

Reassessing the Public Goods Theory of Alliances

Joshua Alley*

October 2, 2019

Abstract

The public goods theory of alliances exerts a large influence on scholarship and policy, but it has not received sufficient empirical scrutiny. The key claim of public goods models is that smaller alliance participants can free-ride on larger partners because security is a public good. Prior statistical tests of this exploitation hypothesis suffer from a mix of identification and generalizability problems, so this study addresses those limitations in two different tests. First, it examines whether small states are more likely to respond to greater allied capability by reducing military spending. Second, it estimates whether shares of allied GDP are positively correlated with percentage changes in military spending within 285 offensive and defensive treaties. Results from the analysis contradict the expectations of the public goods model that alliance participation allows smaller states to allocate fewer resources to defense. Therefore, the argument that alliances provide a public good and generate free-riding incentives should be treated with more skepticism.

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

1 Introduction

Olson and Zeckhauser (1966) argue that international alliances generate a collective action problem. According to their theory, security from an alliance is a public good, so smaller alliance participants “free ride” on the contributions of larger members. Free-riding is reflected by disproportionate allocations of resources to defense, where smaller alliance spend a lower share of their national income on the military. This view of alliances and defense effort is quite influential in scholarship (Walt, 1990; Mearsheimer, 1994; Goldstein, 1995; Sandler and Hartley, 2001; Garfinkel, 2004; Walt, 2009; Norrlof, 2010; Barrett, 2010; Plümper and Neumayer, 2015).

In addition, policy discussions of alliances rely on collective action ideas. Commentators and American policymakers often refer to allied “free-riding.” Thinking of alliance security as a public good and beset by collective action problems generates fear that the United States is “being taken advantage of” by junior partners. US policymakers make frequent use of free-riding and disproportionate contributions to the common good to criticize lackluster allied defense expenditures. For example, Barack Obama complained in 2016 that “Free riders aggravate me” and US allies “have to pay your fair share.” Donald Trump has implied the US will not protect allies who spend too little on defense. Such exhortations for allies to contribute more go back to the Eisenhower administration (Lanoszka, 2015).

The North Atlantic Treaty Organization (NATO) is the epicenter of free-riding discussions. Following Olson and Zeckhauser’s emphasis on military spending as a share of GDP, accusations of free-riding emphasize that NATO allies have a lower defense burden than the United States. Pundits and policy makers now employ spending 2% of GDP on defense as a threshold for adequate effort. While NATO members made a formal commitment to spend at least 2% of GDP on defense by 2024 at the 2014 Wales Summit, by 2017 only four members besides the United States met that threshold, furthering concerns and rhetoric about free-riding (Economist, 2017).

The failure of many NATO members to bear an adequate defense burden is presented as ev-

idence of free-riding. US policymakers assert that low allied defense burdens are an inadequate contribution to collective defense. However, European allies reject this characterization of their defense spending, arguing they make other crucial contributions (Boyer, 1993). Academic studies echo this debate about free-riding. Scholars have spent decades arguing over whether defense burdens of NATO members reflect free-riding (Sandler and Forbes, 1980; Palmer, 1990; Gates and Terasawa, 1992; Sandler and Hartley, 2001; Lanoszka, 2015; Plümper and Neumayer, 2015).

Olson and Zeckhauser's argument has shaped a generation of policy and scholarly debate about alliances and defense spending. Increasing great power competition makes understanding the value of the public goods model of alliances even more important. Competing visions of future US grand strategy hinge in part on the explanatory power of the public goods model. Advocates of retrenchment and "restraint" use the public goods logic to claim that US allies free-ride and do not contribute enough to their own defense (Preble, 2009; Posen, 2014). As a result, they argue that the United States should withdraw from its alliances. Others assert that the extent of allied free-riding is overstated and alliances do not provide a public good (Brands and Feaver, 2017).

In this paper, I assess the public goods model by deriving and testing two observable implications of free-riding. Revisiting the public goods model helps clarify its value to contemporary scholarship and policy and addresses an empirical shortcoming. 53 years after the publication of "An Economic Theory of Alliances," there is little reliable and general evidence for or against Olson and Zeckhauser's prediction that small alliance members free-ride on larger partners. Therefore, it is still unclear whether the public goods logic is a useful model of alliance politics.

Existing scholarship leaves the debate unsettled because studies suffer from two key limitations. First, many empirical estimates of free-riding within alliances are unidentified. Olson and Zeckhauser measure defense burdens using military spending as a share of GDP and state size based on GDP. While this approach is widely emulated, it creates an identification problem because GDP is present on both sides of the equation. In Olson and Zeckhauser's research design, changes in GDP impact both the independent and dependent variable. There is a deterministic

component in the relationship between GDP and the defense burden— changes in GDP shift the defense burden.¹

One noteworthy paper addresses the identification problem, but may not produce generalizable findings. Plümper and Neumayer (2015) examine how military spending by NATO members responds to changing US and Soviet spending. This approach removes GDP from the dependent variable and focuses on how states respond to changes in potential threat. In their framework, unresponsiveness to US and Soviet military spending is evidence of free riding by non-US NATO members. They find no correlation between NATO member size and free-riding, which they argue contradicts Olson and Zeckhauser.² But this brings me to the second limitation: a lack of generalizability.

Most studies of the public goods model focus on NATO, but NATO is a difficult case from which to draw general conclusions because it is exceptional. NATO is now 70 years old, contains 29 members and includes many of the most capable states in the international system. Most alliances are less durable and aggregate less capability (Leeds et al., 2002). Explaining dynamics within NATO is important, but a robust model of alliance politics should apply to other treaties as well.

We do not know whether the public goods theory of alliances applies to treaties besides NATO. My survey of the literature on alliance participation and military spending found seven tests of alliance free-riding outside of NATO (Russett, 1970; Starr, 1974; Reisinger, 1983; Thies, 1987; Conybeare and Sandler, 1990; Oneal and Whatley, 1996; Siroky, 2012). Six of these studies include GDP in the independent and dependent variable and the seventh uses case comparisons. All seven studies examine only a few treaties at a time. Thus, there is still need for a reliable and general test of the public goods logic.

¹The only exception occurs when military spending changes in such a way that defense spending's share of GDP is unchanged. Such changes are highly unlikely. Moreover, ratio dependent variables such as military expenditures as a share of GDP often generate spurious results (Kronmal, 1993).

²However, they do not include the United States in their sample, which omits a crucial data point for testing the size argument. Given the focus on responses to US spending, this decision is understandable.

In this paper, I address the identification and generalizability issues in two tests of the public goods model of alliances. Both tests examine what some observers describe as Olson and Zeckhauser's "exploitation hypothesis" in different ways. The tendency of small states to free-ride could result from allied capability or disparities in economic resources. Therefore, the first test looks at general free-riding between alliance participants, while the second looks for specific within-alliance free-riding. Implementing two tests creates diverse evidence and two opportunities for the public goods model to demonstrate explanatory power.

