model.

If the scores are to be included as a control variable, the concerns with this metric are reduced. There remains the risk that CFscores may underspecify the role of gubernatorial ideology on the outcome of interest. To adjust for this, scholars may wish to incorporate gubernatorial partisanship as a second control, or scale CFscores with another measure of ideology as data availability allows (see other SM).

Another option, if gubernatorial ideology is a key variable in the analysis, is to reduce the number of timepoints under study to account for the sparsity of CFscores and “even the playing field” between covariates. I used this strategy to a small degree in the article by studying policy every other year; a more aggressive strategy would be to limit the analysis to gubernatorial election years in each state. This has the effect of keeping variation in other covariates (which may change each year) from overwhelming that in CFscores, which may only change once or twice per decade in a given state.

A final option is to present fixed-effects models alongside other choices, explaining that the former is likely to understate the role of gubernatorial ideology by quite a lot, while the latter may overstate it somewhat if state-specific considerations are not adequately controlled for. Analysts and readers can then deduce for themselves how meaningful an effect they find gubernatorial ideology to have on the outcome in question. I take this approach in a reanalysis of the article’s central model below, in which I compare the outputs of random- and fixed-effects models.

*Comparing random- and fixed-effects models*

In the main text of this article, after validating gubernatorial CFscores against other measures of preferences, I set up four models to explain state policy liberalism as a function of gubernatorial, legislative, and citizen ideology in each state.

Those models using random effects. Here, I present their results alongside those of fixed-effects models. In Table A6, I focus on the relationship between actor ideology and policy when Democrats are in office. Table A7 studies the same relationships in state-years with GOP governors.

The general pattern observed is that fixed effects nullify the relationship between gubernatorial ideology and policy. Among control variables, it also reduces the effect of public opinion to zero. The one exception we see is among Democratic governors on economic policy, where even in a fixed-effects model, these actors appear to be more influential over others.

Again, with a median value for λ of 0.85 across models and units, the random effects models control for unit-specific effects to a very strong degree. That even the small amount of across-unit variation allowed in these models results in a strong effect for executive ideology suggests the problems theorized above: that CFscores may introduce dynamic misspecification and be relatively underpowered compared to other covariates.

But, if we assume away these concerns to an extent, another explanation may be plausible. Perhaps executive influence is highly (even overwhelmingly) contingent on state-specific features. Although I have argued extensively here that random effects will pick up on some of these features—and that fixed effects will inordinately underestimate the effect of CFscores—this alternative is one that converges with research that describe governors as institutionally constrained (Kousser and Phillips 2010, Cooper et al. 2018). Hopefully, with its overview of measurement options and validation of CFscores, this piece can facilitate further research into the constraints that governors face.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table A6. The effects of executive, legislative and citizen ideology  on policy under Democratic governors** | | | | |
| *Method* | **Random Effects** | | **Fixed Effects** | |
| *Policy Area* | (1)  Social Policy | (2) Economic Policy | (3) Social Policy | (4) Economic Policy |
|  |  |  |  |  |
| Executive Ideology | 0.28\* | 0.43\* | -0.07 | 0.18\* |
|  | [0.08] | [0.09] | [0.08] | [0.10] |
| Legislative Ideology | 0.13\* | 0.16\* | 0.11\* | 0.10\* |
|  | [0.04] | [0.04] | [0.03] | [0.03] |
| Citizen Ideology | 0.12\* | 0.09\* | 0.05 | 0.01 |
|  | [0.04] | [0.04] | [0.04] | [0.03] |
| Constant | -0.25\* | -0.33\* |  |  |
|  | [0.11] | [0.12] |  |  |
| N (n) | 143 (40) | | 143 (40) | |
| Adj. R2 | 0.32 | 0.25 | 0.24 | 0.12 |
|  |  |  |  |  |
| Notes: \* p < 0.05, one-tailed. Brackets contain standard errors. Policy and ideology scores standardized to mean = 0 and standard deviation = 1, with positive scores indicating greater liberalism. Model estimated with random effects. Data represent policy, politics, and public opinion in 50 states, measured biannually from 1991 to 2013. | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table A7. The effects of executive, legislative and citizen ideology  on policy under Republican governors** | | | | |
| *Method* | **Random Effects** | | **Fixed Effects** | |
| *Policy Area* | (1)  Social Policy | (2) Economic Policy | (3)  Social  Policy | (4) Economic Policy |
|  |  |  |  |  |
| Executive Ideology | 0.17\* | 0.25\* | -0.11 | -0.01 |
|  | [0.10] | [0.11] | [0.09] | [0.10] |
| Legislative Ideology | 0.28\* | 0.22\* | 0.16\* | 0.06 |
|  | [0.05] | [0.05] | [0.05] | [0.05] |
| Citizen Ideology | 0.06 | 0.03 | -0.06 | -0.05 |
|  | [0.05] | [0.04] | [0.04] | [0.03] |
| Constant | 0.08 | 0.10 |  |  |
|  | [0.13] | [0.13] |  |  |
| N (n) | 174 (41) | | 174 (41) | |
| R2 | 0.21 | 0.13 | 0.09 | 0.03 |
|  |  |  |  |  |
| Notes: \* p < 0.05, one-tailed. Brackets contain standard errors. Policy and ideology scores standardized to mean = 0 and standard deviation = 1, with positive scores indicating greater liberalism. Model estimated with random effects. Data represent policy, politics, and public opinion in 50 states, measured biannually from 1991 to 2013. | | | | |

1. It is worth noting that data availability in the opposite direction—for governors without CFscores but with other measures—is minimal in the case of SOTS and voter placements. The years for which these measures are available are also years for which CFscores provide near-total coverage. There are a handful of governors with legislative scores but not CFscores, mostly in the 1990s before campaign finance reporting was required in all 50 states (for state-level elections). [↑](#footnote-ref-1)
2. When additional models are run using social policy as the dependent variable, the estimates obtained using CFscores and combined measures are the same in all cases. [↑](#footnote-ref-2)
3. When the inverse occurs—a “sticky” independent variable is used to explain a fluid dependent variable—the authors continue to recommend fixed effects. To give a state politics example, this may be the case when electoral institutions (which change infrequently over time) are used to explain a variable such as voter turnout. [↑](#footnote-ref-3)