Appendix: Reassessing the Public Goods Theory of Alliances

Joshua Alley*

May 15, 2019

This document contains supporting materials for the tests of Hypothesis 1 and 2. The first section includes material for the test of Hypothesis 1. The second section describes the priors, convergence diagnostics and results from an alternative sample for the multilevel model I used to test Hypothesis 2.

1 Other Estimates of Panel Data Interaction

1.1 Alternative Estimators

Robust regression is appropriate for residuals from the growth variable. Several observations of states during war see gigantic increases in spending— the largest value is 140, relative to a median of .063. This generates extremely heavy-tailed residuals, so OLS is inefficient.

Even after applying the inverse hyperbolic sine (IHS) transformation, the residuals in Figure 1 deviate strongly from normality. I use the IHS transformation because it accomodates positive, negative and zero values. Figure 1 shows the residuals from models 2 and 3 of Table 1, which reports some robustness checks.

Although OLS is not the best estimator for the spending growth data, it produces similar inferences. Table 1 summarizes estimates from OLS models with and without transforming growth in

*Graduate Student, Department of Political Science, Texas A&M University.

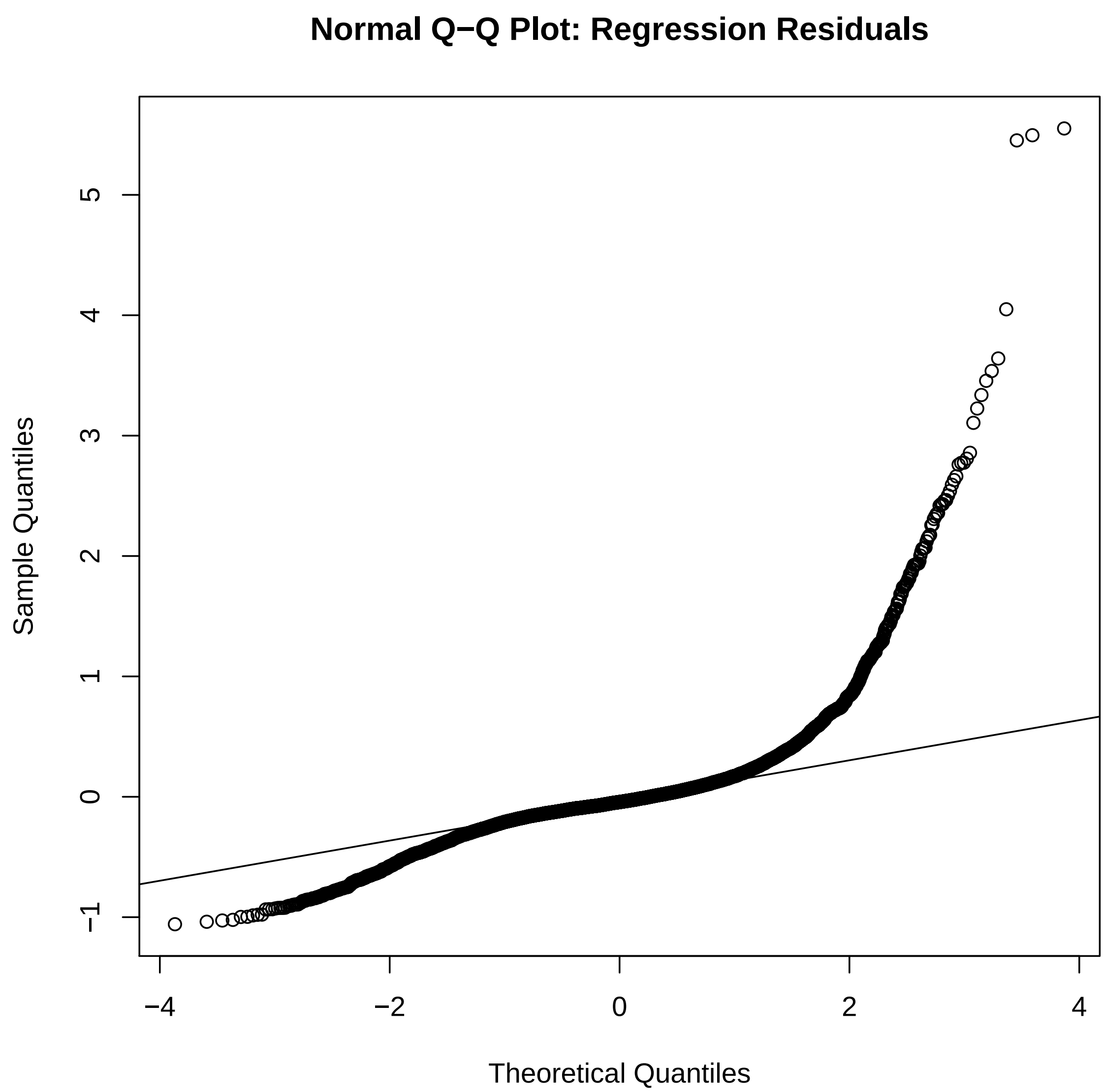

Figure 1: Plot of residuals against normal quantiles. Deviations from the straight line are deviations from the normal distribution.

spending as well as robust regression with transformed spending.

The OLS model is a poor fit for the dependent variable. The R^2 is miniscule, and the standard error of the residual is much larger than the robust regression. Transforming the DV improves the performance of OLS, but leads to similar inferences. Despite the lackluster fit, I present these results to show that OLS generates similar inferences.

Rather than interpret the coefficients, I plot the marginal effect of allied spending across the range of GDP for these three models. Figure 2 compares these results with the marginal effects plot from the robust regression in the paper. Three of the four estimation strategies produce a positive marginal effect, along with broad enough confidence intervals to suggest a conditional relationship with GDP is unlikely.[1] OLS without a transformed DV finds no positive impact of changing allied capability on growth in spending, but this result should be taken with great caution due to poor model fit. Regardless, none of the results matches the expectation of a negative effect which increases in state size as articulated by Hypothesis 1.