The first design uses an aggregate measure of alliance participation to examine whether state size modifies the impact of changing allied capability on percentage changes in defense spending. I interact GDP and changes in allied spending to see if small states reduce military spending in response to greater allied military spending. If alliances provide a public good, small states should be more inclined than large states reduce defense spending in response to increasing allied capability. This test examines broad patterns in alliance participation and state size in a sample of non-microstates from 1816 to 2007.

The second design employs a Bayesian model to estimate the association between a state's share of allied economic resources and percentage changes in military spending within 285 alliances. This test examines whether a greater share of total alliance GDP is positively correlated with percentage changes in military spending. If smaller states free ride on larger partners, states with a smaller portion of allied GDP will have lower military expenditures through alliance participation. Unlike the first design, this approach distinguishes between individual alliance treaties and estimates within-treaty correlations.

Neither approach finds evidence to match Olson and Zeckhauser's prediction that small states free-ride on larger partners. I find little evidence economic size modifies the impact of allied capability on military spending. As for the second approach, no treaties have a reliably positive correlation between a share of allied GDP and military spending.

Many scholars have questioned and revised the public goods theory of alliances on theoretical

grounds (Palmer, 1990; Sandler and Hartley, 2001; Norrlof, 2010). But the public goods model could have some empirical validity and theoretical revisions may be premature without clear evidence that a parsimonious public goods model is inappropriate. Therefore, revisiting the pure public goods logic and testing it is a useful preliminary step for refining our theories of alliances. Moreover, popular and policy discussions of alliance politics implicitly depend on the public goods model of alliances. Policy efforts to address low allied military spending could fail if they are based on an inaccurate model of alliance politics.

The paper proceeds as follows. First, I summarize the public goods theory of alliances and derive two testable predictions of its logic. Then, I describe the results from testing the first prediction. The third section details the model and results from testing the second prediction. The final section assesses aggregate support for the public goods logic, as well as implications for scholarship and policy.

2 Observable Implications of Free-Riding

In deriving testable predictions, I use the public goods logic, but make slightly different predictions than Olson and Zeckhauser. As noted above, predictions about the correlation between GDP and military spending as a share of GDP are difficult to test. Therefore, I use percentage changes in military spending as the outcome and alternative indicators of state size. Note that this theoretical exercise does not critique the public goods logic.³ Instead, I treat the public goods logic as given and express the key prediction that alliance participation allows small states to free-ride in two different ways.

Why might alliances suffer from a collective action problem? The aggregate military capability of an alliance provides security for members, so states contribute by investing in their military. Because a treaty cannot exclude members without undermining its purpose, treaty security is a public

³There are many useful theoretical critiques of the public goods logic, and this paper is focused on what the public goods logic can contribute.

good. Alliance members receive security regardless of their individual contribution.⁴ Therefore, states have incentives to rely on their alliance partners to contribute military capability.

Olson and Zeckhauser expect that larger members of the alliance will bear a higher defense burden, because these states value security from the alliance more.⁵ Small alliance members rely on the contributions of larger partners for security and bear a lower defense burden. As a result, smaller states free-ride on larger alliance participants. Moreover, smaller states have greater bargaining leverage, because a large state cannot credibly threaten to reduce their contribution and has “relatively less to gain than its small ally from driving a hard bargain” (Olson and Zeckhauser, 1966, pg. 274).

According to Olson and Zeckhauser, GDP will be positively correlated with the percentage of national income spent on defense within alliances. Using military spending as a share of GDP as the dependent variable places GDP on both sides of the equation, so I do not test this exact prediction. The same argument applies to percentage changes in military spending, however (Plümper and Neumayer, 2015). If smaller states have an advantage in alliance politics, alliance participation should lower their percentage changes in military spending more than large states.

The first prediction examines how absolute state size modifies the impact of total allied capability on military spending. If the free-riding hypothesis holds, small states will be more likely to reduce military spending in response to greater allied military spending. Increasing allied capability increases the ability of small states to free ride. As their allies provide more security, small states can lower their own investment in defense.

Larger states will be less responsive to increased allied capability. These states are usually the senior partner in their treaties and according to the public goods model, they value security from their alliances more. Thus, large states will be less prone to reduce military spending because it

⁴The marginal costs and benefits of participation depend in part on group size, but Olson and Zeckhauser’s model shows free-riding even in a bilateral alliance.

⁵Olson and Zeckhauser do not define size precisely in the model, so size could refer to capability, geography, power or economic characteristics of states. Following Olson and Zeckhauser, I retain their emphasis on economic size, measured using GDP.

undermines their security commitments. For large states, greater allied capability does not provide the same opportunity to free-ride.

This logic implies a conditional relationship between allied spending, state size, and defense expenditures. Small states should respond to greater allied capability by reducing percentage changes in defense spending. But as state size increases, this negative effect of allied capability on defense spending will diminish.

HYPOTHESIS 1: The marginal effect of increasing allied military spending on annual percentage changes in military spending will be negative for small states, and this effect will become smaller as GDP increases.

The first prediction aggregates alliances to look for general patterns, but Olson and Zeckhauser use the relative size of alliance members to explain changes in military spending. Therefore, I also assess whether the public goods argument makes accurate predictions about differences in military spending within individual treaties. If smaller alliance members are able to free ride on larger partners, they will have lower percentage changes in military spending.

Within alliances, size is a function of relative economic resources. Aggregate economic resources provide the basis for defense spending. One state could have different economic weight, depending on the alliance. For example, Turkey has most of economic resources in an alliance with Armenia, but has a much smaller share of NATO's total economic resources. As a result, which states are large or small varies across treaties, leading to heterogeneous effects on military spending.

Olson and Zeckhauser expect that larger states contribute disproportionately to the alliance. Large states will spend more of their resources on military capability so that the alliance provides adequate security. Relative wealth, or a state's share of the total economic resources in an alliance, will shape what they spend on defense. The larger a state's share of economic resources, the more likely other states are to free-ride.

Inasmuch as larger members of a treaty value security from an alliance more, they will invest in military capability to maintain the treaty. Smaller members can free-ride through lower military spending. As a result, larger members will have higher military expenditures. Within an alliance, the share of total allied GDP and percentage changes in military spending will be positively correlated.

HYPOTHESIS 2: Within individual alliance treaties, as a state's share of total allied GDP increases, annual percentage changes in military spending will increase.

Treaties with a positive correlation between GDP share and military spending conform to Olson and Zeckhauser's expectations. A negative correlation between economic weight and military spending implies that larger members have lower spending from alliance participation. Therefore, I will examine how many alliances have a positive correlation between shares of total allied GDP and military spending.

I now test Hypotheses 1 and 2 with corresponding research designs in a sample of all non-microstates from 1816 to 2007.⁶ State-year observations are the unit of analysis in both tests. The next section examines Hypothesis 1 by interacting changes in total allied spending and GDP. This approach aggregates multiple treaties into a single measure of alliance participation.

