Measuring State Preferences: Using Networks, Combining Indices[☆]

Max Gallop[a], Shahryar Minhas[b]

[a]*Departments of Political Science, University of Strathclyde, Brexit territory*
[b]*Department of Political Science, Duke University, Durham, NC 27701, Trumpland*

Abstract

We do things.

[☆]Thanks to people.

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 belies the importance of preferences in our theories of international processes. A number of formal theories of international relations require measures of preferences to be tested: the expected utility theory proffered by (Bueno de Mesquita, 1983) has similarity of preferences as an important input, and attempts to expand studies of crisis bargaining to include coalitional dynamics (Wolford, 2014), mediation (Kydd, 2003), or the possibility of additional disputants (Gallop, 2017) require a measure of state preferences in order to predict whether war will be the result of bargaining failure. Preferences have been used in empirical studies predicting bilateral trade, foreign aid, stability of international institutions and the incidence of conflict (Kastner, 2007; DeRouen and Heo, 2004; Stone, 2004; Gartzke, 2007; Braumoeller, 2008).

One of the most important reasons we need a good measure of preferences is to correctly understand the democratic peace.It is 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. Farber and Gowa (1995) argued that democracies were only peaceful during the Cold War period because they had similar preferences and alliance structures. Similarly, Gartzke (1998) argues that

dissimilar preferences are a necessary condition for conflict. Oneal and Russett (1999) responded by arguing that democracy has both a direct inhibiting effect on conflict, and an indirect one through influencing state preferences, though Gartzke (2000) argued that even though democracies might have similar preferences, the residual of preferences from democracy explains conflict much better than the residual of democracy from preferences. While there has been some impressive development with our measures of preferences in recent years, a more accurate measure would really help us to disentangle the extent to which peace is the product of shared preferences, and the extent to which institutions and norms are driving peace.

While we would like an accurate measure of state preferences as an independent variable in our theory, measures of preferences can yield insights as a dependent variable in their own right. They could be used to see if the election of Donald Trump 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 near-abroad states and the European Union.

2.2. Sources of Preference Measures: Alliance Portfolios and UN Voting

Given that we cannot directly observe state preferences, scholars have attempted to ascertain it using two main behavioral indicators: who states choose to ally with, and how states vote at the United Nations. 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." 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 Häge (2011) points out, the fact that links are so rare creates artificial similarity of alliance portfolios.

We also 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 (Gartzke, 1998). This is especially fortuitous because the methodology of inferring preferences from voting in a legislature is relatively advanced. A few issues with these measures are that the potential benefit of winning UN votes is low, and so states might have incentives to vote against their preference as they are not costly signals, and the distribution of UN voting is weird and prone to large supermajorities of the type rarely seen in "ordinary" legislatures.

2.3. Current measures of preferences: S-Scores

The initial measures used to measure preference similarity based on alliance portfolios was Kendall's τ_B **?**. This measure is:

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \tag{1}$$

where n_c is the number of pairs where both actor i and j have the same rank ordering (for example both the UK and the US are more closely allied to Israel than to Iran), n_d is the number of pairs where they have discordant rankings (the US is more closely allied to Saudi Arabia than to Russia, Syria is more closely allied to Russia than to Saudi Arabia). The denominator attempts to adjust the total number of pairs with

the number of ties: n_0 is the total number of pairs ($n(n-1)/2$), n_1, n_2 are measures for ties in both i and j's rankings respectively.

Signorino and Ritter (1999) 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. In addition, if we add additional strategically irrelevant states, we will create artificially high τ_B statistics. Thus, Signorino and Ritter introduce the S score, which has since been the most widely used alliance similarity measure.[1]

The equation for the S score is:

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

Where:

$$d(P^i, P^j, W, L) = \sum_{k=1}^{N} \frac{w_k}{\Delta_k^{\max}} |p_k^i - p_k^j| \tag{3}$$

Where w_k is the k'th 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 Singer and Small (1995).

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

[1](Bennett and Rupert, 2003) also find a stronger relationship between theoretical predictions and results when using S-scores than when using τ_B

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). Gartzke (1998) attempted to apply a similar S-score methodology to UN voting data and created the "Affinity of Nations" index.

