Cooperation in Good Times: Are Democracies Really Different?*

Douglas M. Gibler† and Mark David Nieman‡

December 28, 2018

Abstract

Democracies are supposed to behave differently than other states when cooperating in alliances, organizations, trade, or a host of other international institutions. We contend, however, that these democratic differences largely depend upon geopolitical environments that make cooperation possible. Though studies have demonstrated endogeneity between democracy and peace, few analyze the effects of this joint relationship on democratic foreign policies related to cooperation. We use the alliance literature as our example and find that democracies in alliances are no more reliable than other regime types, once the threat environment of states is jointly modeled. We find that alliances formed during times of conflict are particularly unreliable "scraps of paper", and that the general reliability of alliances is concentrated to those coordination alliances existing in already-peaceful environments. Our argument has important ramifications for a host of literatures focused on regime type as well as current debates over the effectiveness of democratic deterrence.

*Gibler thanks the National Science Foundation (Awards #0923406, #1260492, and #1729244) for their generous support on research related to this project. Nieman thanks the National Science Foundation (Award #1728395) for their support for research. For questions and comments, please contact dmgibler@ua.edu.

†Institute for Social Science Research, University of Alabama, Tuscaloosa, AL.

‡Department of Political Science, Iowa State University, Ames, IA.

Since the establishment of the empirical finding that democracies rarely fight each other, many studies have sought confirmation of democratic differences in other types of empirical relationships. Democracies are thought to trade more with other democracies, are more likely to form and cooperate in intergovernmental organizations, and are more likely to ally with each other and be reliable partners. These are just a few examples of the larger democratic peace research program.

We believe these types of inferences are unwarranted, however, since the explanations often miss an important point: democracies tend to be clustered across the globe in mostly peaceful regions. As democratization is more likely in peaceful environments, analyses examining any type of democratic difference must be careful to separate the independent causal effect of democracy on political outcomes from that of the political environment. Since democracy is itself at least partially determined by the political environment, a failure to model both the direct and indirect effect of the political environment on policy outcomes will incorrectly attribute the indirect effect of the political environment to democratic institutions. This, in turn, makes it easier to find statistically significant differences in foreign policy behaviors across regime types. We argue that once the political environment is accounted for, and this source of bias is eliminated, many of the differences in foreign policy behavior between regime types will be eliminated.

We focus our argument on the large body of research that examines democracies in alliances and build on a recent study that found democracies to be more reliable partners. We use a split-population logit with an instrumental variable to model the likelihood of conflict and peace when alliances are formed and maintained, and introduce this likelihood into the study of democratic reliability. The split-population model allows us to statistically account for the degree of threat within the political environments, while the intrumental variable allows us to account for any possible endogeneity between alliance reliability and threat environment. We find that any association between democracy and reliable alliances appears to actually be a function of the political environment facing the state. Further, our analyses suggest the combined, threat-and-reliability model better fits the data provided by the study than the reliability model alone.

Our argument should apply to other ancillary findings suggesting democratic differences as well, since we find that democracies cooperate more easily during periods of peace. Any time peace affects cooperation—whether it is trade, institutions, and similar types of cooperation—the endogeneity we document will pose problems for confirming that democracies behave differently. We begin our argument in the next section with a brief review of the democracy and alliance literature.

Alliance behavior

Traditional alliance theory is replete with arguments that threats to the state cause alliance making in order to deter aggressors. Morgenthau (1960) called it external balancing: faced with threat and unable to respond quickly enough with internal capability changes, leaders sought partners in other states to help them avoid or survive against external threats to their sovereignty. Alliance-making in this manner forms a key component of traditional, mostly realist theory (see also, Waltz 1979; Walt 1985), and most tests of alliance formation find that threats do matter in determining whether alliances form (Siverson and Emmons 1991; Lai and Reiter 2000; Johnson 2017). Of course, the implication of these arguments and findings is that alliances covary with threat, and, when threat diminishes, the need for the alliance does as well. Alliances are, as the famous phrase puts it, "scraps of paper" to be torn as situations change.

Not all alliances, however, are responses to threat. Instead, alliances may be formed to facilitate a number of different tasks (Schroeder 1976; Altfeld 1984). Alliances, for example, may be an instrument used to resolve contentious issues (Gibler 1996, 1997; Weitsman 2004; Mattes and Vonnahme 2010). Alliances may also be used by powerful states as a method of gaining control over smaller states (Morrow 1991; Lake 1996). In these cases, major powers provide security to smaller states in exchange for basing rights, support for specific foreign policy interests, geopolitical access, or other goods (Lake 2009; Johnson 2015; Nieman 2016).

Consistently identifying these different types of alliances *ex ante*, however, has proven difficult. The empirical focus has been on either specific issues, such as territorial settlement treaties, or the presence of asymmetric capabilities within the alliance. Classifying alliances in this way, however, negates cases where alliances are mutually implicated, at least to varying degrees. The former requires that alliances that resolve issues, such as the 1887 pact between Prussia and Russia and the 1960 USSR and China alliance, are not also, at least in part, power aggregating.[1] The latter forces an assumption that some trade-off between capabilities and autonomy is the primary reason for why major powers would partner with minor powers that offer little in the way of additional military capabilities.[2]

Democracies in alliance

Related to the two general alliance types—traditional power aggregating alliances and those formed for other purposes—has been the growth of studies associating democracies with alliance behavior that is quite different from traditional alliance theory expectations. Democracies may engage in deterrent alliances, for example, but their commitments are seemingly not scraps of paper. Their alliances commitments are more likely to deter other states and, when threatened, to be honored by the democracies involved. Democracies also tend to engage in non-traditional alliances of control, and their alliances are more likely to become institutionalized and have provisions for cooperation that are different from other alliances.

The theoretical argument for this has focused on the idea that cooperation is more likely among similar types of states. Leeds (1999), for example, develops a model where coop-

[1]The Prussia-Russia alliance addressed disputes in the Balkans and the Dardanelles, while the USSR-China alliance resolved a border dispute between the two parties (Gibler 1997).

[2]While some asymmetric alliances are often an effort to "buy influence" from the perspective of major powers, they can also provide access and basing rights necessary to confront distant adversaries. Additionally, from the perspective of minor powers, these arrangements do supplement their security and enhance the effectiveness of other major power signals (McManus 2018), as well as increase the minor power's bargaining position vis-à-vis rivals in disputes (Langlois 2012).

eration is essentially a method of policy coordination, and leaders consider the likelihood of agreement fulfillment—foreign policy changes—when forming or proposing cooperation. Without the likelihood of fulfillment by the other actor, then there is little incentive to alter state policies when it will not be reciprocated. These audience costs seemingly make it more likely that democracies make better alliances and have longer-lasting cooperation.

A number of studies have empirically analyzed whether pairs of democracies tended to "flock together". Siverson and Emmons (1991) found some evidence that democracies were more likely to form alliances with other democracies, but there were strong period effects (see also Kimball 2006). The finding was consistent in the post-World War II era data but not during the 1920 to 1939 period. Lai and Reiter (2000) revisited this empirical claim and found a strong relationship with multivariate controls for threats to the dyad. Jointly-democratic dyads were more likely to be involved in defense pacts than mixed or fully non-democratic dyads, and regime similarity systematically predicted both defense pacts and other types of commitments in the dyad.[3]

Alliances composed of jointly democratic states also appear to last longer than alliance comprised of other types of states (Gaubatz 1996; Bennett 1997). Gaubatz (1996) dichotomized dyadic democracy, using data from Doyle (1986), and found that jointly democratic alliance dyads lasted twice as long as other alliance dyads. There was no empirical difference, however, between mixed-regime and non-democratic dyads. This led Gaubatz (1996, 135) to conclude, "democracy by itself does not appear to either increase or decrease the ability of a state to make commitments to nondemocracies." Bennett (1997), also using Doyle's data, used the average number of liberal regimes in an alliance and found a positive, statistically significant effect for increased liberalness in the alliance. The substantive effects were especially strong, since all-liberal alliances increase the duration of an average alliance by almost fifteen years.[4]

[3]McManus and Yarhi-Milo (2017) suggest that while democracies are more likely to engage in public acts of support, cooperation with autocracies tends to occur in less public ways in order to avoid domestic backlash.

[4]It is worth noting that only a handful of alliances (less than 1% of the data) were comprised solely of

Finally, democratic states have also been found to be more reliable alliance partners than nondemocracies. When confronted with threats to the alliance, democracies are more likely to honor the provisions of their alliances because their leaders risk sanction by their publics when reneging. Thus, among all the states in alliances, democracies are expected to be better partners—more reliable, less likely to terminate their alliances, and less likely to violate the terms of the agreement (Leeds 2003; Leeds and Savun 2007; Leeds, Mattes and Vogel 2009).

Do political institutions explain these differences?

There is evidence, however, that suggests that democratic differences in alliance behavior may be less well understood than it appears. For example, regarding alliance formation, Gibler and Wolford (2006) questioned the research design used by both Siverson and Emmons (1991) and Lai and Reiter (2000). Gibler and Wolford argued that these studies were not technically examining alliance formation but were instead identifying whether dyads were allied. By switching the analysis from whether a dyad was allied in a given year to focusing on a dyad at the time of alliance formation, Gibler and Wolford demonstrated that democracies were not more likely to form alliances; instead, states were becoming democratic *after* having formed an alliance. The peace provided by large, regional defense pacts promoted the development of democracy, and, indeed, over 90% of jointly-democratic alliance-dyads were in three broad, regional defense pacts: NATO at 55%, OAS at 29%, and the WEU at 7%. This clustering confirmed a more complicated relationship between democracies and alliance-making and also hints at the likelihood that the distribution of democracies in alliance is at least partially determined by some omitted factor.

This finding raises questions about the reliability of other alliance outcomes associated with democracy. For example, given the logic of cooperation among similar regime types outlined by Leeds (1999) and Lai and Reiter (2000), while it is clear why democracies may be more reliable allies with other democracies, it is less clear why democracies would unilaterally

democratic states at the time of alliance formation.

restrict their options when interacting with nondemocracies, especially given the expected higher degree of defection among non-democratic states. While it may be the case that democratic states are simply less willing than nondemocracies to break their international commitments, this commitment does not appear to extend to all policy areas; for example, democracies do not appear more reliable than other regime types to honor their monetary commitments (Simmons 2000) or territorial treaties (Chyzh 2014). Moreover, in a study looking at whether partners are more likely to intervene in response to an alliance obligation, Gartzke and Gleditsch (2004) found that democracies were less reliable than other states.