1.2 Continuous Modifying Variable

Hainmueller, Mummolo and Xu (2019) show linearity assumptions and a lack of support in the range of the modifying variable can generate misleading inferences in interactive models. They suggest estimating associations with binning and kernel estimators to check for non-linearity and adequate support. Figure 3 plots the results of the binning estimator.

The binning estimator suggests that the linearity assumptions of the robust regression are acceptable. There is still no evidence state size modifies the impact of allied spending on growth in military spending, however. The marginal effect of allied spending cannot be distinguished from zero across the range of GDP.

The kernel estimator also generates little evidence of a conditional relationship. As Figure 4

[1]The robust regression results are so similar because even after transforming spending growth, the robust estimator heavily down-weights the most unusual observations.

	Dependent variable:		
	Spending Growth	IHS(Spending Growth)	
	OLS	*OLS*	*Robust Reg.*
	(1)	(2)	(3)
Change Allied Spending	−0.235	−0.099	−0.026
	(0.648)	(0.087)	(0.047)
ln(GDP)	−0.038**	−0.008***	0.0003
	(0.017)	(0.002)	(0.001)
Change Allied Spending × ln(GDP)	0.011	0.005	0.002
	(0.027)	(0.004)	(0.002)
Average Alliance Size	−0.003	−0.0001	0.0003*
	(0.002)	(0.0003)	(0.0002)
Average Alliance Democracy	−0.001	−0.001	−0.0003
	(0.008)	(0.001)	(0.001)
International War	0.476***	0.228***	0.096***
	(0.141)	(0.019)	(0.010)
Civil War Participant	0.206**	0.025*	0.001
	(0.104)	(0.014)	(0.008)
Polity	0.001	0.0001	−0.0001
	(0.005)	(0.001)	(0.0004)
External Threat	0.107	0.080***	0.042***
	(0.152)	(0.021)	(0.011)
Cold War	0.032	0.034***	0.047***
	(0.058)	(0.008)	(0.004)
Constant	1.074***	0.259***	0.026
	(0.413)	(0.056)	(0.030)
Observations	9,139	9,139	9,139
R^2	0.003	0.025	
Residual Std. Error (df = 9128)	2.649	0.358	0.157
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1: Robustness checks of conditional relationship between allied spending, GDP and growth in military spending.