One advantage, however, of using UN General Assembly Voting, is that it allows you to take advantage of methodological advances in the study of legislatures. Bailey and Voeten (2015) do so by using an Item Response Theory model on UNGA voting. This model seeks to place states on a unidimensional latent preference space using their voting behavior. The assumption of this model is that states' votes on a resolution are a function of states' ideal points, characteristics of the vote, and random error. In particular, for each bill v, a states vote will be based on the latent variable Z_{itv}

$$Z_{itv} = \beta_{iv}\theta_{it} + \epsilon_{iv} \tag{4}$$

such that the state will vote yes if $Z_{itv} < \gamma_{1v}$, no if $Z_{itv} > \gamma_{2v}$ and otherwise abstain. Here, θ_{it} is state i's ideal point at time t, and β_{iv} is the discrimination parameter of a particular bill v. When β_v is positive, states with high ideal points will be more likely to vote no. When it is negative, they will be more likely to vote yes.

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 θit. With these constraints, they solve for the ideal points and outpoints using a Metropolis Hastings MCMC.

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.[2] However, the dearth of contentious UN votes makes it difficult to add additional dimensions.

2.4. 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

[2]You 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.

insights, in particular, when it comes to predicting and explaining interstate conflict.

3. Methodology

3.1. AME, Why Network stuff matters

Figure 1: Tensor representation of longitudinal dyadic, representational measures. The green and blue colors represent different relational measures and darker shading indicates later time periods. Specifically, we show a tensor with dimensions of $4 \times 4 \times 2 \times 3$, where 4 represents the number of actors, 2 the number of relational measures, and 3 the number of time points.

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 Hoff (2005). The model is an extension of the Social Relations Model:

$$f(Y_{i,j}) = \beta^{'} \mathbf{x_{i,j}} + \alpha_i + b_j + \epsilon_{i,j} \tag{5}$$

where $f(.)$ 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 $\mathbf{M} = \mathbf{UDV}^{'}$ such that $\mathbf{U}$ and $\mathbf{V}$ are are n x n matrices with othonormal columns, and $\mathbf{D}$ is an n x n diagonal matrix. This is called the singular value decomposition of $\mathbf{M}$.

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

$$\text{logit}(P(Y_{i,j} = 1|x_{i,j}) = \beta^{'}\mathbf{x_{i,j}} + \alpha_i + b_j + \mathbf{u_i D v_j^{'}} + \epsilon_{i,j} \tag{6}$$

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:

$$\text{logit}(P(Y_{i,j_j} = 1) = \mu + \alpha_i + b_j + \mathbf{u_i D v_j^{'}} + \epsilon_{i,j,t} \tag{7}$$

What is particularly useful here is the eponymous multiplicative effect $\mathbf{u_i D v_j^{'}}$. This effect not only helps to account for homophily and stochastic equivalence, it also places each state in a latent space. What is key to understand about this latent space is that it is non-euclidian. Rather than have states behave similar to the states which are close to them, this latent space is a two dimensional representation of a hypersphere, and thus states are apt to behave similarly to the states that are placed in the same

direction on said sphere. Thus, if two states vectors (from the center of the space) are in the same direction, they are apt to send and receive both alliances and co-voting to similar targets. The way we measure this similarity in dimension is by looking at the absolute distance of the angles created by each states position and the center of the latent space.

3.2. 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 Bailey and Voeten (2015) 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.[3] 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 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 includ-

[3] D = 1 - S

ing 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 their 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.

4. Constructing Latent Angle Measure

4.1. Face Plausibility

An important question for these different measures of preferences, are whether they give results that "make sense" for certain prominent pairs of countries. In particular we would hope that the measures both give sensible levels to relationships – sorting states into friends and foes effectively – but also that when these measures change, they do so in sensible ways, and in ways that correspond to changes in the world. We thus present all three measures' accounts of eight dyadic relationships over time.[4] A note is that each of these measures is on a different scale, and so just because one measure has a higher value, does not necessarily mean that it posits more dissimilarity of preferences.

We first look at three close relationships where we'd expect to see similar preferences: as depicted in figure 2 the relationships between France and Germany, the US and Israel, and North Korea and China. In all three cases our measure using Latent

[4]For the sake of having each measure operate in the same direction, we transform the S-score such that the value shown is 1 - S. Thus for all three variables, 0 implies perfect symmetry of preferences.

