Summary of Previous Empirical Work on the Arms-Alliances Tradeoff

Joshua Alley[*]

July 10, 2018

Introduction

Military spending and alliance behavior the two of best-studied security policies in international relations. Despite the individual salience of arms spending and alliances, we know much less about how these policies impact one another. This shortcoming is the result of limitations in theory and research design. Although current scholarship has defects, those defects can be overcome, and it also provides material to build on.

States can use their own or allied military resources as a source of capability. Because military spending and allies provide military capability, scholars believe these two policies affect one another, creating an arms-alliances tradeoff. To assess research on this tradeoff, I undertook a comprehensive review of scholarship on the association between alliance behavior and domestic defense effort. The review cataloged the primary claim, theory, research design, and evidence of 27 papers.[1]

Theory on the arms-alliances tradeoff has made some valuable inroads during a debate over whether alliance membership increases or decreases military spending. There are two competing perspectives— one sees arms and allies as substitutes, while the other argues that these policies are complements. Neither perspective incorporates the key insights of the other, which has stunted theoretical progress. Substitution theory omits foreign policy goals besides security, while complementarity theories ignore domestic political incentives to limit military spending.

Serious empirical flaws in many studies have constrained our ability to assess the relative merits of complementarity and substitution theories. Many scholars test their claims with limited samples and misspecified models. Evidence from a few reliable studies suggests that the association between alliance membership and military spending may be conditional.

Although the relationship between arms and alliances has multiple components, most inquiry has focused on how alliance membership affects military spending. In turn, most work on alliance membership and military spending addresses the prevalence of free-riding. The emphasis

---

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

[1]In the interest of parsimony, I did not comprehensively address free-riding in NATO. Instead, I focused on work in that literature made a major methodological or theoretical contribution.

on free-riding created serious research design problems, and left several other research questions in the arms-alliances tradeoff unanswered. We have no theory on how the international context affects the arms-alliances tradeoff, and no general evidence of how the costs of alliance formation affect domestic military spending. Addressing these and other research questions will sharpen understanding of the arms-alliances tradeoff, as well as the relative merits of complementarity and substitution theories.

Clarifying the relationship between arms and alliances is important. The common perception that alliances lead to reduced defense spending guides policy, as US policymakers attempt to redress perceived opportunism. But inasmuch as free-riding is a poor explanation of the arms-alliances tradeoff, it is unlikely to inform effective policy.

This paper summarizes prior scholarship and suggests some directions for future inquiry. I start by examining the theoretical logic behind theories of arms and allies as substitutes and complements. Then I assess the validity of prior research design and empirical evidence. The last section concludes by discussing open issues in the arms-alliances tradeoff literature and considerations for future research.

Arms and Allies as Substitutes

If arms and alliances are substitutes, then additional alliances or capability from alliance partners should lead to decreased domestic military spending.[2] Two distinct theories predict this inverse association between alliance membership and domestic arms. Substitution theories of the arms-alliances tradeoff and the economic theory of alliances focus on how arms and allies both produce security, leading states to replace domestic military spending with allied spending.

Both approaches motivate substitution through the opportunity costs of military spending. Although military spending can have positive economic and technological spinoffs (Deger and Sen, 1995; Whitten and Williams, 2011), allocating funds to the military takes resources away from other social and economic policies. Leaders can use these public and private goods to ensure their survival in office (Bueno de Mesquita et al., 2002). Therefore if states have ample domestic political motives to reduce spending and rely on their allies to provide security.

Predictions of substitution between arms and alliances rely on the insight that states can use multiple policies means to achieve their foreign policy goals (Most and Starr, 1989). Morrow (1993) argues that domestic arms and alliances are both sources of military capability, allowing states to substitute between the two in their search for security. Instead of internal balancing through spending more on domestic arms, states can rely on external balancing through allied capability (Conybeare, 1992).

The marginal costs of arms and alliances determine which policy states use to increase their security (Sorokin, 1994). For instance, some evidence suggests that as the opportunity cost of military spending increases, states are more likely to form alliances (Kimball, 2010; Allen and DiGiuseppe, 2013). Measuring the foreign policy costs of a potential alliance is the main barrier to examining whether higher costs of an alliance lead states to rely more on military spending.

---

[2]Most empirical tests focus on this prediction, although substitution logic also predicts that when alliances become more costly, states will turn to domestic arms.

Because the costs of arms are financial, while the costs of alliances are measured in foreign policy freedom, arms and alliances are not perfect replacements.

As a source of capability, arms and allies are imperfect substitutes. States can rely on their own arms in any contingency, but domestic military capabilities take a long time to develop. Because members have divergent foreign policy interests, alliances are a less reliable source of capability than domestic arms, but provide immediate capability gains. The moment a treaty enters into force, alliance members gain their partner's support.

More reliable alliances are more like domestic arms as a source of capability. Therefore, credible alliance treaties are a better substitute for domestic arms. DiGiuseppe and Poast (2016) expand on this insight by arguing that because defense pacts with a democracy are more credible, these alliances will lead to reductions in spending.

The theoretical logic of substitution theory is clear, but we have little evidence for these claims. Only four studies test substitution of alliances for arms using military spending or personnel as the dependent variable. This lack of attention is due to the provocative claim that alliances lead to free riding, which has received more attention.