Finally, McDonald (2015) suggests that democratic alliances are more likely to be non-power aggregating than other states. He contends that major powers structure the international order in their own image and that the last two hegemeons (UK and US) have been democracies. These dynamics, he argues, result in minor powers forming asymmetric alliances with the hegemon, democratizing under their influence, and having peaceful relations with other states aligned with the hegemon (see also Lake 2009; Nieman 2016). That is, rather than alliance being more reliable among democracies, it may instead be the case that alliances among states aligned with the hegemon are non-power aggregating in nature and not intended to deter external threats. Consistent with this account, Digiuseppe and Poast (2018) find that pairs of allied democracies tend to reduce their military spending, which would not be expected by the power aggregating model. Taken together, these alternative explanations suggest that differences in alliance behavior often attributed to domestic institutions may instead be driven by some omitted factor.

Peaceful environments and democracy

One potentially omitted factor is the political environment around a state: specifically, how threatening the political environment is helps to shape a state's foreign policy (Vasquez 2009). For example, alliances are a behavior that is most likely observed by states in order to counteract a threat (Kimball 2006; Johnson 2017). Yet, peace—or the lack of threat—encourages or even causes democracy (Cederman and Gleditsch 2004; Gleditsch and Ward 2006; Gleditsch 2009; Gibler 2012). This creates a puzzle in terms of alliance behavior because, without a threat, democratic states should have no need for alliances. Nevertheless, democracies do make and maintain alliances.

Peace causing democracy is not a new argument, of course, and has developed over time and been integrated into the larger democratic peace project. Russett and Oneal (2001, 37), for example, contend in their seminal work that "Democracy is easier to sustain in a peaceful environment," and "external threats become reasons or justifications for suspending normal civil liberties, elections, and constitutional government." Their model of a Kantian peace recognizes the endogenous "feedback loops" from peace to democracy, trade, and international organization, so there is an explicit recognition that peace at least partially causes democracy even among some of the staunchest democratic peace advocates.

The problem for those who study democratic differences is that *any* degree of endogeneity between the political environment and democracy will introduce bias into additive models of international conflict. Even a weakly-endogenous relationship, such as that suggested by Russett and Oneal (2001), may bias estimates of the standard error of any variable that is correlated with both peace and democracy. Since properly specified models will include only variables directly related to conflict, by definition, and democracy is a common predictor of the lack of conflict, the likelihood of bias will turn mostly on the correlation between a variable of interest and democracy. Given the large amount of effort that has been used to determine that democracies are different from other states in their relations, this implies far-reaching concerns for studies examining the ancillary properties of the democratic peace.

A strongly endogenous relationship—or the case where democracy is a result of a peaceful environment—may have even more troubling implications for existing studies. Should democracy be dependent in some fashion on the establishment of peace (see arguments by Thompson (1996) or Gibler (2012), for examples), then democratic differences in a particular variable may simply underscore a more pervasive sample-selection process that made these cases observable. Alliances with democracies would, almost by definition, be more likely to be those that were formed or continue after a threat has subsided. In other words, if peace is causally related to both democracy and cooperation, studies that fail to explicitly model this when looking at the effect of democracy on cooperation will suffer from a specific form of omitted variable bias: functional form misspecification, where the omitted variables represent the nonlinearities—such as those introduced by selection processes—between the dependent and independent variables (Heckman 1979; Signorino and Yilmaz 2003).

We contend that this is a problem facing many studies of the axillary effects of democracy, such as the finding that democracies are more reliable alliance partners. Democracies exist within more peaceful political environments than other states (Ward and Gleditsch 2002). Without a clear, immediate threat, alliances formed by democratic states are unlikely to serve the purpose of power aggregation to ward off potential enemies. Alliances, after all, are not costless; well-functioning alliances, in particular, require information sharing and perhaps even joint military exercises (Morrow 1994; Fearon 1997). Given these costs, alliances formed or maintained by states in peaceful environments are likely to serve purposes other than power aggregation, such as helping to resolve contentious issues (Gibler 1996, 1997) or facilitating foreign policy coordination in other issue areas.

These alliances that are designed, or re-tasked, to primarily serve purposes other than aggregating power in order to deter aggression—such as those that address other contentious issues or powerful states acquire influence over smaller states—are more likely to be formed and maintained in less threatening political environments than power aggregating alliances. An implication is that these types of alliance run in direct contrast with the (often implicit) assumption made in most studies of alliance reliability: that the conflict environment facing alliance partners does not vary substantially from the conflict environments of other types of allies.

Because they are in peaceful environments, non-power aggregating alliances are less likely to be tested. This, in turn, makes them less likely to be abrogated, more likely to last longer, and more likely to be institutionalized over time. So, when we observe democracy correlated with these tendencies, we are really observing the effects of peaceful environments on both democracies and alliance behavior.

The result of the two distinct types of political environments implies that there are two distinct types of alliance within the overall sample of alliances: a subsample of alliances that act as traditional, power aggregating alliances—those within a more threatening environment—and a second subsample that primarily serve other purposes, such as cooperative or legitimizing alliance—those in more peaceful environments. Those alliances drawn from the former are significantly more likely to be invoked, and constitute the sample that most accurate tests whether democracies are more reliable alliance partners, than those alliances drawn from the latter. Our primary research hypothesis is that, once the peaceful environment alliances are accounted for, democratic states will be no more reliable alliance partners than other states.

Research design

In order to analyze the effect of a state's conflict environment on its propensity to violate alliance agreements, we analyze all members of bilateral alliances formed during the period 1919-1989 using the Alliance Treaty Obligations and Provisions dataset (Leeds et al. 2002; Leeds and Mattes 2007), which consists of 234 bilateral alliances. Our unit of analysis is the directed alliance member-year. This is appropriate as we are interested in which party broke an alliance, rather than just that one is broken. The temporal domain of our analysis is 1920-2001. We use a split-population logit estimator to probabilistically separate the alliances that exist in peaceful political environments—i.e. those that are unlikely to be invoked—from those alliances in threatening political environments. We employ an instrumental variable approach to account for the endogeneity between alliance reliability and militarized conflict.

Methodology

We expect alliances to separate into two stylized types—those that are "at-risk" of being invoked and those that are not—based on the level of external threat environment affecting a state. States in more threatening environments are significantly more likely to have their alliance invoked, providing opportunities to violate their alliances terms, while states in safer environments have fewer opportunities to commit violations. If our argument is correct, then full-sample estimates of the predictors of alliance violation, which ignore these different types of environments, will recover biased estimates.

Ignoring the conditioning effect of threat environments and treating all alliance observations as equally at-risk of entering the population of states that may violate their alliance commitments—which is true for traditional addative binary-choice estimators, such as logit or probit—is a type of model misspecification (Heckman 1979; Signorino and Yilmaz 2003). Unfortunately, we cannot definitively know *ex ante* with certainty which alliances are unlikely to be invoked; leaders do not often volunteer whether they considered violating their alliance obligations. Instead, we only have data on whether an alliance was violated, but not direct data on the degree that the alliance is "at-risk."

To address these data limitations, we use a split-population logistic regression model (Xiang 2010; see also Beger et al. 2011). A split-population logit is a type of mixture model, where an outcome variable is a function of two processes.[5] The logic of the estimator is that there are two distinct populations in the data, and entry into each population can be estimated *probabilistically.* Though the structure of the alliance data does not let us directly observe which cases are actually in the not reliable, or "at-risk" sample, we can estimate them. The estimator does this by using two equations: one equation that functions as

[5] All selection and zero-inflated models are types of mixture models, with the familiar censored probit-types of selection models (e.g., Heckman 1979; Sartori 2003) including data on the outcome of the first stage, and more recent extensions modeling selection when there are not data available on the outcome for the first stage (Xiang 2010; Nieman 2015, 2018; Bagozzi 2016; Bagozzi and Mukherjee 2012). The relationship between selection and mixture models is such that the probit variant of the split-population model is mathematically equivalent to Poirier's (1980) bivariate probit with partial observability (Xiang 2010, 488).

the selector, identifying *relevant* observations to include in the at-risk sample, and a second equation that estimates the outcome of interest on these relevant observations. The *relevance* equation affects the *outcome* equation probabilistically: some cases are treated as more "at-risk" than others, and this probability conditions estimates of the outcome equation.[6]

More formally, the estimator treats the outcome variable as a function of two processes:

$$Y_i = 0 \text{ with probability } (1 - R_i) + (R_i)(1 - V_i) \tag{1}$$

$$Y_i = 1 \text{ with probability } R_i V_i \tag{2}$$

where R and V are cumulative distribution functions (CDF) of a binary choice model (see Xiang 2010, 487-488). R_i represents the probability that a case is *relevant* to the sample—that the observation should be in the outcome equation, i.e. an "at-risk" state[7]—and conditions V_i, which represents the probability of violating an alliance. These probabilities can be specified as

$$\Pr(Y_i = 0) = [1 - \Lambda(Z_i\gamma)] + [(\Lambda(Z_i\gamma))(1 - \Lambda(X_i\beta))] \tag{3}$$

$$\Pr(Y_i = 1) = (\Lambda(Z_i\gamma))(\Lambda(X_i\beta) \tag{4}$$

where Z and X are vectors of covariates associated with the relevancy and outcome equations, respectively, γ and β the accompanying parameter estimates, and Λ is the link function for the cumulative logistic distribution. Equation 3 can, of course, be simplified as $\Pr(Y_i = 0) =$

[6] Partial observability models, of which split-population logit and any other mixture/zero-inflated model is a type, have been shown to correctly recover the sign and significance for parameters, even if variables are specified to the wrong equation, permitting accurate hypothesis testing (Nieman 2015, 438-439), though there are some criticisms of these models in terms of their overall reliability (Rainey and Jackson 2017). In this particular case, of course, the alternative to using a partially observed model is to simply ignore the potential bias induced by two populations of political environment entirely, which would result in inaccurate hypothesis testing and substantive effects (Xiang 2010). We formally assess model fit and robustness in the Appendix in Tables A.2, A.3, and A.4 on pp. 4, 6, and 7, respectively.

[7] This implies, of course, that the inverse, $1 - R$, is the probability of an observation selecting out of the "at-risk" subsample—i.e. being identified as "not at-risk."