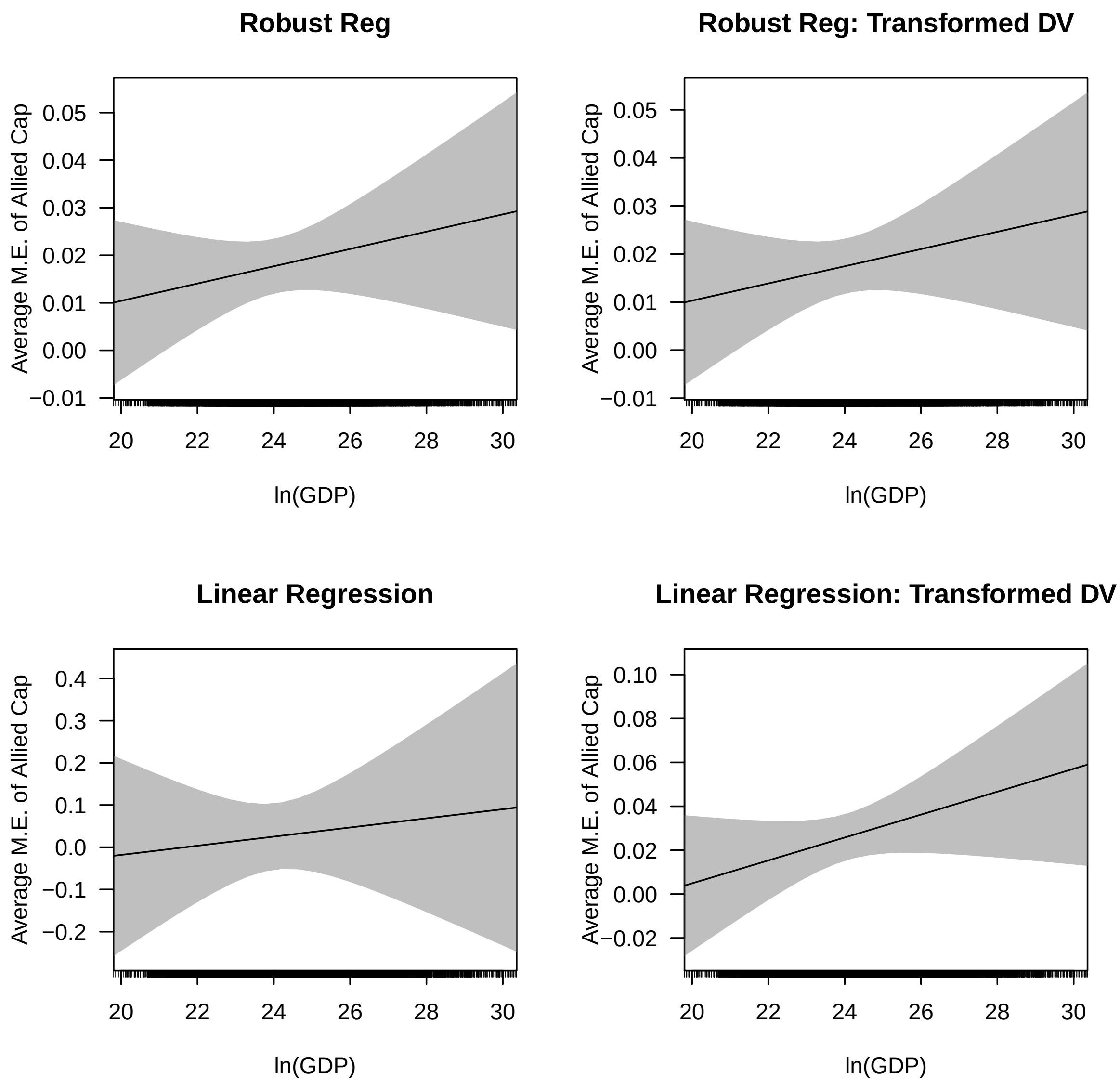

Figure 2: Comparison of marginal effect of changing allied spending on growth in military spending across the range of GDP. Each plot corresponds to an estimation strategy.