Free Riding and the Economic Theory of Alliances

The economic theory of alliances predicts that alliances lead to reduced domestic arms spending, but relies on a different mechanism. Olson and Zeckhauser (1966) treat security from an alliance as a public good, which generates a collective action problem for members. Each member's military spending is a contribution to the public good, but all members have incentives to free ride on their partners' expenditures. As a result, they predict that larger members of an alliance will bear a higher defense burden, because these states have a higher absolute value of the public good. Olson and Zeckhauser interpret a positive correlation between GDP and the ratio of defense spending to GDP as evidence of free-riding. Under this theory, alliance membership leads smaller states to reduce defense effort.

Olson and Zeckhauser's claims sparked extensive debate about free riding by alliance members. Much of the discussion focused on NATO, but other scholars checked the correlation between defense burdens and GDP in other alliances, with mixed results (Reisinger, 1983; Thies, 1987; Gates and Terasawa, 1992; Chen, Feng and Masroori, 1996; Oneal and Whatley, 1996; Siroky, 2012). The best evidence of free-riding is work by Plümper and Neumayer (2015), who show that many NATO allies are unresponsive or even reduce spending when growth in Soviet spending exceeds growth in US spending, and that the degree of free-riding depends on distance from the USSR. As with substitution theories, free-riding may only occur when treaty commitments are sufficiently credible (Gates and Terasawa, 1992).

A theoretical extension of the economic theory of alliances relaxes the public goods treatment of military spending. The joint product model treats military spending as having public and private benefits (Conybeare and Sandler, 1990), which lead to different predictions. As the ratio of private to public benefits from military spending increases, we should observe less free-riding (Murdoch, 1995; Sandler and Hartley, 2001). Differentiating between public and private benefits is difficult, however. Some studies divided types of military spending according to the benefit they provide, especially nuclear and non-nuclear defense effort (Hansen, Murdoch and Sandler, 1990). Nuclear

arms provide a pure public good through deterrence, while conventional spending has private benefits such as increased border security.

Emphasizing free riding creates theoretical and empirical issues. The concept of free-riding itself is problematic. Reductions in spending by NATO members could reflect decreased threat perceptions or a bargain with the United States (Lanoszka, 2015). Distinguishing between free-riding and exchange is difficult, especially if smaller alliance members trade security for autonomy in asymmetric alliances (Morrow, 1991).

The focus on burden sharing in the economic theory of alliances then leads to research design problems complications. The standard indicator of free riding is a state's defense burden— military spending as a share of GDP. It is difficult to identify a model of this ratio variable, especially with GDP on both sides of the equation, casting serious doubt on many of the studies that rely on correlations between defense burdens and economic size as evidence (Plümper and Neumayer, 2015).

Despite problems from the focus on free-riding in the economic theory of alliances, substitution theories of alliances have a clear internal logic and mechanism. Because of this clarity, substitution theory is the best known theoretical paradigm in the arms-alliances tradeoff literature. Policymakers complaints about free-riding further reinforced the perception that alliances lead to reduced defense effort.

The major weakness of substitution theories is their emphasis on security as the only output of an alliance. Substitution theory acknowledges that multiple policies can advance a state's security, but it does not incorporate other foreign policy goals, such as autonomy. Therefore, this perspective cannot capture the full range of how alliances may affect domestic arms.

States have a range of foreign policy goals. Even security focused paradigms like neorealism admit states may have different preferences over whether to maintain or change the status quo (Schweller, 1994; Walt, 2009). Substitution theory addresses some of the domestic political incentives states face in making security policy, but it provides little detail about possible differences in foreign policy. By contrast, theories of arms and alliances as complements include diverse foreign policy aims.

Arms and Allies as Complements

The literature on arms and alliances as complements is less well developed. These theories expect that arms and alliances are positively correlated— alliance membership leads to increased military spending. There are several tentative explanations for this prediction.

Diehl (1994) argues that arms and alliances are not comparable policies and have multiple purposes besides deterrence, so substitution is unlikely. He then lists several possible mechanisms behind complementarity between arms and alliances, but does not articulate which is most important. These mechanisms include the possibility that alliances require increased foreign policy activity, or that major powers increase military spending to cover free-riding by junior alliance partners. The most general argument asserts that autonomy benefits from an alliance increase the marginal value of military spending, which is the main argument of complementarity theories.

Morgan and Palmer (2006) also use an increase in the marginal benefit of military spending to explain complementarity between arms and alliances. Their theory starts with the premise that states have two foreign policy goals— maintenance and change. States then allocate all their foreign policy resources to pursue a mix of maintenance and change across a range of issues. Due to budget constraints, there is a tradeoff between maintenance and change. Within this framework, Morgan and Palmer examine substitution between policies by starting with the assumption that states choose the most efficient policy tool to realize their goals.

Given the efficiency assumption, there are three reasons states might increase their use of a particular policy. The first is an increase the overall resources available to the state, and the second is changes in relative preferences for maintenance and change. The last path is changes in the efficiency of different policies, which is the only path that leads to substitution. In this model, substitution is motivated by states moving resources from less to more efficient policies, which they claim is uncommon.