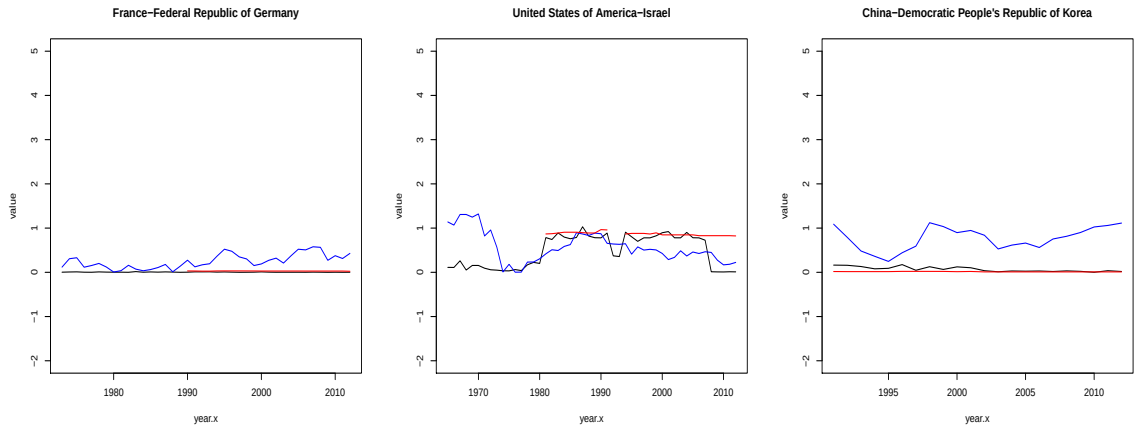

Figure 2: Dyadic relationships between 3 pairs of countries according to the different preference measures. Latent angle distance is in black, ideal point distance based on UN voting is in blue. S-score has been rescaled such that it works in the same direction as the other measures (1-S-score, so that the best score is 0), and is in red.

Angle distance has consistently low values – nearly 0 in the case of France and German, and China and North Korea, and low but less stable values for the US/Israel relationship. Alliance S-Scores correctly classify Franco-German and Sino-North Korean friendship, while the US/Israel relationship is characterized as being relatively neutral. The measure based on UN voting is most out of step in terms of characterizing these relationships, with notable divergence in the preferences of many of these pairs.

In figure 3 we look at the United States' relationship with three countries that are characterized by change and major events. The biggest difference between the measures is that alliance S-scores are the only measure that does not detect a marked improvement in the US/Russian relationship at the end of the Cold War. Both the UN ideal points and Latent Angle space find significant improvements followed by a drift toward enmity, whereas S-scores have a constant (though slightly improving) neutral relationship. For the US and Iraq no measure depicts more similar preferences following the US occupation, though only Latent Angle distance finds an increase in enmity in the run-up to the US's invasion. Finally, for the US and China, the S-score has a consis-

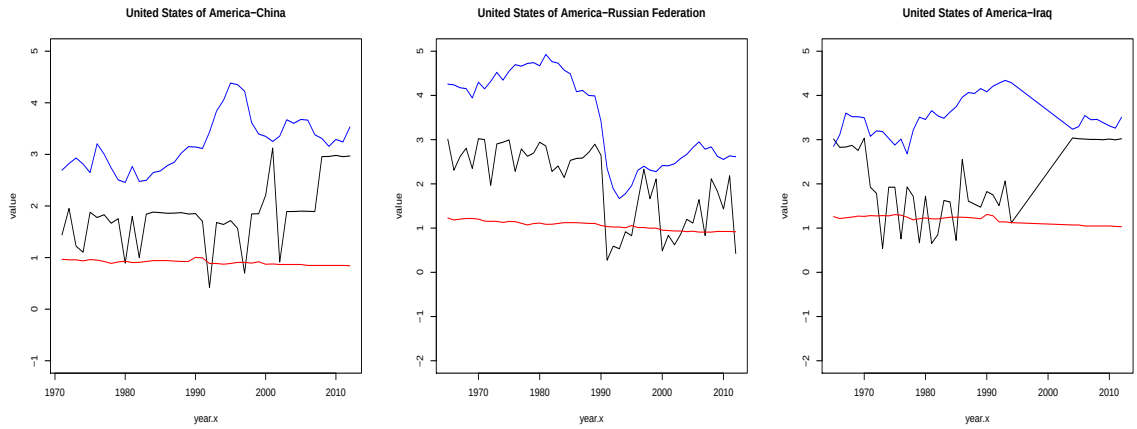

Figure 3: Dyadic relationships between 3 pairs of countries according to the different preference measures. Latent angle distance is in black, ideal point distance based on UN voting is in blue. S-score has been rescaled such that it works in the same direction as the other measures (1-S-score, so that the best score is 0), and is in red.

tent neutral relationship, while both UN ideal points and Latent Angles have more both more negative relationships, and more variance, with Latent Angles finding particularly large spikes at the beginning of the Bush and Obama administrations.

