

Appendix: Democracy and Alliance Treaty Depth

This appendix contains a series of robustness checks for the findings in the manuscript. In the first section, I examine results with an alternative measure of democracy when the alliance formed: the proportion of democracies in the alliance. Then I look at results from two alternative measures of treaty depth. In the third section, I use hurdle models to Last, I consider how uncertainty in the latent treaty depth measure affects my inferences.

1 Proportion of Democracies

In the paper, I use the average polity score of all alliance members when the treaty as the key independent variable. The proportion of democracies in the alliance is an alternative measure of how much heft democracies have in negotiations Chiba, Johnson and Leeds (2015). Rather than average the POLITY scores of alliance members, this variable is the share of alliance members with a POLITY score greater than 5, which is a common threshold for democracy. My argument anticipates that the proportion measure should lead to similar inferences about the connection between democracy and alliance treaty depth.

First, the descriptive statistics for the proportion of democracies across treaty depth and unconditional military support are as expected. Figure 1 shows the average proportion of democracies in four types of alliances. Deep, conditional alliances have the highest average proportion of democratic members, although it is not much higher than the shallow and conditional average.

Results from a generalized joint regression model of treaty depth and unconditional military support match the findings in the paper. In Figure 2, there is a no substantive association between

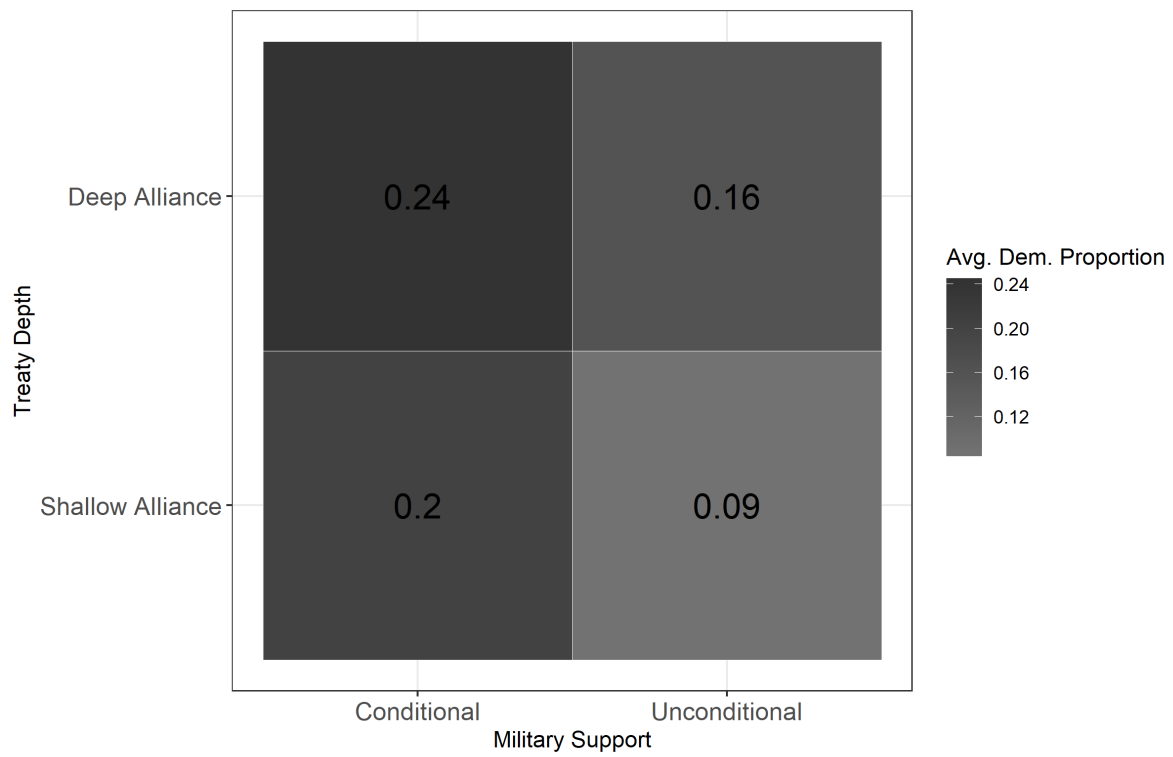


Figure 1: Average proportion of democratic members in four groups of alliances. Alliances are grouped based on treaty depth and unconditional military support.