$[1 - (\Lambda(Z_i\gamma))(\Lambda(X_i\beta))]$. The likelihood function of the split-population logit is written as

$$\mathcal{L} = \prod_{i=1}^{n}[(\Lambda(Z_i\gamma))(\Lambda(X_i\beta))]^{y_i}[1 - (\Lambda(Z_i\gamma))(\Lambda(X_i\beta))]^{1-y_i} \tag{5}$$

and estimates of β and γ are recovered via maximum likelihood estimation.[8]

In other words, the estimator treats cases where $Y = 0$ in the data as being the outcome of either (1) not at-risk, or (2) being at-risk, but not abrogated, whereas $Y = 1$ is the outcome of being both at-risk and abrogated. This modeling approach allows us to statistically separate alliance members in non-threatening environments, which are unlikely to have their alliance enacted, from the more traditional conception of alliance membership, where threats to member states will lead to the possibility of alliance terms being invoked. As an example, suppose an alliance is formed during a relatively high-threat time period in which the likelihood of conflict in that dyad-year is as high as 35%. The split-population logit would then assign 35% of the estimation to the relevance equation since it is part of the at-risk population of alliances. The remaining percentage of the estimation would be considered not at-risk and would be grouped with the alliances formed during more peaceful periods. The result of this weighting is analogous, in a sense, to including an interaction term, since the model corrects for the conditional effect of the sample-selection process. However, rather than interacting two variables, the interaction is between the full set of variables from the outcome and relevance equations.

Though this process does essentially weigh each observation by its political threat environment, the test remains quite conservative. The likelihood of threat in any given dyad-year is most often a small fraction of the 35% figure we use in our example, so for each alliance we are only assigning a small portion of its effect on the overall model to the relevance equation. The current baseline within this research tradition is to conflate such cases as peace settlements or trade pacts that have alliance clauses with the offensive and defensive pacts

[8]As the estimator is not included as an 'off-the-shelf' option with most statistical software, we include sample stata code on p. 8 of the Appendix for interested readers.

formed in the years prior to major wars. As we demonstrate below, this baseline assumption controls whether several key variables predict alliance reliability or failure.

Data

Our dependent variable is *alliance violation*, which captures whether a state abrogates its alliance commitments. We follow Leeds, Mattes and Vogel (2009, 469-470) and code *alliance violation* as 1 if state A violated the terms of an alliance. They treat an alliance as abrogated if (1) a major provision is violated and governments do not agree to continue with the alliance or (2) that one government unilaterally ends the alliance prior to its terms (see also Leeds and Savun 2007, 1124). There are 74 instances of *alliance violation* among the 234 alliances in the data set, roughly 32% of total alliances.[9]

We argue that not all alliances have an equal probability of being invoked, and we account for this non-random process using a split-population logit. To estimate the split-population logit, we must specify both the relevancy and outcome equations (Xiang 2010). Consistent with our theory, we specify the *relevance equation* with predictors related to a state's geopolitical threat environment. We expect that the absence of a threatening environment is associated with fewer opportunities for a state to violate the terms of its alliances, as there are fewer reasons to invoke an alliance. Alliances that are formed or renewed in such environments are likely for peaceful, cooperative purposes, rather than power-aggregation. Alliances formed in more threatening environments, however, are more likely to invoked. Such alliances are also more likely to be broken, as states called upon to assist an ally may determine that it is no longer rational for them to honor their agreed terms. We specify the *outcome equation*, with predictors that have previously been identified with affecting alliance commitment, namely the probability of a violation. We build on Leeds, Mattes and Vogel (2009) and include variables such as joint democracy, whether a state has had a change in

[9] After accounting for missing data, there are 70 violations in the exact replication of Leeds, Mattes and Vogel (2009) reported in Table 1 Model 1, 68 violations in the samples reported in Table 1 Models 2 and 3, and 34 observations in the post-1949 sample reported in Models 4 and 5. See Table A.5 on p. 10 of the Appendix for additional information.

their leader's societal coalitions, as well as dyadic and alliance-specific features.

Relevance equation

Our primary independent variable of interest is *territorial threat*. We conceptualize *territorial threat* as the maximum level of cross-border threat that a state faces. We operationalize this as the maximum predicted probability of a fatal militarized interstate dispute for state A for all contiguous neighbors. This value provides a continuous, latent measure of *territorial threat*. We construct a time-varying measure of *territorial threat* for each observation in the data set.

The predicted probability of a fatal militarized interstate dispute is estimated using data from Gibler and Tir (2014, Table 1). Gibler and Tir emphasize *territorial* predictors of conflict among contiguous neighbors, such as previous peaceful and violent transfers of territory within a dyad, the highest level of militarization of a state's neighbors, previous territorial MIDs within a dyad, and the age of the dyad's border. They include controls for whether there is a shared colonizer, the presence of a civil war for either state within a dyad, or defense pacts with neighbors. The results for the logit model used to construct the *territorial threat* measure are presented in Table A.1 on p. 1 of the Appendix.

We use this measure for two reasons: first and foremost, it best captures the theoretical concept of a dangerous neighborhood. As we describe later, we do include other possible sources of a dangerous political environment, but we expect that *territorial threat* is likely to be the best identifier of "at-risk" alliance observations. There is strong evidence that territorial disputes are the single best predictor of militarized conflict (Bremer 1992; Vasquez 1995, 2009; Reed and Chiba 2010). Yet, it is the threat of conflict—rather than just the realization of that threat—that is likely to spur the formation of power aggregating alliances. Neighbors with territorial disputes are likely to appear as more stressful and immediate threats than other potential sources, as neighboring states are usually able to quickly project power to their borders. Thus, we expect territorial threats to be the mostly likely source for the type of power aggregating alliances described by traditional alliance theory.

Second, using an instrumental variable—such as the predicted probability of conflict—rather than looking at observed militarized conflict is advantageous in methodological terms, in that it helps avoid issues related to possible endogeneity between alliance reliability and militarized conflict.[10] We account for uncertainty in our estimate of the instrumental variable by taking 10 draws from the estimated distribution of the maximum predicted territorial threat and use these to calculate the point estimates and standard errors, using Rubin's (1987) formula for multiple imputation (see Boehmke, Chyzh and Thies 2016, for a similar re-sampling approach).[11]

As initial evidence of a relationship, we find that 53 of the 68 alliance violations ($\approx 78\%$) have a *territorial threat* above the median territorial threat for all allied states.[12] Figure 1 displays the kernel density of *territorial threat* for the observations within the sample. It also reports the frequency of *alliance violations* at differing threat levels. The figure shows that at low threat levels, alliances are violated less than expected by chance, while at high threat levels there are more violations than would be expected. These results provide preliminary support for our expectations.

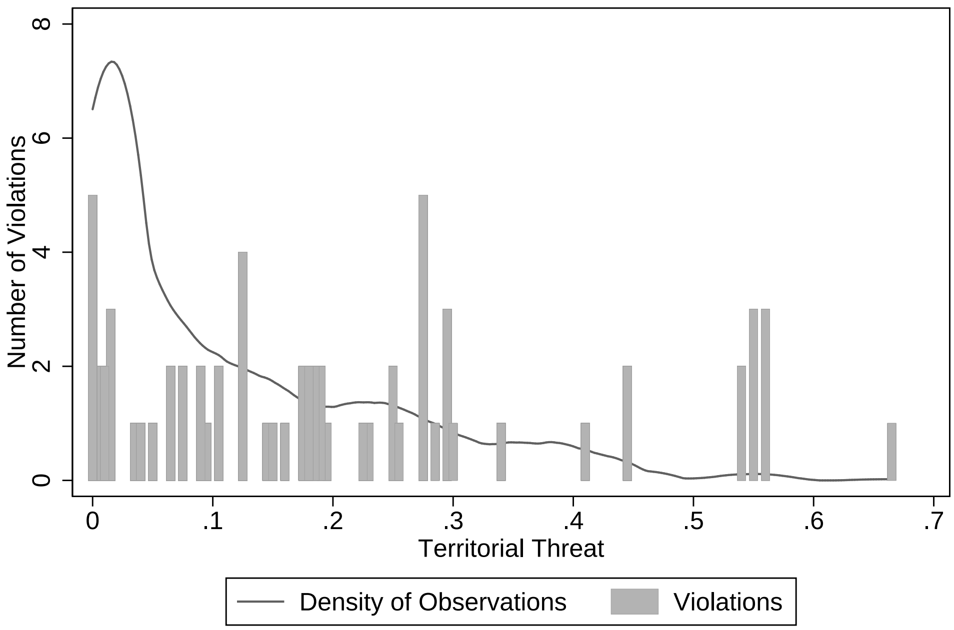
We also control for a number of other factors that may also influence the level of threat in a state's political environment. We account for the *number of borders* and *proportion of democratic borders*. States with more boarders have been shown to be more conflictual (Vasquez 1995, 2009), though the effect is mitigated as a greater proportion of a state's

[10] Whether an alliance is violated is unlikely to be related to the exogenous variables used to construct the instrumental variable: the correlation between *alliance violation* and each exogenous variable to construct the instrument is $r < |0.1|$. Moreover, our instrument appears to be strong; the difference in the F-statistic between nested logit models is 19.12, well above the threshold of 10 that is frequently used to indicate that an instrument does not suffer from weak instrument bias (Stock and Watson 2011). To conduct the F-test, we estimate a logit model where *alliance violation* is treated as a function of the independent variables from the relevance equation.

[11] The point estimate for each parameter is the mean from the 10 draws, or $\frac{1}{10}\sum_k^{10} \beta_k$, while the standard error is the average of the estimated variances within the datasets plus the variance in the point estimates across datasets, or $\sqrt{\frac{1}{10}\sum_k^{10} s_k^2 + (1+\frac{1}{10})\sigma_\beta^2}$, where s_k^2 is the standard error for dataset k and σ_k^2 is the variance in β between datasets. See Rubin (1987) As few as 5 draws from the estimated distribution is sufficient to incorporate uncertainty (Mislevy 1991).

[12] If we expand to the entire data set, this figure is 57 out of 74 ($\approx 77\%$).

Figure 1: Density of Territorial Threat and Frequency of Violations.