3 Testing Hypothesis 1

To test Hypothesis 1, I employ conventional measures of state size and aggregate alliance participation. Olson and Zeckhauser use GDP to measure state size, so I construct a measure of GDP using data from the Maddison Project, which provides longer historical coverage (Bolt et al., 2018). I then take the natural log of GDP to address the variable's skewed distribution. I measure changes in allied capability by differencing the military spending of all ATOP defensive

⁶Including states with small GDP values makes estimating the interaction in Hypothesis 1 difficult. These observations create a large range of the modifying variable with no support for interactive estimates.

or offensive alliance partners of a state.⁷ I use military spending data from the Correlates of War Project (Singer, 1988). All alliance membership data comes from version 4 of the Alliance Treaty Obligations and Provisions (ATOP) data (Leeds et al., 2002).

Percentage changes in military spending is the dependent variable. Olson and Zeckhauser use defense spending as a share of GDP as their dependent variable, which is the source of previously described identification problems (Kronmal, 1993; Plümper and Neumayer, 2015). I use percentage changes instead of the defense burden because this measure gives a sense of burdens from changing defense budgets with a lower risk of spurious inferences. Annual percentage changes in spending is the change in military spending as a share of the previous year's budget:

$$\% \text{ Change Military Spending} = \frac{\text{Mil. Ex.}_t - \text{Mil. Ex.}_{t-1}}{\text{Mil. Ex.}_{t-1}} = \frac{\Delta \text{Mil. Ex.}}{\text{Mil. Ex.}_{t-1}} \quad (1)$$

Measuring percentage changes in spending matches Olson and Zeckhauser's emphasis on how alliance participants allocate resources to the military. Positive changes in spending imply an expanding defense budget and higher defense burden, all else equal.⁸ While defense budget levels only fall given negative percentage changes in military spending, the accumulation of differences between higher and lower percentage changes in spending over time can lead to divergences in military spending levels and burdens.

Moreover, using percentage changes in spending mitigates the risk of spurious inferences due to non-stationarity in panel data. The log-level of military spending is not mean-reverting in long panels. A differenced military spending variable has increasing variance over time, as budgets expand and generate larger changes. Modeling the DV in levels or changes might create spurious inferences (Granger and Newbold, 1974).

Using percentage changes in military spending as the dependent variable benchmarks changes to budget size. This facilitates comparisons across states and years. A 2% change is an equally

⁷This also corresponds to a key independent variable in the best test of the public goods model to date (Plümper and Neumayer, 2015).

⁸Including GDP in the model controls in part for changes in national income.

burdensome increase in the defense budget for large and small states, all else equal.

Because Hypothesis 1 implies a conditional relationship, I interact allied spending and GDP to predict percentage changes in military spending in the following linear model:

$$\% \text{ Change Mil. Ex.} = \beta_1 + \beta_2 \ln(\text{GDP}) + \beta_3 \text{Ally Spend} + \beta_4 \ln(\text{GDP}) \times \text{Ally Spend} + \beta \mathbf{X} + \epsilon \quad (2)$$

β_2 and β_3 are the constituent terms of the interaction and \mathbf{X} is a matrix of control variables with coefficient vector β . Hypothesis 1 implies β_2 will be positive because larger states place a higher value on security, which results in higher spending when allied capability is zero. There are cases with no allied capability—Costa Rica and Iceland are two examples of states with no defense spending. Predictions about β_3 are irrelevant, because this represents the effect of increasing allied capability when GDP is zero and no such cases exist. The interaction term β_4 should be positive because if the impact of allied capability is negative for small states and that effect is decreasing over the range of GDP, this implies positive changes.

I estimate this model with a robust regression estimator—M estimation with Tukey's Biweight function. Percentage changes in military spending produces heavy-tailed regression residuals, so OLS is inefficient (Rainey and Baissa, 2018). Given heavy-tailed residuals, non-linear estimators such as robust regression can outperform OLS. Robust regression re-weights observations using the size of the residual, making it less sensitive to large deviations than OLS.

Besides the interaction terms, I control for other variables that are correlated with alliance participation and military spending. I adjust for international war (Reiter, Stam and Horowitz, 2016), civil war participation (Sarkees and Wayman, 2010), and a count of annual MIDs (Gibler, Miller and Little, 2016). Other controls measure salient alliance characteristics, including average democratic membership (DiGiuseppe and Poast, 2016) and average number of members across a state's alliance treaties. Last, I control for regime type, external threat (Leeds and Savun, 2007),

and the Cold War era.

3.1 Results

Table 1 summarizes results from the robust regression estimator. The constituent terms for the log of GDP and changes in allied spending are statistically insignificant. The interaction term is also statistically insignificant.

I cannot infer the absence of a conditional relationship from the coefficients alone, however (Brambor, Clark and Golder, 2006). Figure 1 plots the marginal effect of allied military spending across the range of GDP. The marginal effect of allied military spending can only be distinguished from zero for positive values. Although the estimated marginal effect is increasing across the range of GDP, at no point are the effects statistically distinguishable from one another.

Parts of the interaction are consistent with substantively meaningful effects of increasing allied capability. Median percentage change in military spending in this sample is .0603, so an effect of 0.02 adds roughly 33% to typical spending changes. At the highest values of GDP, much smaller and larger effects are plausible.

I draw similar conclusions from a wide range of estimation strategies, and report all those results in the appendix.⁹ Because the modifying variable is continuous, I employ kernel estimation of the interactive relationship (Hainmueller, Mummolo and Xu, 2019). In another check, I use a robust regression estimator with random effects for states and years (Koller, 2016). I also transform the military spending variable using an inverse hyperbolic sine to reduce the heavy-tails in positive and negative values, and used the transformed variable in OLS and robust regressions.

These results do not match the expectations of Hypothesis 1. The point estimate is uniformly positive and although the confidence intervals include negative values at very small levels of GDP, the effect of increasing allied capability is reliably positive for most of the range of GDP. There is little evidence that increasing allied capability decreases percentage changes in spending for small

⁹One of those checks estimates the model with OLS instead of robust regression.

	% Change Military Spending
Change Allied Spending	−0.026 (0.047)
ln(GDP)	0.0003 (0.001)
Change Allied Spending \times ln(GDP)	0.002 (0.002)
Avg. Alliance Size	0.0003* (0.0002)
Avg Alliance Democracy	−0.0003 (0.001)
International War	0.097*** (0.010)
Civil War Participant	0.001 (0.008)
POLITY	−0.0001 (0.0004)
External Threat	0.043*** (0.011)
Cold War	0.047*** (0.004)
Constant	0.026 (0.030)
Observations	9,139
Residual Std. Error	0.157 (df = 9128)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 1: Robust Regression Estimates: Association of allied spending, GDP, and their interaction with percentage changes in military spending 1816-2007.

