Advance Access publication February 12, 2008

Political Analysis (2008) 16:226–233
doi:10.1093/pan/mpm031

The Search for Comparability: Response to Binder

Fang-Yi Chiou
Institute of Political Science, Academia Sinica, 128 Academia Road, Taipei, Taiwan
e-mail: fchiou@gate.sinica.edu.tw

Lawrence S. Rothenberg
Department of Political Science, University of Rochester, Rochester, NY 14627
e-mail: lrot@mail.rochester.edu (corresponding author)

Binder (n.d., Taking the measure of Congress: Reply to Chiou and Rothenberg. *Political Analysis*. Forthcoming) highlights areas of agreement and disagreement with our discussion of preference measurement and legislative gridlock. We now both agree that W-NOMINATE scores—employed in Binder (1999, The dynamics of legislative gridlock. *American Political Science Review* 9:519–33) to measure key independent variables, including bicameral differences—should not be used when examining multichamber legislatures over time. We continue to disagree over whether Common Space scores or Binder's conference vote measure is superior. In this response, we show that, although several of the theoretical and statistical objections that Binder (n.d.) raises to our Common Space measure do not apply, they are all relevant for her conference vote analog. Additionally, we detail how, despite protests to the contrary, the conference vote measure is plagued by insufficient data. Finally, we demonstrate how new efforts to show that Binder's (1999) results continue to hold are not robust.

1 Introduction

We thank Sarah Binder for her efforts to clarify the key issues involved in measuring preferences when examining multichamber legislatures over time. Focusing on bicameral differences, her response crystallizes her original critique of our work (Binder 2003) concerning our claim that using Poole's (1998) Common Space scores is desirable compared to available alternatives, defends her conference vote measure in response to the critique that we offer in our analysis, and concludes that her new measure outperforms our Common Space alternative.

We also appreciate that we have arrived at consensus between us on some key issues. We now agree that W-NOMINATE scores are inappropriate for measuring the four relevant independent variables in Binder (1999) and, therefore, the empirical results in this original work are problematic. More generally, we concur that a better measure than W-NOMINATE is imperative for testing the kinds of theoretical conjectures that Binder (1999, 2003) has made in past work. Finally, we both acknowledge that no existing measure, be it Common Space estimates or conference votes, is perfect.

Where we continue to disagree is which of these two alternative measures under consideration is relatively better. We believe that applying Common Space scores is clearly

© The Author 2008. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

superior, whereas Binder maintains that her conference vote approach is preferable. Here, we will concentrate on making our case by addressing Binder's most recent assertions directly.

Specifically, we believe that utilizing Common Space scores is far better than relying on Binder's measure for two reasons. First, the key elements of her critique regarding Common Space scores—principally that they ignore changes in dimensionality, that they involve implausible methodological devices for "gluing" legislators across Congresses and chambers, that they are corrupted by strategic voting, and that they are negatively impacted by the distribution of cutpoints—are either unfounded or apply to the conference vote measure as well, often to a considerably greater degree than for Common Space scores. Indeed, several of these criticisms only apply to the conference measure. Second, as we detail and despite protests to the contrary, the conference vote measure is further undermined by assuming that voice votes can be treated as unanimous votes and by an insufficient number of nonvoice votes, whereas Common Space scores suffer from no analogous data deficiency.

Substantively, whether we accept our claims or Binder's critique is quite important. As we will demonstrate, Binder's new efforts to show that her original (1999) results continue to hold are not robust.

2 Binder's Critique of Common Space Scores

Binder has a host of criticisms of Common Space scores. Although some have merit, they do not undermine the superiority of employing such scores relative to relying on conference votes.

One critique that she expounds on considerably involves discomfort with assuming that preferences and ideological dimensions are fixed over time. For example, it seems that the meaning of antagonism to, and support of, civil rights on the second dimension changes at different points in time and, indeed, that this dimension's meaning can vary in different periods.