Rather, Morgan and Palmer argue that alliances increase the amount of resources available to the state, which frees up additional resources to pursue other goals. Alliances expand the foreign policy possibilities for a state, which makes them more likely to increase military spending. As an illustrative example, they argue that an alliance with the US gave Britain and France extra resources to pursue change in the Middle East during the 1957 Suez crisis.

Because states do not expend the same kinds of resources on domestic arms and alliances, the concept of a foreign policy "budget" is nebulous, and it does not incorporate domestic policy at all. Neither Diehl nor Morgan and Palmer adequately explain the process by which joining an alliance increases military spending. Greater clarity about the mechanism by which alliances increase the marginal benefits of military spending is necessary. Otherwise, we are left with theories that have minimal process, and merely assert that arms and alliances have complementary characteristics.

A different complementarity mechanism is cooperation and policy coordination between alliance members. The cooperation thesis acknowledges collective action problems among alliance members, but asserts that these problems can be overcome. Palmer (1990) conceptualizes relations among alliance members as an iterated prisoner's dilemma, where states can communicate and use side issues to overcome collective action problems. Quiroz Flores (2011) attributes a positive spatial correlation in the defense burden of allied states to increased cooperation between allies.

There is evidence that alliances facilitate cooperation in other issue areas (Gowa, 1995; Gowa and Mansfield, 2004; Poast, 2012, 2013). But payments on side issues will not necessarily lead to increases in spending. Small states can use concessions on other issues as part of a bargain with their patron that allows for reduced military spending— this is the same kind of exchange that makes the concept of free-riding problematic. Intra-alliance cooperation is an insufficient explanation of complementarity between arms and allies because it avoids dealing with exchange and elides domestic political incentives to reduce defense spending.

The challenge for complementarity theory is clear— scholars must develop better theoretical explanations for their prediction that alliance membership leads to increases in military spending. The shortcomings of current theories do not imply that scholars should rule out complementarity between arms and alliances, however. Theories of arms and allies as complements still make a valuable contribution to knowledge of the arms alliances tradeoff by considering diverse foreign

policy goals.

The State of Arms-Allies Theory

Given arguments that arms and alliances are complements or substitutes, what is the overall theoretical development of the literature? The two perspectives have distinct strengths and weaknesses, but debate over whether arms and allies are substitutes or complements treats the two perspectives as incompatible. Framing substitutes and complements as competing hypotheses has stunted theoretical progress, as neither perspective incorporates the key insight of the other.

Substitution theory has not engaged with the claim that states have diverse foreign policy goals. Complementarity theories ignore domestic political incentives for security-seeking states to rely on allied capability. Because the strengths and weaknesses of these perspectives mirror each other, it may be possible to combine the two. Rather than pitting complementarity and substitution theories against each other, a general approach should consider whether each applies in different circumstances. This combination implies a conditional association between arms and alliances.

DiGiuseppe and Poast (2016)'s theory is a useful first step towards a conditional theory. They argue that states will only reduce substitute alliances for arms if the alliance is sufficiently credible, and that democracies make more credible alliance promises. By acknowledging heterogeneous effects of alliances, their theory and results suggest that arms and alliances are not uniformly substitutes or complements.

Regardless of how theories of the arms-allies tradeoff evolve, they must offer more precise hypotheses. Many theories do not specify clear observable implications of an arms-alliances tradeoff, beyond some advice to focus on marginal changes (Morrow, 2000; Starr, 2000). Others are less clear about the scope conditions of their theories. Without precise theoretical predictions, the risk of spurious results from poor research designs increases.

The debate over free-riding in alliances is an example of the research design problems that result from inadequate attention to scope conditions. Most scholars ignore possible differences between major and non-major powers. The economic theory of alliances expects greater free riding by smaller alliance members, but many tests of this theory examine alliances between major powers (Thies, 1987; Conybeare and Sandler, 1990; Siroky, 2012), which may not have large enough size differences.

The scope conditions problem points to another major challenge in scholarship on the arms-alliances tradeoff— many empirical results have limited viability. Absent reliable empirical results, assessing the relative value of different theories is impossible. Despite the general claims in different theories, most studies of the arms-allies tradeoff provide poor evidence for their claims. Many researchers use limited samples of states or alliances, or fail to account for critical empirical problems.

Empirical Challenges

Estimating the association between arms and alliances is difficult. The following section summarizes research design in studies of the arms-allies tradeoff, and whether these studies addressed key problems. Any research design must consider causes of simultaneous change in arms and allies, measurement, and scope conditions. Besides these theoretical aspects of research design, military spending is a challenging dependent variable. National military spending is highly autocorrelated, and its skewed distribution is problematic for conventional regression estimators.

Few studies estimate the general association between alliances membership and arms. Only six researchers use military expenditures as the dependent variable, and account for the presence of multiple alliances. Five of these six models are untrustworthy, as they do not control for key correlates of spending and use limited samples.

For example, Morgan and Palmer (2003) regress changes in military expenditure on indicators of state and alliance capabilities. Their sample is the year after a state's alliance portfolio changed, and the model includes no control variables, raising the risk of selection and omitted variable bias. Other studies focus only on great power alliances (Conybeare and Sandler, 1990; Conybeare, 1994; Diehl, 1994; Most and Siverson, 1987), with little attention to possible differences in how these states make defense allocations. While great powers face salient security threats, a better empirical approach is to control for differences in threat.