Allied Spending and Growth in Military Spending

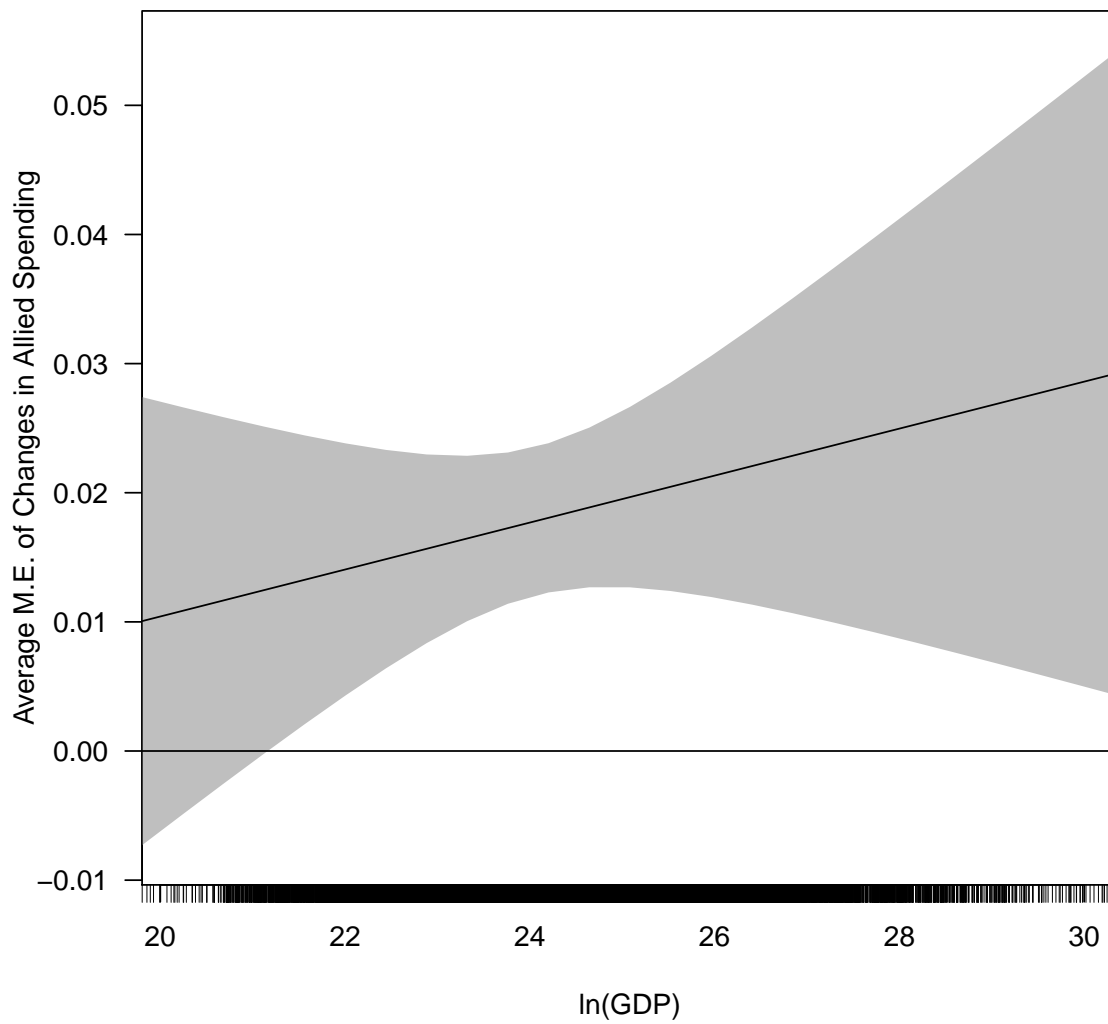


Figure 1: Average marginal effect of increasing allied military spending on percentage changes in military spending across the range of GDP. The rug plot on x-axis summarizes the distribution of GDP.

states.

However, it is possible that this aggregate analysis obscures variation across individual treaties. Free-riding may be more obvious by comparing percentage changes in spending within alliances. The next section tests Hypothesis 2 by examining how many treaties have a positive correlation between economic weight and percentage changes in military expenditures.

4 Testing Hypothesis 2

Testing Hypothesis 2 requires estimating the association between economic weight and military spending within each alliance. Unlike the test of Hypothesis 1, this process disaggregates alliances to examine relative state size. For each of the 285 alliances that promise military support,¹⁰ I estimate a parameter measuring the impact of increasing economic weight on military expenditures. This model examines free-riding within alliances using evidence from many treaties. As a result, the estimates give a general picture of the within-alliance differences Olson and Zeckhauser emphasize. Bayesian estimation regularizes estimates with many parameters, so I fit the following model using STAN (Carpenter et al., 2016).

The model starts with state-year percentage changes in military spending y_{it} . I model this outcome with a t-distribution to account for heavy tails. ν is the degrees of freedom parameter—as ν increases, the t distribution becomes more like a normal distribution.

$$y_{it} \sim student_t(\nu, \mu_{it}, \sigma) \quad (3)$$

Most of this model follows standard panel data designs. The expected value of the outcome μ_{it} depends on a constant α , state and year varying intercepts α^{st} and α^{yr} and control variables $X_{it}\beta$. In this specification, I include all controls besides the alliance portfolio averages from the test of

¹⁰ATOP offensive and defensive treaties. Other alliances do not provide security through promises of intervention in conflict. Thinking about the economic resources as a possible contribution to the aggregate capability of a non-aggression pact would be odd, for example.

Hypothesis 1 in X .

$$\mu_{it} = \alpha + \alpha^{st} + \alpha^{yr} + \mathbf{X}_{it}\beta + \mathbf{Z}_{it}\gamma \quad (4)$$

The $\mathbf{Z}_{it}\gamma$ term captures the impact of multiple alliances. \mathbf{Z} is a matrix of state participation in alliances— columns are alliances, rows are state-year observations. If a state is not in an alliance, the corresponding element of the matrix is equal to zero. If a state is part of an alliance, the corresponding element of the matrix is equal to a state’s GDP as a share of total GDP in the alliance.¹¹ The elements of \mathbf{Z} range between zero and one.

\mathbf{Z} is a quasi-spatial approach to capturing the impact of participation in multiple alliances. The “weight” of an alliance for military spending depends on the relative economic resources a state contributes. As a state’s share of total allied GDP increases, so does its potential military spending contribution to allied security. A larger share of allied GDP will increase demand for military spending, according to the public goods model. Smaller states will have lower percentage changes in spending, because they can free-ride on their partners. Thus, alliance participation impacts military spending through economic weight. This correlation between economic weight and military spending for each alliance is expressed in the γ parameters.

γ is a vector of 285 alliance-specific parameters. Because \mathbf{Z} contains each state’s share of allied GDP in different alliances, these coefficients estimate the association between economic weight and military spending. When a state is not in an alliance, the corresponding γ is multiplied by zero, and has no impact.

Each alliance has a separate impact on military spending. While these alliance parameters are distinct, they are related through partial pooling— they have a common prior distribution. Partial pooling estimates the dispersion of the alliance parameters from the data, so the prior for γ is normally distributed with mean θ and variance σ_{all} . θ is the mean hyperparameter of the alliance coefficients, so each alliance parameter deviates from that value. The extent of those deviations

¹¹ All GDP values are logged.

from θ depend on the variance hyperparameter σ_{all} . Every alliance estimate holds the impact of other treaties constant. A positive γ implies that alliance members with more economic weight have higher percentage changes in military spending, as Hypothesis 2 predicts.

