
Brett V. Benson*    Joshua D. Clinton†

May 25, 2012

**Keywords:** Alliances; Measurement; Item Response Theory

*Assistant Professor of Political Science, Vanderbilt University, E-mail: brett.benson@vanderbilt.edu
†Associate Professor of Political Science and Co-Director of the Center for the Study of Democratic Institutions, Vanderbilt University. E-mail: josh.clinton@vanderbilt.edu. PMB 505, 230 Appleton Place, Nashville TN, 37203-5721.
Abstract

Alliances play a critical role in the international system and understanding the determinants and consequences of their strength is an important task. Even though many have argued that the strength of an alliance is theoretically determined by both the power of the signatories involved and the formal terms of the agreement, using these insights to measure the strength of alliances is difficult for many reasons. Using a statistical measurement model, we estimate the strength of all alliances signed between 1816-2000 along two theoretically derived dimensions: the strength of the signatories involved, and the strength of the formal terms of the alliance. In addition to estimating the strength of every alliance in these two dimensions, our Bayesian latent variable model also allows us to: characterize the relationship between the two dimensions of alliance strength, identify how observable characteristics relate to each of the recovered dimensions, and, perhaps most importantly given the difficulty of the task, document how precisely we are able to use the measures to estimate alliance strength. By generating estimates of every alliance signed between 1816 and 2000, we not only provide scholars with an empirical characterization of alliance strength in two dimensions of conceptual interest that should prove vital for future studies of international alliances, but the flexibility of our measurement model also offers additional opportunities to refine and extend our measure.
Understanding alliances is critical for understanding the international system and a great deal of effort has been devoted to their study. The strength of an alliance is central to many theories about alliances, including: the capabilities aggregation theory of alliance formation (Morgenthau 1978, Organski 1968, Waltz 1979), balance-of-power theory and bandwagoning (Waltz 1979; Walt 1987), chain ganging and buck passing (Christensen and Snyder 1990), burden sharing (Olson and Zeckhauser 1966; Sandler 1977), coalition formation (Riker 1962; Niou and Ordeshook 1994), alliance reliability (Sabrosky 1980, Siverson and King 1980, Smith 1996, Leeds et al 2002, Leeds 2003), deterrence and war (Gartner and Siverson 1996, Altfeld and Bueno de Mesquita 1979, Morrow 1994, Smith 1995, Vasquez 2008), asymmetrical alliance formation (Morrow 1991), moral hazard (Yuen 2009, Benson 2012, Benson et al 2012), and deterrence versus restraint (Fearon 1997; Snyder 1997; Snyder 1984, 1997; Zagare and Kilgour 2003, 2006; Pressman 2008; Benson 2012). Because alliances differ in many respects, measuring the nature of the differences is critically important for understanding their consequences and determinants. Without a theoretically defensible measure of alliance strength, it is difficult to empirically characterize alliances and investigate their potential causes and consequences.

Because of the vast amount of information that has been used to empirically describe alliances (see, for example, Singer and Small 1966; Small and Singer 1969; Leeds et. al. 2002; and Gibler and Sarkees 2004) the question of how best to use the available information in theoretically informed ways to empirically characterize the strength of alliances is difficult to resolve. At issue is the fact that while we can observe the formal terms of an alliance agreement and measure characteristics of the countries involved in an alliance, it is not entirely clear how best to use the wealth of available information to characterize the strength of an alliance along these dimensions while also accounting for the inherent uncertainty that must be present in such a determination. Similar to the task of measuring the ideology of elected and unelected officials (e.g., Poole and Rosenthal 1997; Martin and Quinn 2002; Clinton, Jackman and Rivers 2004), the positions of a political party in an underlying policy
space (Budge et. al. 2001), the extent to which a country is democratic (Pemstein, Meserve, and Melton 2010), the positions taken by a country in the United Nations General Assembly (Voeten 2000), or the significance of a piece of legislation (e.g., Mayhew 2005), the essential measurement question that scholars interested in studying alliances face is: how can the observable characteristics of alliance agreements best be used to characterize the strength of alliances in the international system?

We propose a statistical measurement model based on theoretical arguments about the correlates of alliance strength that uses the observable characteristics of an alliance and the associations between these characteristics along with theoretically motivated arguments about the dimensions of alliance strength to estimate the strength of alliance agreements at the time of their formation. In so doing, we make several important contributions to the study of alliances. First, we show how a Bayesian latent trait model (Quinn 2004) can account for scholars’ theoretical arguments about the correlates and determinants of alliance strength and recover a theoretically informed, multi-dimensional estimate of alliance strength that reflects both the terms of the formal alliance agreement and the characteristics of the signatories. Not only do our estimates reflect theoretically relevant dimensions of alliance strength, but they also reflect the information contained in multiple observable characteristics while making relatively few assumptions compared to available alternatives.

Second, because any assessment of alliance strength is inherently ambiguous given the nature and difficulty of the task, our measurement model enables us to quantify how certain we are about the resulting estimates (or even any function of the estimates). Quantifying uncertainty is a critical task for any science – perhaps particularly when we are dealing with a concept such as “alliance strength” that is inherently difficult to measure (Jackman 2009b).

Third, our method is sufficiently general that we can extend the method to all alliances and treaties – including multi-level alliances and alliances without a target – signed between 1816 and 2000. (The period covered by the databases collected by Leeds et. al. (2004), Gibler and Sarkees (2004), and integrated by Bennett and Stam (2000) in EUGene v3.204.)
Finally, although we think our estimates are based on strong theoretical foundations and possess strong conceptual validity, the statistical measurement model we employ empowers scholars to construct their own measures if their questions of interest are sufficiently different or if they choose to make alternative assumptions about the underlying relationships. The statistical measurement model we use is sufficiently general so that scholars can adjust the model to include additional dimensions or characteristics.

To be clear, our assessment of alliance strength is not based on the consequences of the alliance. That is, we do not use the success or failure of the alliance in the international area to determine the strength of an alliance – such an approach would be clearly problematic and provide no insights into how various features of an alliance may or may not be related with outcomes. Instead, we use the formal terms of the alliance agreement and characteristics of the countries involved at the time of the signing to construct a measure of alliance strength based on the underlying associations of the observable measures. The critical assumption that is required to do so is that strong alliances share similar features in terms of the characteristics of their signatories and their formal alliance terms and that by statistically analyzing the associations among observable characteristics we can extract the latent dimensions of strength that structure the variation in alliances that we observe.

The outline of the paper is as follows. Section 2 briefly recaps the extensive literature dealing with the strength of international alliances to extract the primary dimensions that scholars have identified as influencing the strength of an alliance – the terms of the alliance and the characteristics of the signatories. Section 3 describes the Bayesian latent variable model we use to measure alliance strength and it describes the observable characteristics we use to estimate our two-dimensional estimate of alliance strength. Section 4 describes the resulting estimates and discusses their validity. Section 5 concludes by discussing the possible uses and extensions of both our estimates of alliance strength and also the Bayesian latent variable measurement model we employ.
1 Conceptualizing Alliance Strength

The strength of an alliance is central to many questions related to the study of conflict. It underlies inquiries about the effectiveness of alliances in achieving their security objectives and the reasons states form alliances. Many early studies argued that governments formed alliances to aggregate capabilities in an effort to offset external threats (Morgenthau 1978, Organski 1968, Waltz 1979). Balance of power research and derivative studies built on the idea that governments pool capabilities to form balancing or bandwagoning alliances (Walt 1987). Christensen and Snyder (1990) argued that there may be externalities to strengthening alliances because of “chain-ganging” and “buck passing” between alliance members. The notion that there are constraints on the optimal strength of an alliance also related to research on burden-sharing and moral hazard. Some argue that governments make decisions about how much to contribute to an alliance given alliance members’ incentives to free ride on one another (Olson and Zeckhauser 1966, Sandler 1977, Sandler 1993, and Conybeare 1994). Others claim that governments form alliances to build strength through coalitions subject to natural constraints on the size of the alliance or the number of allies to be included (Riker 1962; Niou and Ordeshook 1994). Research on moral hazard within an alliance focuses on the effect that strengthening an alliance has on the behavior of alliance members (Snyder 1984, 1997; Yuen 1997; Benson et al 2012; Benson 2012). Still more studies argue that the content of an alliance agreement might restrain alliance members by creating uncertainty about whether an alliance member will intervene or limiting the domain in which alliance members’ military obligations are relevant (Snyder 1984, 1997; Pressman 2008; Zagare and Kilgour 2003, 2006; Benson 2012). An implication from all these studies is that governments design the content of an agreement to titrate the strength of an alliance, and a major goal of alliance formation is to combine signatories’ military resources to achieve a security goal.