Eighteen other studies examine dynamics within specific alliances. These research designs address alliance-level variation in more detail, looking at specific changes in the political context or nature of the alliance itself. Seven of these studies examine NATO or the Warsaw Pact only. Much of the work on specific alliances tests whether larger members have a higher defense burden, in accordance with the primary prediction of the economic theory of alliances. Early studies of free-riding in specific alliances rely on correlation coefficients or regressions with 20 or fewer observations (Olson and Zeckhauser, 1966; Starr, 1974; Reisinger, 1983; Siroky, 2012), where there is a high risk of spurious results.

The focus on specific alliances is driven by data availability— NATO members and great powers have more data on military spending. Even so, some focused studies provide plausible estimates. Plümper and Neumayer (2015)'s quasi-spatial model of how NATO members respond to US and Soviet spending is the best evidence of free-riding by NATO members. Sorokin (1994) takes a similar approach to estimating the responses of four states to the military spending of a key ally with a two-stage least squares GLS estimator.

Most studies, whether specific or general, ignore autocorrelation in the data-generating process of military spending. Past military spending is an excellent predictor of future spending. Therefore time-series observations within states are not independent, but many studies treat these observations as independent. Four of the 26 studies use an estimator or specification that incorporates autocorrelation in military spending. Four others use a differenced dependent variable.

Even when accounting autocorrelation, the standard linear regression estimator may be inefficient. When the regression residuals are not normally distributed, OLS is still unbiased, but it is not the minimum variance estimator (Rainey and Baissa, 2017). As Figure 1 shows, military spending from 1950-2001 is not normally distributed, even after applying a log transformation. As a result, regression models that use this data will likely have skewed residuals as well, making their
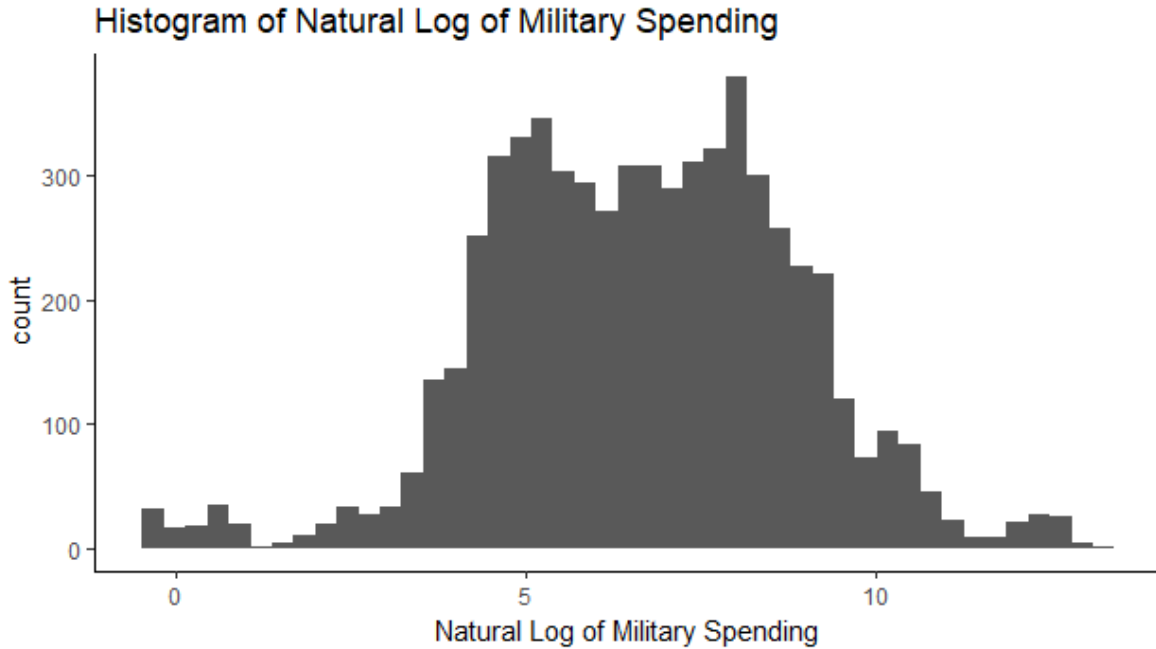
Figure 1: Distribution of Annual Military Spending by all States: 1950-2001

estimates inefficient.

Given all of these concerns, the best model of the arms alliances tradeoff is the work of DiGiuseppe and Poast (2016), who find that substitution of alliances for arms is conditional on the credibility of the alliance. They argue that that democracies make more credible commitments, and estimate regression models of the association between defense pacts with a democracy and the natural log of military spending, using a lagged dependent variable to account for autocorrelation. This research design is a marked improvement over prior tests of how alliances affect domestic military spending.[3]

Arms Costs and Alliance Formation

Only two studies consider the another facet of the arms-alliances tradeoff— how the opportunity costs of military spending affect alliance formation. Kimball (2010); Allen and DiGiuseppe (2013) both use panel data, and expect that states with high social demand or costs of sovereign borrowing will be more likely to form alliances. While they find the expected positive correlation between the key independent variable and a binary indicator of whether a state joined an alliance in a given year, neither model fits the data well.[4]

One way to assess model fit, especially with a binary dependent variable, is compare the predicted probabilities of an event from the the model to the observed values of the dependent variable.