4.1 Results

If the public goods theory of alliances is correct, we should observe many positive γ parameters. Because I employed Bayesian modeling, each γ has a posterior distribution.¹² I focus interpretation on the posterior mean and 90% credible intervals.¹³ The posterior mean is the expected value of γ , while the credible intervals capture uncertainty around that estimate.

Figure 2 summarizes the posterior mean of the 285 alliance coefficients. Most means are close to zero. All 285 alliances have a negative posterior mean. This alone is evidence against Hypothesis 2.

However, the posterior means do not convey the full range of possible effects. Positive mass in the posterior of the γ parameters matches Hypothesis 2. Figure 3 plots the relative positive and negative posterior probability of each γ parameter.

In all 285 alliances, the majority of the posterior mass is negative. Three alliances have 37% positive posterior mass, and all other treaties are more negative. There is no alliance where the preponderance of evidence suggests that relative economic weight increases military spending.

The credible intervals of each γ parameter are another source of information about the likely effect of economic weight.¹⁴ These intervals help determine if the correlations between the share of allied GDP and percentage changes in military spending can be distinguished from zero. Figure 4 plots the γ parameter for each alliance against the start year of the treaty. Points mark the posterior mean. The error bars encapsulate the 90% credible interval.

¹²See the appendix for a full summary of priors, convergence and model fit. I also show that the model can recover known parameters from simulated data.

¹³I use 90% credible intervals because inferences around 95% intervals can be less stable.

¹⁴The 90% credible interval falls between the 5% and 95% quantiles of the posterior.

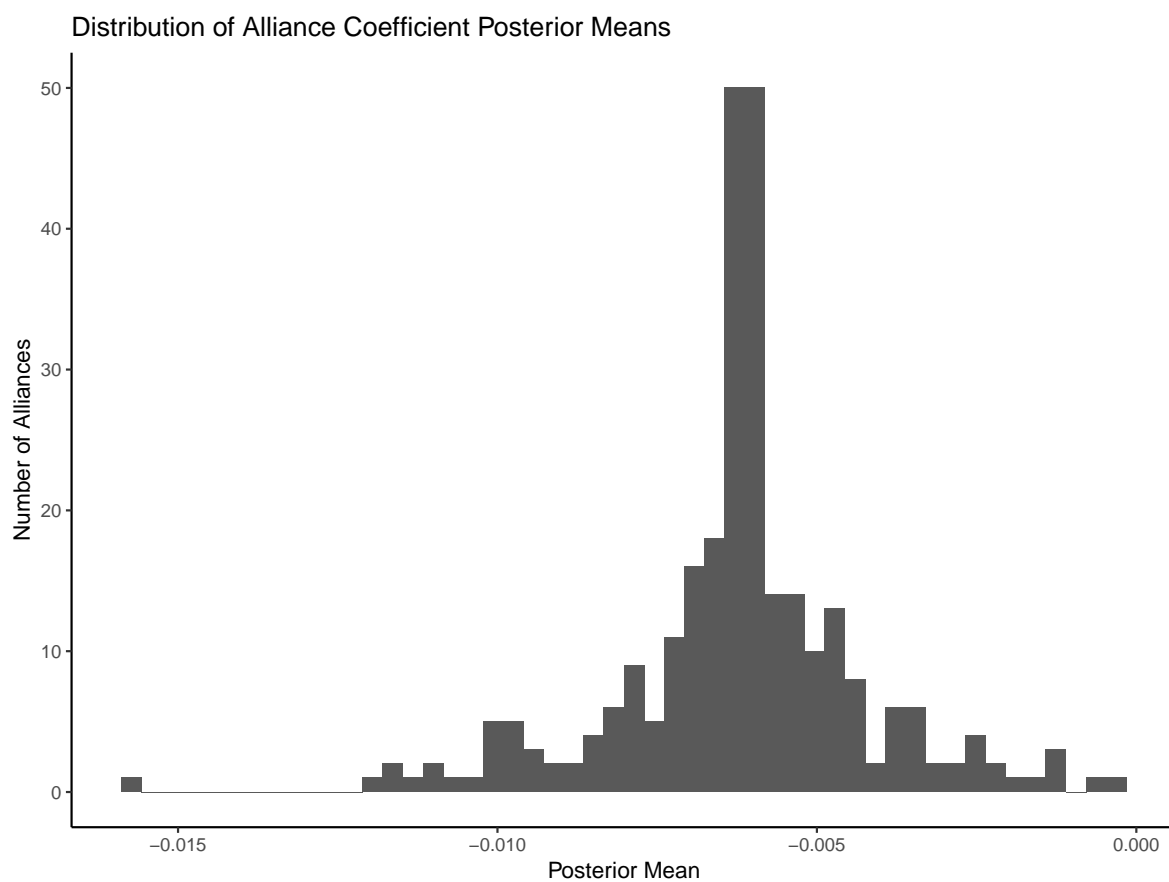


Figure 2: Posterior mean of association between share of total allied GDP and military spending in 285 defensive and offensive alliances from 1816 to 2007.