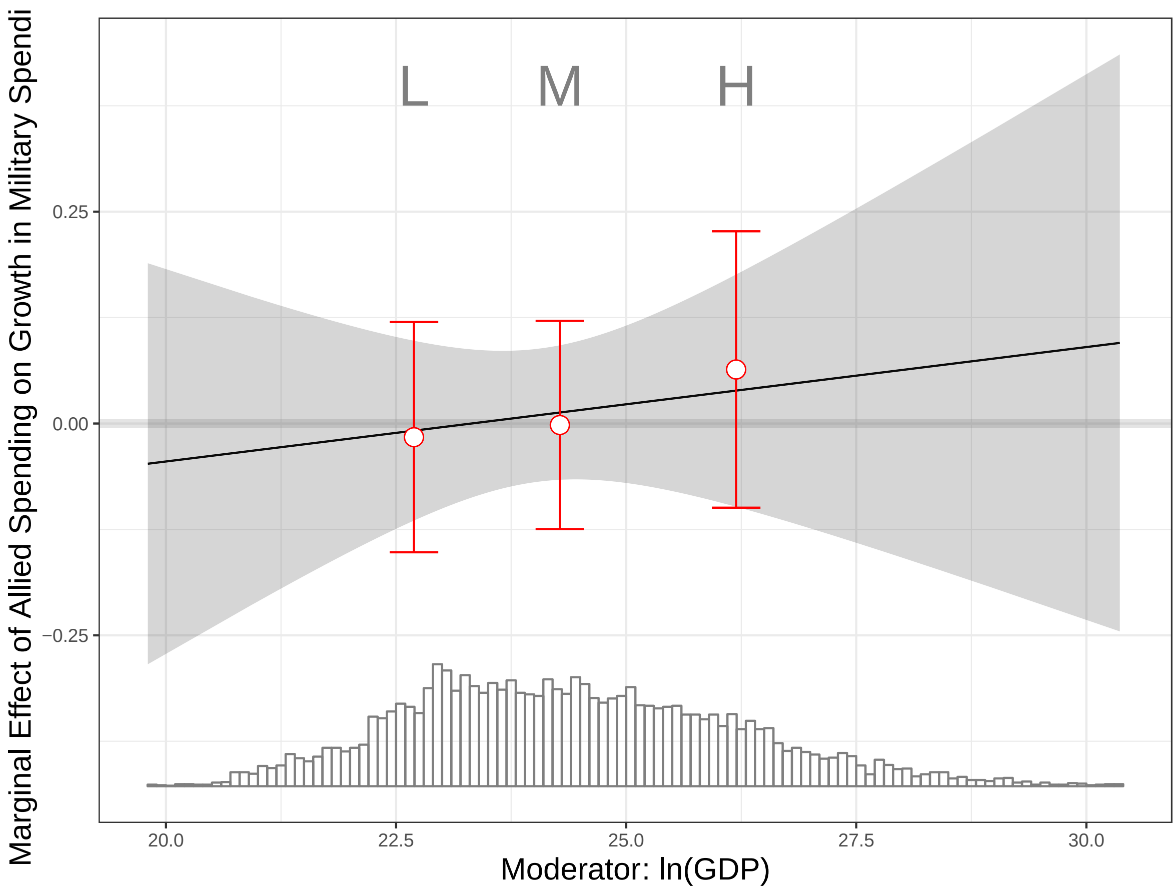

Figure 3: Binning estimates of interaction between changes in allied spending and GDP.