---

[3]While their regression estimator does not account for heavy tails in the dependent variable, the results are robust to using a robust regression estimator.

[4]Allen and DiGiuseppe's model fits better than Kimball's model.

Both models assign low predicted probabilities of alliance formation to many observations where formation occurred, as shown by the separation plots in Figure 2. Poor model fit casts doubt on the validity of inferences about the covariates (Montgomery, Hollenbach and Ward, 2012; Muchlinski et al., 2015). Even in state-year panel data, alliance formation is rare, making it more difficult to predict.[5]

Besides poor model fit, these two research designs indirectly test the logic of substitution. The key independent variable differentiates between states with high opportunity costs of military spending. But this does not account for differences in national defense burden, or the absolute costs of defense.

Using panel data with a binary indicator of whether a state formed an alliance in a particular year ignores heterogeneity among alliances. Alliances make different promises, and provide different amounts of capability to support those promises. A defense pact with the United States provides more security than a neutrality pact with Senegal, for instance. The choice of alliance types and partners is neglected by these theories and research designs.

Overall Empirical Assessment

Although theoretical progress depends on the accumulation of evidence, we have little reliable evidence from 27 studies. What we know about the arms-alliance tradeoff depends on a few studies. Other research designs are narrow, flawed, or both.

Many of the research design problems in past studies are the result of a levels of analysis problem. There are two distinct levels of analysis in the arms-alliances tradeoff, because international organizations such as alliances are separate from states (Mattes, 2012; Chiba, Johnson and Leeds, 2015). This difference creates serious measurement problems.

Measurement of alliance membership is inconsistent across studies. In studies of alliance membership and military spending, scholars have relied on crude summaries of the overall alliance portfolio, or only estimated the impact of a few alliances. Some scholars use a dummy indicator of whether a state has at least one alliance, while others use allied expenditures or capability. Most studies of specific alliances use allied capabilities or expenditures, while general studies use dummy variables.

State-level dummy indicators of alliance membership are the most common independent variable, but this measure can create a mismatch between the theory and empirical test. For instance, DiGiuseppe and Poast (2016) emphasize differences in alliances, but then examines differences between states. The binary independent variable captures the presence of a defense pact with a democracy, and coupled with a control for defense pacts with non-democracies, they compare states that have a defense pact with a democracy to states with no defense pacts. Their main regression coefficient does not compare alliances, as the theory would lead you to expect.

Because they average out between-alliance variation, state-level dummies also remove critical details. For instance, democracies tend to form more limited alliance commitments (Chiba, Johnson and Leeds, 2015), so democratic membership and conditionality are correlated. Other factors such as superpower membership, institutionalization, military aid are also correlated with key al-

---

[5]Only 4% of Kimball's observations and 7% of Allen and DiGiuseppe's observations see alliance formation.
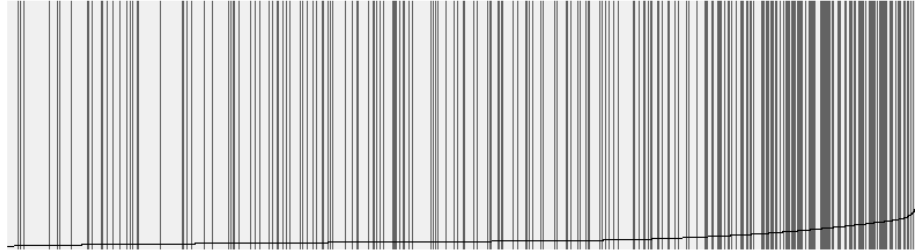
Kimball 2010

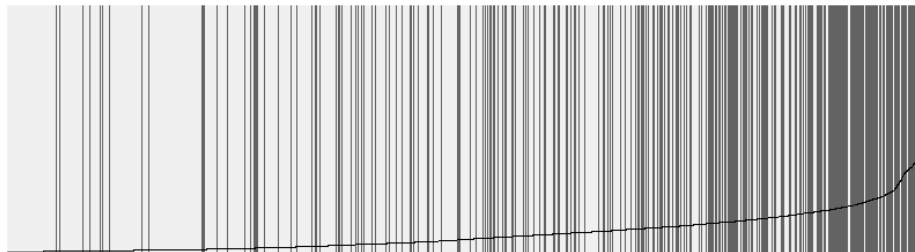**Allen and Digiuseppe 2013**

Figure 2: Separation plots of primary models from Kimball 2010 and Allen and DiGiuseppe 2013. A separation plot summarizes model fit by ordering observations according to the predicted probability assigned by the model. Dark lines mark observations where the observed value of the dependent variable is one. Better fitting models assign higher predicted probabilities to those observed realizations of the dependent variable, which will be reflected by clustering on the right-hand side of the plot. Greater dispersion of the dark lines indicates that the model does a poor job predicting the outcome of interest.

liance conditions. Currently, these alliance characteristics are omitted variables, which could affect our estimates of how alliance membership affects military spending.