Many theories focus on the security benefits of strong alliances. A common feature of formal theories of alliances is a focus on the effect of an alliance on an alliance member’s
war payoff (Morrow 1994, Smith 1995, Zagare and Kilgour 2003/6, Yuen 2009, Benson 2012). This approach directly connects the strength of an alliance to its deterrence benefits as well as to its potential moral hazard effects, and theorists have analyzed governments’ decisions about the optimal strength of an alliance. Snyder (1997), for example, argues that prospective allies bargain over how strong to form an alliance given the objectives of alliance members and their risks, and he uses Kennedy’s (1984) estimates of European country’s power to show that the pre-World War I alliances emerged as a result of bargaining over how strong to design their alliances given the strength of opposing alliances, government’s individual risks, and the capabilities of respective states. Benson, Meirowitz and Ramsay (2012) propose a model that endogenizes the choice of how strong of an alliance to form so as to deter a third party threat. The strength of an alliance also figures prominently in Morrow’s (1991) theory of asymmetrical alliances, which argues that alliances form as a result of a mutually beneficial exchange between states that have comparative advantages in security and states that have comparative advantages in supplying political influence. Relatedly, the diversity of risk theory of alliances concentrates on the returns and risks associated with a government’s investment in an alliance, where the return is the sum of the military efforts of the allies (Conybeare 1994a, 1994b). Finally, alliance strength is a key factor that determines whether an alliance is reliable. Central to the question about the reliability of an alliance is a calculation on the part of an alliance member about whether an alliance is strong enough to influence the behavior of a third-party challenger. The third party may not challenge the alliance if it perceives the alliance to be sufficiently powerful that honoring the alliance is cost efficient even in war for the alliance members. Thus, challengers are likely to target weak alliances. Accordingly, prospective alliance members might prefer to join an alliance if they believe the combined strength of the alliance members committed toward a specified objective laid out in an alliance agreement will induce the desired response in a third party (Smith 2005, 2006; Leeds et al 2002). In all these theories, the strength of an alliance is critical for understanding whether an alliance gets formed, how it is designed, and whether
Empirically assessing these theories is difficult because the strength of an alliance is not directly measurable and scholars are forced to infer strength based on observable variables such as the aggregate capabilities of the alliance members (Bueno de Mesquita 1983, Morrow 1991, Smith 1996, Poast et al 2012). These measurement choices are consequential. For example, the aggregate capabilities of the members of an alliance approximates alliance strength only if the act of forming an alliance results in an unconditional guarantee of automatic and unlimited support in any war involving an ally, and resources are efficiently transferred across alliance members without cost. However, few, if any, agreements satisfy these conditions, and there are many factors that limit how much of the overall potential capabilities may actually be mobilized for the benefit of the alliance. The terms of the alliance treaty often specify what alliance members are required to contribute and what circumstances alliance members may gain access to those resources (Leeds et al 2002, Benson 2011, Benson 2012) and affect alliance strength. Additionally, factors such as transportation of military resources over distances (Holsti et al 1973; Bueno de Mesquita 1983) and coordination among allies (Riker 1962) may degrade raw capabilities because of inefficiency costs. Other circumstances may also affect the reliability of an alliance and whether alliance members can expect to receive any of the benefits of the alliance at all. For example, it may be that alliances among countries with convergent preferences or government systems (Smith 1996) are perceived to be stronger because there is a higher likelihood allies will turn out to support one another.

While there is no thermometer we can use to directly measure the strength of an alliance – we do possess multiple observable measures that are presumably related to the strength of the alliance. So long as the observed variables are related to the strength of an alliance, we can use modern statistical measurement models to use the relationship between the observed measures to make inferences about the underlying unobserved (latent) features of an alliance.

Consider first how we might assess the raw potential power of the alliance. Aggregate
military capabilities are an obvious starting point for assessing combined capabilities, since it estimates the upper bound of the amount of resources an alliance can muster. Researchers typically sum the capabilities of alliance partners using the Composite Index of National Capabilities (CINC scores) and then estimate the power of the alliance relative to an external threat (Bueno de Mesquita 1983; Reiter 1996; Wagner 2007). The sum of the allies’ capabilities over the total capabilities of the alliance members and the external threat gives an estimate of the probability the alliance will prevail over the threat. Thus, the sum of allies’ combined CINC scores represents a measure of the potential resource capacity of an alliance and is a good starting point for estimating the strength of an alliance. Another method of assessing alliance strength involves identifying alliances with a major power (Morrow 1991, Benson 2011). Both features are potentially informative – summing capabilities might yield a measure of the total potential resource capability of an alliance while flagging alliances with major powers might indicate which alliances possess partners with particularly important qualities, such as nuclear power, that make those powers unique in the international system.

Aggregate strength based on joint capabilities only describes a partial picture of the strength of an alliance, however, as some factors limit alliance members’ full access to the raw potential capabilities of the alliance. Many scholars claim that distance degrades strength (Bouldning 1962; Starr and Most 1976; Bueno de Mesquita 1983; Bueno de Mesquita and Lalman 1986; Smith 1996; Weidmann et al. 2010) because of increased difficulty of projecting forces and coordinating military actions. Problems coordinating access to and mobilization of pooled resources may also affect the strength of an alliance. One such constraint is the size of an alliance. The more members there are involved in an alliance the more difficult it may be to coordinate the mobilization of military resources across multiple organizations. One the other hand, the coordination problems may be more than offset by the increased military capacity of involving more countries in an alliance (Riker 1962). The difficulty of determining a priori whether more alliance partners increases or decreases the strength of an alliance highlights the importance of a statistical model that provides a principled manner of
inductively determining the relationship between observable characteristics and the strength of an alliance.

The effect of shared domestic political regimes is also possibly ambiguous. Similar regimes reflect similar preferences for outcomes in the international area, and jointly democratic alliances may also generate domestic audience costs that cause alliance commitments to be more credible (Lai and Reiter 2000; Leeds 1999; Leeds et al. 2002; Gibler and Sarkees 2004; Leeds et al. 2009; Mattes 2012). Others, however, argue that democracies may prefer not to ally with one another (Simon and Garzke 1996; Gibler and Wolford 2006), and it is also possible that the constraints and veto-points created by domestic political institutions may create difficulties for taking action (e.g., Tsebelis 2002), or that high leadership turnover due to elections may make them unreliable allies (Gartzke and Gleditsch 2004). It is difficult to specify the precise relationship of the net effect between regime type and alliance strength given these considerations.

In addition to characterizing the strength of an alliance in terms of the actual and potential military capacities of the signatories involved, scholars have also argued that the formal terms of the alliance matter (Leeds et al. 2002). Accounting for the content of alliance agreements and the obligations of alliance members established by the alliance (e.g., when and how they are expected to provide assistance) is likely equally critical for measuring the strength of an alliance. Consider, for example, the Helsinki Final Acts signed in 1975 by nearly every major power in the international system. Despite the strength of the signatories involved, it is doubtful that may would argue that the treaty was strong because the agreement was not a formal treaty, the terms did not require any active military assistance on the part of the signatories, and the member states merely made a non-binding promise not to use military action against each other (Abbott and Snidal 2000; Abbott et al. 2000).

In considering which observable aspects are likely related to the strength of the formal terms of an alliance, there is again a wealth of available scholarship. Many scholars, for example, suggest the type of alliance signifies its strength. For example, an offensive alliance
may be stronger than a non-provocation defensive alliance because it charges alliance members with a greater obligation to mobilize their pooled military strength in a greater range of circumstances. By contrast, a non-provocation defensive alliance limits military action to specific military engagements. Thus, the offensive alliance might be a stronger signal of allied military strength because it is limited by fewer conditions (Smith 1995). Others have also argued that the type of alliance affects the likelihood alliance members will intervene (Sabrosky 1980; Siverson and King 1980; Smith 1996). Benson’s (2011) typology of alliance commitments provides a related schematic for thinking about how alliance commitments might relate to alliance strength by emphasizing the importance of conditional versus unconditional terms of the alliance for the strength of the alliance.

