

MEASURING STATE PREFERENCES: USING NETWORKS, COMBINING INDICES

ABSTRACT.

2. INTRODUCTION

2.1. Why we care about preferences. Compared to other concepts in the study of international conflict and cooperation, state preferences have been understudied. Perhaps this is a consequence of black box theories of international relations, where it is assumed that all state preferences can be reduced to an automatic desire for more power. Alternatively, it may be because it is more difficult to measure something like preferences, as compared to more tangible material or institutional factors.

Yet this relative dearth of attention should not be assumed to be caused by the relative unimportance of preferences in the study of international processes. A number of formal theories of international relations require measures of preferences to be tested: the expected utility theory proffered by (?) has similarity of preferences as an important input, and attempts to expand studies of crisis bargaining to include coalitional dynamics (?), mediation (?), or the possibility of additional disputants (?) require a measure of state preferences in order to predict whether war will be the result of bargaining failure. As ? has pointed out, preferences have been used in empirical studies predicting bilateral trade, foreign aid, stability of international institutions and the incidence of conflict (?????). Preferences similarly pose the risk of omitted variable bias in a number of issues of paramount concern: without a good measure of preferences, it would be difficult to entangle whether democracies avoid war with other democracies because of the intrinsic nature of democracy, or simply because they happen to have common ends; similarly attempting to assess the impact of crisis behavior on a state's reputation requires us to determine how acceptable an outcome was to the state, as (?) does in his study of interdependence and resolve. Finally, measures of preferences can yield insights as an independent variable in their own right: they could be used to see if Edward Snowden's revelations about the United States spying caused states to move their preference away from the US's, to see how the United States's preferences towards states in the Middle East changed after 9/11, or to see the impact of Russia's annexation of Crimea on their relations with their satellite states and Western Europe.

2.2. Current measures of preferences: S-Scores. The idea behind alliance portfolio measures is that we can infer a state’s foreign policy by looking at the states they choose to align with. In the extreme case, if two states have all of the same allies, it is likely that their foreign policy goals are quite similar. Conversely, if all allies of one state are not allied to another, and vice versa, our best guess is that these states would have different aims and desires in foreign policy. ? encapsulate the logic when they note that ”alliance commitments reflect a nation’s position on major international issues.” While the initial measures used to get at this intuition were Kendall’s τ_B , ? convincingly pointed to flaws in this measure, notably its focus on rank-ordering as applied to a context where we instead care mostly about the presence or absence of an alliance. Thus, Signorino and Ritter introduce the S score, which has since been the most widely used alliance similarity measure.

The equation for the S score is:

$$(1) \quad S(P^i, P^j, W, L) = 1 - 2w_k \frac{d(P^i, P^j, W, L)}{d^{\max}(W, L)}$$

Where:

$$(2) \quad d(P^i, P^j, W, L) = \sum_{k=1}^N \frac{w_k}{\Delta_k^{\max}} |p_k^i - p_k^j|$$

Where w_k is the k’t h element of a weight matrix, $d^{\max}(W, L)$ is the maximal distance on a given dimension, and Δ is a normalizing constant. For the weight matrix, generally analysis has used S scores calculated with a weight matrix of ones—giving each potential ally equal weight—though the other plausible choice would be to weight states by import, for example using their share of world military capability, as calculated by ?.

One important distinction for these scores is that they are purely dyadic. One can look at the S-score between two states, but one cannot look at a state’s preferences in comparison to a larger cluster, or note the movement a states preferences made over time. In monadic analysis, these score measures are not even available, and once we are dealing with situations involving

more than two states, the number of S-scores necessary to fully characterize the preferences balloons quickly (it is the number of actors choose two).

These measures of alliance behavior also suffer from the relatively glacial movement and sparsity in these relationships. Formal alliances are relatively constant over time, whereas in many cases state preferences will be more fluid, and therefore these scores will be at best a lagging indicator of preferences. Furthermore, as ? points out, the fact that links are so rare creates artificial inflation of S scores. Signorino and Ritter propose a method to, somewhat account for this by weighting S scores using national capabilities, but these measures of capabilities are so skewed that this version of the measure is rarely used and would put undue emphasis on the alliances of the United States.