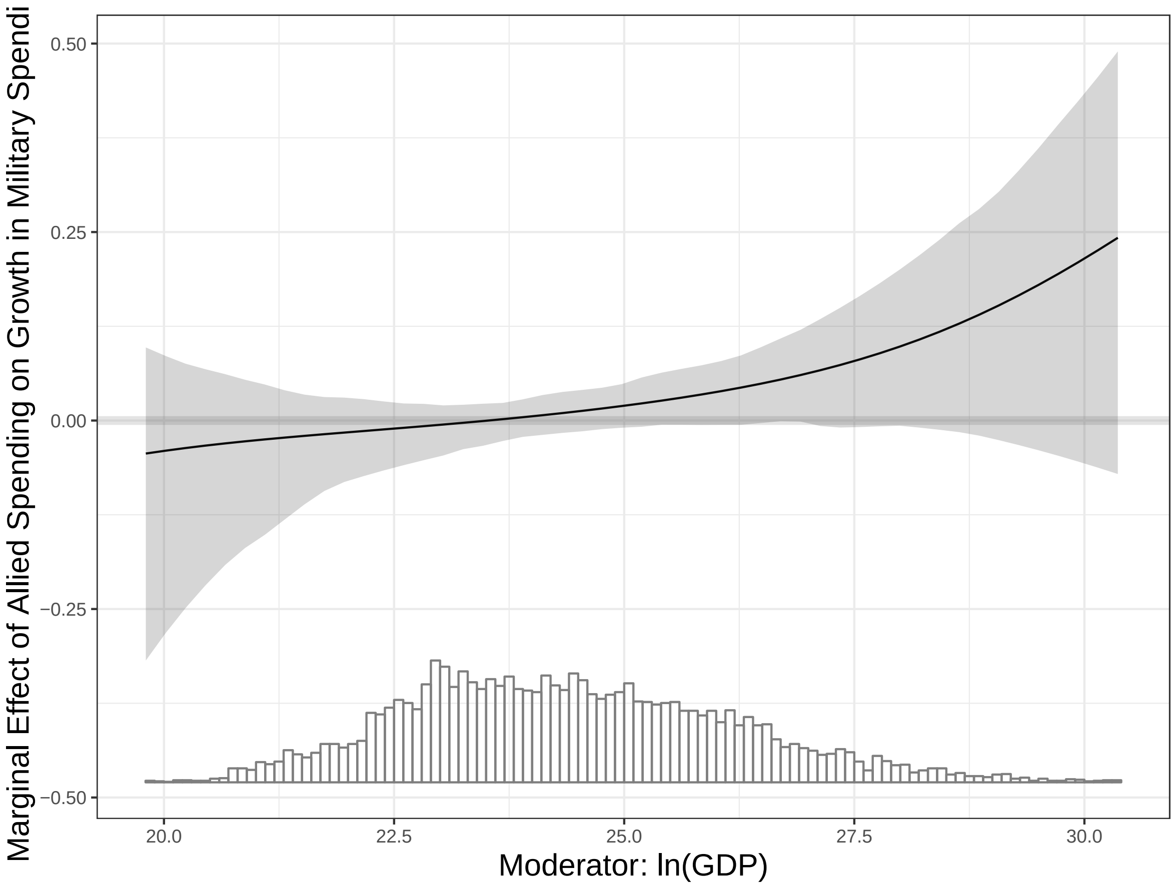

Figure 4: Kernel estimate of interaction between changes in allied spending and GDP.

$$
\begin{gathered}
p(\alpha) \sim N(0,1) \\
p(\sigma) \sim \text{half-}N(0,1) \\
p(\alpha^{yr}) \sim N(0,\sigma^{yr}) \\
p(\sigma^{yr}) \sim N(0,1) \\
p(\alpha^{st}) \sim N(0,\sigma^{st}) \\
p(\sigma^{st}) \sim \text{half-}N(0,1) \\
p(\gamma) \sim N(\theta,\sigma^{all}) \\
p(\theta) \sim N(0,.5) \\
p(\sigma^{all}) \sim \text{half-}N(0,1) \\
p(\beta) \sim N(0,1) \\
p(\nu) \sim gamma(2,0.1)
\end{gathered}
$$

Table 2: Summary of Priors in Multilevel Model

demonstrates, the marginal effect of spending is mostly positive. Negative point estimates of the marginal effect are indistinguishable from zero. Allowing non-linear changes in the marginal effect of allied capability does little to change substantive inferences about the presence of a conditional relationship.

2 Multilevel Model

This section describes the priors on the multilevel model, convergence diagnositics for the Hamiltonian Monte Carlo, and results from running the same model on a sample of only states with at least one alliance.

2.1 Priors

All priors are specified to be weakly informative relative to the scale of the data (Gelman, Simpson and Betancourt, 2017). I summarize the prior distributions for each set of parameters in Table 2. $p(\nu)$ is a well-behaved prior for the degrees of freedom in a t-distribution (Juárez and Steel, 2010). Given that median growth in military expenditures is 0.06, the priors are quite diffuse.

To facilitate estimation, I use a non-centered parameterization for the state and year varying intercepts, as well as the γ parameters (Betancourt and Girolami, 2015). The prior is equivalent, but a non-centered parameterization decouples the mean and variance, which makes sampling easier. I also employ a sparse matrix representation of the alliance membership matrix $\mathbf{Z}$ to speed up estimation.

2.2 Convergence

There were no divergent iterations in sampling, which would have invalidated the results. However, there are other threats to inference from the posterior samples. Given heavy tails in military spending growth, STAN might have struggled to explore the posterior distribution.

Energy plots can diagnose this problem. Figure 5 plots the marginal energy distribution and the first differenced distribution. If the two histograms do not overlap, sampling was impeded by heavy tails. The substantial overlap in the distributions for all four chains in Figure 5 indicates this was not a problem.