2 Measurement Model

As the prior section makes clear, the strength of an alliance likely depends on both the strength of the signatories involved and the formal terms of the alliance agreement. Providing an empirical characterization of alliance strength that reflects this insight is difficult. The number of variables available for assessing alliance strength and the correlates of alliance strength signify the amount of work scholars have dedicated to measuring various aspects of alliances, but the sheer number of these variables also complicates efforts to tackle standard questions in alliance research. Most of the collected factors have some bearing on alliance strength, but the precise relationship is unclear. Controlling for every possible variable related to alliance strength reduces the degrees of freedom, and makes model specifications unwieldy to interpret or even estimate (Ray 2003; Achen 2005). Most critically, it is not clear how the numerous variables relate to one another and whether their combined impact on variables of interest is additive, multiplicative, or something else. Without such guidance, however, it is difficult to assume that the regression specifications correctly account for the relationships of interest. Moreover, combining one or more of these variables often
necessitates making tradeoffs on other variables – aggregating capabilities across alliance partners and then determining how distance between those partners degrades aggregate strength can be complicated when the number of alliance members is greater than two and often results in the listwise deletion of multilateral alliances.¹

The issues confronting scholars interested in characterizing the strength of an alliance are endemic to social sciences. Similar issues arise, for example, when using observed characteristics to measure how “democratic” a country is at a given point in time, or how liberal a member of the US Congress is. We observe characteristics that are related to the concept of interest, and we must use the observed characteristics and a statistical measurement model to make inferences about the latent traits.

A Bayesian latent variable model provides a framework for measuring alliance strength that uses the information contained in the many measures that researchers have already collected that are plausibly related to the strength of an alliance while also allowing researchers to make weaker assumptions about the precise relationships involved. To motivate the proposed measurement model and more fully explicate the issues scholars confront when trying to measure alliance strength using conventional means, consider the hypothetical task of measuring the strength of observed alliances – hereafter denoted by the latent variable \( x^* \) – using the observed measures \( x_1 \) and \( x_2 \).

One possibility is to chose a single observable characteristic as a proxy for alliance strength. Using either \( x_1 \) or \( x_2 \) to measure \( x^* \) is problematic for several reasons. First, insofar as there are multiple dimensions of alliance strength – perhaps depending on both the formal alliance agreement (say \( x_1 \)) and the characteristics of the signatories involved (say \( x_2 \)) – relying on a single observable characteristic will fail to account for this complexity unless the dimensions (and measures) are perfectly correlated. Second, relying on a single measure ignores the information contained in the other characteristic that is related to the

¹Particularly because the number of allies may be an important determinant of alliance strength, dropping multilateral alliances risks omitting critical information.
strength of an alliance. Finally, characterizing the strength of an alliance is obviously a difficult endeavor and formally acknowledging the inherent imprecision of our efforts is desirable to highlight which distinctions are statistically meaningful. If one alliance is “5 units” more than another in the space spanned by $x_1$ does that reflect a meaningful difference in alliance strength? How certain are we that the difference is meaningful in terms of affecting the strength of the alliance? Our measure of alliance strength should ideally account for the uncertainty that we have about the estimated strengths.

Combining multiple observable characteristics into a single measure to try and reflect the information contained in multiple measures faces similar difficulties. Consider, for example, the task of creating an index of alliance strength using the observed (binary) characteristics $x_1$ and $x_2$. There is no theoretical basis or guidance for determining how to combine measures to create a single index of alliance strength. Using the characteristics to create an additive index by adding the characteristics together makes extremely strong assumptions about the meaning of the scale that is not obviously plausible. Even if the binary measures $x_1$ and $x_2$ are both related to the strength of an alliance, on what basis can we conclude that an alliance possessing only characteristic $x_1$ as strong as the alliance that possess only characteristic $x_2$? Moreover, is the alliance containing both $x_1$ and $x_2$ twice as strong as an alliance containing one feature but not the other? It seems difficult to rationalize the relationships that are assumed by an additive index (and such assumed equivalences increase as the number of variables used to construct the measure increase).

If the goal is to predict the effects of alliance strength on an outcome of interest (say $y$) we are in a slightly better position in terms of measuring alliance strength because we can use the regression specification to control for multiple features of an alliance.\footnote{While concerns about the possible endogeneity of the relationship may emerge in such a specification, we are focused on the more limited question of measuring the strength of an alliance.} For example if we are predicting the effect of alliance strength on outcome $y$, the typical regression specification is $y = \alpha + \beta_1 x_1 + \beta_2 x_2$. The specification allows the terms $\beta_1 x_1 + \beta_2 x_2$ in the specification
to both measure alliance strength – as a linear function of $x_1$ and $x_2$ – and measure the correlation with $y$. Even though we can account for multiple measures, measurement issues persist and become specification issues for the regression of interest. For example, is the strength of an alliance a linear function of observable characteristics? Might the relationship depend on functional transformations of the observed characteristics (e.g., logs and power transformations) or depend on potential interactions between the characteristics? Moreover, given the number of potential indicators of alliance strength, the degrees of freedom that analysts have may be quickly reduced depending on the number of assumed interactions. Including multiple characteristics of alliance strength in a regression specification may allow us to account for the multiple characteristics that are related to the strength of an alliance, but difficulties remain in correctly specifying the nature of the relationship. Moreover, it may also be difficult to interpret effects in a saturated regression model (Achen 2005).

Regardless, of the approach used, a shortcoming of all three approaches is that they fail to account for the uncertainty we inevitably have about how the observed concepts relate to the underlying dimensions of alliance strength and the precision with which we are able to estimate the strength of an alliance.

Bayesian latent variable models have been developed to address precisely the issues we face when measuring the strength of an alliance. Bayesian latent variable models are able to extract the latent dimensions that are assumed to be responsible for generating

---

3We can treat $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ as accounting for the regression of $y$ on the unobserved alliance strength $x^*$ given the true specification $y = \alpha_0 + \beta_0 x^*$ if we can assume that $x^*$ is a linear function of $x_1$ and $x_2$. If, for example, $x^* = \gamma_1 x_1 + \gamma_2 x_2$ and $y = \alpha_0 + \beta_0 x^*$ the regression of $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ is equivalent to the regression of $y = \alpha_0 + \beta_0 (\gamma_1 x_1 + \gamma_2 x_2)$ because $\alpha = \alpha_0$, $\beta_1 = \beta_0 \times \gamma_1$, and $\beta_2 = \beta_0 \times \gamma_2$ even though we do not observe $x^*$! Note however, that this decomposition relies on the extremely strong – and implausible – assumption that $x^* = \gamma_1 x_1 + \gamma_2 x_2$. This requires that not only must the latent trait be a function of observables that are correctly specified in the regression specification, but also that the relationship between $x^*$ and the observables is without error. If there is error in this relationship – say $x^* = \gamma_1 x_1 + \gamma_2 x_2 + \epsilon$ – then we are in a classic error-in-variables situation and the estimated regression coefficients are inconsistent (see, for example, the nice review by Hausman 2001).

4One can certainly employ non-Bayesian methods to do so (e.g., Bollen 1989), but for both theoretical (see the arguments of Gill 2002 and Jackman 2009a) and practical reasons we adopt a Bayesian approach for measuring latent traits (e.g., in contrast to traditional frequentist approaches, measures of uncertainty emerge directly from the Bayesian estimation procedure).
the association between and within the distribution observed characteristics (see, for example, Quinn 2004; Jackman 2004; Jackman 2009b), and scholars have used such models to measure critical latent concepts in the studying the politics of the United States (e.g., Clinton and Lewis 2008; Levendusky and Pope 2010) and comparative politics (e.g., Rosenthal and Voeten 2007; Rosas 2009; Pemstein, Meserve and Melton 2010; Treier and Jackman 2008; Hoyland, Moene and Willeumsen 2012), scholars have only recently begun to apply the models to questions of international relations (see, for example, Schakenberg and Fariss (2009) and Gray and Slapin (2011)). Because the problem we confront is to use the observable characteristics of alliances to measure the strength of the alliance which is presumably responsible for structuring the association among the observed variables given theoretical expectations about the meaning of the dimensions of interest, a Bayesian latent variable model provides an ability to not only use all available and relevant information to reduce the parameter space via a parsimonious model, but it also provides an assessment of how certain we are about the recovered estimates.

