Appendix to Paper: Arms, Alliances and Alliance Treaty Design

Joshua Alley

June 19, 2019

1 Priors

Table 1 summarizes the prior distributions in the multilevel model. All these priors are weakly informative relative to the scale of the data. ν is the degrees of freedom for the t-distribution, and the gamma prior is the recommended default prior for STAN.

2 Hamiltonian Monte Carlo Diagnostics

There were no divergent iterations running 4 chains for 2,000 iterations in either sample. The $\hat{R}$ is less than 1.1 for all parameters in both samples.

$$
\begin{gathered}
p(\alpha) \sim N(0,1) \\
p(\sigma) \sim \text{half-}N(0,1) \\
p(\alpha^{yr}) \sim N(0,\sigma^{yr}) \\
p(\sigma^{yr}) \sim N(0,1) \\
p(\alpha^{st}) \sim N(0,\sigma^{st}) \\
p(\sigma^{st}) \sim \text{half-}N(0,1) \\
p(\sigma^{all}) \sim \text{half-}N(0,1) \\
p(\beta) \sim N(0,1) \\
p(\gamma) \sim N(0,1) \\
p(\nu) \sim gamma(2,0.1)
\end{gathered}
$$

Table 1: Summary of Priors in Multilevel Model

Trace plots in Figure 1 and Figure 2 indicate good mixing of the chains for the alliance-level parameters.

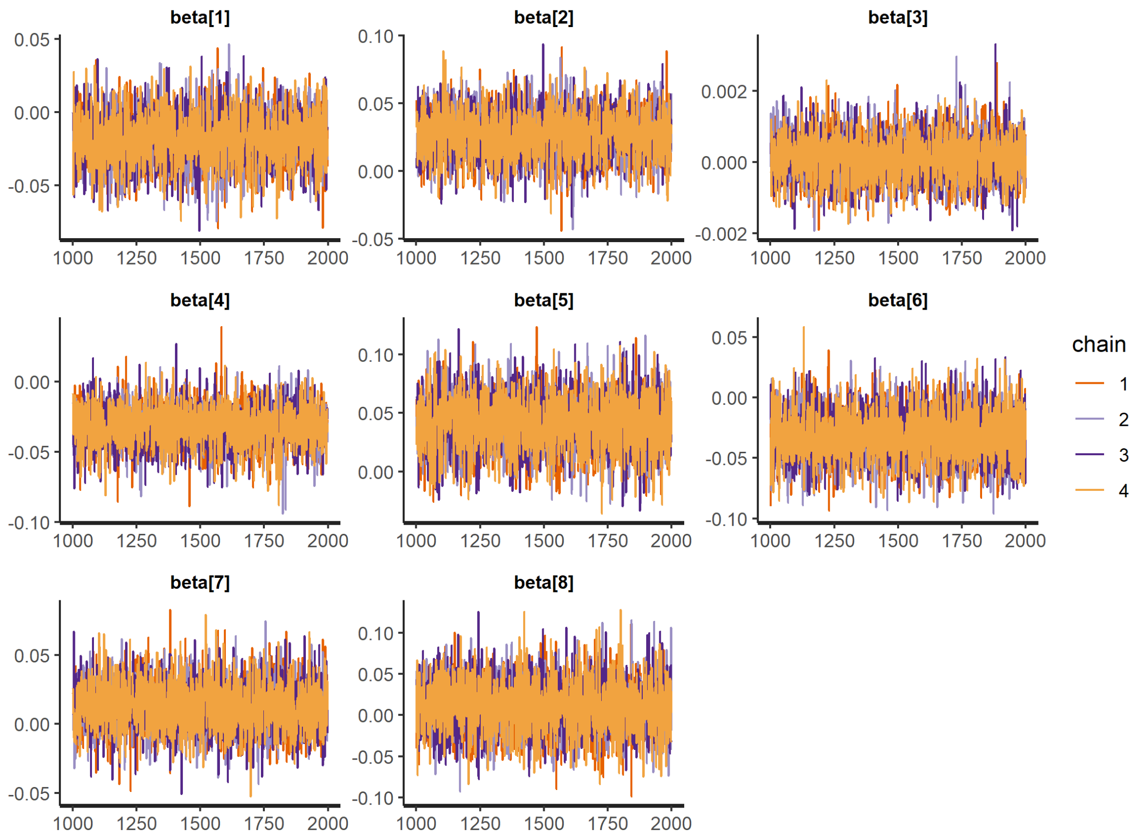

Figure 1: Traceplot of alliance level parameters in the non-major power sample.