However, although perhaps intuitively appealing, this critique does not undermine the usefulness of Common Space scores because it does not recognize what such scores do. The Common Space method involves uncovering the underlying *latent* variable, member preferences, and basic dimensions mapped from multiple issue dimensions (Poole 1998). Whereas the mapping, or the meaning, of such dimensions may change, the basic dimension can remain stable. Although developing a systematic means of testing whether latent dimensions and ideal points shift over time is an interesting question for future research, that one dimension, or two legislators with similar coordinates on one dimension, may be interpreted differently at varying points in time provides no evidence that preferences and ideological dimensions are not fixed.[1] Put another way, the underlying concern which we should focus on is whether the bicameral difference measure for one Congress is comparable with that for another Congress, rather than whether Common Space scores are comparable over time. As long as most legislators move in one direction in lockstep, the

[1]Given the assumption that the two dimensions of the Common Space scores are equally salient and lie within a unit circle, Binder also worries that the second dimension's reduced salience after the 1970s may distort second-dimension Common Space scores. However, as we mention in our original discussion, since the coordinates of the two dimensions are extracted independently, the second dimension's importance will not affect the comparability of the first-dimension coordinates. As we use the latter to measure bicameral difference, no problem exists. Were we to employ both dimensions jointly, for example, in computing two-dimensional Euclidean distances (which we do not do), then comparability would be an issue because we would not know the relative weights of the dimensions.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

latent variable remains stable and the Common Space measure of bicameral difference is temporally comparable.

Furthermore, even should we believe that Common Space scores are undermined because latent dimensions and ideal points shift over time, the conference vote measure would not be a superior alternative because it suffers from the same malady to an extreme extent. As Fig. 2 of our original analysis shows graphically, and as our Appendix to that analysis proves formally, the conference measure essentially has no chance of being comparable. Given that it is constructed by taking the average of absolute differences in vote margins of two chambers for each conference report, the conference measure is only comparable if a series of knife-edge conditions about the curvature of cutpoint distributions are met at each point in time.[2]

Thus, whether or not comparability issues plague Common Space scores, the likelihood of comparability is far higher than for the conference vote measure. Whereas Common Space measures control for cutpoints by including item parameters capturing the characteristics of cutpoints of individual votes, the conference measure does nothing analogous. This dramatically increases the probability that Common Space scores are temporally comparable relative to the conference vote measure.

A second criticism of Common Space scores, about which we are sympathetic, is that assuming that legislators moving from the House to the Senate have stable preferences is unrealistic. Even though many legislators ascending to the upper chamber come from small states, implying that their electoral constituencies are fairly stable, we would certainly prefer a measure having a less restrictive assumption.

However, in trying to establish evidence of member preference change, Binder pushes this critique in ways lacking foundation. First, she argues that cutpoint distributions may be dissimilar due to different rules in the two chambers and, therefore, that cross-chamber comparability is undermined. However, such chamber differences are irrelevant as comparability of Common Spaces scores does not require similar or identical cutpoint distributions. Also, she maintains that the different party loyalty scores of members as they rise from the House to the Senate constitute evidence of member preference change. However, because cutpoints or item parameters are not estimated and controlled for in calculating loyalty scores, changes in such scores need have no relationship to changes in preferences. Therefore, that a House member has a different loyalty score when she moves to the Senate is not evidence of preference change.[3]

Also, to repeat a common theme, the conference measure suffers from a related, but likely a more severe, problem that undermines the claim that it is measuring bicameral difference. Whereas Common Space scores may be subject to the critique that an unrealistic assumption is adopted about how ideal points vary for members serving in both chambers, the conference vote measure is plagued by not knowing the distribution of the ideal points of representatives and senators in a comparable matter, which is essential for it being a measure of agreement between the chambers.[4] As Fig. 2 in our original analysis illustrates, comparing vote margins to one another requires knowing the distribution of

[2]Here, cutpoints are referred to as the midpoints of bills and their associated status quos.

[3]Moreover, if Binder believes that fixed reference points (i.e., bills) for Common Space scores are required for chamber comparability, the same problem applies for party loyalty scores. As such, examining party loyalty score changes is an inappropriate means of investigating whether legislators' preferences change.