To motivate the approach taken by a Bayesian latent variable model, consider the problem of uncovering a single latent dimension from a series of observable characteristics. To solidify the exposition, suppose that we are interested in measuring the strength of alliance \( i \) at the time of its founding and let \( x^*_i \) denote the unobserved (latent) strength of alliance \( i \). Our measurement task is to use observable characteristics that are theorized to affect the strength of alliance \( i \) to construct an estimate of \( x^*_i \) that not only describes the relative strength of the alliance relative to other alliances, but it also how much uncertainty that we have regarding our estimate of alliance strength. Suppose further that we have \( k \in 1...K \) observable measures of alliance strength, and let the observed value for variable \( k \) for alliance \( i \) be denoted by \( x_{ik} \). As will be the case given the invaluable data collected scholars interested in alliances (e.g., Leeds et. al. 2004; Gibler and Sarkees 2004), our observed measures include continuous, binary and ordinal measures so our measurement model needs to be able to reflect
the different data generating processes.\footnote{Technically, if variable $k$ is a discrete variable, the observed value for alliance $i$ in variable $k$ is the category $c$ which is generated according to $x_{ik} = c$ if $x_{ik} \in (\gamma_{k(c-1)}, \gamma_{kc})$ and where $\gamma_k$ is the vector of cut points for the $C$ categories in variable $k$ (Quinn 2004).}

Figure 1 provides a graphic representation of the measurement model for the case of 3 observable characteristics. As Figure 1 makes clear, the critical assumption in the model is that (unobserved) alliance strength $x_i^*$ is related to the observed variables $x_{i1}$, $x_{i2}$, and $x_{i3}$ across all alliances. However, the relationship may differ between the observables. For example, $x_{i1}$ and $x_{i2}$ may not be identically related to $x_i^*$, and the differences are captured by the parameters: $\beta_1$ and $\beta_2$ or $\sigma_1^2$ and $\sigma_2^2$.

![Directed Acyclic Graph: Bayesian latent variable model](image)

Figure 1: \textbf{DIRECTED ACYCLIC GRAPH: BAYESIAN LATENT VARIABLE MODEL}: Circles denote observed variables, squares denote parameters to be estimated.

Given the number of parameters to be estimated, recovering the latent measure of alliance strength ($x^*$) from the matrix of observed characteristics $x$ clearly requires some additional
structure on the problem. The structure we use is provided by a Bayesian latent variable specification (see, for example, Jackman 2009a,b). That is, for all alliances \( i \in 1...N \) we assume:

\[
\mathbf{x}_i \sim N(\beta_{k0} + \beta_{k1}x^*_i, \sigma^2_k).
\] (1)

Intuitively, the measurement model given in equation (1) assumes that the observed correlates of alliance strength \( \mathbf{x} \) are all associated with the latent level of alliance strength in identical ways across the \( N \) alliances, but the relationship may differ across measures in a few specified ways. Not only may the mean value of the \( x_k \) and \( x^* \) differ (as will be reflected in the estimate of \( \beta_{k0} \)), but the the scale of the observed variable and the latent variable may also differ (as is reflected in \( \beta_{k1} \)). For example, \( \beta_{k1} > 1 \) means that a one-unit change in the latent scale of \( x^* \) corresponds to more than a one-unit change in the observed measure \( x_k \), \( \beta_{k1} < 1 \) implies that a one-unit change in the latent scale corresponds to less than a one-unit change in the observed measure, and \( \beta_{k1} < 0 \) implies that the orientation of the observed and unobserved measures are “flipped” (i.e., positive values of \( x_k \) correspond to negative values of \( x^* \)). Moreover, if an included measure is unrelated to the latent trait revealed in the other included measures, the model can also account for that possibility – \( \beta_{k1} = 0 \) means there is no relationship between \( x^* \) and \( x_k \). Finally, the model also allows the relationship to be more or less precise through the \( \sigma^2_k \) term to account for varying amounts of error in the mapping between the observed and unobserved variable. Finally, because we estimate a version of equation (1) for each of the \( K \) observed measures, we allow for the relationship to vary across observed traits, and we can therefore use all available measures to help uncover the latent trait without having to specify the precise terms of the relationship.

A strength of this approach is that we can use the observed data and the specification of equation (1) to recover estimates of both the latent trait \( x^* \) (sometimes called the “factor score”), but also the extent to which the observed matrix of variables \( \mathbf{x} \) are related to the latent trait \( \mathbf{\beta}_k \) (i.e., the coefficient matrix \( \mathbf{\beta} \) sometimes called the “factor loadings”). As a
result, we can characterize both the latent strength of alliances as is revealed in the matrix of observable characteristics, and also which of the observed characteristics are most influential for structuring the latent trait that is recovered.

Given the the unknown parameters $x^*$ and $\beta$ that are to be estimated from the observed covariate matrix $x$, the likelihood function that is to be maximized is given by:

$$L(x^*, \beta) = p(x|x^*, \beta) \propto \sum_{i=1}^{N} \sum_{k=1}^{K} \phi \left( \frac{x_i - (\beta_{k0} + \beta_{k1} x^*_i)}{\sigma_k} \right)$$

where $\phi(\bullet)$ is the pdf of the normal distribution. To complete the specification and form the posterior distribution of the (latent) factors $x^*$ and factor loadings $\beta$, we assume the typical diffuse conjugate prior distributions.\(^6\)

As specified, the model is unidentified. Examining equation (2), for example, reveals that every parameter except for $x_i$ has to be estimated and it is therefore possible to generate an infinite number of parameter values that yield the same likelihood by appropriately adjusting $\beta_{k0}$, $\beta_{k1}$, $x^*_i$ and $\sigma_k$. Identifying the model therefore requires imposing additional assumptions about the scale and location of the latent policy space $x^*$. As Rivers (2003) shows, in one dimension, two constraints are required to achieve local identification and fix the scale and location of the space – the orientation of the space can be fixed by constraining a factor to be positively or negatively related to the latent trait. Typically, this involves assuming that the mean of $x^*$ is 0 and the variance of $x^*$ is 1 (see, for example, Clinton, Jackman and Rivers 2004). In multiple dimensions the number of required constants increases to $d(d + 1)$ where $d$ denotes the dimensionality of the latent space of interest. Given the discussion of Section 2 which suggests theoretical reasons to suspect that the strength of an alliance varies in two dimensions – the formal terms of the alliance and the strength of the involved signatories – to estimate our two-dimensional latent model requires at least 6 parameter constraints.

\(^6\)Specifically, the prior distribution of $\beta_k$ conditional on $\sigma_k^2$ is normally distributed and the prior distribution for $\sigma_k^2$ is an inverse-Gamma distribution (Jackman 2009).
3 The Measurement Model Applied & Identified

We are interested in estimating a two-dimensional measurement model of alliance strength where the first dimension reflects the strength of the signatories of the alliance and the second dimension captures the effect of the formal terms of the alliance. Because scholars have theorized that these dimensions determine the true strength of an alliance ($\mathbf{x}^*$), our measurement model will recover estimates of alliance strength $\hat{\mathbf{x}}^*$ in $\mathbb{R}^2$. For clarity, let $x[1]_i^*$ denote the latent strength of the alliance in the first dimension – estimated by $\hat{x}_1^*$ – and let $x[2]_i^*$ denote the latent strength in the second dimension (estimated by $\hat{x}_2^*$). Given the identification concerns noted above, we identify the two dimensions of alliance strength by imposing appropriate and theoretically informed constraints on the characteristics of the alliances.

To identify the center of the latent parameter space, we assume that the mean of $x[1]_i^*$ and the mean of $x[2]_i^*$ are both 0. This means that the latent traits are centered at $(0, 0)$. This assumption is completely innocuous and it reflects an arbitrary centering of the unobserved latent space. To fix the scale of the recovered space, we assume that the variance of $x[1]^*$ and $x[2]^*$ are both 1. While we are assuming that the scale of the latent space in the first and second dimensions is identical, the assumption is again without loss of generality because it is just defining the arbitrary scale of the dimensions that are to be estimated. To fix the rotation of the policy space and prevent “flipping”, we assume that higher values of the summed capacity of signatories correspond to positive values in the first dimension, and alliances that are both offense and defense receive positive values in the second dimension.\footnote{\textsuperscript{7}To be clear, none of these assumptions affect the likelihood function being maximized. These choices only define the scale of the estimates we recover, and different choices will produce different values of the estimates but the same likelihood.}

To account for theoretical determinants of alliance strength noted in Section 1 and to identify the meaning and interpretation of the two dimensions we recover, we impose a series of assumptions on how the characteristics of alliances and signatories relate to the latent
dimensions. Given the dimensions of alliance strength we are interested in recovering – the formal agreement and the characteristics of the signatories – the implied constraints are straightforward. For every characteristic pertaining to the written terms of the alliance we assume that $\beta[1]=0$, and for every characteristic related to the alliance partners we assume that $\beta[2]=0$. This implies that characteristics related to the signatories themselves only affect the first dimension, and characteristics of the formal agreement are only allowed to affect alliance strength in the second dimension.\(^8\)

To be clear, we are not assuming about how alliances are located within the two dimensions of alliance strength. In fact, a question of empirical interest is how $x^*[1]$ and $x^*[2]$ are related (which is why we identify the dimensions by placing constraints on $\beta$ rather than by making assumptions about the relationship between $x^*[1]$ and $x^*[2]$). Because we identify the latent dimensions using characteristics of the alliances rather than an assumption about the relationship between the latent dimension, our measurement model can shed important insights into the relationship between the formal terms of an alliance and the characteristics of the signatories.