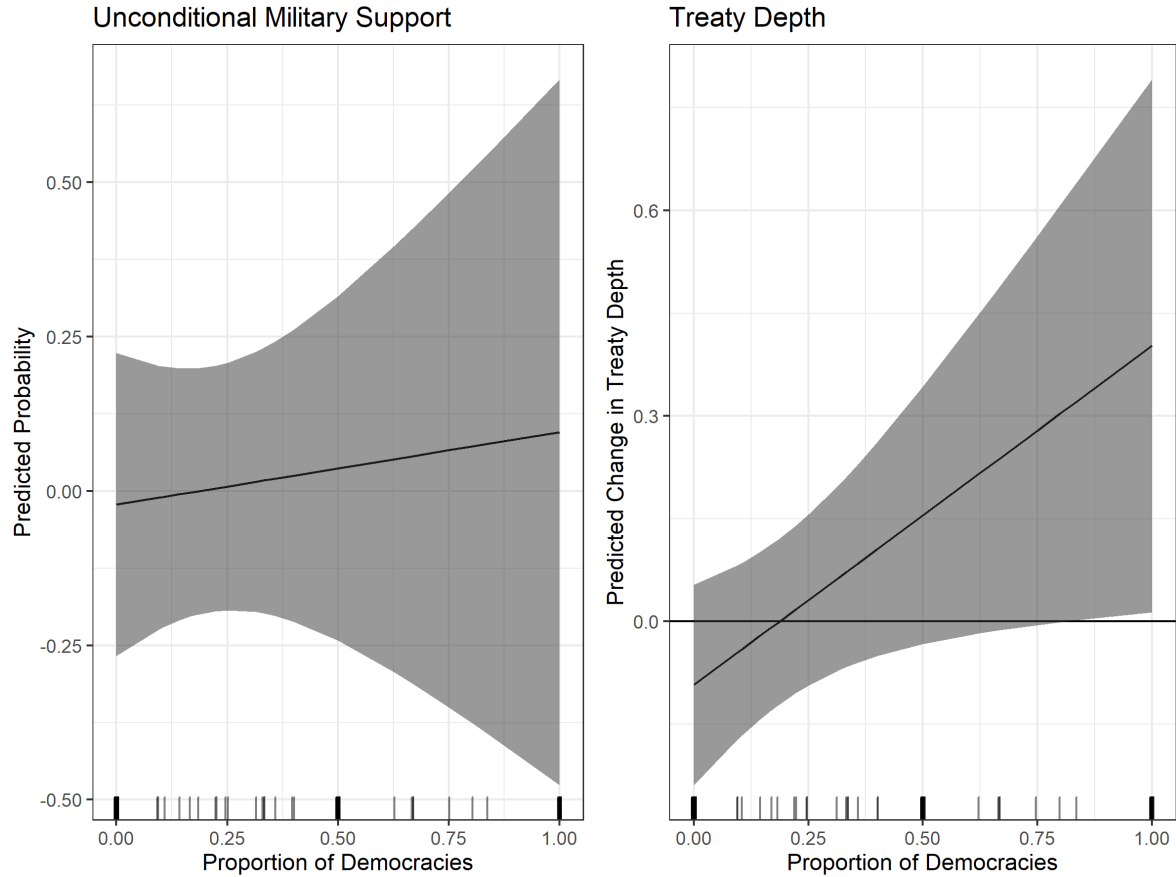


Figure 2: Predicted probabilities of unconditional military support and predicted changes in treaty depth by the the proportion of democratic alliance members. The line marks predicted values, and the shaded areas encapsulate the standard errors. The rug plot on the x-axis marks observed values of allied democracy. Predictions based on the smoothed terms from the joint generalized regression model.

the proportion of democracies in an alliance and the probability of unconditional military support. Alliances between democracies are more likely to have high treaty depth, however. Therefore, I find little evidence that democracies are more likely to form alliances with conditional obligations, but I do find a positive relationship between democratic alliance membership and treaty depth.