Empirical work on the arms-alliances tradeoff faces similar challenges to research on the consequences of membership in international organizations. Selection bias is a ubiquitous concern in that literature (Chaudoin, Hays and Hicks, 2016), but is not acknowledged in past research designs. And while theory recognizes possible simultaneity between arms and alliances, only one paper ((DiGiuseppe and Poast, 2016)) uses a simultaneous model. As with theories of the arms-allies tradeoff, there is ample need for improvement in research design.

Conclusion

What can we conclude about our understanding of the arms-alliances tradeoff? Despite substantial attention to this question, we know surprisingly little about the connection between alliance behavior and military spending. Substitution and complementarity theories provide some useful insights, but most research designs suffer from serious shortcomings. As a result, there is ample room for theoretical and empirical progress in this literature.

Progress will require less emphasis on free-riding. While free-riding is a useful rhetorical and analytic device, its applicability to alliances is uncertain at best. Distinguishing between exchange among alliance partners and free riding is difficult theoretically and empirically. Scholars would be better served by considering how alliance membership affects defense spending, without the distraction of trying to identify models of relative defense burdens. Given the policy salience of free-riding claims, scholars should not ignore this concept, but it creates more problems than it solves for scholarship on the arms-alliances tradeoff.

To improve our understanding of the arms-alliances tradeoff, scholars should improve existing theory and research design in work on how alliance membership affects military spending, and then consider novel research questions. To start, theory must move past the complements-substitutes debate.

Theory

The relationship between arms and alliances is probably conditional— arms and alliances are not uniformly complements or substitutes. The best research design finds only credible alliances are a substitute for domestic arms, which suggests that a conditional explanation is plausible. Scholars can use components from complementarity and substitution theories identify when alliance membership decreases or increases spending. This endeavor should incorporate domestic incentives to reduce military spending and diverse foreign policy goals, and can build on a shared aspect of the two theories.

Complementarity and substitution theories use an income effect to connect alliances with changes in military spending. Income effects address how an actor changes the mix of goods it consumes after an increase in its budget. Both paradigms expect the capability gains from an alliance to lead to consumption changes, the difference is that they focus on changes in of different

goods. Substitution theory emphasizes shifts in domestic consumption, while complementarity theories consider changes in foreign policy.

To identify when arms and allies are complements or substitutes, the key task is articulating when states emphasize domestic or international gains. Alliance treaty design may reflect the focus of their members. Alliances can produce security or autonomy for members, which are distinct purposes (Morrow, 1991). Security reflects interests in maintaining the international status quo, while autonomy is the ability of states to seek change. Security-producing alliances will facilitate domestic gains, while autonomy-generating alliances will lead to foreign policy gains. If not the security-autonomy distinction, alliance heterogeneity will play an important role in theory.

In addition to differences between alliances, scholars should address changes in the nature of military spending. The price and complexity of conventional weapons has increased over time, especially in the 20th century (Bitzinger, 2003). The marginal capability purchased by a dollar of military spending is decreasing, which affects the purchasing power of small states, and may alter their mix of security policies. And as the time needed to build of complex weapons systems increases, domestic arms will become a more imperfect substitute for alliances, because capability gains from arms investment will take longer to realize.

Last, theory must provide better guidance for research design. Clear scope conditions are indispensable. If smaller states are more likely to use alliances as a source of security, rather than to increase their foreign policy autonomy (Morrow, 1991), then autonomy-seeking major powers have less incentive to substitute alliances for arms. Other theories beyond the economic theory of alliances must consider these and other differences between major and minor powers.

Besides articulating scope conditions, theory must clarify whether alliance membership or capability impacts domestic arms. It is reasonable to expect that more capable alliances are more valuable to their members. So the capability an alliance brings is a key source of alliance heterogeneity, and research design should reflect this.

Research Design

Future research designs should connect to the implications of new theory developments, and chart a middle course between the specific and general studies in previous work. General studies of how the association between alliance behavior and military spending covered a wide universe of cases, but ignored alliance heterogeneity. Studies of specific alliances sacrificed generality for detailed examination of particular treaties.

The gap between specific and general studies reflects difficulties from moving between different levels of analysis. (Chaudoin, Milner and Pang, 2015) provide a theoretical and empirical framework of how scholars can consider the role of different levels of analysis in international relations. While they focus on the association between systemic and domestic factors, their approach can be extended to cover membership in international institutions like alliances.

Multilevel models can compromise between the specific and general approaches by incorporating the alliance and state levels of analysis. Modeling variation at the alliance and state levels will allow for more detailed comparisons of states and alliances. These comparisons are essential for testing any conditional theories that ascribe differences in the arms-alliances tradeoff to different alliance characteristics.

For modeling how the cost of arms affect alliance formation, it is unclear whether the panel data research designs of Kimball (2010); Allen and DiGiuseppe (2013) are the best approach. Other approaches to modeling alliance formation, including dyadic, k-adic, or network analysis, could capture strategic elements of alliance formation, which are otherwise omitted. But the panel design can be incorporated into a simultaneous model with alliance formation.