Because our unit of analysis is the strength of the alliance at the time of the signing on the dimensions related to the formal terms of the alliance as well as the characteristics of the signatories, we are able to rely upon the impressive efforts of *The Correlates of War Formal Alliance (v 3.03)* data set of Gibler and Sarkees (2003) with the *Alliance Treaty Obligations and Provisions* data of Leeds et. al. (2002) as recorded in *EUGene v3.204* (Bennett and Stam (2000)).\(^9\) Our unit of observation is therefore the strength of every alliance at the time

---

\(^8\)Because we impose theoretically derived parameter constraints to identify the dimensions being estimated, our estimator is similar to “confirmatory” factor analysis where theoretical insights are used to define the dimensions of interest *a priori*. “Exploratory” factor analysis places fewer constraints on the measurement model and lets (possibly spurious) relationships present within the data to define the recovered dimensions.

\(^9\)Gathering these measures into a dataset with the alliance agreement as the unit of analysis required some manipulation. Many of the variables, such as distances between countries and s-scores are directed data. To gather such data into a dataset with the alliance at the time of formation as the unit of measure we combined COW data with ATOP data (Leeds et al 2002) in directed dyad format. Then we merged EUGene COW measures of CINC scores, s-scores, distances, and major power status into the directed dyad data. Then we transformed those data into agreement level data and merged with Benson’s (2011) data on
of its signing as determined by Leeds et. al. (2002) and, because of data availability, which was signed between 1816 and 2000.

We do not need to know the precise nature of the relationship between the observed characteristics and the strength of the alliance to implement the model, but we do need to identify which measures are, and are not, related to each of the two dimensions to impose the identification assumptions discussed above. To identify the first dimension we use characteristics related to the capabilities of alliance members at the time the alliance is formed. To do so, we use the the sum of alliance members’ military capacity based on CINC scores (Bueno de Mesquita 1983; Morrow 1991; Smith 1996; Poast et al 2012) contained in the COW data available in the EUGene project.\textsuperscript{10}

Studies have also included variables indicating whether alliances have major powers (Morrow 1991, Benson 2012) because such countries may possess distinctive characteristics (e.g., the possession of nuclear weapons or influence in international institutions such as the United Nations Security Council) so we also include an indicator for whether at least one major power is a signatory. There is some disagreement as to whether alliance strength increases as the number of alliance partners increases (Riker 1962, Niou and Ordeshook 1994) or whether there may be diminishing returns to the numbers of allies in a coalition (Riker 1962) but rather than impose an assumption on the relationship is between alliance size and alliance strength, we include a variable for the number of allies and a separate indicator for bilateral alliances and let the data determine the appropriate relationship.

We also include measures for the distance between alliance members because the distance between countries may degrade the strength of an alliance (Boulding 1972, Holsti et al 1973, Starr and Most 1976, Bueno de Mesquita 1987, Diehl 1985). The literature is mixed, however, about how exactly distance degrades strength. Some divide the capabilities by the distance alliance commitment types.

\textsuperscript{10}We are not interested in knowing how aggregate capabilities relate to an alliance member’s threat, because our objective is to develop a measure of the strength of the alliance at the time that the alliance is formed. If interested, however, scholars can extrapolate from our measure of strength to estimate the strength of an alliance relative to a specific in particular crises across multiple dyad years.
between countries (Smith 1996), and others degrade strength linearly according to available freight indices (Bueno de Mesquita 1987). However, we do not know whether strength degrades linearly, nor do we know whether the appropriate measure of distance should be capital-to-capital, border-to-border, or largest city-to-largest city. Moreover, is the rate of degradation sensitive to the technological sophistication and geography of a country? We make no strong assumptions about the precise relationship between distance and strength when implementing our measurement model. We simply include a simple measure of the mean distance between every unique pair of alliance partners and the standard deviation of the average distance and let the relationship in the data determine the implied relationship between distance and strength.\(^\text{11}\)

Whether signatories share similar preferences may also affect the strength of the alliance. One measure of preference similarity as well as credible commitment might be the similarity of political systems and, in particular, whether alliance members are jointly democratic (Lai and Reiter 2000; Leeds et al. 2002; Gibler and Sarkees 2004; Gibler and Gleditsch 2004; Leeds et al. 2009). To estimate the effect of alliance members’ political systems, we include measures of whether all signatories are considered to be democratic, the average Polity score of alliance members, and the standard deviation of the mean Polity score. We also include Signorino and Ritter’s (1999) “S” score to measure alliance portfolio similarity. We include the mean “S” score between all unique pairs of alliance signatories as well as the standard deviation of the average scores in an alliance. To be clear, our model only assumes that these characteristics are potentially related to the strength of an alliance in terms of the signatories involved – the precise relationship between each measure and our estimate of alliance strength is something that is estimated using the likelihood of equation (2) and it is possible that at least some of the included measures are not statistically related to our estimate of alliance strength.

To estimate the second dimension we rely on the insights of scholars who argue that

\(^{11}\)For bilateral agreements where there is a single pair of countries we set the standard deviation to 0.
some types of alliances are stronger than others either because different types have more or less impact on the behavior of alliance members (Leeds 2003, Benson 2011, Benson et al. forthcoming) or because the type of agreement effects the likelihood signatories will intervene (Sabrosky 1980, Siverson and King 1980, Smith 1996). To measure the influence of the formal terms of an alliance on alliance strength we use ATOP data (Leeds et al. 2002) and Benson’s (2011) typology. Alliance agreements are coded in the ATOP data as being offensive, defensive, neutrality, consultation, and non-aggression. An alliance agreement can contain multiple provisions, allowing it to be classified as more than one type of alliance. Benson’s (2011) typology creates more granular categories depending on the expressed objective of the provision to provide military assistance and whether the obligation to deliver military assistance is guaranteed and conditioned on an action in a dispute. His typology includes unconditional compellent, conditional compellent, unconditional deterrent, conditional deterrent, and probabilistic deterrent commitments. We include both the ATOP and Benson (2011) categories. We impose some minor restrictions on the ordering of alliance strength according to theory about how different types of alliances should be ranked (Smith 1995; Sabrosky 1980; Siverson and King 1980; Smith 1996; Benson 2012).12

In addition to alliance type, the *Alliance Treaty Obligations and Provisions* data of Leeds et. al. (2002) measures many terms that are likely related to the strength of the formal terms of an alliance (see also, Leeds and Anac 2005). In terms of the institutionalization of the alliance, we include whether there are mentions of the possibility of conflict between the

12Using the ATOP coding, we assume that alliances that are both offensive and defensive are more powerful than alliances that are just offensive, which are more powerful than those that are just defensive. For Benson’s coding, we assume that conditions limit the application of military force to specified situations. Thus, an unconditional guarantee of military support in any circumstances is a stronger commitment than a commitment of support only if an adversary attacks and an alliance member did not provoke, which is in turn a stronger commitment still than a promise maybe to intervene. Following the rank ordering of alliance types suggested by Benson (2012), we assume that commitments containing both compellent and deterrent objectives are more powerful than those than contain just compellent objectives, which are stronger than those containing just deterrent objectives. Within those general ranking rules, commitments that promise military support without conditions are more powerful than those that stipulate conditions for causus foederis. The final ranking rule is that any agreement containing a probabilistic commitment, or a provision that creates uncertainty about or whether active military assistance is guaranteed.
members of the alliance (CONWTIN), an integrated military command (INTCOM), the promise of active military support (ACTIVE), the exchange of economic aid (ECAID), the exchange of military aid (MILAID), provisions for an increase or reduction of arms (ARMRED), and joint troop placements (BASE). We also account for whether the formal obligations vary across the alliance partners (ASYMMETRY), whether it was formed in secret (SECRECY), whether it allows a signatory to renounce obligations under an alliance agreement during the term of the agreement (RENOUNCE), and whether the alliance provided for a specific term (SPECLGTH). When the variables were coded to contain multiple categories, we often collapsed the categories to a binary measure to denote the presence or absence of each feature because it was often unclear how the categories should be ordered.\footnote{For example, the coding of MILAID is “if there are no provisions regarding military aid, the variable is coded 0. If the agreement provides for general or unspecified military assistance, the variable is coded 1. If the agreement provides for grants or loans, the variable is coded 2. If the agreement provides for military training and/or provision or transfer of technology, the variable is coded 3. If the agreement provides for both grants and/or loans and training and/or technology, the variable is coded 4” (p. 27). It is unclear whether terms that denote specific loans or grants (MILAID=2) are stronger than terms that provide for unspecified military assistance (MILAID=1), or half as strong as terms that include both grants and/or loans and training and/or technology (MILAID=4). As a consequence, we use whether there are any provisions for military aid or any kind (i.e., if MILAID ≥ 1).}