The split $\hat{R}$ statistic is another way to assess convergence. $\hat{R}$ compares the behavior of each chain by measuring the ratio of the average variance of draws within each chain to the variance of the pooled draws across chains. When $\hat{R}$ is close to 1, all the chains have similar variance, and are therefore in equilibrium.

The standard heuristic is that an $\hat{R}$ greater than 1.1 is problematic. Figure 6 plots the $\hat{R}$ statistic for every parameter in the model. No parameters generate concern, even at a more conservative threshold of 1.05.

2.3 Alternative Sample

It is possible that estimating a model on the full sample of states makes misleading comparisons by including states with no alliances. This adds many zeros to the membership matrix. To check whether inferences are sensitive to including states with no alliance participation, I re-fit the

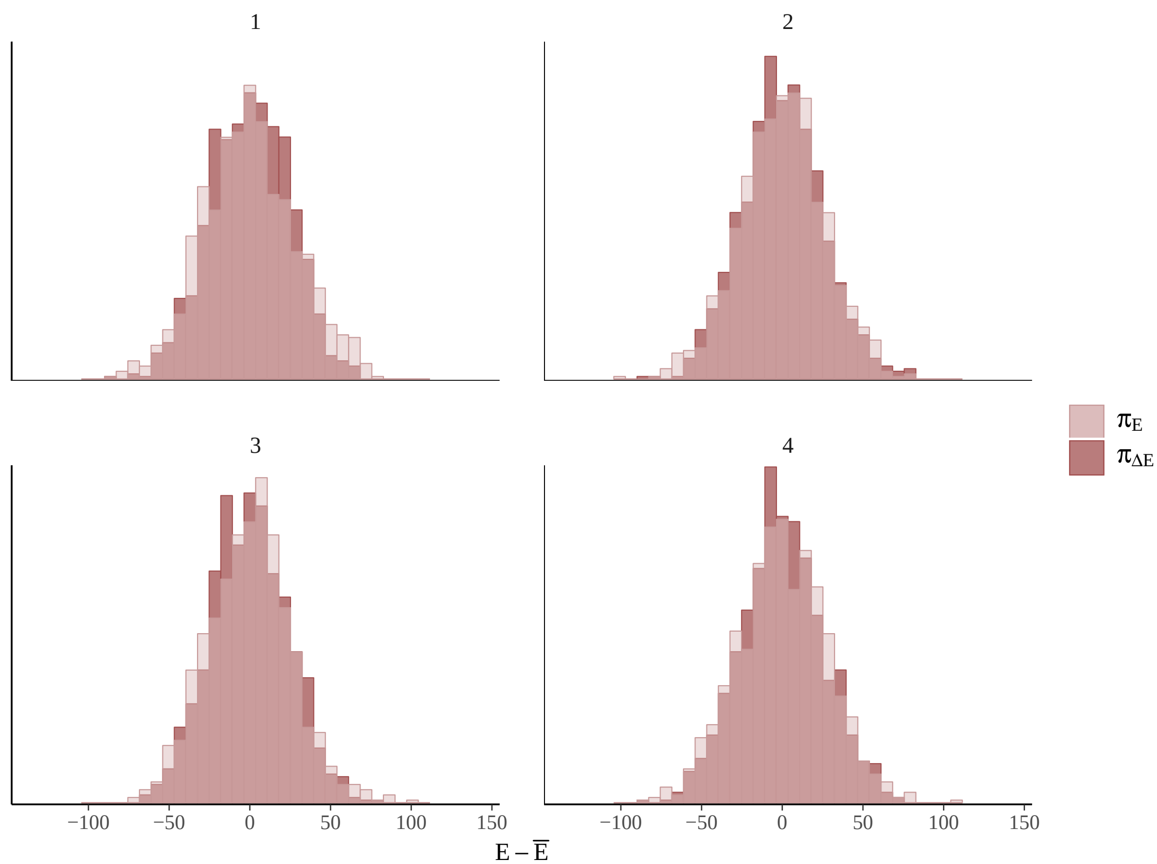

Figure 5: Energy plot of multilevel model results. Greater overlap in the two histograms indicates adequate exploration of the posterior distribution.