Simultaneity and selection bias are critical issues, but there are several possible solutions. In single-equation models, theory can provide guidance about observable correlates of selection that would otherwise generate omitted variable bias. Scholars could also check for selection on unobservables in these models (Chaudoin, Hays and Hicks, 2016). Simultaneous models of arms and alliances may address selection bias by modeling the alliance formation process and the correlation between the two policies. Even as work to improve research design continues, scholars should examine other open research questions.

Open Research Questions

Some open research questions are simple empirical tests of other predictions from substitution and complementarity theories, besides the association between alliance membership and military spending. Testing these other predictions will provide additional evidence about the accuracy of these theories. For instance, we know very little about how the cost of domestic arms affects alliance formation, though substitution theory expects a positive relationship between these variables. This test will need a direct measure the costs of defense. The defense burden or the costs of leading weapons systems are two possible measures.

Other research questions are more novel. Do changes in the rate of alliance fulfillment over time affect domestic military spending? Do other characteristics of an alliance besides democratic membership affect its credibility? What is the role of arms imports for protege states? How does the international context alter the arms-alliances tradeoff?

Understanding the role of the international context in the arms-alliances tradeoff is especially important. Between nuclear weapons, changing norms about international conquest (Fazal, 2011), and economic interdependence (Frieden, 2006), the international environment has changed in dramatic ways, especially after 1945. So far, no theories of the arms-alliances tradeoff have addressed whether these shifts in the international environment alter the value of alliances or military spending, and whether context is a control or modifying variable.

These research questions have tremendous academic and policy importance. The academic rhetoric of free-riding has an outsized impact on popular perceptions of how alliance membership affects defense spending. But the true association may be more nuanced, which should affect our view of the value and purpose of alliances.

Despite the importance of arms and alliances, there is much we do not know about the relationship between these policies. Improving theory and research design to address how alliance membership affects military spending is the first task. Then scholars can address other open research questions to increase our understanding of the arms-alliances tradeoff.

Appendix

Visual Summary of Prior Results

	Substitutes	Complements	Null
Most and Siverson (1987)			X
Morrow (1993)	X		
Conybeare (1994)	X		
Diehl (1994)		X	
Goldsmith (2003)			X
Morgan and Palmer (2006)		X	
Quiroz Flores (2011)		X	
DiGiuseppe and Poast (2016)	X		
Horowitz, Poast and Stam (2017)		X	

Table 1: General Findings of Association Between Alliance Membership and Military Spending

	Free-Riding	Alliance Scope
Olson and Zeckhauser (1966)	X	NATO 1964
Starr (1974)	X	Warsaw Pact 1967-1971
Reisinger (1983)		Warsaw Pact 1970-1978
Thies (1987)	Mixed	Seven pre-1945 Alliances
Conybeare and Sandler (1990)		Triple Alliance & Triple Entente 1880-1914
Palmer (1990)		NATO 1950-1978
Chen, Feng and Masroori (1996)		Arab League 1950-1988
Oneal and Whatley (1996)	X	NATO, Rio Pact & Arab League 1953-1988
Siroky (2012)		Quintuple Alliance Members 1820
Plümper and Neumayer (2015)	X	NATO 1956-1988

Table 2: Findings on Free-Riding in Alliances. Most of these studies uses military expenditures as a share of GDP for the dependent variable, and assess whether it is correlated with economic size. A positive correlation between GDP or GNP and the defense burden among alliance members is the standard evidence of free-riding.

References

Allen, Michael A and Matthew DiGiuseppe. 2013. "Tightening the Belt: Sovereign Debt and Alliance Formation." *International Studies Quarterly* 57(4):647–659.

Bitzinger, Richard A. 2003. "Towards a Brave New Arms Industry?" *The Adelphi Papers* 356(1):63–79.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2002. *The Logic of Political Survival*. Cambridge: Mass: The MIT Press.

Chaudoin, Stephen, Helen V Milner and Xun Pang. 2015. "International Systems and Domestic Politics: Linking Complex Interactions with Empirical Models in International Relations." *International Organization* 69(2):275–309.

Chaudoin, Stephen, Jude Hays and Raymond Hicks. 2016. "Do We Really Know the WTO Cures Cancer?" *British Journal of Political Science* pp. 1–26.

Chen, Baizhu, Yi Feng and Cyrus Masroori. 1996. "Collective action in the Middle East? A study of free-ride in defense spending." *Journal of Peace Research* 33(3):323–339.

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.

Conybeare, John A. C. 1992. "A Portfolio Diversification Model of Alliances." *Journal of Conflict Resolution* 36(1):53–85.

Conybeare, John AC. 1994. "Arms versus Alliances: The Capital Structure of Military Enterprise." *Journal of Conflict Resolution* 38(2):215–235.

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.

Deger, Saadet and Somnath Sen. 1995. Military Expenditure and Developing Countries. In *Handbook of Defense Economics*, ed. Keith Hartley and Todd Sandler. Vol. 1 Elsevier pp. 275–307.

Diehl, Paul F. 1994. "Substitutes or Complements?: The Effects of Alliances on Military Spending in Major Power Rivalries." *International Interactions* 19(3):159–176.

DiGiuseppe, Matthew and Paul Poast. 2016. "Arms versus Democratic Allies." *British Journal of Political Science* pp. 1–23.