3 Fake Data Simulation Check

Another way to ensure the hierarchical model generates reasonable estimates is simulating fake data and seeing if the model can recover known parameters. Fake-data simulation is an essential aspect of model checking. This section summarizes results from fitting the multilevel model to fake data.

I simulated a dataset of 2000 t-distributed observations with 50 states observed for 200 years

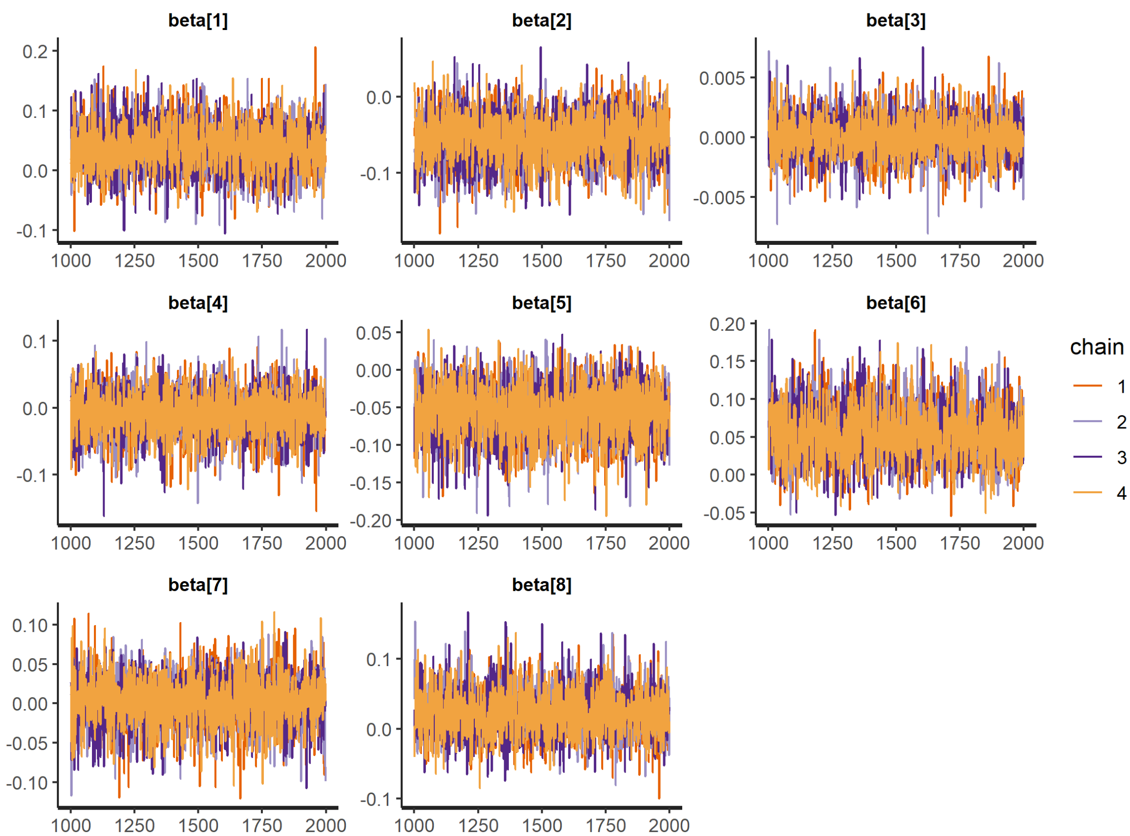

Figure 2: Traceplot of alliance level parameters in the major power sample.

and 100 alliances. The outcome has a slightly different scale than the military spending growth variable in the paper, so coefficient values will be different. I then simulated two state and alliance level variables and a sparse matrix of state membership in alliances. Last, I ran the model without evaluating the likelihood, generating a posterior prediction of the outcome based on the fake data.