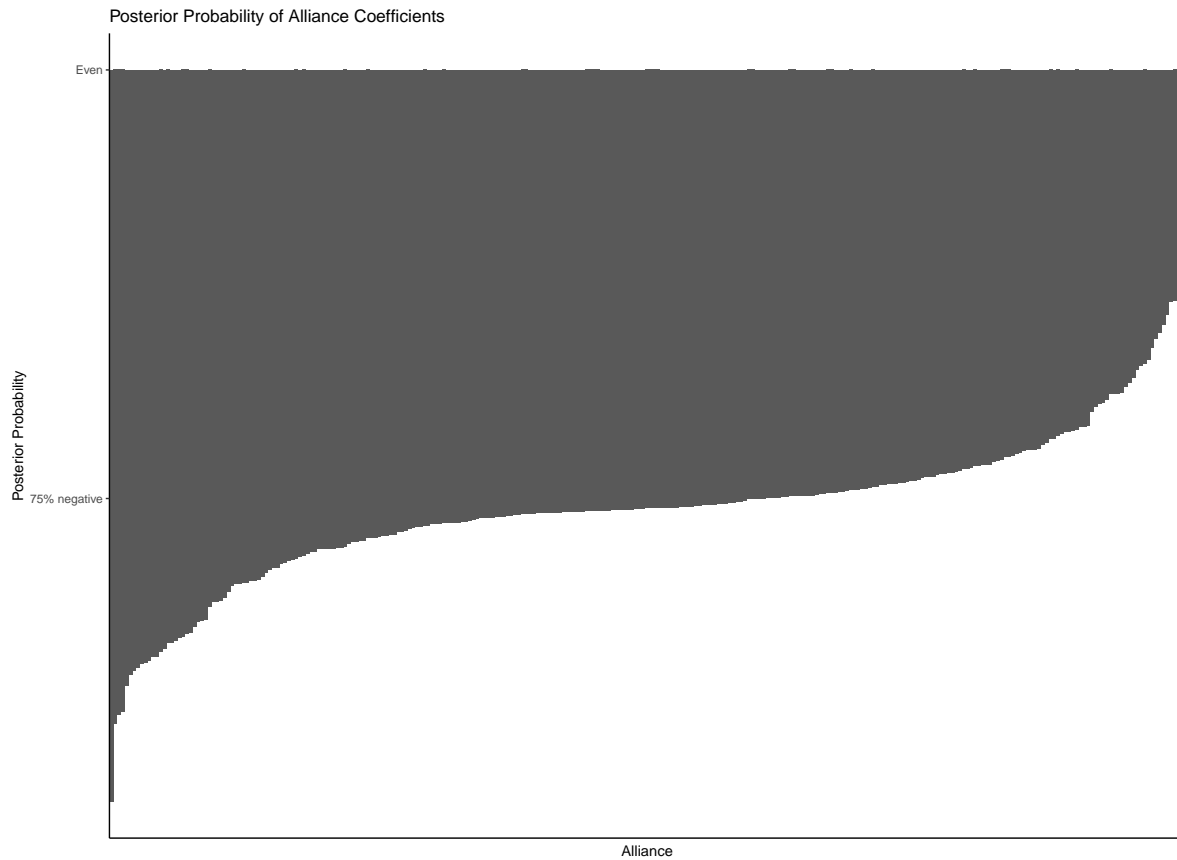


Figure 3: Positive and negative posterior probability of γ parameters, sorted by posterior mean. Each column corresponds to an alliance and expresses the relative positive or negative posterior probability. Even posterior probability indicates 50% positive and negative posterior mass— a γ parameter with mean 0 and perfectly symmetric uncertainty. Deviations from this add positive or negative posterior mass.

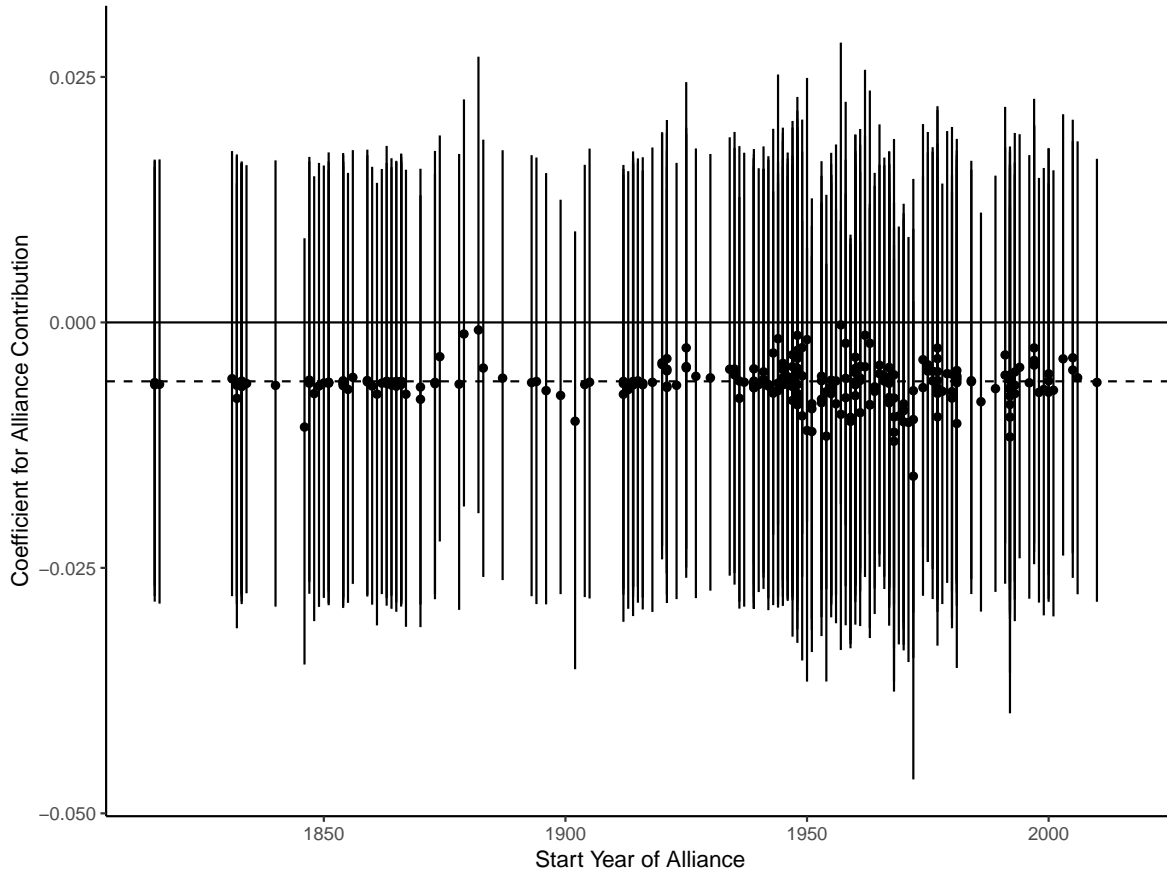


Figure 4: Estimated association between share of total allied GDP and defense spending in 285 defensive and offensive alliances from 1816 to 2007. Points represent the posterior mean and the error bars cover the 90% credible interval. The dashed line marks the posterior mean of the θ parameter, which estimates the average effect of increasing economic weight on percentage changes in military spending.

All 285 credible intervals include zero. Most credible intervals are consistent with effects ranging from roughly -0.02 to 0.02. No approach to examining the association between GDP as a share of allied GDP and military expenditure shows a preponderance of positive coefficients. Most alliances have no clear association between economic weight and percentage changes in military spending.

As Figure 4 shows, partial pooling of the γ parameters strongly regularizes these estimates. θ has a posterior mean of -0.006, which is close to zero. The variance hyperparameter σ_{all} has a posterior median of 0.01. As a result, the data indicates that the effects of increasing economic weight γ are tightly clustered around the overall mean. Under these circumstances, it is difficult to distinguish any estimates from zero, even as partial pooling shares strength across coefficients and reduces uncertainty.

The small scale of the γ parameters is reflected in predictions about military spending. To assess whether increasing a state's share of allied GDP leads to higher defense spending, I simulated the effect of changing GDP shares on percentage changes in military spending. In the simulated data, I used the full posteriors of the, intercept α , all the β coefficients, and one γ value. I selected the γ parameter with the most positive posterior mass, so this test is the *best case* for the public goods model. I set the state-level regressors at their median value and varied economic weight between minimum, median and near maximum values.

In Figure 5, I summarize predicted changes in military spending at different levels of alliance contribution. In this figure, the point marks the mean and the error bars summarize the 90% credible interval. Though moving from the minimum share of allied GDP to the high value of .70 slightly decreases the median value, the three predictions are impossible to distinguish from one another. Changing a state's relative economic contribution to an alliance leads to no perceptible differences in defense spending.