2 Alternative Measures of Treaty Depth

There are two measures of treaty depth that operationalize similar concepts. Leeds and Anac (2005) develop an ordinal measure of military institutionalization. Benson and Clinton (2016) use a latent variable model to model treaty depth. While my measure is a better fit for my purposes in this paper, I find somewhat similar results with these two models. The null finding about unconditional military support holds. The Leeds and Anac measure of military institutionalization produces similar results in a GRJM model. The Benson and Clinton measure is harder to compare, because it covers fewer alliances and has more outliers that may impact inferences.

Figure 3 plots predictions from the average democracy smooth term for the two alternative measures. For the military institutionalization measure, I could not find an ordinal model in GJRM. Therefore, I created a dummy variable which is equal to one if the alliance had an institutionalization score of 1 or 2. I then modeled the military institutionalization dummy with a probit distribution. The results of this analysis match the treaty depth hypothesis, as the probability of military institutionalization is increasing in average democracy.

Findings about the association between treaty depth and Benson and Clinton's measure are more mixed. The relationship is highly non-linear, and high average democracy is not positively correlated with Benson and Clinton's measure of depth.

3 Adjusting for Alliance Formation

Observed alliances are the result of negotiations between potential members. Some alliance negotiations fail, however, so the set of observed alliances is not a random sample (Poast, 2019). In this section, I account for alliance formation, and show that a hurdle model of alliance treaty depth and conditions on military support produces similar inferences.

My research design for this robustness check follows existing procedures for dealing with selection into alliances and treaty design. To construct the dataset, I followed the process of Poast