Fazal, Tanisha M. 2011. *State death: the politics and geography of conquest, occupation, and annexation*. Princeton: Princeton University Press.

Frieden, Jeffrey. 2006. *Global Capitalism: Its Fall and Rise in the Twentieth Century*. New York: W.W. Norton.

Gates, William R and Katsuaki L Terasawa. 1992. "Commitment, Threat Perceptions, and Expenditures in a Defense Alliance." *International Studies Quarterly* 36(1):101–118.

Goldsmith, Benjamin E. 2003. "Bearing the Defense Burden, 1886-1989: Why Spend More?" *Journal of Conflict Resolution* 47(5):551–573.

Gowa, Joanne. 1995. *Allies, Adversaries, and International Trade*. Princeton University Press.

Gowa, Joanne and Edward D. Mansfield. 2004. "Alliances, Imperfect Markets, and Major-Power Trade." *International Organization* 58(4):775–805.

Hansen, Laurna, James C Murdoch and Todd Sandler. 1990. "On distinguishing the behavior of nuclear and non-nuclear allies in NATO." *Defence and Peace Economics* 1(1):37–55.

Horowitz, Michael C, Paul Poast and Allan C Stam. 2017. "Domestic Signaling of Commitment Credibility: Military Recruitment and Alliance Formation." *Journal of Conflict Resolution* 61(8):1682–1710.

Kimball, Anessa L. 2010. "Political Survival, Policy Distribution, and Alliance Formation." *Journal of Peace Research* 47(4):407–419.

Lanoszka, Alexander. 2015. "Do Allies Really Free Ride?" *Survival* 57(3):133–152.

Mattes, Michaela. 2012. "Reputation, Symmetry, and Alliance Design." *International Organization* 66(4):679–707.

Montgomery, Jacob M, Florian M Hollenbach and Michael D Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20(3):271–291.

Morgan, T Clifton and Glenn Palmer. 2003. "To protect and to serve: Alliances and foreign policy portfolios." *Journal of Conflict Resolution* 47(2):180–203.

Morgan, T Clifton and Glenn Palmer. 2006. *A Theory of Foreign Policy*. Princeton: Princeton University Press.

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.

Morrow, James D. 1993. "Arms versus allies: trade-offs in the search for security." *International Organization* 47(2):207–233.

Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3:63–83.

Most, Benjamin A and Harvey Starr. 1989. *Inquiry, Logic and International Politics*. Columbia: University of South Carolina Press.

Most, Benjamin A. and Randolph M. Siverson. 1987. Substituting Arms and Alliances 1870–1914: an Exploration in Comparative Foreign Policy. In *New Directions in the Study of Foreign Policy*, ed. Charles F Hermann, Charles W. Kegley Jr and James N. Rosenau. Boston: Allen & Unwin pp. 131–160.

Muchlinski, David, David Siroky, Jingrui He and Matthew Kocher. 2015. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24(1):87–103.

Murdoch, James C. 1995. Military Alliances: Theory and Empirics. In *Handbook of Defense Economics*, ed. Keith Hartley and Todd Sandler. Vol. 1 Elsevier pp. 89–108.

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.

Poast, P. 2012. "Does Issue Linkage Work? Evidence from European Alliance Negotiations, 1860 to 1945." *International Organization* 66(1):277–310.

Poast, Paul. 2013. "Can Issue Linkage Improve Treaty Credibility? Buffer State Alliances as a Hard Case." *Journal of Conflict Resolution* 57(5):739–764.

Quiroz Flores, Alejandro. 2011. "Alliances as Contiguity in Spatial Models of Military Expenditures." *Conflict Management and Peace Science* 28(4):402–418.

Rainey, Carlisle and Daniel K. Baissa. 2017. "When BLUE Is Not Best: Non-Normal Errors and the Linear Model.".

Reisinger, William M. 1983. "East European military expenditures in the 1970s: collective good or bargaining offer?" *International Organization* 37(1):143–155.

Sandler, Todd and Keith Hartley. 2001. "Economics of alliances: The lessons for collective action." *Journal of Economic Literature* 39(3):869–896.

Schweller, Randall L. 1994. "Bandwagoning for Profit: Bringing the Revisionist State Back In." *International Security* 19(1):72–107.

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.

Sorokin, Gerald L. 1994. "Arms, Alliances, and Security Tradeoffs in Enduring Rivalries." *International Studies Quarterly* 38(3):421–446.

Starr, Harvey. 1974. "A Collective Goods Analysis of the Warsaw Pact after Czechoslovakia." *International Organization* 28(3):521–532.

Starr, Harvey. 2000. "Substitutability in Foreign Policy: Theoretically Central, Empirically Elusive." *Journal of Conflict Resolution* 44(1):128–138.

Thies, Wallace J. 1987. "Alliances and Collective Goods: A Reappraisal." *Journal of Conflict Resolution* 31(2):298–332.

Walt, Stephen M. 2009. "Alliances in a Unipolar World." *World Politics* 61(01):86–120.

Whitten, Guy D. and Laron K. Williams. 2011. "Buttery Guns and Welfare Hawks: The Politics of Defense Spending in Advanced Industrial Democracies." *American Journal of Political Science* 55(1):117–134.