To check whether the model could recover known parameters, I took the 12th draw of the posterior distribution. This draw included a simulated outcome for each observation and the associated coefficients. I then fit the multilevel model to this draw of the simulated posterior, and checked whether the credible intervals contained the corresponding parameter values. Credible interval coverage shows whether the model captures the parameter.

The model recovers known parameters with a high degree of accuracy. As shown by Figure 3, the two credible intervals of the alliance-level regression include the known values. Credible interval coverage for the hyperparameters and γ parameters is also acceptable.

The 100 λ parameters are harder to plot, so I offer a descriptive summary here. Among the λ parameters, 93 of 100 intervals contain the known λ value. Given the large number of parameters and smaller sample, this is acceptable accuracy. Even the 7 inaccurate confidence intervals were quite close— all missed the true parameter by .015 or less.[1]

In summary, convergence diagnostics and fake data fitting both suggest that the multilevel model is working well. No convergence diagnostics indicate problems exploring the posterior. Just as importantly, the model can recover known parameters from fake data. The next section provides more detail on results from the major and non-major power samples.

4 Posterior Intervals

I do not present tabular summaries of all the alliance-level parameters in the manuscript for parsimony. The next two tables summarize the posteriors of the alliance-level parameters. Using

[1]Fine margins around these intervals implies that accuracy in coverage of λ are very sensitive to simulation variance.