Note: The frequency of violations is overlayed with the kernel density of Territorial Threat within the sample from Table 1 Model 3.

neighborhood is democratic (Kadera, Crescenzi and Shannon 2003). These variables are based on Stinnett et al. (2002) and Marshall and Jaggers (2014). We control for *rivalry*, as states within a rivalry are already in hostile environments (Diehl and Goertz 2000; Rasler and Thompson 2006; Mitchell and Thies 2011), and whether a state is a *major power*, as these states are more generally more active and attractive alliance partners (Morrow 1991; Chiba, Martinez Machain and Reed 2014). Alliances with major power, of course, are also the most likely to be asymmetric alliances, functioning for purposes other than just power aggregation (Morrow 1991; Lake 2009). Data on rivalries and major powers are obtained from Klein, Goertz and Diehl (2006) and the Correlates of War Project (2016). We include a dummy variable for *Cold War* to account for systemic effects owing to a bipolar system (Bennett and Stam 2004). Economically developed states, often clustered geographically, may have complex economies that create norms that constrain state behavior (Mousseau 2003). We operationalize *economic development* as the log of energy consumption per capita

(Singer, Bremer and Stuckey 1972).[13] Finally, we control whether a state is an *oil producer*, as these states are more conflict prone (Colgan 2013).

Outcome equation

We rely on Leeds, Mattes and Vogel (2009, Table 1, Model 1) to specify our outcome equation. Leeds, Mattes and Vogel (2009) find that changes in a leader's core constituency and (the absence of) democratic institutions are highly correlated with alliance violations. Their conclusion affirms previous studies that find democratic governments to be seemingly more reliable alliance partners.

Changes in a leader's societal coalition are expected to increase the likelihood that a state fails to honor the terms of an existing alliance (Leeds, Mattes and Vogel 2009). *Change in leader's societal coalition* is measured as a binary variable that is 1 if there is a change in the core domestic supporting coalition of state A in a year. *Democracy* is expected to increase the reliability of an alliance, and is coded 1 if state A has a score ≥ 6 on the -10 to 10 Polity IV index (Marshall and Jaggers 2014).

Leeds, Mattes and Vogel (2009) include several dyadic measures expected to decrease the reliability of international commitments. *Change in international power* is a binary variable measured 1 if there is a change of >20% in either state since the alliance was formed. *Change in political institutions* is a dichotomous variable coded as 1 if either state experiences a change in political institutions since the alliance was formed. *Change in external threat* is a binary variable coded as 1 if the level of external threat between the current year and the start year of the alliance changed by 30%.[14] *Formation of new outside alliance* is a binary

[13]Economic consumption is available for a broader time frame than GDP/capita. Energy consumption and GDP/capita are correlated at $r = 0.7$.

[14]This measure differs substantially from Gibler and Tir's measure in terms of both composition and by focusing on whether there is a *change* in environment. The measure used by Leeds, Mattes and Vogel (2009) is based on a measure from Leeds and Savun (2007, 1127), which represents the sum of the capabilities (Correlates of War CINC scores) for politically relevant states (neighbors and major powers) of states that do not share an alliance and and have a foreign policy affinity score (S score, a similarity score based on alliance portfolio) below the median value in their sample (median = .775). The measure of *Territorial Threat*, however, is a latent measure of threat focused on territorial determinants of conflict from contiguous neighbors.

measure coded as 1 if state A formed a new alliance.

Leeds, Mattes and Vogel (2009) also include four alliance-specific variables in their analysis. Each of these are expected to reduce the risk that an alliance is abrogated. *Asymmetry* is a dichotomous variable equal to 1 if an alliance includes a major and minor power. *Non-military cooperation* is a binary variable coded as 1 if an alliance has provisions linking nonmilitary issues to the alliance. *Ratification* is a dichotomous variable measured as 1 if an alliance was formally ratified. *Military cooperation* is a binary variable coded as 1 if an alliance includes provisions related to peacetime military cooperation. Lastly, we include cubic polynomials to account for temporal dependence (Carter and Signorino 2010).

Empirical analysis

Table 1 presents the results comparing logit and a split-population logit. We estimate five models: the first two are estimated with a traditional logit and the third model is the full model estimated with a split-population logit. Model 1 provides an exact replication of Leeds, Mattes and Vogel (2009) for estimating alliance violations. Model 2 re-estimates Leeds, Mattes and Vogel (2009) but restricts the sample to only those observations that are common to both the original and full model. This ensures that parameter estimates in the outcome question are comparable once we account for the relevance equation, as both samples contain the exact same cases. Model 3 reports the estimates of the full model using the split-population logit regression, which includes both *relevance* and *violation* equations. Finally, Models 4 and 5 reports estimates of the 1950-2001 period for when relevance is ignored and accounted for, respectively.[15] The top of the table reports the *outcome* (violation) equation, and the bottom of the table reports the *relevance* (whether an observation is "at-risk") equation.

The results are interpreted in a relatively straightforward way: positive coefficients in-

[15]We look at a subsample since alliance compliance rates decline dramatically pre- and post-WWII (Berkemeier and Fuhrmann 2018).

dicate that increases in a variable make the outcome for that equation more likely. Hence, positive coefficients for variables in the *relevance* equation indicate an *increase* in the probability of being in the "at-risk" sub-sample of violating an alliance, while negative coefficients indicate a *decreased* likelihood of being in the "at- risk" sub-sample.[16] Similarly, positive coefficients for variables in the *outcome* equation indicate an effect that suggests an increased likelihood of an alliance violation.

Again, the first model is an exact replication of the original Leeds, Mattes and Vogel (2009, Table 1, Model 1) study, and the results are consistent with their earlier findings: all coefficients and standard errors remain the same as in the original study. Model 2 restricts the Leeds, Mattes and Vogel (2009) sample of cases to only those observations included in both the exact replication and the full split-population model. All of the parameter estimates are approximately the same, and all relationships are in the same direction and have the same level of significance as the original analysis. Model 2 thus provides a baseline from which to compare our full split-population model.

Model 3 estimates a split-population model where external threat and other factors related to the political environment are treated as part of the *relevance* equation, which identifies cases that select into the cases at-risk of violating their alliances in the *violation* equation. As expected, the coefficient on *territorial threat* is positive and statistically significant in the *relevance* equation, which indicates that states are more likely to enter the at-risk population for alliance abrogation when external territorial threats are high.

Turning to the *violation* equation, we see that, after accounting for the underlying sample-selection process, the sign on *democracy* is now positive, though statistically insignificant. This result is suggestive that the previous finding of a negative and significant effect associated with *democracy* in Models 1 and 2 may, in fact, have indicated that democracies were

[16] We focus on "opting into," rather than "opting out of," the relevant sub-sample that has the opportunity to violate an alliance—i.e. those that are in dangerous neighborhoods. Our focus on observations being treated as "at-risk" or "opting in" to the outcome equation, of course, is the mathematical inverse of identifying the "zero-inflated" observations that "opt out." Reporting results in this manner is consistent with previous studies using this estimation technique (e.g., Xiang 2010, 2017; Bagozzi 2016; Nieman 2015, 2018).

Table 1: Political Environment, Democracy, and Alliance Violations.

Time	LMV Original 1920-2001	LMV Reduced 1920-2001	Full 1920-2001	LMV Reduced 1950-2001	Full 1950-2001
Outcome Equation					
Change in Leader's Societal Coalition	0.889*	0.910*	5.477	1.772*	11.247
	(0.436)	(0.444)	(5.055)	(0.528)	(9.586)
Democracy	-1.322*	-1.341*	1.336	-1.813*	0.856
	(0.393)	(0.401)	(1.841)	(0.556)	(2.517)
Change in International Power	0.803*	0.877*	3.018	0.727	2.222*
	(0.330)	(0.340)	(2.247)	(0.464)	(0.928)
Change in Political Institutions	0.131	0.195	1.703	0.010	2.651
	(0.308)	(0.308)	(1.294)	(0.465)	(2.365)
Change in External Threat	0.421	0.425	0.765	-0.069	-0.673
	(0.270)	(0.278)	(1.261)	(0.445)	(1.859)
Formation of New Outside Alliance	1.070*	1.064*	2.452	1.145*	2.985*
	(0.253)	(0.261)	(1.628)	(0.380)	(1.298)
Asymmetry	-0.408	-0.503	-1.625	-0.510	-1.696
	(0.259)	(0.265)	(0.909)	(0.424)	(1.329)
Non-military Cooperation	-0.746*	-0.746*	-2.684*	0.055	-2.784
	(0.257)	(0.261)	(1.190)	(0.498)	(3.363)
Ratification	-0.083	-0.134	1.513	-0.265	1.172
	(0.354)	(0.355)	(2.299)	(0.466)	(4.741)
Military Cooperation	0.557*	0.575*	4.978*	1.043*	5.794*
	(0.180)	(0.186)	(2.224)	(0.280)	(2.552)
Time	-0.086	-0.069	-0.124	-0.000	-0.065
	(0.077)	(0.076)	(0.276)	(0.114)	(0.526)
Time Squared	0.001	-0.000	-0.003	-0.002	-0.002
	(0.004)	(0.004)	(0.010)	(0.006)	(0.022)
Time Cubed	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-4.448*	-4.464*	-3.210*	-5.660*	-3.412
	(0.440)	(0.455)	(1.447)	(0.901)	(3.267)
Relevance (At-risk) Equation					
Territorial Threat			4.198*		6.412*
			(1.126)		(2.620)
Proportion of Democratic Borders			-0.299*		-0.072
			(0.121)		(0.246)
Number of Borders			0.119		0.012
			(0.062)		(0.125)
Major Power			0.446		-0.291
			(0.511)		(1.087)
Rivalry			-0.009		-0.140
			(0.399)		(0.477)
Cold War			0.424		2.195*
			(0.309)		(1.043)
Economic Development			-0.150		-0.163
			(0.083)		(0.137)
Oil Producer			0.675		1.333*
			(0.513)		(0.540)
Constant			-4.564*		-5.178*
			(0.403)		(1.406)
Log-likelihood	-352.394	-339.988	-307.564	-177.293	-156.730
Observations (Alliances)	6612 (223)	6395 (223)	6395 (223)	4952 (139)	4952 (139)

Note: *p <0.05, two-tailed. Standard errors in parentheses. Point estimates and standard errors in models 3 and 5 were calculated from 10 draws using Rubin's (1987) formula for multiple imputation to account for uncertainty in the *territorial threat* instrumental variable.

less likely to violate their alliances not because they are more reliable, but because they exist in peaceful neighborhoods and were less likely to have military provisions invoked. Thus, the endogeneity between peace and democracy seems to explain why democracies behave differently when engaging in alliance politics. Similarly, *changes in leader's societal coalition* also fails to reach any traditional level of statistical significance.