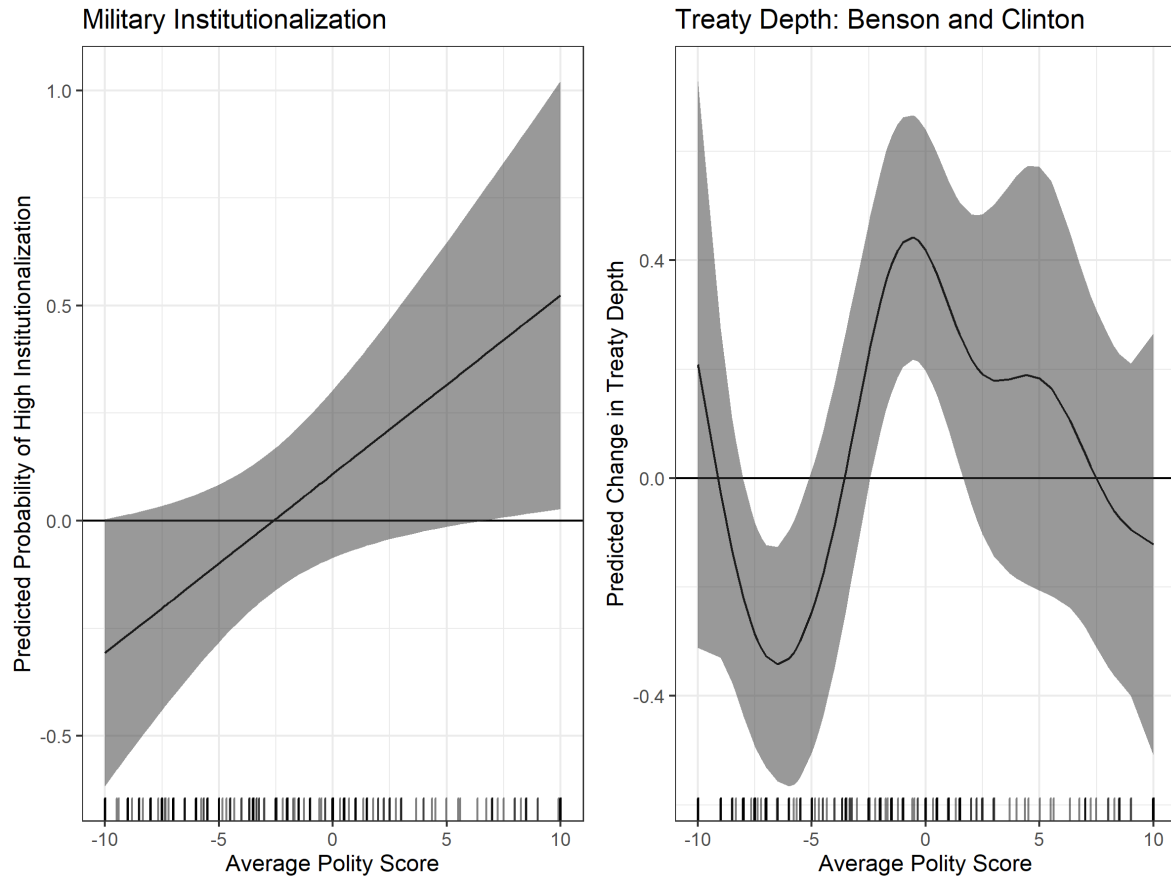


Figure 3: Predicted probabilities of unconditional military support and predicted changes in treaty depth by the the proportion of democratic alliance members. The line marks predicted values, and the shaded areas encapsulate the standard errors. The rug plot on the x-axis marks observed values of allied democracy. Predictions based on the smoothed terms from the joint generalized regression model.

(2010) for dealing with k-adic data, because alliances can have more than two members. First I constructed a random sample of groups of states that could have formed an alliance in each year, but did not (Fordham and Poast, 2014). I then took a stratified sample of the non-allied k-ads to include five times as many non-allied observations as alliance observations for each observed value of alliance size. For example, there are 215 bilateral alliances, so I included 1075 non-allied bilateral groups. There is only one alliance with 34 members, so I include five non-allied k-ads with 34 members. Finally, I summarized the characteristics of these non-allied k-ads and merged them with the observed alliance data.

To model this process, I built on the research design of Chiba, Johnson and Leeds (2015). They used a hurdle model to assess whether after accounting for non-random selection into alliances, democracies were more likely to offer conditional obligations. Hurdle models have two stages or parts. A first stage model predicts which observations clear the hurdle to a second stage where the outcome is observed. Unlike a sample selection model, the second stage in a hurdle model is logically undefined, which is how alliance treaty design works.¹ After overcoming the barriers to alliance formation, states then specify treaty content.

Unlike Chiba, Johnson and Leeds (2015) Therefore, I jointly estimate two hurdle models- one for the depth and the other for unconditional military support. To have zero values for the outcome at the hurdle stage, I adjusted the outcomes in these models. First, I rescaled latent depth by adding one, which shifted the distribution onto uniformly positive values. I then applied a gamma hurdle model to this shifted depth outcome and used the same covariates as the main model in the manuscript to predict depth. For unconditional military support, I replaced the dummy indicator in the manuscript with an ordinal indicator. This ordinal variable captures the relative strength of the military support promise in the alliance. ATOP alliances can be conditional on four things: adversaries, locations, a particular conflict, the number of opponents, a specific demand or non-provocation in the case of defense treaties. The maximum number of observed conditions in an

¹Hurdle models also do not require an exclusion restriction.

alliance is four, which is the most conditional and limited alliance in the data. I therefore invert the number of conditions and assign unconditional alliances a value of four. Essentially, the measure then captures the number of circumstances where an alliance could have applied conditions but did not. To avoid including zeros in the measure, I add one to it, because promising military support adds some baseline strength to any alliance (Morrow, 2000). Thus, the final measure of military support commitment strength ranges between 1 and 5 for observed alliances, but is zero for non-alliances. In both models, I use average democracy, mean threat, asymmetric capability and wartime variables in the hurdle equation.