Given these measures and identification constraints, we use the Bayesian latent factor model that can accommodate both continuous and ordinal measures described by Quinn (2004) and implemented via \texttt{MCMCpack} (Martin, Quinn, and Park 2011). 500,000 estimates were used as “burn-in” to find the posterior distribution of the estimated parameters, and we used one out of every 1,500 iterations of the subsequent 750,000 iterations to characterize the estimates’ posterior distribution.

## 4 Estimates of Alliance Strength

Our Bayesian latent variable model of alliance strength produces two sets of results – estimates about how the various observable features are related to the latent dimensions that we recover, and estimates about the strength of alliances in each of the two theoretically derived
dimensions. The latter are of most interest because they describe the relative strength of various alliances over time in terms of the strength of signatories and the formal terms of the alliance, but investigating the relationship between the included observables and the latent space reveals the features that are and are not important for structuring our estimates. They are also a useful check on the plausibility of the estimates by revealing which features affect our estimates.

Table 1 reports the relationship between each of the observed measures that we include in our statistical measurement model and its relationship to the underlying latent strength of an alliance in each dimension. Because we identify the space by assuming that characteristics of alliance signatories are only related to the first dimension of alliance strength and the formal terms of an alliance are only related to the second dimension, blank entries indicate instances where the coefficient is constrained to be 0. For descriptive purposes we also report the mean of each variable and its scale.

Table 1 reveals relationships that reassuringly comport with expectations. Factors such as the summed military capacity of signatories (Sum of Capacities) and whether a major power is involved (Major Power) are positively related with the strength of an alliance in the first dimension, whereas alliances comprised of only two counties (Bilateral Agreement) or among only democracies (All Democracies) are slightly weaker. In terms of the factors that affect the strength of the formal terms of an alliance, alliances that are both offensive and defensive are stronger alliances in the second dimension (ATOP ordering), and alliances with probabilistic commitments and commitments stipulating conditions according to the Benson ordering are weaker. Alliances consisting only of non-aggression treaties (Non-Aggression Agreement) are estimated to be the weakest. Moreover, a strength of a Bayesian latent variable model is that we can also assess the precision of these estimated relationships. As Table 1 makes clear, some of the included measures cannot be statistically distinguished from zero. For example, there is no obvious relationship between the average Polity IV score of alliance signatories (Avg. Polity IV Score) and the strength of an alliance in the first
<table>
<thead>
<tr>
<th>Factor</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
<th>Mean</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of Capacities</td>
<td>1.01*</td>
<td></td>
<td>.097</td>
<td>[0.001, 0.649]</td>
</tr>
<tr>
<td></td>
<td>(.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Power</td>
<td>.658*</td>
<td>.341</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.040)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ally Count</td>
<td>.332*</td>
<td>2.79</td>
<td></td>
<td>[2.42]</td>
</tr>
<tr>
<td></td>
<td>(.041)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral Agreement</td>
<td>-.457*</td>
<td>.846</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.069)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Distance</td>
<td>.156*</td>
<td>1179</td>
<td></td>
<td>[0.7773]</td>
</tr>
<tr>
<td></td>
<td>(.045)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev. of Distance</td>
<td>.287*</td>
<td>39.58</td>
<td></td>
<td>[0.1953.76]</td>
</tr>
<tr>
<td></td>
<td>(.045)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Democracies</td>
<td>-.307*</td>
<td>.194</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.083)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Polity IV Score</td>
<td>-.068</td>
<td>3.59</td>
<td></td>
<td>[0.10]</td>
</tr>
<tr>
<td></td>
<td>(.044)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev. of Polity IV</td>
<td>.068</td>
<td>1.37</td>
<td></td>
<td>[0.5]</td>
</tr>
<tr>
<td></td>
<td>(.044)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. SGLO</td>
<td>-.136*</td>
<td>.691</td>
<td></td>
<td>[-269.1]</td>
</tr>
<tr>
<td></td>
<td>(.046)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev. of SGLO</td>
<td>.088</td>
<td>.021</td>
<td></td>
<td>[0.312]</td>
</tr>
<tr>
<td></td>
<td>(.048)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATOP ordering</td>
<td>24.64*</td>
<td>1.67</td>
<td></td>
<td>[0.3]</td>
</tr>
<tr>
<td></td>
<td>(3.36)</td>
<td>(2.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benson ordering</td>
<td>4.89*</td>
<td>2.42</td>
<td></td>
<td>[0.7]</td>
</tr>
<tr>
<td></td>
<td>(0.506)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consultation Agreement</td>
<td>-.056</td>
<td>.547</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Agression Agreement</td>
<td>-.958*</td>
<td>.455</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrality Agreement</td>
<td>-.240*</td>
<td>.455</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymmetry</td>
<td>.650*</td>
<td>.080</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.126)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secrecy</td>
<td>.429*</td>
<td>.088</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.105)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conwtin</td>
<td>-.954*</td>
<td>.406</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specified Length</td>
<td>-.299*</td>
<td>.718</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renounce</td>
<td>-.046</td>
<td>.114</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military Aid Provisions</td>
<td>.494*</td>
<td>.103</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.103)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic Aid Provisions</td>
<td>-.105</td>
<td>.112</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.084)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provisions for Military Bases</td>
<td>.666*</td>
<td>.107</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.122)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provisions for Arm Reductions</td>
<td>-.018</td>
<td>.027</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing Active Conflict</td>
<td>41.18*</td>
<td>.436</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(9.77)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Military Command</td>
<td>.865*</td>
<td>.044</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.196)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conditio</td>
<td>24</td>
<td>5.05</td>
<td></td>
<td>[0.1]</td>
</tr>
<tr>
<td></td>
<td>(.066)</td>
<td>(.031)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Estimated Correlates of Alliance Strength, 1816-2000
dimension, and explicit provisions in the treaty for a reduction of arms (Provisions for Arm Reductions) has no obvious effect on the strength of an alliance in terms of the agreement terms. The characterizations that are evident in Table 1 are noteworthy because while we imposed a restriction as to which dimension was relevant for each characteristic, we imposed no assumptions about the nature of the relationship between the measure and the strength of the alliance. To be clear, the measurement model imposes absolutely no constraints on the coefficients reported in Table 1.

While the relationship between observable characteristics and the dimensions of alliance strength are interesting and they reveal useful insights about the correlates of alliance strength, presumably we care more about what the relationships imply about the strength of the alliances themselves. Figure 2 presents this characterization and plots the distribution of estimated alliance scores in the dimensions defined by the strength of the signatories (x-axis) and the strength of the formal terms of the treaty (y-axis). We estimate a score for each of the 525 alliances signed between 1816 and 2000 for which we have data on the observable characteristics (reported in the Appendix), but we focus our attention on a few selected alliances to illustrate the face validity of our estimates.

Figure 2 presents the position of all the treaties with atopIDs in the two-dimensional space defined by the strength of the signatories and the formal terms of the alliance agreement. The positioning of many of the alliances comports with our intuition about the relative strength of those alliances, despite the fact that we have not imposed any constraints on the scores themselves. The relative position of alliances relative to one another in the two-dimensional space, as well as the consistency of the scores with our a priori beliefs about how these alliances might be scored relative to one another, is driven by the characteristics of the alliance agreements and the signatories rather than assumptions (or priors) used when estimating the measurement model.