2.3. Current measures of preferences: UN voting data. An alternative way to measure preferences would be to look at behavioral indicators of these preferences. For most behavioral indicators this is impossible to do in a systematic measure as the contexts in which behaviors arrive differ between states, and it would be difficult to classify two behaviors as equivalent. However, we have a relatively large corpus of somewhat behavioral information in UN Voting Records. The cost of voting in the UN is low, and so, scholars have argued that measures of affinity based on UN voting are relatively representative of the underlying distribution of preferences (?). This is especially fortuitous because the methodology of inferring preferences from voting in a legislature is relatively advanced. ? classify different types of votes in the UN and use a model where both the type of vote, and states innate ideal points determine whether a state votes Yes, No, or Abstains on an issue. This spatial model leverages the fact that certain resolutions in the United Nations are voted on multiple times in order to identify situations where states change their ideal points.

Voeten and his coauthors use an MCMC algorithm to iteratively estimate both the ideal points of each state as well as the cutpoints which would cause states with a certain ideal point to vote for, against, or abstain.

The basic idea behind this model is that state voting behavior is determined by an unobserved ideal point, such that on a given vote, anyone who's ideal point is past a certain outpoint will vote in a particular way. In particular, we call the spatial of state i in year t on vote v Z_{itv} , and we assume that:

$$(3) \quad Z_{itv} = \beta_{iv}\theta_{it} + \epsilon_{iv}$$

where θ_{it} is State i 's unidimensional ideal point in year t and β_{iv} is the effect of that ideal point on a given type of vote (some types of votes, high θ states will be inclined to vote yes, and so β_v will be positive, whereas when it is negative states with high values of θ will be inclined to vote "No". The magnitude of β determines how well the vote distinguishes between states with different ideal points. The state then has vote $y_{itv} \in \{1, 2, 3\}$, where $y_{itv} = 1$ implies a no vote, 2 implies an abstention, and 3 implies a yes vote. The spatial preference then maps onto voting behavior, such that if $Z_{itv} < \gamma_{1v}$, $y_{itv} = 1$, if $\gamma_{1v} < Z_{itv} < \gamma_{2v}$, $y_{itv} = 2$, else $y_{itv} = 3$.

Thus the probability that a state has vote k is:

$$(4) \quad P(y_{itv} = k) = \Phi(\gamma_{kv} - \beta_v\theta_{it}) - \Phi(\gamma_{k-1v} - \beta_v\theta_{it})$$

The authors specifically fix the parameters γ_{1v} and γ_{2v} such that the same bill will have the same value in different years, and they standardize and normalize θ . They also use θ_{it-1} as a prior on θ_{it} . With these constraints, they solve for the ideal points and outpoints using a Metropolis Hastings algorithm.

The issues here are that the voting behavior, especially in the EU general assembly, is not well behaved in the way that voting in the US Congress is. We actually can get some sense of it by the existence of multiple identical resolutions: UN resolutions have no legal force, and so most votes are symbolic. Thus UN voting is rife with nearly unanimous voting and other super-majorities, which means that the requirements to distinguish between state preferences

are more onerous. Another issue here, not necessarily with use of voting in general, but with this application, is the limitation to one dimension: it could be that two states which have very similar preferences on issues of trade – say the United States and Saudi Arabia – might differ mightily on questions related to the Middle East, and in particular Israel.¹ However, the dearth of contentious UN votes makes it difficult to add additional dimensions, and therefore ...

3. OUR APPROACH: USING EVENT DATA

We propose that one can get a useful proxy of state preferences is a more fluid and behavioral approach—using event data on cooperation and conflict to measure amity and enmity between states. Event data is (generally) machine coded data which uses news articles to capture the type and tenor of relationships between actors. The goal would be to take a measure of cooperative actions, and conflictual actions, and use these two counts to estimate each state’s position in a two dimensional latent space in a given year.

The intuition behind this use of event data is that states are more likely to cooperate with states with similar preferences, and they are more likely to conflict with states that have divergent preferences. Similarly, we can also infer similarity by how states act towards common third parties. The reason we use event data is that it is more dynamic and representative of current preferences than data on alliance or international organization comembership (which becomes quite sticky and might represent states preferences at some point in the distant past), while at the same time it is more revealing than behavior at the United Nations, because these conflictual and cooperative actions are actually consequential, in contrast to non-binding UN resolutions.