I then estimate the two hurdle models simultaneously using the BRMS package for R, which employs STAN for fully Bayesian inference (Bürkner, 2017). Unlike the GJRM models in the manuscripts, there is no correlation between the residuals of the separate hurdle models, but Instead, I connect the two models with a correlated set of year varying intercepts. The complexity of this joint hurdle model and limited sample size in the second stage means results should be treated with caution.²

Even after accounting for non-random alliance formation with the joint hurdle model, I find similar results to the appendix. Figure 2 plots the results of the joint hurdle model. As in the sample with only observed alliances, greater democracy among alliance members has no clear relationship with conditionality, but increases treaty depth.

4 Uncertainty in Latent Treaty Depth

In a separate model, I consider how measurement uncertainty shapes inferences about the connection between non-major power alliances and treaty depth. The latent measure of treaty depth has some uncertainty. This is a reasonable approximation of alliance politics, because alliance treaty depth is not observed with certainty. There are perceptible differences in treaty depth, espe-

²I do find similar results with separate hurdle models for depth and conditions on military support.

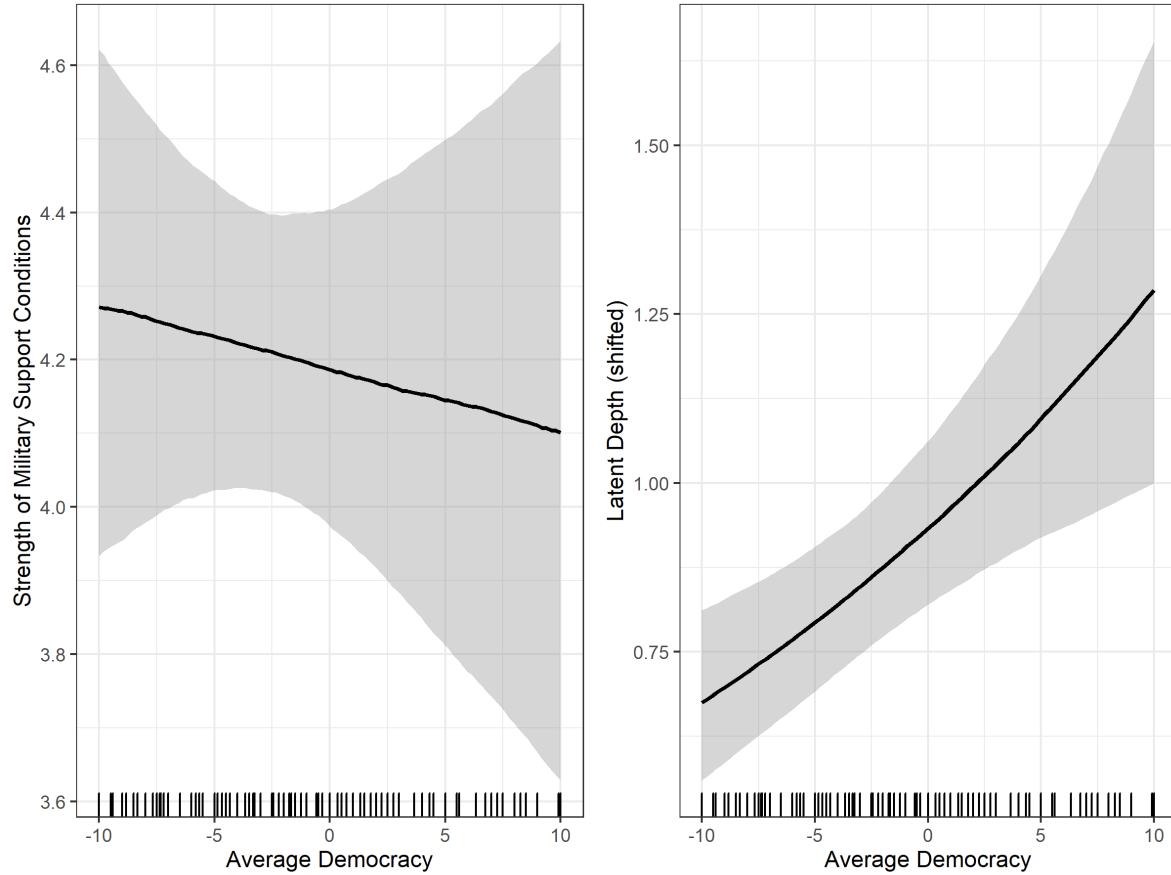


Figure 4: Predicted strength and predicted changes in treaty depth by the the average POLITY score alliance members when the treaty formed based on a joint hurdle model of alliances. The line marks predicted values, and the shaded areas encapsulate the 90% credible interval. The rug plot on the x-axis marks observed values of allied democracy. Predictions holding all other variables.

cially once states add substantial depth to the treaty. Even so, the results from the analysis of mean treaty depth may overstate the effect of non-major power membership.

To incorporate uncertainty over treaty depth, I fit a modification of the joint model. First, I created 1,000 datasets, one for each draw of the posterior distribution of the latent measure. Then I fit the model of mean treaty depth to 500 randomly sampled datasets from those 1,000 to facilitate computation. For models of depth with uncertainty, I use BRMS (Bürkner, 2017). BRMS is an interface to STAN, a probabilistic programming language for Bayesian estimation (Carpenter et al., 2016). Joint Bayesian estimation has the flexibility to incorporate the probit and beta models and can be easily extended to account for uncertainty in the depth measure, but it does not allow correlated errors. Fitting the model sequentially to each dataset produces 500 separate models, which I combine into a single model by aggregating the posterior draws in a single posterior that accounts for uncertainty in the treaty depth measure.