[4]This, for the moment, sets aside the issue of exactly what concept of bicameral difference is behind the conference measure: the median difference or the difference between the tails of the distribution in each chamber. For the Common Space measure, bicameral difference is defined quite precisely as the difference between chamber medians.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

ideal points. But, although conference votes serve as reference points in determining bicameral difference, the vote margin in each chamber only reveals information about the proportion of legislators locating on each side of the cutpoint, and the absolute difference between the vote margins of the two chambers only reveals information about proportion differences.[5] Additionally, as almost all conference votes are lopsided, their usage at most reveals information about the two tails of each chamber's preference distributions. We are only finding the number and distribution of rather extreme legislators in the two chambers. Clearly, using conference votes does not measure the distribution of members' preferences in a manner that can accurately capture bicameral difference across chambers and over time.

A third criticism regarding the use of Common Space scores is that they are polluted by strategic voting.[6] This critique is certainly a reasonable objection to consider. Furthermore, although the conference vote measure would not be immune from the same problem, as conference reports might be accepted or rejected in anticipation of what future conference votes might hold (unless Congress was about to adjourn), strategic voting would certainly impact Common Space scores more if it was a corrupting force. However, as even Binder herself acknowledges in her response to us (p. 17), contrasting NOMINATE estimates using different subsets of votes, which may contain different degrees of strategic voting, matters very little. Indeed, the preponderance of evidence suggests that the kind of preference-based ideology scores that concern us varies little regardless of what kinds of votes are used for estimation (for a recent discussion, see Poole and Rosenthal 2007).

The final criticism offered of our Common Space measure is that different cutpoint distributions in both chambers will hinder across-chamber and pre- and post-1970s comparability of Common Space scores. However, such a critique is not a problem. Common Space measures are generated by adjusting W-NOMINATE scores, whose estimates do not require identical or similar distributions across chambers or between the periods pre- and post-1970s because item parameters are included and estimated to control for cutpoint distributions.[7] Ironically, this critique is relevant for the conference vote measure. As there is no ability to incorporate cutpoint distribution controls, comparability issues will be great.

3 Lacking Data: Is Unanimity Really Enough?

Whereas Common Space scores are estimated with a huge array of data, the conference measure is not. In our original discussion, we maintain that treating voice votes as unanimous votes, which is crucial for the conference measure given the unavailability of a substantial number of nonvoice conference votes, is almost certainly incorrect. If true, this would be one more reason why the conference vote measure would be flawed and less viable than the Common Space alternative.

In her response, Binder defends assuming that voice votes are evidence of true unanimity on the grounds that the requirement for requesting a roll call vote by a majority

[5]Furthermore, taking the average of the absolute difference imposes the implausible assumption that each conference vote should be equally weighted when measuring bicameral difference.

[6]Binder seemingly misinterprets our objection to including voice votes as having to due with strategic voting. As we will discuss shortly, our objections are based on the mismeasurement of preferences produced by including voice votes.

[7]Binder's (n.d.) finding that there is a strong correlation between estimated chamber medians using Common Space scores based solely on final passage votes and those using Common Space scores based on all roll call votes illustrates this point. Although the two sets of votes have different cutpoint distributions, the estimates generated are not impacted.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

hoping to embarrass a minority is low. Therefore, dissenters could be heard if they wished to be.

Yet, although the hurdle for requesting a vote may not be onerous, there may be many deterrents to having a roll call as well. Most obviously, if there is no additional political gain because the proposal in question is not salient for external constituencies, there may be no incentive for either a minority or a majority to demand a roll call on a conference report with a supermajority or overwhelming support or opposition as the ultimate result is foreseen. This is not tantamount to unanimity. Furthermore, even a majority wishing to embarrass a minority may decide not to do so if there is a cost involved in having such a vote in terms of time and effort, for example, if the floor's docket is full at the time in question (especially before the introduction of electronic voting in the 1970s) or these members are trying to adjourn.[8] This may be particularly true if there have already been recorded votes on similar issues, so that the value of any embarrassment from position taking is small. Thus, despite Binder's claim that majority incentives will induce only unanimous voice votes (assuming that the majority does not get tarred as well by a recorded ballot), this will certainly not be the case if there are opportunity costs attached to having a roll call.

Hence, although there may be a few voice votes that reflect true unanimity, it is almost certainly unreasonable to view the vast majority of such actions as unanimous rather than lopsided votes. However, we do not know who or how many dissenters there are for a given voice vote. Without a method for imputing who would dissent, including voice votes is unjustified. By extension, besides the theoretical problems associated with the conference measure, it simply suffers from insufficient data.