3.1. ICEWS Event Data. To estimate state preferences. The ICEWS data were collected via a Defense Advanced Research Project Agency (DARPA) funded project that created a

¹You can see issues like this on legislative positions in the United States: during the post WW2 era, two democrats who might agree on the need to expand the social safety net might be diametrically opposed on issues of civil rights.

dataset of over two million machine-coded daily events occurring between relevant actors within twenty-nine countries in the the Asia-Pacific region. ICEWS utilized news articles from over 75 electronic regional and international news sources and machine coded these events, using the Penn State Event Data Project’s TABARI (Text Analysis By Augmented Replacement Instructions) software program(?) and a commercially developed, java variant (JABARI). TABARI and JABARI used sparse parsing and pattern recognition techniques to machine-code daily political events based primarily on a categorical coding scheme developed by the Conflict and Mediation Event Observation (CAMEO) project(?). This ICEWS dataset is the current gold- standard for event data(?, p.4).

For an example of how a conflictual event was coded in the data, consider the example:²

Israel today mounted its long-threatened invasion of South Lebanon, ploughing through the United Nations lines on the coast of south of Tyre and thrusting forward in at least to inland areas.

In this sentence, Israel is coded as the source of the event, and South Lebanon is coded as the target. Then South Lebanon is aggregated with events involving the entire country of Lebanon. The event type coded in this story is “occupy territory,” coded as such because of the use of the verb “mounted” and the noun “invasion.” Events of the type “occupy territory” are then put into the category “material conflict.”

Similarly, this story from a major news service:

The Afghan foreign ministry announced that Kabul and Tehran have agreed to a prisoner exchange, a move seen by many analysts as yet another sign of warming relations between the two neighbors ahead of the planned withdrawal of foreign combat forces from Afghanistan in 2014.

is coded as an action between Afghanistan and Iran, based on the nouns “Kabul” and “Tehran,” and is of the event type “express intent to release persons or property” based on the phrase “agreed to a prisoner exchange.” This type of event is coded as verbal cooperation.

²Taken from the CAMEO code book, (?).

We aggregate this data to the yearly level, combined based on the concept of Quad Categories—a two by two that differentiates verbal and material acts, as well as conflictual and cooperative ones.

3.2. Synthesizing Measures of State Preference. We propose that preferences and ideal points can be better measured by combining multiple proxies, and accounting for network interdependencies. Obviously, the idea of using multiple metrics to get a better handle on preferences is not new, in fact Signorino and Ritter suggested it in introducing S scores, which were designed to allow for aggregation of similarity on multiple dimensions (such as alliances and UN voting). What we propose, is to combine the dyadic measures of state similarity created using S-scores for alliances, and using voting models for UN data, in a manner that is both principled, and allows us to account for interdependencies. In particular, we use these two measures of state preference in a network model, in order to ascertain the state positions that best explain not only states dyadic similarity and dissimilarity on both measures, but also why states form the clusters they form. Our hope, is that by combining different measures of state preferences, and better accounting for spatial dependencies, we are able to generate a measure for preference that maintains the insights of both UN voting scores and S-scores, but which can also yield some new insights, in particular, when it comes to predicting and explaining interstate conflict.

4. METHODOLOGY

4.1. AME, Why Network stuff matters. We want a model that takes a set of actions between countries, and infers each countries position in a latent preference space, such that those countries close to each other are likely to have similar preferences and therefore have similar alliances and UN voting records. We would like this methodology to be able to, in a principled way combine different sources of data, for example imputing ideal points based on both alliance behavior and behavior at the UN. Finally, and importantly, this method should be able to account for interdependencies: similarity in preferences should be transitive (if the

US has similar preferences to the UK, and the UK to France, the US's preferences should be relatively close to France's) and should allow for clusters of states with similar preferences.

The Additive and Multiplicative Effects model (AME) model is a relatively new technique that is a generalization of the Generalized Bilinear Mixed-Effects model from ?. The model is an extension of the Social Relations Model:

$$(5) \quad f(Y_{i,j}) = \beta' \mathbf{x}_{i,j} + \alpha_i + b_j + \epsilon_{i,j}$$

where $f(\cdot)$ is a general link function corresponding to the distribution of Y , $\beta' \mathbf{x}_{i,j}$ is the standard regression term for dyadic and nodal fixed effects, α_i, b_j are sender and receiver random effects, and $\epsilon_{i,j}$ is an IID error term. The AME model further decomposes the error term as follows. If we assume the matrix representation of deviation from the linear predictors and random effects is \mathbf{Z} , then $\mathbf{Z} = \mathbf{M} + \mathbf{E}$ such that the matrix \mathbf{E} represents noise, and \mathbf{M} is systematic effects. By matrix theory, we can decompose $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}'$ such that \mathbf{U} and \mathbf{V} are $n \times n$ matrices with orthonormal columns, and \mathbf{D} is an $n \times n$ diagonal matrix. This is called the singular value decomposition of \mathbf{M} .