One big difference between these measures is that Latent Angle distance has a more full time series than the other two measures, this is particularly relevant when looking at the preferences of certain rogue states, as we do in figure 4. In both the case of Iran/Iraq and North Korea/South Korea there is no data for UN voting (and for Iran/Iraq missing data for S-scores). For the Korean relationship, Latent Angle distance better characterizes the relationship as enmity, while S-scores treat the two as having very similar preferences. Whereas for Iran and Iraq's relationship S-scores give a consistent close preferences (where data exists), while Latent Angles show more instability, and notable elevation (though not as large as we would expect) during their war.

As can be seen from these relationships, the measure of preferences based on Latent Angle distance is in many cases as plausible or more than incumbent measures of preferences. While the measure has more temporal instability than S-scores or UN

Figure 4: Dyadic relationships between 2 pairs of countries according to the different preference measures. Latent angle distance is in black, ideal point distance based on UN voting is in blue. S-score has been rescaled such that it works in the same direction as the other measures (1-S-score, so that the best score is 0), and is in red.

ideal points, in these 8 cases it often does better, and rarely worse than the other measures at conforming to our expectations of the relationships. Of course, this could just be a case of us picking particularly propitious cases. These 8 dyads show the plausibility of our measure of state preferences, but the real test is in the large-N analysis of conflict in the next section.

5. Model Competition

To evaluate different measures of state preferences, we compare them in a model of interstate disputes. Here we look at four non-nested models: a model using no measures of state preferences, one using an S-Score based on similarity in alliance portfolio (as in **??**, one using the ideal points determined by UN voting (as in **??**), a model using both UN ideal points and alliance S-scores, and finally, a model using our latent angle approach to combine data from UN voting and alliances. We evaluate the models on two criteria: whether state preferences have a consistent effect in the predicted direction, and how well each model does at predicting disputes on out of

sample data.

5.1. Data, Controls

In each of these models, we look at a logistic regression of Militarized Interstate Dispute participation on measures of state preferences and a vector of control variables. These control variables are most of the standard ones used most famously in O'Neal and Russett's work on the democratic peace (Oneal and Russett, 1997).[5] In particular, we include a binary measure of joint democracy (whether both states have Polity IV scores $geq7$), whether the states are contiguous, and the ratio of state capabilities as measured by the Correlates of War Project's Composite Index of National Capabilities (CINC). We also account for temporal interdependence using a peace year spline, as in **??**.

5.2. In sample explanation

As detailed in figure 5, each measure of state preferences performs as we would expect them to. Our measure of state preference using latent angle difference is highly significant and positive: states with more dissimilar preferences will have greater difference in their latent angles, and this is highly associated with a greater risk of conflict. Similarly, both incumbent measures of preferences pass this test. The measure using UN voting ideal points is positive and clearly distinct from 0, indicating that states with more distant ideal points, and thus more dissimilar preferences, are more likely to find themselves in conflict. Similarly, higher alliance S-scores are consistently associated with lower probabilities of conflict – so states with more similar preferences as measured using alliance portfolios are less likely to quarrel. These results hold when the measure of preferences is used in isolation, or in tandem.

[5]The exception is that our models ignore trade interdependency, as including that data drastically decreases the number of observations.

The models have one major difference in terms of the controls: in the model using latent angle distance, joint democracy is indistinguishable from 0. This is particularly interesting because of one of the major criticisms of democratic peace theory is that, for one reason or another, democracies have similar preferences, and this is what actually causes peace among democracies. Despite this dispute, most attempts to include preferences in the standard democratic peace regressions still found a consistent pacifying effect of democracy. (as do those models with UN voting and S-scores presented here). With our measure of preferences, however, democracy's effect is negligible.