4 Robustness Checks

In her new analysis (see her table 1), Binder resorts her conference votes and claims to find evidence that her original (1999) analysis is robust. Specifically, she drops two variables, ideological diversity and filibuster threat, from her analysis on the grounds that W-NOMINATE scores should not be used to measure variables requiring comparability across chambers and Congresses.[9] However, she retains the rest of the independent variables from her original analysis, including percentage of partisan moderates, which is measured with W-NOMINATE. She then finds that divided government, percentage of partisan moderates, and bicameral difference have statistically significant effects on legislative gridlock with the expected signs but uncovers no strong evidence for the importance of party mandates.

Although we appreciate the effort that Binder has taken to examine robustness, we do not believe that this is a convincing check. First, W-NOMINATE scores are still being used to measure percentage of partisan moderates. Although doing so is justified by claiming that our original work finds that using W-NOMINATE or Common Space scores does not affect the measurement of this variable, this is misleading because the variable itself has been redefined. We compare Common Space and NOMINATE measures of Binder's (1999) definition of the variable (the average proportion of moderates between the party medians in the two chambers). However, following Binder (2003), Binder (n.d.) redefines this variable by dividing it by the distance between party medians. Whereas the former specification might not require preferences that are comparable across chambers and Congresses, the latter certainly does, given that the new definition incorporates the measure of

[8]Such costs may explain why there were less than six nonvoice conference votes in the Congresses before the 1970s (except in the 81st Congress).

[9]If Binder's previous (1999, 2003) theoretical conjectures are correct that ideological diversity and filibuster threat matter, dropping them likely biases the estimates for her other independent variables.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

Table 1 Robustness of policy gridlock results (grouped-logit coefficients with SEs in parentheses)

Variables	*Binder analysis*	*Robustness Checks*		
		With Common Space measure of partisan moderates	*With Bailey (2007) measure of bicameral difference*	*With Bailey (2007) measure of partisan moderates and bicameral difference*
Divided government	0.443** (0.157)	0.326 (0.180)	0.300* (0.167)	0.427 (0.195)
Partisan moderates (W-NOMINATE)	−0.015** (0.005)		−0.023** (0.008)	
Partisan moderates (Common Space)		−1.581* (0.774)		
Partisan moderates (Bailey)				−0.004 (0.010)
Time out of majority	−0.050 (0.035)	−0.043 (0.039)	−0.077 (0.045)	−0.003 (0.042)
Bicameral distance (Conference measure)	7.739* (3.706)	4.386 (4.521)		
Bicameral distance (Bailey measure)			0.655 (0.874)	−1.111 (0.707)
Budgetary situation	0.017 (0.011)	0.007 (0.014)	0.004 (0.010)	−0.009 (0.021)
Public mood (lagged)	−0.004 (0.015)	−0.010 (0.016)	−0.006 (0.016)	−0.009 (0.021)
Constant	−0.116 (0.995)	0.628 (1.214)	0.609 (1.060)	0.291 (1.343)
Adjusted R^2	0.415	0.240	0.298	0.028

Note: Number of cases = 24; one-tailed *t*-test, *$p < .05$, **$p < .01$.

party median distance. As such, if it is unreasonable to use Common Space scores, we should eliminate this variable as well. Doing so leads to three of the six principal independent variables of interest to Binder being dropped, with the results for two of the remaining variables, divided government and bicameral difference, gaining empirical support.[10]

Alternatively, we could adopt a Common Space analog to measure the new definition of percentage of partisan moderates and retain the conference vote measure of bicameral difference; however, bicameral difference then becomes statistically insignificant (compare the first two columns of results in Table 1).[11] In the same spirit, if we substitute Bailey's (2007) measure of preferences to calculate bicameral difference (keeping the W-NOMINATE measure for moderates), we get results that contrast with Binder's results and confirm our findings when we use a Common Space measure (see the third column of results in Table 1).[12] Moreover, as with our Common Space results, not only is the coefficient for bicameral difference using the Bailey preference measure insignificant, but also it is negative if we replace W-NOMINATE scores with Bailey's preference estimates to measure percentage of partisan moderates (see the fourth column of results in Table 1).[13] Although Binder interprets the Common Space scores' relationship with gridlock as evidence of their poor measurement and the break in the correlation between the conference vote measure and the Common Space measure as reflecting changes in how much the second dimension biases the Common Space estimates, the fact that the same results are found with Bailey's measure indicates that this view is problematic.[14] These results are not surprising as the correlation between the conference vote measure with either that using Bailey's estimates (.05) or that employing Common Space measures is (−.01) is virtually 0. Thus, we cannot conclude that results are strictly a function of the pecularity of our Common Space measure.