The hypothetical scenario in Figure 5 is the best case for the public goods models predictions. There is no evidence that increasing a state's share of total allied GDP leads to free-riding. As a

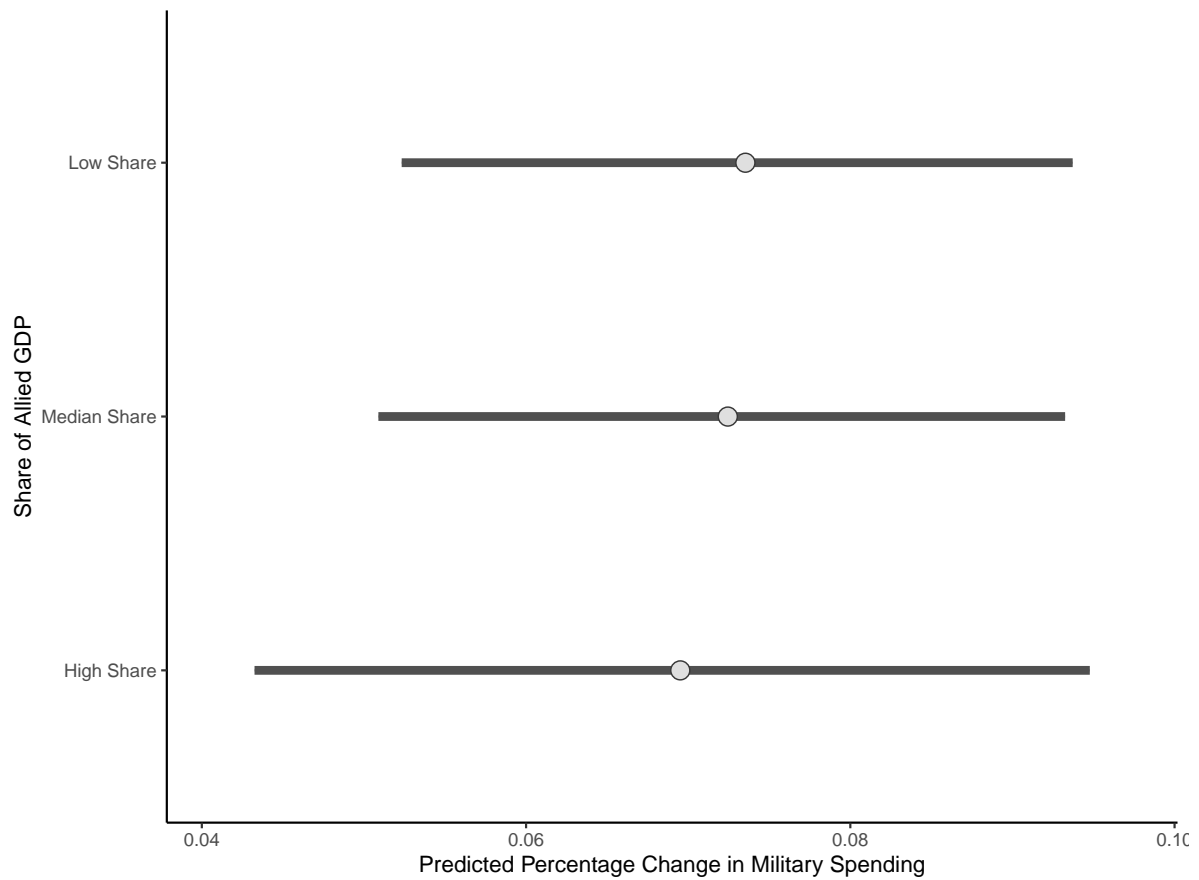


Figure 5: Predicted percentage changes in military spending for a simulated state with minimal, median and high shares of total allied GDP. Points mark the median value, and the error bars summarize the 90% credible interval.

final step, the γ estimates allow for inferences about individual alliances, so I now highlight a few noteworthy cases. One alliance has greater than 90% negative posterior probability: a 1972 pact between Hungary and Romania (ATOPID 3680). Several other bilateral alliances among Warsaw Pact members have large negative posterior means (ATOPID 3605 and ATOPID 3610).

The estimated γ for NATO offers no support for the public goods theory of alliances. The posterior mean is -0.01 , and the credible interval ranges from $-.03$ to $.01$. A greater share of NATO's total GDP has no clear association with percentage change in military spending. This finding corroborates the results of Plümper and Neumayer (2015).¹⁵

In summary, this test of the second hypothesis turns up minimal that evidence smaller alliance members have lower military spending due to treaty participation. Instead, all treaties have a negative mean association between an increasing share of alliance GDP and percentage changes in military spending. Few treaties have a preponderance of evidence for a positive or negative correlation between economic weight and percentage change in expenditures.

5 Conclusion

Taking the results from testing Hypothesis 1 and 2 together, there is little evidence for the prediction of the public goods theory of alliances that small alliance members are able to free-ride on their partners. I do not find that economic size modifies the impact of alliance participation on percentage changes in military spending. Moreover, there is no clear association between economic weight and military spending. No matter how it is expressed, the free-riding hypothesis that smaller states are able to spend less on the military by relying on larger partners has scant support. These findings should increase our skepticism about the public goods model of alliances. Although Olson and Zeckhauser's model is parsimonious, it lacks explanatory power. Better identified empirical

¹⁵But it does not rule out the possibility that NATO members spend less on the military thanks to allied capability, which a great deal of evidence suggests is the case (George and Sandler, 2017). Instead, it shows that greater economic weight has an unclear impact on NATO members military spending.

models and estimates from multiple treaties show little evidence to match predictions from a public goods logic. My results reinforce theoretical skepticism about the public goods model. While I have shown little general evidence of collective action problem, pieces of the public goods model may still be useful. Critiques by Palmer (1990) and Sandler and Hartley (2001) merit further attention, but scholars should be cautious about using modified public goods models.

A plausible theoretical alternative to the public goods approach emphasizes exchange between alliance members and intra-alliance bargaining (Norrlöf, 2010; Brooks, Ikenberry and Wohlforth, 2013; Kim, 2016). Perhaps large and small states use alliances for distinct purposes, so members reap specific gains from treaty participation (Morrow, 1991; Johnson, 2015). Small states might use alliances to defend their homeland. Large states could be more focused on using alliances to defend other states and expand their influence abroad.

Scholars and policymakers should reassess the way they use “free-riding” to describe alliance politics. Free-riding is inextricable from a public goods understanding of alliances. But if key predictions of the public goods model have little empirical support, free-riding is an inaccurate description of reduced defense effort by alliance participants. Again, charges of free-riding may mask exchange between alliance members (Lanoszka, 2015). Junior partners in an alliance may still lower defense spending, but this reflects mutually beneficial exchange, not free-riding.

Though policymakers may still find free-riding rhetoric useful, the idea is charged. Substantial anecdotal evidence suggests people dislike being taken advantage of. Talk of free-riding could activate these concerns and undermine the alliance itself. Policymakers might undermine support for alliances by accusing allies of free-riding.