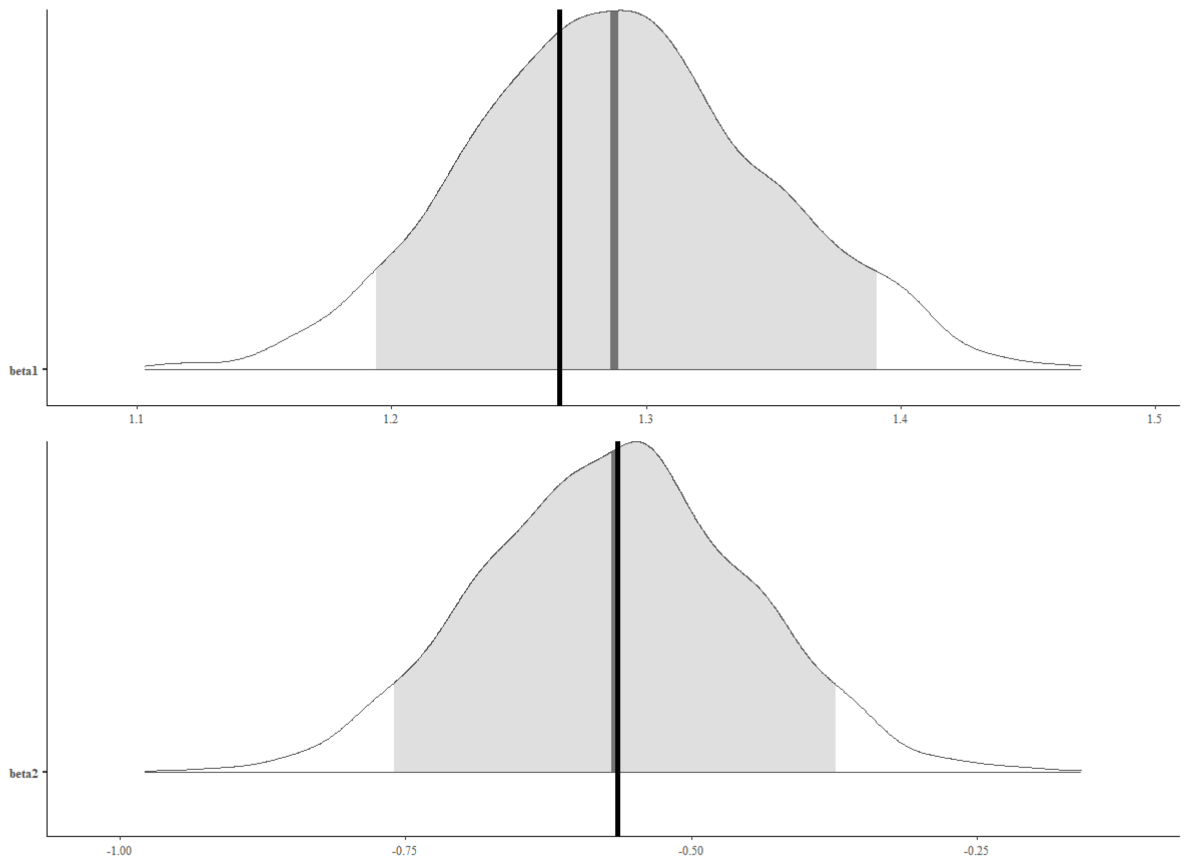

Figure 3: Posterior distributions of β parameters from fitting multilevel model to fake data. The black vertical line marks the known parameter value, and the grey area is the 90% credible interval.

a 90% credible intervals implies there is a 90% chance the coefficient is between the 5% and 95% values. Because Hypotheses 1 and 2 are directional, I report the positive and negative posterior probabilities in the manuscript.

4.1 Major Powers

Table 2 summarizes the 90% credible intervals for the alliance parameters in the major power sample, as well as the number of effective samples and $\hat{R}$ for each marginal posterior.[2] σ Alliances is the variance hyperparameter for the λ estimates. The $\hat{R}$ statistics are all close to one, indicating convergence. The number of effective samples is adequate for most parameters.

	mean	S.D.	5%	95%	n_eff	$\hat{R}$
Constant	0.038	0.038	-0.025	0.102	3380.954	1.000
Latent Str.	-0.054	0.031	-0.107	-0.005	3278.923	1.000
Number Members	0.000	0.002	-0.003	0.003	4000.000	0.999
Democratic Membership	-0.009	0.033	-0.065	0.042	4000.000	1.000
Wartime	-0.057	0.035	-0.115	-0.001	4000.000	1.001
Asymmetric	0.053	0.035	0.001	0.115	2218.509	1.000
US Member	0.002	0.031	-0.051	0.051	4000.000	1.000
USSR Member	0.023	0.033	-0.028	0.079	4000.000	1.000
σ Alliances	0.066	0.029	0.019	0.117	599.081	1.007

Table 2: 90% Credible intervals for major power alliance-level parameters

4.2 Non-major Powers

Table 3 summarize the 90% credible intervals for the alliance-level regression parameters in the non-major power sample. The $\hat{R}$ statistics are all close to one, indicating convergence. Again, the number of effective samples is adequate for all parameters.

[2] I report 90% credible intervals because 95% interval estimates can be unstable.

	mean	sd	5%	95%	n_eff	$\hat{R}$
Constant	-0.018	0.018	-0.047	0.012	2211.374	1.000
Latent Scope.	0.026	0.017	-0.002	0.054	2191.382	1.000
Number Members	0.000	0.001	-0.001	0.001	4000.000	1.000
Democratic Membership	-0.031	0.015	-0.056	-0.009	3213.621	1.000
Wartime	0.041	0.023	0.002	0.078	4000.000	1.000
Asymmetric	-0.031	0.021	-0.065	0.003	4000.000	0.999
US Member	0.013	0.018	-0.016	0.042	2895.419	1.000
USSR Member	0.011	0.031	-0.041	0.062	4000.000	1.000
σ Alliances	0.014	0.009	0.002	0.030	1254.268	1.001

Table 3: 90% Credible intervals non-major power alliance-level parameters

5 Varying Slopes Model

Splitting the sample is a simple way to capture differences between major and non-major powers in the alliance level regression. Estimating varying slopes across major and non-major power observations is another way to express the same distinction. I do not report the varying slopes specification in the paper because it is more complicated and computationally intensive.[3]

Letting the β and λ parameters of the multilevel model vary across major and non-major powers is not a trivial problem. The challenge is that some alliances only have major power members, and others only have non-major power members. As a result, a $2 \times a$ matrix of λ parameters where there are two groups of state-year observations and a alliances includes parameters with no underlying variation. Such a model cannot be estimated.

To allow the β parameters to vary across groups, I constructed two separate local models, one for major powers, the other for non-major powers. The two models are connected through shared parameters, so they share information. This is just an alternative way of expressing a multilevel model Gelman and Hill (2007, pg. 263).