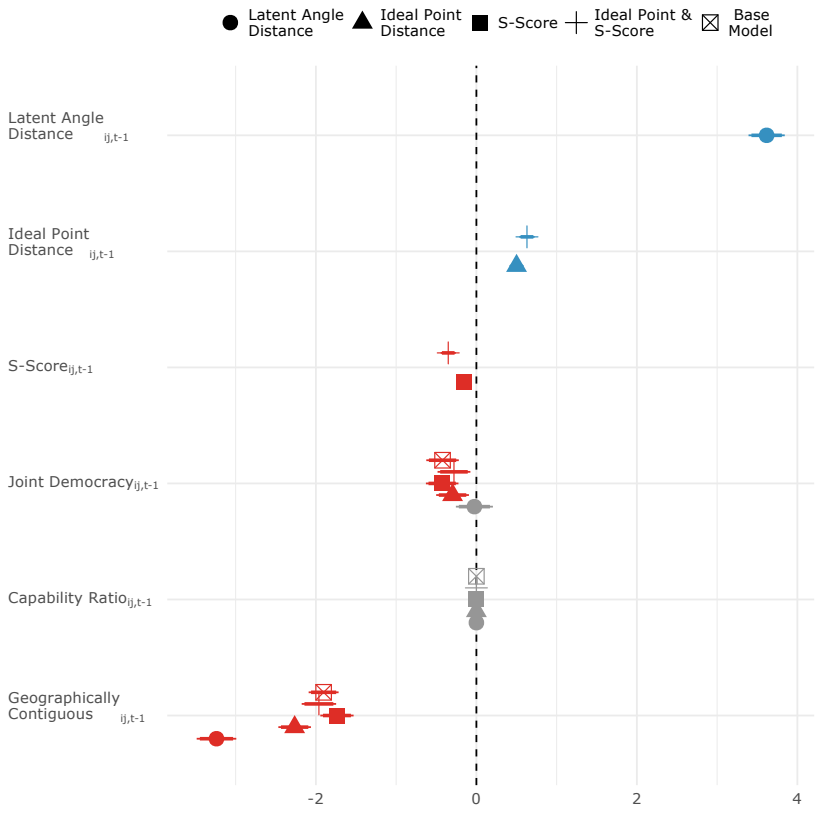

Figure 5: Parameter estimates from models with different measures of state preference. Point represents average estimate, line through the point represents the 95% confidence interval.

5.3. Out of sample prediction

Given these results, we can say that all of the measures of preferences behave as we would expect. To adjudicate which measures best capture state preferences, and what we should make of the differing effect of democracy, we turn to out of sample prediction. The way we do this is by partitioning our data into 30 different folds, and for each fold we generate a predicted value for each case the other 29 folds for each model. This leaves us with a set of predicted values generated entirely out of sample. We then compare the models performance using the area under both the Receiver Operator Characteristic Curve (AUC ROC) which examines the tradeoffs between true positives and false positives, and the area under the Precision Recall Curve (AUC PR) which looks at the tradeoffs between making only correct predictions, and predicting all the disputes which occurred. In general, the AUC ROC will disproportionately reward those models that predict 0 well, and we can interpret the AUC ROC as the likelihood a prediction is correct, the numeric value for the AUC PR has less of a clear interpretation, but models with a higher AUC PR do a better job of predicting when events actually occur.

As shown in figure 6, the model using latent angle distance decisively outperforms all the other models. While the AUC ROC is somewhat higher with the Latent Angle model, the real difference between the measures shines through in the AUC PR, where the model using this measure performs twice as well as the base model. In contrast, models using other measures of state preferences yield only minimal improvements in prediction over the base model. Thus we have real reason for confidence in both the usefulness of this measure of state preferences and renewed reason for skepticism in the effect of joint democracy once we control for state preferences.