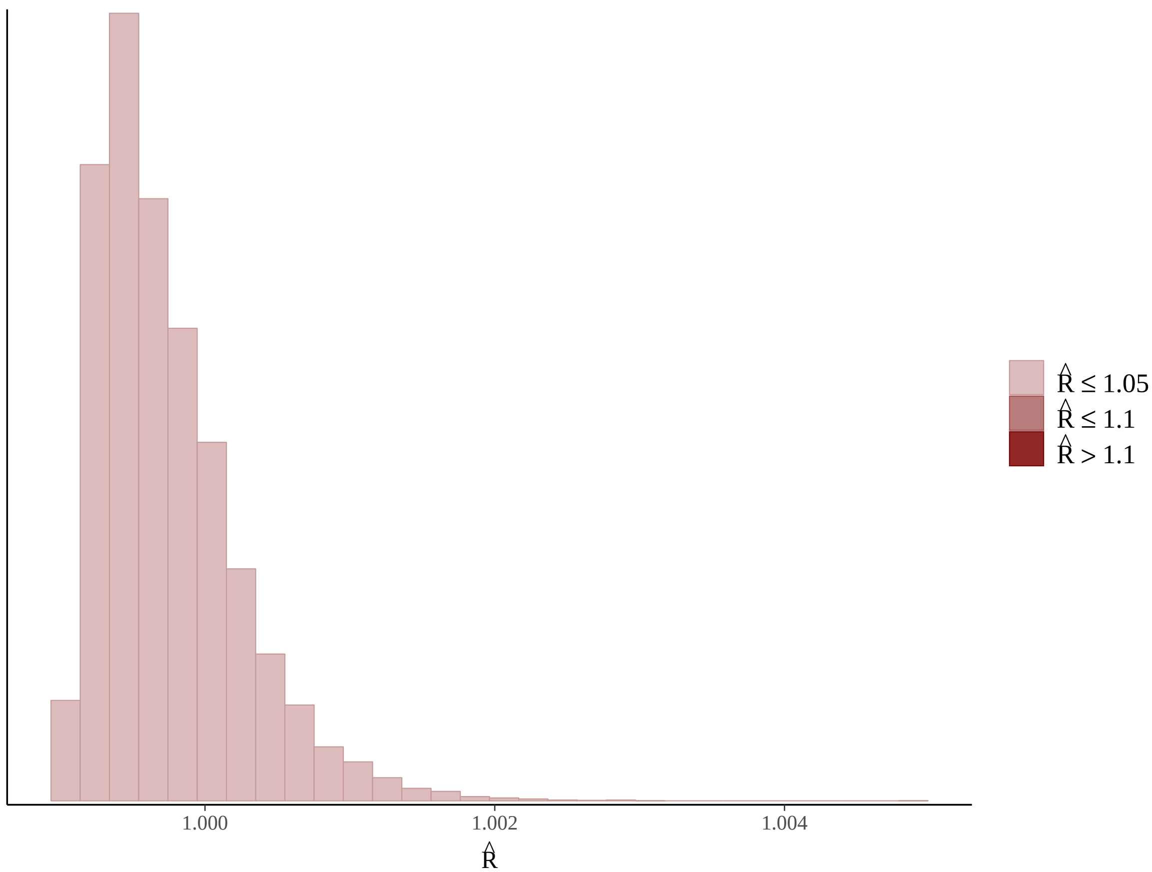

Figure 6: Histogram of split $\hat{R}$ statistic for all parameters in the multilevel model.