Consistent with expectations, the strongest alliance both in terms of the signatory characteristics as well as the terms of the formal agreement is the Allied agreement in World War
II. The alliance is a joint declaration by 39 countries, including the United States, Russia, the United Kingdom, and China, to devote their full resources, military or economic, against those members of the Tripartite Pact and its adherents with which such government is at war. There are no conditions or termination dates imposed on the terms of the agreement. It is a sweeping declaration of war, offensive and defensive, by the most powerful coalition in the international system to defeat Germany, Japan, and Italy.

The weakest alliance both in treaty terms and based on the characteristics of the signatories, is the Belarus-Bulgaria alliance in 1993. The agreement is a bilateral treaty reaffirming the nonaggression promise made in the Helsinki Final Act. In addition to promising not
to be aggressive toward each other, the signatories also pledge to refrain from using force in their international relations, to consult with one another when their security has been breached, and to remain neutral in hostilities directed at the other alliance member. The treaty states that it is to be in effect for a period of 20 years, and members may unilaterally terminate the agreement with advance notice of one year.

There are also alliances that are strong on one dimension but weak on the other. For example, the 1958 United Arab Republic (UAR) is an example of a strong formal agreement among weak signatories. The UAR was formed as an effort to unite the Arab community against the expansion of communism in Syria and elsewhere in the Arab world (Walt 1987, pp. 71-80). It included Egypt, Syria, and Yemen, though Yemen’s inclusion was regarded as merely a cosmetic gesture (p. 72). The agreement called for the integration of the allies militaries and unified command over those forces. Gamal Abdel Nassar, former President of Egypt and the President of the United Arab Republic, insisted on a full union and control over both countries in exchange for his agreement to halt the rising influence of the Syrian Communist Party. Consequently, the terms of the agreement granted full military power of the signatories to a unified command and authorized a Commander-in-Chief to pursue the unified foreign policy drawn by the Union, which could extend to both defensive and offensive campaigns. Nassar responded immediately by seizing control of Syria and banning all political parties.\textsuperscript{14}

On the opposite end of the two-dimensional space is the Helsinki Final Acts signed in 1975. It is the most striking outlier, given that it one of the five or six strong alliances when measured by signatory characteristics, but for such a potentially powerful alliance its formal

\textsuperscript{14}Even though the alliance was weak in terms of the signatory characteristics, it was powerful enough to satisfy the primary objective of the Union, which was to crack down on the Syrian Communists. The larger hope, which was explicitly expressed in the alliance, was that other Arab states would also join to enhance the combined influence of the Arab community. Instead, Jordan and Iraq felt threatened by the UAR and formed the short-lived Iraq-Jordan Federal Union (atopID 3340). Walt argues that the Federal Union was designed to balance against the UAR. At least in terms of our measure of overall alliance strength, the UAR and Federal Union appear to be matched. The Federal Union lies within the confidence interval of the UAR, implying that the strength of the Federal Union is statistically indistinguishable from the UAR.
terms are weak. Thirty-five countries signed the Accords, including the United States, the USSR, the UK, and most of Europe. The combined strength of the members comprised the preponderance of military strength on the earth at the time. Yet, the military obligations as laid out in the terms of the agreement are loose. The main objective of the agreement is to set forth a bargain between respecting territorial boundaries and human rights. However, the terms of the signatories, and especially the military obligations, are nonbinding, as the agreement lacks the legal status of a formal treaty. As an agreement that the parties agreed would not be governed by international law, scholars classify the Helsinki Final Acts as an example of soft law (Abbott and Snidal 2000; Abbott et al 2000). On our dimension of the strength of the terms of the military alliance, the agreement is not the weakest agreement in the data because it does not specify an end date for the alliance. However, it ranks low on the continuum because it is a nonaggression pact that makes no promise or provisions of military aid, basing or active military assistance under any circumstances. Hence, the Helsinki Final Acts is unique in that it combines the military strength of most of the powerful countries of the world under the promise to refrain from threatening or using force against each other.

The large break in the distribution of alliances in the second dimension of Figure 2 indicates that our measure clearly divides alliances into different “bins” based on the formal terms of the agreement. The largest determinant of the ordering of alliances along the second dimension is whether the alliance agreement promises active military support by any alliance member, and a promise of active military assistance is a clear indicator of the strength of an alliance agreement, naturally distinguishing alliances formed for compellent and deterrent purposes from consultation, neutrality, and non-aggression pacts that contain no provisions for an alliance member to use its military. The gap reveals that the alliances that promise active military support are reassuringly estimated to be very different than those that do not.

While the location of these alliances are where we might expect they would lie in the dimensions of alliance strength we estimate, the empirical characterization of alliance strength
depicted in Figure 2 reveals many other important and interesting insights that would not be evident without the method and measure we propose. First, there is a great deal of heterogeneity between alliances – some have strong terms between weak signatories (e.g., the alliance between the United Arab Emirates and Yemen in 1958 that effectively combined the militaries of the two countries), some involve weak terms among weak signatories (e.g., the nonaggression pact signed between Belarus and Bulgaria in 1993), some involve strong signatories agreeing to terms that are extremely weak (e.g., the Helsinki Final Acts signed in 1975 by many of the world’s leading powers), and some involve strong signatories agreeing to strong terms (e.g., the alliance formed by the Allies in World War II). Our estimates of alliance strength show clear variation, and the variation corresponds to plausible expectations about the strength of various alliances.

Second, because we identify the dimensions of alliance strength by constraining observable characteristics to affect a single dimension, we can explore the relationship between the two dimensions of alliance strength. While we find that there is a positive relationship – implying that as the strength of the signatories involved increases so too does the strength of the formal terms of the alliance – the relationship is relatively modest (correlation of .256). Moreover, it is possible to make interesting comparisons by using variation in the two dimensions. For example, although the strength of the signatories of the Helsinki Final Acts are stronger than the NATO alliance (because of the addition of Warsaw Pact countries such as the USSR in the Helsinki Final Acts), the terms of the NATO alliance are far stronger as we might expect given the divergent preferences of the United States and USSR.

Third, a strength of our measurement model is the ability to account for the uncertainty that we have about the estimates themselves. For every alliance, we can quantify how certain we are about the estimated strength of the alliance in each of the dimensions we estimate. Moreover, the uncertainty may vary across alliances and across dimensions. For example, as Figure 2 illustrates, while we can certainly distinguish the 1993 nonaggression alliance between Belarus and Bulgaria from the 1975 Helsinki Final Acts in terms of the strength of
the signatories because the 95% regions of highest posterior density do not overly, we cannot be certain that the formal terms of the two alliances are distinguishable. The score of the Helsinki Final Acts is slightly higher in the second dimension, but the size of the confidence ellipses almost completely overlap in the second dimension. Because our Bayesian latent variable model recovers how precisely we are able to estimate the strength of an alliance, we can use such information to characterize the statistical confidence we have in our assessments.

Fourth, and related to this point, is the fact that we have more confidence in our ability to distinguish alliances according to the strength of the signatories than we do using the terms of the agreement. This is evident by the fact that the length of the plotted ellipses denoting the 95% regions of highest posterior density are far larger than the width.\textsuperscript{15} Moreover, the separation in the estimated strength of alliances in the second dimension is related to the fact that alliances signed during active conflicts are also likely to involved both offensive and defensive conditions.

4.1 A Closer Look: Alliances in the World Wars and East Asia

To take a closer look at the estimates and further explore the face validity of our estimates in more detail, we examine the alliances involved in World War I, World War II, and the post-WWII alliances in East Asia. Because these alliances are historically consequential, we have strong priors as to how they ought to be related to one another.\textsuperscript{16} Figure 3 plots the estimated strength of each of the relevant alliances in terms of the first dimension (left graph) and the second dimension (right graph) in temporally order.