In the end, while we applaud Binder's efforts to perform robustness checks, we do not find them ultimately convincing. Indeed, probing further only confirms the conclusion that her original 1999 results are not robust and the inference that using the conference measure does not capture bicameral differences.

5 Final Thoughts

We thank Professor Binder for offering a series of provocative theoretical conjectures and for expending great effort in examining how measurement issues come into play. The end

[10]Binder (2003) measures filibuster threat with DW-NOMINATE rather than W-NOMINATE scores to avoid the comparability problem. However, whether W-NOMINATE, DW-NOMINATE, or Common Space scores are used to measure filibuster threat, it is always insignificant. Furthermore, bicameral difference tends to be insignificant once filibuster threat is included.

[11]Including Common Space measures of the two independent variables that Binder drops, ideological diversity and filibuster threat, produces results roughly comparable to the findings of Chiou and Rothenberg (n.d.) (results available from authors). No independent variable is statistically significant and the adjusted R^2 is only 0.04.

[12]We choose Bailey's measure because Binder relies heavily on Bailey's sharp criticisms of Common Space scores. Bailey, using Bayesian methods, estimates ideal points for the Supreme Court, Congress, and the president. To deal with the inevitable issue of bridging preferences across institutions and legislative chambers, Bailey uses data on presidential and legislator position taking on Supreme Court cases past and present and the relative location of bill and case cutpoints.

[13]Similar results are generated when the two independent variables that Binder drops, ideological diversity and filibuster threat, are measured with Bailey's preference estimates and included in the estimation.

[14]Although Binder (n.d.) suggests that there is no reason to expect bicameral difference to have a negative effect, such a result could be a function of incorrect theoretical conjectures and of ignoring the complex relationship between bicameral difference and gridlock. For example, in earlier theoretical work, we (Chiou and Rothenberg 2003) have shown that the directional impact of bicameral difference on gridlock depends on configurations of pivots.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press

result, we believe, is to clarify a series of more general issues involved with measurement involving multiple institutions and time periods. We are also delighted that there are areas in which we have found common ground.

Nonetheless, we remain convinced that the preponderance of evidence shows that using measures based on Common Space scores is far superior to the alternative that Binder offers. Whereas we continue to recognize that Common Space measures are no panacea and try to be explicit why this is the case, they have a theoretical and statistical grounding that efforts such as the conference vote measure lack. Given that the inherent problems of the conference vote measure are coupled with a lack of available data, it is not surprising that this approach does not capture bicameral difference and that results are not robust.

References

Bailey, Michael A. 2007. Comparable preference estimates across time and institutions for the Court, Congress and presidency. *American Journal of Political Science* 51:433–48.

Binder, Sarah A. 1999. The dynamics of legislative gridlock. *American Political Science Review* 9:519–33.

———. 2003. *Stalemate: Causes and consequences of legislative gridlock.* Washington, DC: Brookings Institution Press.

———. n.d. Taking the measure of Congress: Reply to Chiou and Rothenberg. *Political Analysis.* doi:10.1093/pan/mpm033.

Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2003. When pivotal politics meets partisan politics. *American Journal of Political Science* 47:503–22.

———. n.d. Comparing legislators and legislatures: The dynamics of legislative gridlock reconsidered. *Political Analysis.* doi:10.1093/pan/mpm021.

Poole, Keith T. 1998. Recovering a basic space from a set of issue scales. *American Journal of Political Science* 42:964–93.

Poole, Keith T., and Howard Rosenthal. 2007. *Ideology & Congress.* 2nd rev. ed. New Brunswick, NJ: Transaction Publishers.

https://doi.org/10.1093/pan/mpm031 Published online by Cambridge University Press