To more formally assess whether *democracy* exerts a null or negligible effect once the political environment is accounted for, we use a technique introduced to political science by Rainey (2014). The idea is to identify the smallest 'meaningful effect', and then determine whether the estimated quantity of interest meets this threshold. If the estimate (and its 90% confidence interval) is smaller than the threshold identified as the smallest 'meaningful effect', then the effect is marginal; i.e. there is statistical evidence that the variable has little or no effect on the outcome of interest. If the estimate (and its 90% confidence interval) is larger than the threshold, then the effect is appears to be non-marginal.

Given that estimated parameters in logit-based models are difficult to interpret directly, we follow Rainey's advice and focus on assessing whether the independent variable affects the predicted probability of the outcome of interest. In our case, we are evaluating whether *democracy* exerts a meaningful effect by seeing if it reduces the likelihood of alliance violations by at least one-half of one percent (0.5%); i.e. the effect size should be less than -0.005. To do this, we take the first difference of the parameter of interest, holding all other parameters at the their mean or median values. The 90% confidence interval *democracy* [-0.001, 0.007] is *greater* than -0.005, indicating that the effect of *democracy* on reducing alliances violations is negligible. This result is consistent with our theorized expectations and indicates that the influence of political institutions may actually be attributable to political environment rather than the institutions themselves.

Applying the same test to *changes in leader's societal coalition*, we find different results. We expect a meaningful effect to *increase* the likelihood of alliance violations by at least one-half of one percent when there is a change in a leader's societal coalition. In this case, the 90% confidence interval [-0.001, 0.020] includes 0.005, indicating that, though it is not statistically significant, we cannot rule out a meaningful effect for *changes in leader's societal coalition*.

We re-estimate the baseline and split-population models for the period 1950-2001 in Models 4 and 5. Once again, *territorial threat* is positive and statistically significant, and *democracy* is no longer statistically significant once the sample-selection process of identifying "at-risk" observations are modeled jointly with the likelihood of abrogation. Applying the more formal test of the smallest 'meaningful effect', we again expect a 0.5% reduction in alliance violations as the threshold. We again find that the 90% confidence interval for *democracy* [-0.001, 0.004] is again larger, rather than smaller, than the smallest meaningful effect size. Taken together with the previous results, this indicates that the negative and statistically significant results associated with *democracy* in Models 2 and 4 are likely spurious, and their effect is instead attributable to the political environment in which democracies form and exist. Lastly, and in contrast to estimates for time period 1920-2001 reported in Model 3, *change in leader's societal coalition* is statistically significant at the $p < 0.10$-level in the period after 1950, even after accounting for the sample-selection processes of identifying the "at-risk" observations. This difference suggests that the effect of changes in a leader's societal coalition are stronger in the 1950-2001 than before.

We formally compare and evaluate the model fits for the baseline and split-population models, for each temporal sample, using the Vuong and Clarke distribution-free tests, in Table A.2 on p. 4 in the Appendix. These tests indicate a strong preference for the split-population models over the reduced model.[17] We also assess the reliability and robustness of the estimates from the split-population logit in Tables A.3 and A.4 on pp. 6-7 in the Appendix. The stability of the estimates in the partially observed (relevancy) equation indicate the split-population logit is appropriate.

Overall, the results from these empirical analyzes are consistent with our theoretical expectations. Threatening environments affect the underlying propensity of states to enter the sample of states at-risk of violating their alliance terms. Moreover, once the effects of threat environment are accounted for, democratic institutions do not appear to exert a significant

[17] We apply the Schwarz's correction to each test in order to penalize the split-population models for the additional parameters they estimate.

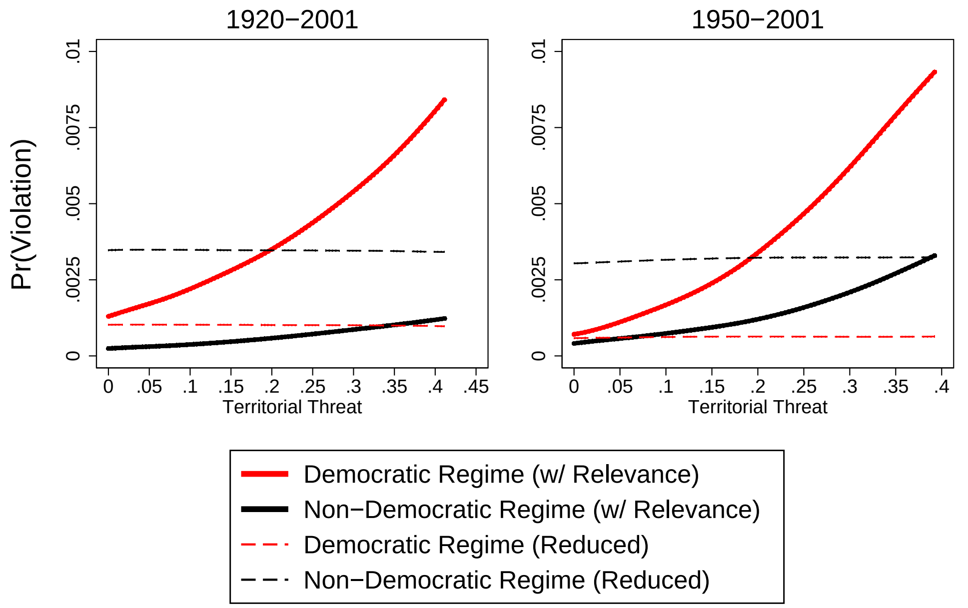
impact on whether an alliance violation occurs. States in non-threatening environments appear less likely to be involved in the traditional, capability-aggregating alliances associated with international conflict. Instead, alliances signed in peaceful environments seem to have been signed for other reasons, such as deepening economic ties (Gowa and Mansfield 2004; Li and Vashchilko 2010) or signaling embeddedness within a political network (Lake 2009; Nieman 2016; Henke 2017). When democracies do form alliances, their alliances are more likely to be among these other alliance types.

Substantive effects

To assess the substantive effect of accounting for threat environment on alliance abrogation, we report predicted probabilities of an alliance violation in Figures 2 and 3. In the first figure, we report predicted probabilities for two conditions: for democracies [red line] and for non-democracies [black line], after accounting for the level of territorial threat affecting the state [solid lines]. As a point of reference, we also compare these predicted probabilities to those from the reduced model [dashed lines], which does not account for the effects of territorial threat on identifying the degree to which an observation is part of the at-risk subsample. In the second figure, we repeat this procedure, reporting a change in the leader's winning coalition [red line] and when there is no change in the leader's winning coalition [black line], after accounting for the level of territorial threat affecting the state [solid lines] and provide a reference to the reduced model [dashed lines]. To make the substantive results more realistic, and to ensure that outliers are not skewing our interpretation, we visualize predicted probabilities of alliance abrogation for the middle 95% of values of *territorial threat* from the estimated sample. Finally, we report predicted probabilities for both the full time period and the 1950-2001 period.

We begin by examining Figure 2, which focuses on the effects of democracy and territorial threat. Figure 2 shows that, while democracy is associated with a lower likelihood of alliance violation in the original analysis, which ignored potential selection effect [red dashed line is

Figure 2: Predicted Probabilities of an Alliance Violation, Democracy, and Territorial Threat.

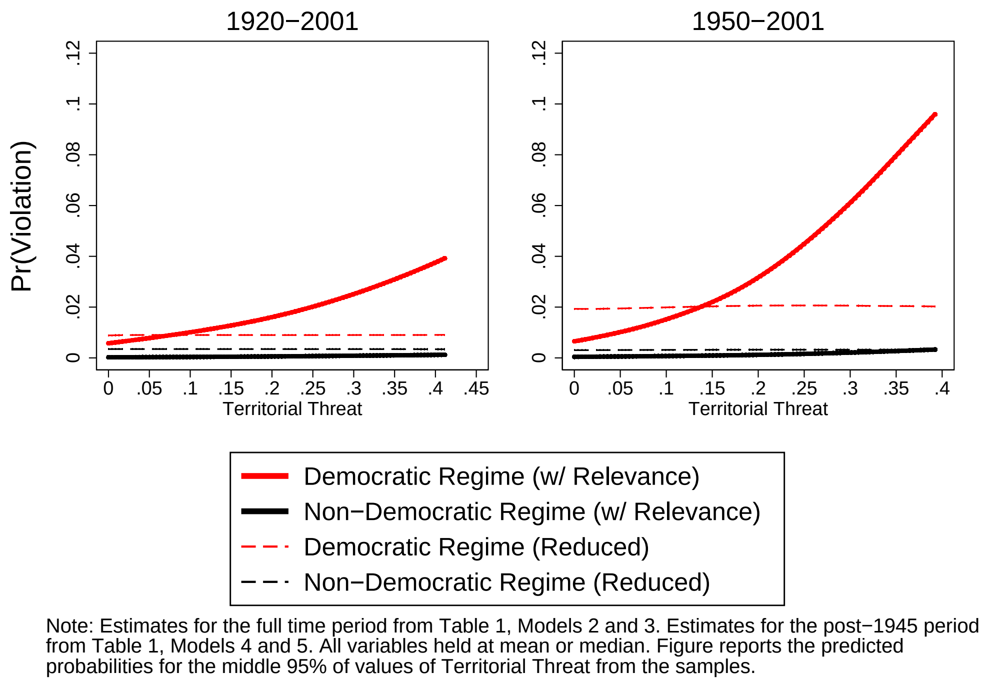


Note: Estimates for the full time period from Table 1, Models 2 and 3. Estimates for the post−1945 period from Table 1, Models 4 and 5. All variables held at mean or median. Figure reports the predicted probabilities for the middle 95% of values of Territorial Threat from the samples.

below the black dashed line], democracies are associated with an increase the probability of an alliance violation [red solid line above black solid line]. The former result runs in contrast to traditional expectations regarding the effects of democratic institutions. Moreover, increases in territorial threat increase the probability of an alliance violation regardless of whether regimes are democratic [solid lines increase], suggesting that it is the threat environment that is driving the change. Comparing the subfigure for the full sample to that for the 1950-2001 sample, it is clear that the effects seen in both the original analysis and the analysis that accounts for threat environment are similar, though the difference is weaker in the 1950-2001 period. It is worth keeping in mind, of course, the difference between democracies and non-democracies were shown to be negligible (or in the wrong direction), according to the test suggested by Rainey (2014).