Consider first the alliances that were involved in World War II and which are plotted in the middle of Figure 3. As Figure 2 revealed above, the strongest alliance in the first dimension involved the alliance formed between the Allies in 1942. Notice that the 1940

\textsuperscript{15}Because of the constraints imposed on the factors used to measure the dimensions of interest, there

\textsuperscript{16}To be clear, although we use a Bayesian framework for theoretical and practical reasons, the priors we use are diffuse and contain no information about the relationships we examine. Our posterior is are almost entirely driven by the likelihood function.
Figure 3: **Selected Alliance Scores** Estimated strengths for alliances involved in World War I, World War II, and post-WWII East Asian security in terms of the first dimension (left) and second dimension (right) are plotted. The points denote estimate in each dimension for each alliance and the lines show 95% regions of highest posterior density for the selected alliances.
Tripartite Alliance (atopID 2515), which was targeted by the Allied Pact, is estimated to be weaker on both dimensions than the alliance formed by the Allies during World War II. This is reassuring given that it is a multilateral defensive pact signed during World War II between countries whose combined capabilities are not as great as the Allied powers. Additionally, the terms of the defensive obligation are conditional upon one of the signatories being attacked by a party not involved in World War II at the time the alliance was signed. However, the antecedent for this defensive pact was the 1939 Pact of Steel between Germany and Italy (atopID 2445). As an unconditional pledge to undertake shared offensive and defensive military campaigns, it is on par with the Allied Pact in the strength of its agreement terms. The 1940 Tripartite Pact was replaced by a more aggressive agreement (atopID 2540), which, like the Allied Pact, is also a wartime alliance containing similar terms. In Figure 3, it is in approximately the same position in the second dimension as the Allied Pact, indicating the similarity in the strength of the terms of the agreements between the opposing World War II alliances. Like the Allied Pact, the three signatories to the Tripartite Alliance pledged to use all their means, offensive and defensive, to pursue the war. The strength score for the terms of the Allied Pact is slightly higher probably because the Tripartite Pact specifies a termination date, which was a stipulation initiated in the Pact of Steel and passed down to the successor alliances. Even though the terms of the formal agreements are on parity, the measure of the combined strength of the Allies according to the members characteristics is greater than that of the Tripartite Pact.

The estimated strength of the alliances involved in World War I reported in the bottom of Figure 3 also comport with prior expectations. There is clear parity in the prewar alliances. The 1882 Triple Alliance between Germany, Italy, and Austria-Hungary was similar to the 1893 Franco-Russian alliance both on alliance terms and signatory strength. The motivation for the Franco-Russia alliance may have been motivated, as Snyder (1997) suggests, by a desire by France and Russia to gain parity of strength with the growing relative strength of the Triple Alliance. The terms of the wartime treaty signed by France, Russia, the
United Kingdom, and Italy (atopID 2025) is similar to the opposing declaration agreed to by Germany, Austria-Hungary, and Bulgaria (atopID 2030). However, the addition of the United Kingdom and Italy to the alliance with France and Russia shifted the signatory strength significantly in favor of the Allies.

Another interesting manifestation of the strength scores is the location of the post-WWII deterrent alliances in Asia. Notice that the alliance terms are roughly similar. The 1950 USSR-China alliance, the 1951 US-Japan alliance, and the 1961 China-North Korea alliance have slightly stronger treaty terms than the others. They obligate alliance partners to defend each other if a fellow ally is attacked while the other alliances stipulate the possibility that an alliance member might be able to escape its defensive obligations if there is war. However, the differences between the strength of these alliances based on the alliance terms dimension are statistically indistinguishable. The parity in signatory strength between rival alliances is also noteworthy. Comparing the 1950 USSR-China alliance to the 1951 US-Japan alliance, we see that the signatory strength of each alliance is approximately the same. The US-South Korea alliance was signed after the Korean War to deter North Korea, China, and the USSR. The signatory strength of the US-South Korea is within the confidence interval of the USSR-China alliance. The 1954 US-ROC alliance was signed during the first Taiwan Strait Crisis. It was also designed to deter China and the USSR while also restraining Chiang Kaishek. Its relatively weaker treaty terms reflect the motivation to restrain an alliance partner, and its signatory strength is on par with the USSR-China.

It is also noteworthy that the parity of alliances remained roughly in tact even as the schism between China and the Soviet Union emerged in the early 1960s. During this period, China and North Korea formed a separate alliance, as did the USSR and North Korea. The US and Japan renewed and revised the terms of their alliance. Even though the signatory strength of all the alliances decreased, the opposing alliances remained approximately similar to each other both in terms of signatory strength and treaty terms. The 1960 US-Japan alliance and the 1953 US-South Korea alliance are roughly on parity with the 1950
USSR-China alliance, the 1961 USSR-North Korea alliance, and the 1961 China-North Korea alliance. The scores, however, in our measure are driven by the characteristics of the alliances and not because we have imposed constraints on the scores themselves. The similarities in characteristic strength and treaty terms of these rival alliances correspond to our intuition that these alliances might be formed to counter each other. These empirical characterizations are consistent with claims that some alliances may be designed to balance threats (Morgenthau 1978 and Waltz 1979).

5 Conclusion, Caveats, and Implications

Scholars of alliances have been blessed with a tremendous amount of data due to the generous and impressive efforts of those who have collected data on both the formal terms of the various alliances as well as the characteristics of the signatories involved. The amount of available data prompts the question: how can we best measure the strength of international alliances given the wealth of available data and our theoretical conceptions of alliance strength while also accounting for the inevitable ambiguity that must necessarily accompany such a determination?

We show how a Bayesian latent variable trait model – a model that has been applied to many other important concepts in political science – can be used to integrate observable measures with theoretical arguments about the determinants of alliance strength to provide estimates of: how the various observable factors relate to the theoretically implied dimensions, how the strength of alliances in the two dimensions relate to one another, and how certain we are about all of the estimated parameters. Applying the model to all alliances in the international system between 1816 and 2000 provides estimates of alliance strength in terms of the strength of signatories and the strength of the formal terms of the alliance agreement. In total, we estimate the strength of 525 alliances – including those that are multilateral and lack a specific target. The estimates have strong face validity – familiar
alliances are located as we would expect, and exploring the post-World War II alliances of East Asia provides reassuringly reasonable estimates.

Not only do our estimates provide scholars with measures that can be used to explore questions such as the determinants of alliance formation and the persistent of alliances, but the estimates also reveal important insights into the nature of alliances. For example, although there is a relationship between the strength of signatories and the strength of the formal terms of an alliance, the relationship is modest (.256) and there is considerable variation of potential interest in the relationship. Moreover, when comparing the estimates of temporally and geographically proximate alliances, the estimates suggest that the alliances that are formed are nearly balanced. Perhaps reflecting the theoretical claims of (Morgenthau 1978 and Waltz 1979), alliances formed in response to one another are very similar in strength. Third, our ability to distinguish between alliances on the basis of the strength of signatories is better than our ability to do so using the formal terms of the alliance – we have more uncertainty about the estimates of the latter than we do about the former. Relatedly, the ability to assess the certainty with which we are able to estimate the location of alliances provides the ability to account for the precision of our estimates when comparing alliances or using the estimates in secondary analyses.

Finally, although we think our estimates are theoretically sound and possess strong face validity, the measurement model we employ is sufficiently general and scholars can use the model to generate their own estimates using different assumptions or different measures.

While we think our measure and the method provide an important advance for scholars interested in alliances and the international system, some caveats are worth noting. First, while our statistical model provides a principled way of extracting information from multiple measures related to theoretically relevant dimensions with a minimal amount of assumptions, the resulting estimates are still dependent on the relationship between the observable measures to extract the latent dimensions. For example, the fact that CINC scores are relative to the global capacity at the time means that comparisons across long periods of time may
be difficult and the estimates are likely most appropriate for temporally bounded comparisons or when temporally-related differences are account for. That said, this concern emerges whenever these scores are used in any regression that explores variation over time.

Second, the ability to use a Bayesian latent variable model to extract the latent dimensions structuring the strength of an alliance does not ameliorate potential concerns about the endogeneity of alliances. While we have been careful not to include variables in our measure of alliance strength that scholars may seek to correlate with the strength of an alliance in the hopes of better understanding alliance formation, nothing in our analysis discounts the fact that the formation of an alliance is presumably a strategic act based on the assessment of expected consequences and those interested in using the estimates should be careful of the potential pitfalls. That said, these concerns can also emerge when using current measures of alliance strength.

While no measure is perfect, our model is able to use theoretical insights to: identify how features are related to the strength of an alliance in theoretically relevant dimensions, measure the strength of alliances formed between 1816 an 2000 along these two dimensions, and provide a measure of our uncertainty about the resulting estimates. Our measurement model and the estimates we produce will hopefully allow scholars to focus on better understanding the determinants and consequences of alliances rather than continually having to grapple with the question of how best to measure the relative strength of international alliances.
References


38


41


6 ONLINE APPENDIX

The table on the following pages report the ATOP identification number for every alliance formed between 1816-2000 for which we possess the requisite characteristics in order to measure the strength of the alliance. The alliances are sorted from the strongest alliance in terms of the military capacities of the initial signatories to the weakest. It also reports the estimated strength in the first dimension (Dimension 1) and the posterior standard deviation for that estimate, as well as the estimated strength in the second dimension (Dimension 2) and the posterior standard deviation.