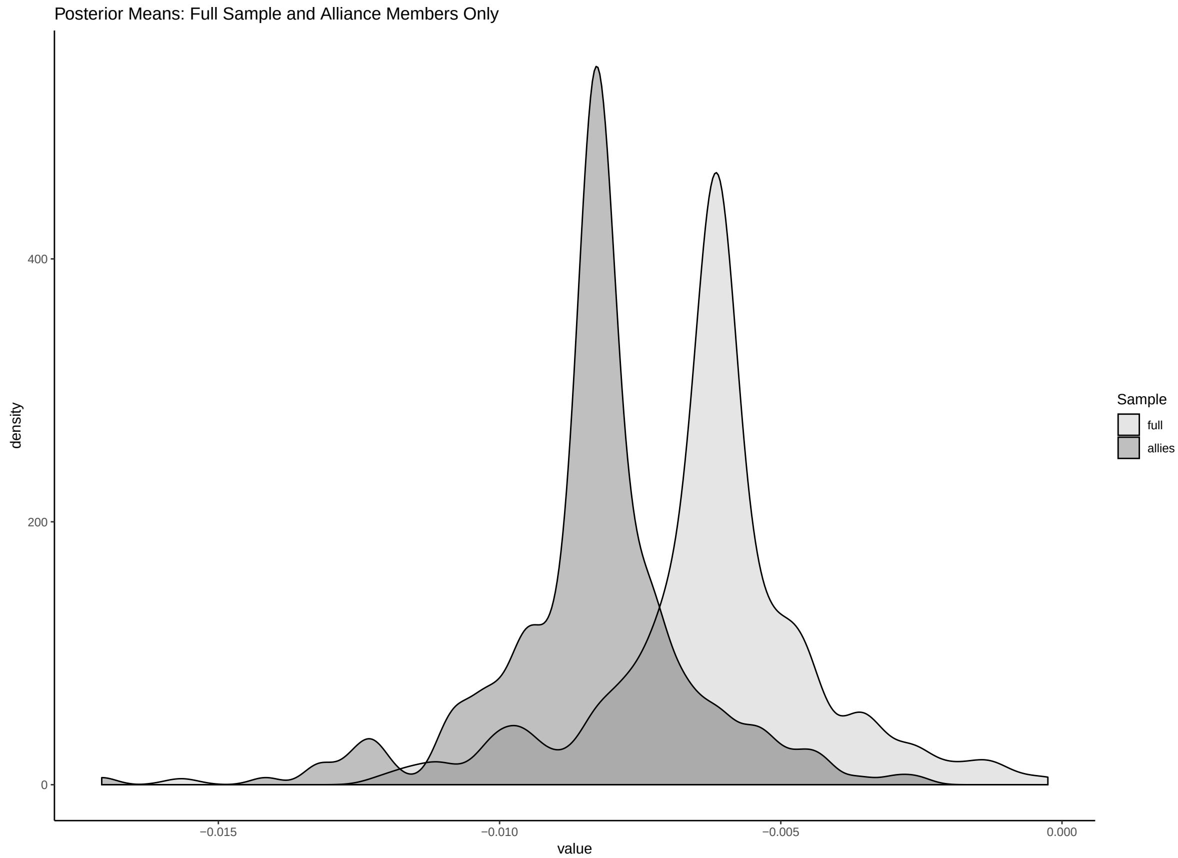

Figure 7: Comparison of the distribution of posterior mean γ parameters in full sample and a sample of only alliance participants. The darker gray distribution is the mean of the γ parameters in the sample of alliance members. The impact of treaty contribution on growth in military spending is more negative in the alliance-members only sample.

multilevel model on a sample of only states with at least one alliance. This reduced the sample size from 9,961 observations to 5,222, but the results are relatively unchanged.

All 285 treaties have a negative mean γ estimate, and none have a 90% credible interval that excludes zero. The overall mean θ is more negative in this sample and the γ estimates remain tightly clustered around that mean. Thus as Figure 7 shows, the distribution of the association between treaty contribution and spending across alliances is more negative in the alliance members sample. This figure overlays the distribution of the mean γ parameters in each sample.

Therefore, inferences about the impact of treaty contribution on growth in military spending are unchanged if the sample is restricted only to alliance members. Increasing a state's share of total allied GDP leads to a more negative effect on treaty spending in expectation. I am still unable to identify any reliably positive γ parameters, which contradicts Olson and Zeckhauser's exploitation hypothesis.

References

Betancourt, Michael and Mark Girolami. 2015. Hamiltonian Monte Carlo for hierarchical models. In *Current Trends in Bayesian Methodology with Applications*, ed. Satyanshu K. Upadhyay, Umesh Singh, Dipak K. Dey and Appaia Loganathan. Chapman and Hall/CRC Press pp. 79–102.

Gelman, Andrew, Daniel Simpson and Michael Betancourt. 2017. "The prior can generally only be understood in the context of the likelihood." *arXiv preprint arXiv:1708.07487* .

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models?: Simple Tools to Improve Empirical Practice." *Political Analysis* .

Juárez, Miguel A and Mark FJ Steel. 2010. "Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions." *Journal of Business & Economic Statistics* 28(1):52–66.