³ This approach is analogous to common techniques for analyzing missing data, where multiple imputation generates uncertainty about the missing values (Hollenbach et al., 2018). After multiple imputation, researchers fit a separate model to each imputed dataset and then combine the results.

After accounting for uncertainty over treaty depth, I find a similar pattern. Allied democracy increases treaty depth, but decreases the probability of unconditional military support.

References

- Benson, Brett V and Joshua D Clinton. 2016. "Assessing the Variation of Formal Military Alliances." *Journal of Conflict Resolution* 60(5):866–898.
- Bürkner, Paul-Christian. 2017. "brms: An R package for Bayesian multilevel models using Stan." *Journal of Statistical Software* 80(1):1–28.
- Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt,

³Standard convergence diagnostics indicate convergence in all 500 models. Diagnostics like \hat{r} are less useful for the full posterior, because some of the chains in the submodels do not overlap.

- Michael A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. “Stan: A probabilistic programming language.” *Journal of Statistical Software* 20:1–37.
- Chiba, Daina, Jesse C Johnson and Brett Ashley Leeds. 2015. “Careful Commitments: Democratic States and Alliance Design.” *The Journal of Politics* 77(4):968–982.
- Fordham, Benjamin and Paul Poast. 2014. “All Alliances Are Multilateral Rethinking Alliance Formation.” *Journal of Conflict Resolution* 58(27):1–26.
- Hollenbach, Florian, Iavor Bojinov, Shahryar Minhas, Nils W. Metternich, Michael D. Ward and Alexander Volfovsky. 2018. “Principled Imputation Made Simple: Multiple Imputation Using Gaussian Copulas.” *Sociological Methods & Research* pp. 1–29.
- Leeds, Brett Ashley and Sezi Anac. 2005. “Alliance Institutionalization and Alliance Performance.” *International Interactions* 31(3):183–202.
- Morrow, James D. 2000. “Alliances: Why Write Them Down?” *Annual Review of Political Science* 3:63–83.
- Poast, Paul. 2010. “(Mis) Using dyadic data to analyze multilateral events.” *Political Analysis* 18(4):403–425.
- Poast, Paul. 2019. *Arguing about Alliances: The Art of Agreement in Military-Pact Negotiations*. Cornell University Press.