Moving to the impact of changes in a leader's winning coalition, Figure 3 demonstrates

Figure 3: Predicted Probabilities of an Alliance Violation, Change in Leader's Winning Coalition, and Territorial Threat.



that a change is associated with increases in the probability of an alliance violation [red lines above corresponding black lines] and also that a change in the winning coalition increases the likelihood of a violation as the degree of territorial threat rises [solid red line]. Increases in territorial threat also increase the probability of an alliance violation when there is no change in a leader's winning coalition [solid black line], though this effect is much smaller.[18] The figure shows that threat environment significantly influences the effect of changes in the leader's winning coalition on the probability of an alliance violation. Moreover, comparing the subfigure for the full sample to that for the 1950-2001 sample, it is clear that this effect is stronger in the 1950-2001 period.

Notably, the risk effect matters at all points along the territorial threat spectrum in both

[18]The predicted probability of an alliance violation in a democratic state, which experiences a change in their leader's winning coalition, is almost identical to the probability of a violation when there is a change in the leader's winning coalition in a non-democracy, at every territorial threat level. This suggests that the interaction of the two variables exerts little substantive impact, once territorial threat is accounted for.

Figures 2 and 3: in low threat environments, the presence of a democracy exerts only a small risk of an alliance violation. Similarly, in the absence of an external threat, a change in a leader's winning coalition have very little effect on alliance abrogation. Instead, it appears abrogation becomes much more likely for democracies as the the level of territorial threat increases. The same holds for a change in the leader's winning coalition: alliance abrogation is more likely as territorial threat increases. In the latter case, these effects appear to be even stronger in the 1950-2001 period. One implication that is evident from the figures is that ignoring threat environment significantly *overestimates* the effect of both democracy and changes in a leader's winning coalition at low levels of threat and significantly *underestimates* these effects at high levels of threat. That is, accounting for territorial threat, and the political environment more broadly, improves our understanding of the roles of political institutions and leadership changes in substantively meaningful ways.

Conclusion

We began this paper by pointing out that endogeneity between peace and democracy will bias additive-model estimates of many other democracy-related arguments, and we have shown that to be the case with regard to international alliances. Democracies in alliances have been thought to be more reliable, but we demonstrate that this result is likely to be spurious. Democracy is more likely to take hold in peaceful international environments, and peaceful environments seldom provoke the type of alliance making associated with aggregating capabilities to defend the state. In other words, democratic alliances are different from other types of alliances, but this has little to do with regime type.

Also noteworthy is our finding that, under some conditions, traditional alliance theories may be correct. Quantitative analyses of alliances and conflict generally pool the sample of all cases to assess conflict-proneness and reliability. Our findings suggest, however, that their are two qualitatively distinct types of alliances: power-aggregating alliances formed in hostile environments and alliances serving other functions formed in more peaceful environments. The former type of alliance is a reaction to threats to the state and is manifestly different in their behavior. These alliances are shorter statements of intentions attempting to ward off potential aggressors, and the commitments expressed in these treaties are much more likely to be abrogated. Alliances that correlate with conflict may not be more than the scraps of paper traditional theories expect. Accounting for the threat environment, therefore, may help to explain why alliances are correlated with peace in some periods, and with conflict, or even the diffusion of conflict, in others (e.g., Levy 1981; Kadera 1998; Senese and Vasquez 2008).

Finally, our argument and results have implications beyond the alliance literature, raising concerns about a number of second-order findings associated with the democratic peace research program more broadly. Current scholarship suggests that democratic states trade more often with other democracies, and democracies may also be more active in international governance. Each of these literatures, however, tend to pool samples without regard to threat environment, potentially biasing results by attributing sole explanatory power to an outcome—political institutions—rather an (at least partial) underlying causal process—peaceful political environments. This criticism, moreover, extends to almost all studies that find some type of democratic difference in state behavior. Ultimately, democratic institutions may still affect state behavior once peaceful environments take hold, but we just do not yet know. Our paper presents an important set of questions for these long-accepted relationships: without control for the effect of dangerous environments, current estimates of the effect of democracy on behavior are biased and may be spurious.

References

Altfeld, Michael F. 1984. "The Decision to Ally: A Theory and Test." *Western Political Quarterly* 37(4):523–544.

Bagozzi, Benjamin E. 2016. "The Baseline-inflated Multinomial Logit Model for International Relations Research." *Conflict Management and Peace Science* 33(2):174–197.

Bagozzi, Benjamin E. and Bumba Mukherjee. 2012. "A Mixture model for Middle Category Inflation in Ordered Survey Responses." *Political Analysis* 20(3):369–386.

Beger, Andreas, Jacqueline H.R. DeMeritt, Wonjae Hwang and Will H. Moore. 2011. "The Split Population Logit (SPopLogit): Modeling Measurement Bias in Binary Data." SSRN Working Paper. http://ssrn.com/abstract=1773594.

Bennett, D Scott. 1997. "Testing alternative models of alliance duration, 1816-1984." *American Journal of Political Science* pp. 846–878.

Bennett, D. Scott and Allan C. Stam. 2004. *The Behavioral Origins of War.* Ann Arbor, MI: The University of Michigan Press.

Berkemeier, Molly and Matthew Fuhrmann. 2018. "Reassessing the Fulfillment of Alliance Commitments in War." *Research & Politics* . doi: 10.1177/2053168018779697.

Boehmke, Frederick J, Olga Chyzh and Cameron G Thies. 2016. "Addressing Endogeneity in Actor-Specific Network Measures." *Political Science Research and Methods* 4(1):123–149.

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36(2):309–341.

Carter, David B. and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271–292.

Cederman, Lars-Erik and Kristian Skrede Gleditsch. 2004. "Conquest and Regime Change: An Evolutionary Model of the Spread of Democracy and Peace." *International Studies Quarterly* 48(3):603–629.

Chiba, Daina, Carla Martinez Machain and William Reed. 2014. "Major powers and militarized conflict." *Journal of Conflict Resolution* 58(6):976–1002.

Chyzh, Olga. 2014. "Can You Trust a Dictator: A Strategic Model of Authoritarian Regimes' Signing and Compliance with International Treaties." *Conflict Management and Peace Science* 31(1):3–27.

Colgan, Jeff D. 2013. *Petro-aggression: When Oil Causes War.* Cambridge University Press.

Correlates of War Project. 2016. "State System Membership List, v2016.". url: http://correlatesofwar.org.

Diehl, Paul and Gary Goertz. 2000. *War and Peace in International Relations.* Ann Arbor: University of Michigan Press.

Digiuseppe, Matthew and Paul Poast. 2018. "Arms versus Democratic Allies." *British Journal of Political Science* 48(4):981–1003.

Doyle, Michael W. 1986. "Liberalism and World Politics." *American Political Science Review* 80(4):1151–1169.

Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41(1):68–90.

Gartzke, Erik and Kristian Skrede Gleditsch. 2004. "Why Democracies May Actually Be Less Reliable Allies." *American Journal of Political Science* 48(4):775–595.

Gaubatz, Kurt Taylor. 1996. "Democratic states and commitment in international relations." *International Organization* 50(01):109–139.

Gibler, Douglas M. 1996. "Alliances that never balance: The territorial settlement treaty." *Conflict Management and Peace Science* 15(1):75–97.

Gibler, Douglas M. 1997. "Control the Issues, Control the Conflict: The Effects of Alliances that Settle Territorial Issues on Interstate Rivalries." *International Interactions* 22(4):341–368.

Gibler, Douglas M. 2012. *The Territorial Peace: Borders, State Development, and International Conflict*. New York, NY: Cambridge University Press.

Gibler, Douglas M. and Jaroslav Tir. 2014. "Territorial Peace and Democratic Clustering." *Journal of Politics* 76(1):27–40.

Gibler, Douglas M. and Scott Wolford. 2006. "Alliances, Then Democracy: An Examination of the Relationship between Regime Type and Alliance Formation." *Journal of Conflict Resolution* 50(1):129–153.

Gleditsch, Kristian Skrede. 2009. *All International Politics is Local: The Diffusion of Conflict, Integration, and Democratization*. University of Michigan Press.

Gleditsch, Kristian Skrede and Michael D. Ward. 2006. "Diffusion and the International Context of Democratization." *International Organization* 60(4):911–933.

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

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–162.

Henke, Marina E. 2017. "The Politics of Diplomacy: How the United States Builds Multilateral Military Coalitions." *International Studies Quarterly* 61(2):410–424.

Johnson, Jesse C. 2015. "The cost of security: Foreign policy concessions and military alliances." *Journal of Peace Research* 52(5):665–679.

Johnson, Jesse C. 2017. "External Threat and Alliance Formation." *International Studies Quarterly* p. sqw054.

Kadera, Kelly M. 1998. "Transmission, Barriers, and Constraints: A Dynamic Model of the Spread of War." *Journal of Conflict Research* 42(3):367–387.

Kadera, Kelly M., Mark J. C. Crescenzi and Megan L. Shannon. 2003. "Democratic Survival, Peace, and War in the International System." *American Journal of Political Science* 47(2):234–247.

Kimball, Anessa L. 2006. "Alliance formation and conflict initiation: The missing link." *Journal of Peace Research* 43(4):371–389.

Klein, James P., Gary Goertz and Paul F. Diehl. 2006. "The New Rivalry Dataset: Procedures and Patterns." *Journal of Peace Research* 43(3):331–348.

Lai, Brian and Dan Reiter. 2000. "Democracy, Political Similarity, and International Alliances, 1816-1992." *The Journal of Conflict Resolution* 44(2):203–227.

Lake, David A. 1996. "Anarchy, Hierarchy, and the Variety of International Relations." *International Organization* 50(1):1–33.

Lake, David A. 2009. *Hierarchy in International Relations.* Ithaca, NY: Cornell University Press.

Langlois, Catherine C. 2012. "Power and Deterrence in Alliance Relationships: The Ally's Decision to Renege." *Conflict Management and Peace Science* 29(2):148–169.

Leeds, Brett Ashley. 1999. "Domestic political institutions, credible commitments, and international cooperation." *American Journal of Political Science* pp. 979–1002.

Leeds, Brett Ashley. 2003. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57(4):801–827.