Scholars should not abandon the public goods model in international politics. Other international organizations such as environmental regimes and climate change suffer from clear collective action problems. Using alliances as exemplars of collective action problems in other international organizations is inappropriate, however.

Although influential, the public goods model of alliances does not make accurate predictions

about state size, allied capability and defense spending. Thinking about alliances through exchange may be more fruitful than the collective action problem envisaged by Olson and Zeckhauser. Policymakers and scholars should exercise caution in using the public goods model to understand alliance politics.

References

- Barrett, Scott. 2010. *Why Cooperate?: The Incentive to Supply Global Public Goods*. Oxford University Press.
- Bolt, J, R Inklaar, H de Jong and JL van Zanden. 2018. "Maddison Project Database, Version 2018." *Rebasing 'Maddison': new income comparisons and the shape of long-run economic development*.
- Boyer, Mark A. 1993. *International Cooperation and Public Goods: Opportunities for the Western Alliance*. Baltimore: Johns Hopkins University Press.
- Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political Analysis* 14(1):63–82.
- Brands, Hal and Peter D. Feaver. 2017. "What are America's Alliances Good for?" *Parameters* 47(2):15–30.
- Brooks, Stephen G, G John Ikenberry and William C Wohlforth. 2013. "Don't come home, America: the case against retrenchment." *International Security* 37(3):7–51.
- Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. "Stan: A probabilistic programming language." *Journal of Statistical Software* 20:1–37.
- Conybeare, John AC and Todd Sandler. 1990. "The Triple Entente and the Triple Alliance 1880–1914: A Collective Goods Approach." *American Political Science Review* 84(4):1197–1206.
- DiGiuseppe, Matthew and Paul Poast. 2016. "Arms versus Democratic Allies." *British Journal of Political Science* pp. 1–23.
- Economist, The. 2017. "Military spending by NATO members." Online.
URL: <https://www.economist.com/graphic-detail/2017/02/16/military-spending-by-nato-members>
- Garfinkel, Michelle R. 2004. "Stable alliance formation in distributional conflict." *European Journal of Political Economy* 20(4):829–852.

- Gates, William R and Katsuaki L Terasawa. 1992. "Commitment, Threat Perceptions, and Expenditures in a Defense Alliance." *International Studies Quarterly* 36(1):101–118.
- George, Justin and Todd Sandler. 2017. "Demand for military spending in NATO, 1968–2015: A spatial panel approach." *European Journal of Political Economy* pp. 1–15.
- Gibler, Douglas M, Steven V Miller and Erin K Little. 2016. "An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001." *International Studies Quarterly* 60(4):719–730.
- Goldstein, Avery. 1995. "Discounting the free ride: alliances and security in the postwar world." *International Organization* 49(1):39–71.
- Granger, Clive WJ and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2(2):111–120.
- 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* .
- Johnson, Jesse C. 2015. "The cost of security: Foreign policy concessions and military alliances." *Journal of Peace Research* 52(5):665–679.
- Kim, Tongfi. 2016. *The Supply Side of Security : A Market Theory of Military Alliances*. Stanford University Press.
- Koller, Manuel. 2016. "robustlmm: an R package for robust estimation of linear mixed-effects models." *Journal of Statistical Software* 75(6):1–24.
- Kronmal, Richard A. 1993. "Spurious Correlation and the Fallacy of the Ratio Standard Revisited." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 156(3):379–92.
- Lanoszka, Alexander. 2015. "Do Allies Really Free Ride?" *Survival* 57(3):133–152.
- Leeds, Brett Ashley and Burcu Savun. 2007. "Terminating Alliances: Why Do States Abrogate Agreements?" *The Journal of Politics* 69(4):1118–1132.
- Leeds, Brett, Jeffrey Ritter, Sara Mitchell and Andrew Long. 2002. "Alliance Treaty Obligations and Provisions, 1815–1944." *International Interactions* 28(3):237–260.
- Mearsheimer, John J. 1994. "The False Promise of International Institutions." *International security* 19(3):5–49.
- Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35(4):904–933.
- Norrlöf, Carla. 2010. *America's global advantage: US hegemony and international cooperation*. Cambridge University Press.

- Olson, Mancur and Richard Zeckhauser. 1966. "An Economic Theory of Alliances." *The Review of Economics and Statistics* 48(3):266–279.
- Oneal, John R and Hugh Carter Whatley. 1996. "The effect of alliance membership on national defense burdens, 1953–88: A test of Mancur Olson's theory of collective action." *International Interactions* 22(2):105–122.
- Palmer, Glenn. 1990. "Alliance Politics and Issue Areas: Determinants of Defense Spending." *American Journal of Political Science* 34(1):190–211.
- Plümper, Thomas and Eric Neumayer. 2015. "Free-riding in alliances: Testing an old theory with a new method." *Conflict Management and Peace Science* 32(3):247–268.
- Posen, Barry R. 2014. *Restraint: A New Foundation for U.S. Grand Strategy*. Cornell University Press.
- Preble, Christopher A. 2009. *The Power Problem: How American Military Dominance Makes Us Less Safe, Less Prosperous, and Less Free*. Ithaca: Cornell University Press.
- Rainey, Carlisle and Daniel K. Baissa. 2018. "When BLUE Is Not Best: Non-Normal Errors and the Linear Model." *Political Science Research & Methods* .
- Reisinger, William M. 1983. "East European military expenditures in the 1970s: collective good or bargaining offer?" *International Organization* 37(1):143–155.
- Reiter, Dan, Allan C. Stam and Michael C. Horowitz. 2016. "A Revised Look at Interstate Wars, 1816–2007." *Journal of Conflict Resolution* 60(5):956–976.
- Russett, Bruce. 1970. *What Price Vigilance?: The Burdens of National Defense*. New Haven: Yale University Press.
- Sandler, Todd and John F Forbes. 1980. "Burden sharing, strategy, and the design of NATO." *Economic inquiry* 18(3):425–444.
- Sandler, Todd and Keith Hartley. 2001. "Economics of alliances: The lessons for collective action." *Journal of Economic Literature* 39(3):869–896.
- Sarkees, Meredith Reid and Frank Whelon Wayman. 2010. *Resort to war: a data guide to inter-state, extra-state, intra-state, and non-state wars, 1816-2007*. Washington, DC: CQ Press.
- Singer, J David. 1988. "Reconstructing the correlates of war dataset on material capabilities of states, 1816–1985." *International Interactions* 14(2):115–132.
- Siroky, David S. 2012. "Each Man For Himself? Rival Theories of Alliance Economics in The Early State System." *Defence and Peace Economics* 23(4):321–330.
- Starr, Harvey. 1974. "A Collective Goods Analysis of the Warsaw Pact after Czechoslovakia." *International Organization* 28(3):521–532.

- Thies, Wallace J. 1987. "Alliances and Collective Goods: A Reappraisal." *Journal of Conflict Resolution* 31(2):298–332.
- Walt, Stephen M. 1990. *The Origins of Alliance*. Cornell University Press.
- Walt, Stephen M. 2009. "Alliances in a Unipolar World." *World Politics* 61(01):86–120.