There are three levels to this varying slopes model— first is major/non-major power status, the second is the alliance level, and the third is state-year observations. My theory does not anticipate differences in state-year factors such as conflict participation, so the local regressions share γ

[3]The varying slopes model takes just over 4 days to run on a desktop with 2 cores and 2GB of RAM.

parameters as well as state and year varying intercepts. The β coefficients in the alliance level regression vary by major or non-major power status.

Formally, I express the varying slopes model as follows. Within each of the j groups of state capability, for i in $1...n_j$:

$$y_i \sim student_t(\nu_j, \alpha_j + \alpha^{st} + \alpha^{yr} + \mathbf{W}_i\gamma + \mathbf{Z}_{ji}\lambda_j, \sigma_j) \tag{1}$$

Where the alliance participation coefficients lambda are equal to:

$$\lambda_j \sim N(\theta_j, \sigma_j^{all}) \tag{2}$$

and

$$\theta_j = \alpha_j^{all} + \mathbf{X}\beta_j \tag{3}$$

I give β_j a multivariate normal prior with prior scale τ:

$$\beta_j \sim MVN(\mu_{\beta_j}, \Sigma_\beta) \tag{4}$$

Major powers and non-major powers have different β estimates, which come from a common distribution. Each set of λ_j parameters has a different number of alliances, so they have separate normal priors. Because the number of alliances varies, I define separate Z matrices for major and non-major powers.

Again, the shared parameters γ, α^{st} and α^{yr} shares information across groups. I implement this model with non-centered priors for the varying intercepts and βs, as well as sparse matrix representations of both alliance membership matrices. Results are based on four chains with 2,000 total samples and 1,000 warmup iterations. The inferences from this varying slopes model are comparable to those from the split samples.

5.1 Varying Slopes Results

In the varying slopes model, the preponderance of evidence matches the predictions of Hypotheses 1 and 2. Figure 4 plots the full posterior distributions of the latent scope coefficients among major and non-major powers. There is a 94% chance that increasing treaty scope lowers the impact of alliance participation on military spending for major powers. Meanwhile, there is a 93% chance that increasing treaty scope raises the impact of alliance participation on military spending for non-major powers. Approximately 8% of the posterior mass overlaps between the two distributions.