We then write the AME model for a given value $Y_{i,j} \in \{0, 1\}$:

$$(6) \quad \text{logit}(P(Y_{i,j} = 1 | x_{i,j})) = \beta' \mathbf{x}_{i,j} + \alpha_i + b_j + \mathbf{u}_i \mathbf{D} \mathbf{v}_j' + \epsilon_{i,j}$$

In estimating preference models, we abstain from using fixed effects save an intercept.

An important innovation with the AME, as compared to previous network estimates is the ability to handle replicated datasets – here we use the replicated dataset to incorporate multiple measures of similarity into a single ideal point estimation. The AME with dyadic data treats each different slice of data as independent, save for those dependencies captured by the nodal and multiplicative random effects, as well as those controlled for by fixed effects. The final estimating equation we use is:

$$(7) \quad \text{logit}(P(Y_{i,j} == 1)) = \mu + \alpha_i + b_j + \mathbf{u}_i \mathbf{D} \mathbf{v}_j' + \epsilon_{i,j,t}$$

What we are particularly interested in, in this equation is the estimates for $\mathbf{u}_i \mathbf{D} \mathbf{v}_j'$. This multiplicative effect is composed of ideal points in a sender and a receiver space.³ States that are close in the sender space (receiver space) are likely to send (receive) the dependent variable⁴ to (from) the same third parties. Thus, states which are close in these spaces can be thought of as having similar preferences, and those that are quite distant will have dissimilar preferences.

4.2. Non-Network Model. Dimension reduction tensor stuff.

4.3. Data Sources, Modeling choices. We use the AME model on the two aforementioned measures of state amity to generate a combined measure of state preference similarity which accounts for network effects. We use the distance between states' ideal points (as calculated by ? using UN data) and S-score for two states alliance portfolios. However, to facilitate comparison between the metrics, we first transform the S-score into a measure of distance between alliance portfolios.⁵ We then standardize and normalize these two measures. This gives us an N by N by Y by 2 array, where the first two dimensions represent countries, the third dimension is the year, and the fourth is the particular measure of similarity. So the item at index (1,2,1,1) would be the transformed value of the S-score for countries XXX and YYY at the first year of our data (YYYY), similarly (1,2,1,2) would be the UN ideal point distance.

Another important question is the amount of temporal aggregation used. In our baseline model, we treat each year as separate and gain a unique observation of each states ideal point in each year. However, this raises a real risk of temporal inconsistency in the values. An alternative approach would be to have a rolling average for the measures of similarity over a number of years. This would allow us to infer a country's relative position not just by their

³Which are identical here as the variables we use are symmetric.

⁴Here similar alliance behavior or UN voting.

⁵D = 1 - S

behavior in a given year, but also their behavior in the past few years. The risk if we use too much temporal aggregation is that we are including data which is no longer relevant to a country's relative preferences. For instance, Turkey and Russia's relationship looks a lot more positive when we look at 2013 and 2014 then when we look at 2015. To that end, in addition to our baseline model where years are seen as independent, we also evaluate models where ...

With this data, we run an AME model with a Gaussian link, and in particular we use the uDv term to estimate each states position in a two-dimensional latent space. We then evaluate whether there is additional utility gained from using this latent position, as compared to the component measures of similarity of alliance portfolio and UN ideal point distance.

- Functional Form

5. EXPLORATORY TESTS

5.1. **Face Plausibility.** Are states close to who we'd expect them to be?

5.2. **Temporal Reliability.** Are our measures consistent?

6. MODEL COMPETITION

6.1. **Data, Controls.**

6.2. **In sample explanation.** Are coeffs starry and in right direction

6.3. **In sample postdiction.** Good model fit?

6.4. **Out of sample prediction.** Are we right?

7. CONCLUSION

YAY I WAS RIGHT! WE AM SMRT!

REFERENCES

- Gartzke, Erik. 1998. "Kant We All Just Get Along? Opportunity, Willingness and the Origins of the Democratic Peace." *American Journal of Political Science* 42(1):1–27.
- Hoff, Peter D. 2009. "Bilinear Mixed Effects Models for Dyadic Data." *Computational and Mathematical Organization Theory* 15:286–295.
- Singer, J. David, Stuart Bremer and John Stuckey. 1972. chapter Capabiltiy Distribution, Uncertainty, and Major Power War, 1820-1965.