Leeds, Brett Ashley and Burcu Savun. 2007. "Terminating Alliances: Why do States Abrogate Agreements?" *Journal of Politics* 69(4):1118–1132.

Leeds, Brett Ashley, Jeffrey M. Ritter, Sara McLaughlin Mitchell and Andrew G. Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28(3):261–284.

Leeds, Brett Ashley and Michaela Mattes. 2007. "Alliance Politics during the Cold War: Aberration, New World Order, or Continuation of History?" *Conflict Management and Peace Science* 24(3):183–199.

Leeds, Brett Ashley, Michaela Mattes and Jeremy S. Vogel. 2009. "Interests, Institutions, and Reliability of International Commitments." *American Journal of Political Science* 53(2):461–476.

Levy, Jack S. 1981. "Alliance formation and war behavior: An analysis of the great powers, 1495-1975." *Journal of Conflict Resolution* pp. 581–613.

Li, Quan and Tatiana Vashchilko. 2010. "Dyadic military conflict, security alliances, and bilateral FDI flows." *Journal of International Business Studies* 41(5):765–782.

Marshall, Monty G. and Keith Jaggers. 2014. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013." http://www.systemicpeace.org/polity/polity4.htm.

Mattes, Michaela and Greg Vonnahme. 2010. "Contracting for Peace: Do Nonaggression Pacts Reduce Conflict?" *Journal of Politics* 72(4):925–938.

McDonald, Patrick J. 2015. "Great Powers, Hierarchy, and Endogenous Regimes: Rethinking the Domestic Causes of Peace." *International Organization* 69(3):557–588.

McManus, Roseanne W. 2018. "Making it personal: The role of leader-specific signals in extended deterrence." *The Journal of Politics* 80(3):982–995.

McManus, Roseanne W and Keren Yarhi-Milo. 2017. "The Logic of "Offstage" Signaling: Domestic Politics, Regime Type, and Major Power-Protégé Relations." *International Organization* 71(4):701–733.

Mislevy, Robert. 1991. "Randomization-based Inference about Latent Variables from Complex Samples." *Psychometrika* 56(2):177–196.

Mitchell, Sara McLaughlin and Cameron G. Thies. 2011. "Issue Rivalries." *Conflict Management and Peace Science* 28(3):1–31.

Morgenthau, Hans. 1960. *Politics Among Nations*. 3rd ed. New York, NY: Knopf. Originally published in 1948.

Morrow, James. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35(4):904–933.

Morrow, James. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38(2):270–297.

Mousseau, Michael. 2003. "The Nexus of Market Society, Liberal Preferences, and Democratic Peace: Interdisciplinary Theory and Evidence." *International Studies Quarterly* 47(4):483–510.

Nieman, Mark David. 2015. "Statistical Analysis of Strategic Interaction with Unobserved Player Actions: Introducing a Strategic Probit with Partial Observability." *Political Analysis* 23(3):429–448.

Nieman, Mark David. 2016. "The Return on Social Bonds: Social Hierarchy and International Conflict." *Journal of Peace Research* 53(5):665–679.

Nieman, Mark David. 2018. "Strategic Binary Choice Models with Partial Observability." *Statistica Sinica* 28(4):2089–2105.

Poirier, Dale J. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics* 12(2):209–217.

Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58(4):1083–1091.

Rainey, Carlisle and Robert A. Jackson. 2017. "Unreliable Inferences About Unobserved Processes: A Critique of Partial Observability Models." *Political Science Research and Methods* Forthcoming. doi:10.1017/psrm.2017.3.

Rasler, Karen A. and William Thompson. 2006. "Contested Territory, Strategic Rivalries, and Conflict Escalation." *International Studies Quarterly* 50(1):145–167.

Reed, William and Daina Chiba. 2010. "Decomposing the Relationship Between Contiguity and Militarized Conflict." *American Journal of Political Science* 54(1):61–73.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Russett, Bruce M. and John R. Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations.* New York, NY: WW Norton & Company.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions." *Political Analysis* 11(2):111–138.

Schroeder, Paul W. 1976. *Alliances, 1815-1945: Weapons of power and tools of management.* Lawrence, KS: University of Kansas Press pp. 227–62.

Senese, Paul D. and John A. Vasquez. 2008. *The Steps to War: An Empirical Study.* Princeton, NJ: Princeton University Press.

Signorino, Curtis S. and Kuzey Yilmaz. 2003. "Strategic Misspecification in Regression Models." *American Journal of Political Science* 47(3):551–566.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Association* 94(4):819–835.

Singer, J. David, Stuart Bremer and John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills, CA: Sage Publications.

Siverson, Randolph M. and Juliann Emmons. 1991. "Birds of a Feather: Democratic Political Systems and Alliance Choices in the Twentieth Century." *The Journal of Conflict Resolution* 35(2):285–306.

Stinnett, Douglas M., Jaroslav Tir, Philip Schafer, Paul F. Diehl and Charles Gochman. 2002. "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Management and Peace Science* 19(2):58–66.

Stock, James H. and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd international edition*. Upper Saddle River, NJ: Prentice Hall.

Thompson, William. 1996. "Democracy and Peace: Putting the Cart before the Horse?" *International Organization* 50(1):141–174.

Vasquez, John A. 1995. "Why Do Neighbors Fight? Proximity, Interaction, or Territory." *Journal of Peace Research* 32(3):277–293.

Vasquez, John A. 2009. *The War Puzzle Revisited*. New York, NY: Cambridge University Press.

Walt, Stephen M. 1985. "Alliance Formation and the Balance of World Power." *International Security* 9(4):3–43.

Waltz, Kenneth N. 1979. *Theory of international politics*. Waveland Press.

Ward, Michael D and Kristian Skrede Gleditsch. 2002. "Location, location, location: An MCMC approach to modeling the spatial context of war and peace." *Political analysis* 10(3):244–260.

Weitsman, Patricia. 2004. *Dangerous Alliances: Proponents of Peace, Weapons of War.* Stanford, CA: Stanford University Press.

Xiang, Jun. 2010. "Relevance as a Latent Variable in Dyadic Analysis of Conflict." *Journal of Politics* 72(2):484–498.

Xiang, Jun. 2017. "Dyadic Effects, Relevance, and the Empirical Assessment of the Kantian Peace." *International Interactions* 43(2):248–271.

Appendix for: Cooperation in Good Times: Are Democracies Really Different?

Construction of the Instrumental Variable, *Territorial Threat*

Table A.1 reports the estimates of the variables used to construct the primary independent variable, *territorial threat*. Data and model specification are from Gibler and Tir (2014, Table 1). Our model differs from Gibler and Tir (2014), however, in that while they use the period from 1816-1999, we use the period 1900-2001. We make this change because the factors that affect conflict in the nineteenth century may not be the same as those in the twentieth (Bennett and Stam 2004). All results are consistent with their findings, with the exception of *Civil War in Either State*, which is significant at the p <0.1-level in our replication, compared to the p <0.05-level in the original analysis.

Table A.1: Predicting Fatal MIDs in Contiguous Dyads, 1900–2001.

	β	S.E.
Same Colonial Master	0.235*	(0.113)
Peaceful Territorial Transfer in Dyad	-0.568*	(0.184)
Violent Territorial Transfer in Dyad	0.514*	(0.116)
Defense Pact with All Neighbors	-0.908*	(0.213)
Civil War in Either State	0.169	(0.102)
Highest Militarization Level Among Neighbors	14.023*	(1.619)
Previous Territorial MID Against Either State	0.424*	(0.091)
Border Age (logged)	0.181*	(0.036)
Peace Years	-0.434*	(0.024)
Peace Years (Squared)	0.012*	(0.001)
Peace Years (Cubed)	-0.000*	(0.000)
Constant	-2.753*	(0.134)
Log-likelihood	-1877.638	
Observations	15058	

Note: *p <0.05, two-tailed. Replication of Gibler and Tir (2014, Table 1) for 1900–2001.

Model Fit Comparisons

We demonstrate that inclusion of the *relevance* equation improves our model fit and the quality of our estimates. To do this, we compare the model fit of models 2 and 3 from Table 1, using Vuong's (1989) and Clarke's (2003, 2007) tests for non-nested models. We use the Vuong and Clarke tests, rather than an F-test or likelihood ratio test because models 2 and 3 are non-nested due to their differing functional forms: model 2 assumes an additive non-linear logit function while model 3 is a mixture of two logistic distributions (for a discussion on differing types of non-nested models, see Clarke 2001).[1] We also compare models 4 and 5 from Table 1, which provided analogous estimates during the 1950-2001 period, as a robustness check.

The Vuong test compares the mean log-likelihood ratios of two models. If the first model is closer to the true specification, then the mean log-likelihood ratio is positive and statistically significant. As is common practice, we apply the Schwarz's correction to the Vuong test. The correction penalizes for the inclusion of additional parameters in a model. That is, the models that include the *relevance* equation are penalized because they estimate more parameters than the reduced model. More formally, the corrected Vuong test is:

$$LR_n\left(\tilde{\theta}_n, \tilde{\gamma}_n\right) - \left[\left(\frac{p}{2}\right)\ln n - \left(\frac{q}{2}\right)\ln n\right] \tag{1}$$

where LR is the log-likelihood ratio, $\tilde{\theta}$ and $\tilde{\gamma}$ are the model estimates, and p and q are the number of estimated parameters for model f and g, which are the two models being compared (Vuong 1989).

Clarke's distribution-free test, meanwhile, tests whether the median logged ratio of the likelihood for the individual observations of two models are equal. If the first model is closer to the true specification, more than half of the individual logged ratios of the likelihoods will

[1]Neither AIC nor BIC are appropriate as they do not include information from the rival theory, nor do they permit probabilistic statements regarding model selection (Clarke 2003).

be greater than zero. More formally:

$$H_0 : Pr_0 \left[\ln \frac{f\left(Y_i|X_i;\beta_*\right)}{g\left(Y_i|Z_i;\gamma_*\right)} > 0 \right] = 0.5 \tag{2}$$

where the numerator is estimated model f, which predicts Y_i from a set of covariates, X_i, and estimated parameters, β_*; the denominator is estimated model g, which predicts Y_i from a set of covariates, Z_i, and estimated parameters, γ_*. The null hypothesis is that the median logged ratio of the likelihoods between the two models is equal to 0, i.e. the probability that the median logged ratio of the likelihoods of f is greater than g is 0.5. If d_i is set equal to $\ln f\left(Y_i|X_i;\beta_*\right) - \ln g\left(Y_i|Z_i;\gamma_*\right)$, the test statistic is:

$$B = \sum_{i=1}^{n} I_{(0,+\infty)}\left(d_i\right) \tag{3}$$

where I is a dichotomous indicator equal to 1 if $n_i > 0$ in Equation 2, and 0 if $n_i \leq 0$. Equation 3 is the sum of positive differences and is distributed according to a Binomial distribution with n trials and a mean equal to 0.5. We apply the average Schwarz correction to Clarke's distribution-free test, adjusting the individual log-likelihoods for model f by a factor $[(p/2n)\ln n]$ and those of model g by a factor $[(q/2n)\ln n]$ (see Clarke 2007, 350).