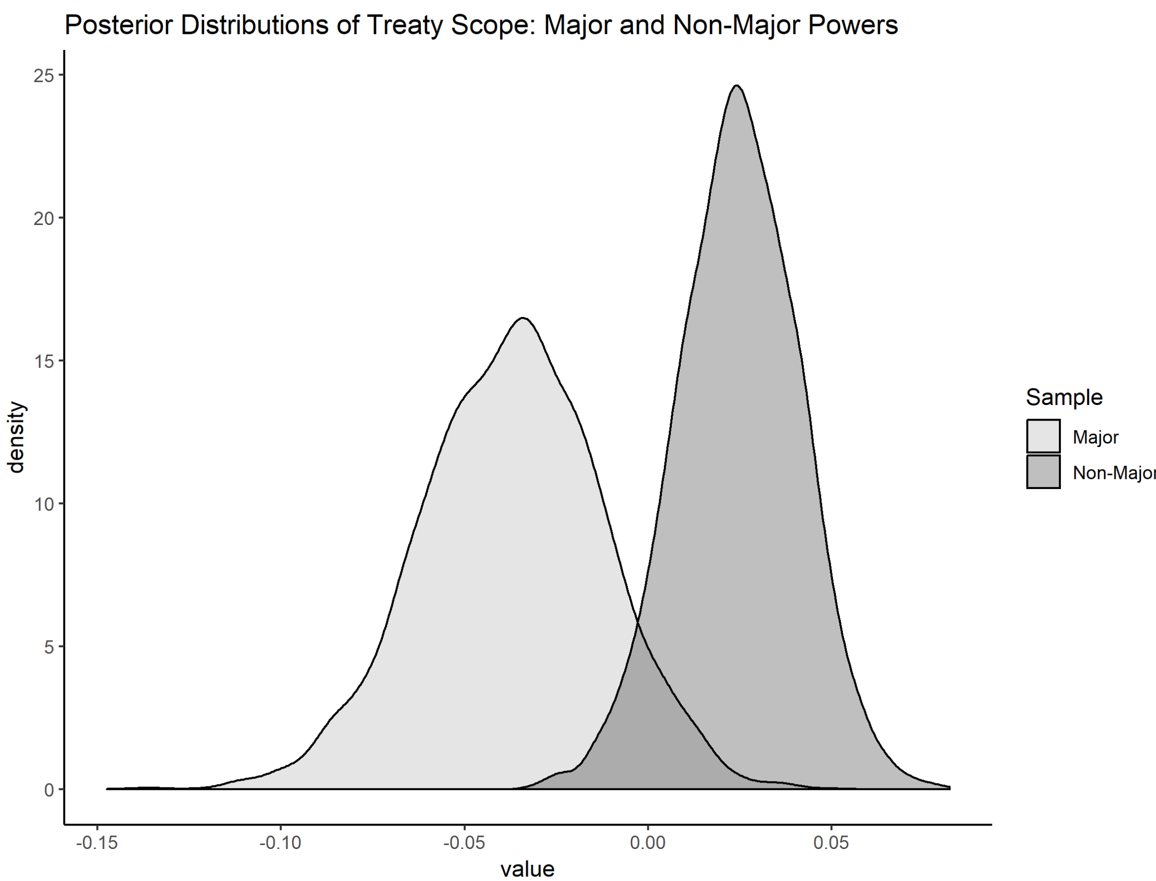

Figure 4: Posterior Distributions of Latent Scope in for major and non-major powers in a varying slopes model of alliance participation and military spending from 1816 to 2007. 94% of the major power posterior mass is negative. 93% of the non-major power posterior mass is positive.

Because the varying slopes model shares information across major and non-major power observations, there is less uncertainty in the major power estimates in this model. Varying slopes also shrinks the posterior mean from -0.06 to -0.04, which indicates slightly smaller substantive effect of treaty scope. Even so, the amount of negative posterior mass is similar for major powers in the varying slopes and split samples.

Table 4 summarizes the 90% credible intervals for all the alliance level parameters. Again, the number of effective samples indicates little autocorrelation in the chains and $\hat{R}$ statistics suggest convergence.

	mean	sd	5%	95%	n_eff	Rhat
Constant: Non-Major	-0.023	0.026	-0.066	0.022	4000.000	1.000
Latent Scope: Non-Major	0.025	0.017	-0.003	0.051	4000.000	1.000
Number Members: Non-Major	0.000	0.001	-0.001	0.002	4000.000	1.000
FP Similarity: Non-Major	-0.011	0.025	-0.053	0.029	4000.000	0.999
Democratic Membership: Non-Major	-0.002	0.001	-0.005	-0.000	4000.000	0.999
Wartime: Non-Major	0.039	0.024	-0.001	0.078	4000.000	1.000
Asymmetric: Non-Major	-0.030	0.022	-0.066	0.007	4000.000	1.000
US Member: Non-Major	0.010	0.018	-0.019	0.040	4000.000	1.000
USSR Member: Non-Major	0.016	0.040	-0.051	0.083	4000.000	1.000
Constant: Major	0.068	0.038	0.005	0.131	4000.000	1.002
Latent Scope: Major	-0.037	0.025	-0.078	0.003	4000.000	1.000
Number Members: Major	-0.000	0.001	-0.002	0.002	4000.000	1.000
FP Similarity: Major.1	-0.059	0.034	-0.116	-0.004	4000.000	1.001
Democratic Membership: Major	-0.001	0.002	-0.004	0.002	4000.000	1.000
Wartime: Major	-0.043	0.028	-0.091	0.003	4000.000	1.000
Asymmetric: Major	0.022	0.023	-0.015	0.061	4000.000	1.001
US Member: Major	-0.009	0.027	-0.056	0.035	4000.000	1.001
USSR Member: Major	0.001	0.023	-0.039	0.039	4000.000	0.999

Table 4: 90% Credible intervals of alliance-level parameters in varying slopes model.

References

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Vol. 1 Cambridge University Press New York, NY, USA.