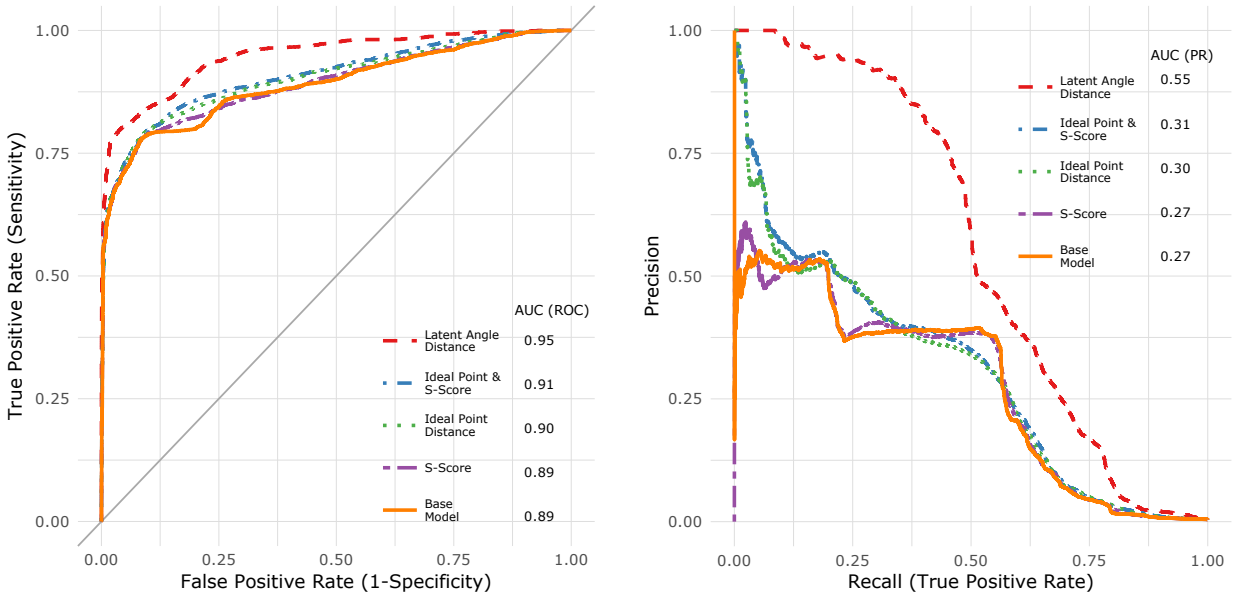

Figure 6: Assessments of out-of-sample predictive performance using ROC curves and PR curves. AUC statistics are provided as well for both curves.

6. Conclusion

YAY I WAS RIGHT! WE AM SMRT!

Bailey, Michael A., A. S., Voeten, E., 2015. Estimating dynamic state preferences from united nations voting data. Journal of Conflict Resolution 61 (2), 430–456.

Bennett, D. S., Rupert, M. C., June 2003. Comparing measures of political similarity: An empirical comparison of S versus τ_b in the study of international conflict. Journal of Conflict Resolution 47 (3), 367–393.

Braumoeller, B. F., 2008. Systemic politics and the origins of great power conflict. American Political Science Review 102 (01), 77–93.

Bueno de Mesquita, B., 1983. The War Trap. Yale University Press, New Haven, CT.

DeRouen, K., Heo, U., 2004. Reward, punishment or inducement? us economic and military aid, 1946âĂŞ1996. Defence and Peace Economics 15 (5), 453–470.

Farber, H., Gowa, J., 1995. Politics and peace. International Security 20 (2), 123–146.

Gallop, M., 2017. More dangerous than dyads: How a third party enables rationalist explanations for war. Journal of Theoretical Politics.

Gartzke, E., 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.

Gartzke, E., 2000. Preferences and the Democratic Peace. International Studies Quarterly 44, 191–212.

Gartzke, E., 2007. The capitalist peace. American Journal of Political Science 51 (1), 166–191.

Häge, F. M., 2011. Choice or circumstance? adjusting measures of foreign policy similarity for chance agreement. Political Analysis, 287–305.

Hoff, P. D., 2005. Bilinear mixed-effects models for dyadic data. Journal of the American Statistical Association 100 (4690), 286–295.

Kastner, S. L., 2007. When do conflicting political relations affect international trade? Journal of Conflict Resolution 51 (4), 664–688.

Kydd, A., 2003. Which Side Are You On? Bias, Credibility, and Mediation. American Journal of Political Science 47 (4), 597–611.

Oneal, J., Russett, B. M., 1999. Is the liberal peace just an artifact of cold war interests? assessing recent critiques. International Interactions 25 (3), 213–241.

Oneal, J. R., Russett, B. M., 1997. The classical liberals were right: Democracy, interdependence, and conflict, 1950-1985. International Studies Quarterly 41 (2), 267–293.

Signorino, C., Ritter, J., 1999. Tau-b or not tau-b. International Studies Quarterly 43 (1), 115–144.

Singer, J. D., Small, M., 1995. National Military Capabilities Data. Correlates of War Project, Ann Arbor, MI.

Stone, R. W., 2004. The political economy of IMF lending in Africa. American Political Science Review 98 (4), 577–591.

Wolford, S., 2014. Showing restraint, signaling resolve: coalitions, cooperation, and crisis bargaining. American Journal of Political Science 58 (1), 144–156.