Table A.2 reports the results of our non-nested model comparisons. The test statistic for the Vuong test is 40.02 for the full model (Table 1, Model 3) compared to the reduced model (Table 1 Model 2). This returns a p-value of <0.001, allowing us to reject the null that the models are equivalent. Using Clarke's test, we find that the split-population model returns a positive log-likelihood ratio for 4468 of the 6395 observations, which generates a p-value of <0.001. We are thus able to reject the null that the models are equal and again find empirical support for the full models. We find similar support for the full split-population model (Table 1, Model 5) versus the reduced model (Table 1, Model 4) in the 1950-2001 sample. In sum, the results indicate that the two split-population models outperform the models which assume all states are initially equally "at-risk" of violating an alliance.

Table A.2: Comparison of Model Fit.

	Full	post-1945
Vuong Test		
Vuong	71.86	58.85
SE	1.80	3.23
t-statistic	40.02	18.23
p-value	$< .001$	$< .001$
Clarke Test		
$\sum_{\mathrm{i}}^{\mathrm{n}}(\mathrm{ll}_{\mathrm{M_{Full}},i} - \mathrm{ll}_{\mathrm{M_{LMV}},i} > 0)$	4468	3559
$\sum_{\mathrm{i}}^{\mathrm{n}}(\mathrm{ll}_{\mathrm{M_{Full}},i} - \mathrm{ll}_{\mathrm{M_{LMV}},i} < 0)$	1927	1393
Positive, one-side test (p-value)	$< .001$	$< .001$

Note: Clarke distribution-free test uses binomial distribution ($p = .5$).

Robustness Checks

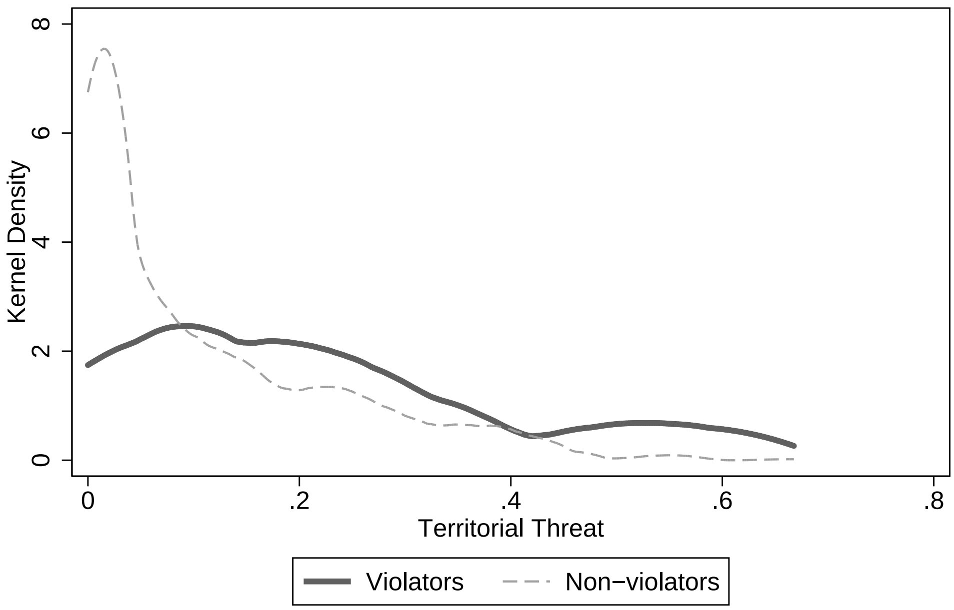
We begin our robustness checks by visualizing the difference in *territorial threat* between states that violate an alliance and the full sample (from Table 1 below). The mean territorial threat for alliance violators is 0.215 with a standard deviation of 0.193 and $N = 74$, while the mean for non-violators is 0.108 with a standard deviation of 0.127 and $N = 6768$. A difference of means between the two samples is statistically significant with $p < 0.001$. As Figure A.1 demonstrates, a large swath of states at low levels of territorial threat comprise a large proportion of the non-violators.

Next, we report several additional models in Tables A.3 and A.4. These are various reduced models that examine the sensitivity of the original model when controlling for the level of external, territorial threat. We also run a model that uses GDP per capita in place of energy consumption per capita in the 1950-2001 period.[2] Note in the relevance equation that territorial threat is stable across all models and is consistently able to differentiate among alliance types—i.e. territorial threat is more likely to lead to alliance violations.

The outcome equation, which is the original model conditioned by relevance, demonstrates that the two primary variables of interest—change in leader coalition and democracy—are not consistent predictors of alliance violation. Democracy is only statistically significant

[2] GDP per capita data are only available from 1950

Figure A.1: Comparison of Kernel Density of Alliance Violators and Sample Average.



Note: The mean territorial threat for alliance violators is 0.215 with a standard deviation of 0.193, while the mean for non−violators sample is 0.108 with a standard deviation of 0.127. A difference of means between the two samples statistically significant at $p < 0.001$.

in one of the models, and it has a positive coefficient, meaning that democracies are more likely to violate their alliances. These results, of course, strongly contradict the argument that democracies are less likely to abrogate their treaties. The generally negligible effects (with the only significant result running in the wrong direction) only emphasizes the conclusion that democracy is not an accurate predictor of alliance violators once the territorial threat environment is considered.

We report results in the manuscript that support the idea that leader change matters for alliance violation, at least among states in the 1950-2001 sample, but our additional models suggest this result may be dependent upon model specification. Once we consider the threat environment affecting the alliance member, the effect of leader change disappears. The original model suggests leader change is statistically insignificant, with an effect size that is quite large, but our re-analyses imply statistical significance at the traditional $p < 0.05$-level in less than half of the conditioned models in Tables A.3 and A.4.

Table A.3: Political Environment, Democracy, and Alliance Violations, Robustness Checks.

Model	Simple	Democratic Development	Full Development	Change in Alliance	Alliance Terms
Outcome Equation					
Change in Leader's Societal Coalition	1.149*	12.506*	13.555*	12.482	15.839*
	(0.475)	(6.005)	(2.552)	(23.077)	(7.443)
Democracy	-0.508	0.566	0.568	1.491	2.311*
	(0.458)	(0.830)	(0.854)	(1.226)	(1.133)
Change in International Power				1.980*	
				(0.725)	
Change in Political Institutions				0.252	
				(0.637)	
Change in External Threat				0.903	
				(0.847)	
Formation of New Outside Alliance				1.578*	
				(0.693)	
Asymmetry					-2.204*
					(0.735)
Non-military Cooperation					-2.204*
					(0.735)
Ratification					2.676
					(1.456)
Military Cooperation					18.621*
					(2.152)
Time	0.028	0.051	0.114	-0.119	0.108
	(0.073)	(0.109)	(0.116)	(0.140)	(0.154)
Time Squared	-0.005	-0.006	-0.009	0.001	-0.010
	(0.005)	(0.006)	(0.007)	(0.007)	(0.009)
Time Cubed	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-1.700	-0.392	-0.703	-2.078*	-1.224
	(0.981)	(0.752)	(0.699)	(0.672)	(1.209)
Relevance Equation					
Territorial Threat	4.686*	4.312*	3.849*	3.664*	4.509*
	(0.791)	(1.025)	(1.045)	(1.014)	(0.969)
Number of Borders		0.083	0.046	0.067	0.124*
		(0.066)	(0.069)	(0.067)	(0.056)
Proportion of Democratic Borders		-0.347*	-0.303	-0.334*	-0.281*
		(0.153)	(0.157)	(0.148)	(0.112)
Major Power			0.574	0.345	0.756*
			(0.381)	(0.387)	(0.374)
Rivalry			0.248	0.171	0.110
			(0.372)	(0.386)	(0.385)
Cold War			-0.048	0.178	0.466
			(0.263)	(0.275)	(0.287)
Economic Development			-0.230*	-0.216*	-0.189*
			(0.078)	(0.075)	(0.075)
Oil Producer			0.374	0.422	0.887*
			(0.362)	(0.376)	(0.372)
Constant	-3.097*	-4.345*	-4.509*	-4.447*	-5.454*
	(0.925)	(0.385)	(0.464)	(0.522)	(0.444)
Log-likelihood	-385.975	-377.079	-364.525	-350.253	-321.929
Observations (Alliances)	6842 (234)	6811 (234)	6618 (234)	6543 (231)	6470 (226)

Note: *$p<0.05$, two-tailed. Standard errors in parentheses. Point estimates and standard errors were calculated from 10 draws using Rubin's (1987) formula for multiple imputation to account for uncertainty in the *territorial threat* instrumental variable.

Table A.4: Political Environment, Democracy, and Alliance Violations, Additional Robustness Checks.

Model	Democratic Development	Democracy Directly	Exclude Democracy	Alt. Econ Dev. Measure
Outcome Equation				
Change in Leader's Societal Coalition	4.443	7.290	5.173	7.535
	(2.821)	(6.123)	(2.774)	(6.987)
Democracy	0.756	2.497	0.542	0.808
	(1.607)	(2.047)	(1.073)	(2.044)
Change in International Power	3.794	2.740*	3.256*	2.281*
	(5.669)	(1.146)	(1.570)	(0.835)
Change in Political Institutions	1.861	1.609	1.598	2.478
	(2.434)	(1.002)	(1.236)	(2.058)
Change in External Threat	1.073	0.896	0.607	-0.637
	(2.608)	(1.087)	(1.128)	(1.488)
Formation of New Outside Alliance	3.069	2.330*	2.491*	2.997*
	(4.513)	(0.833)	(1.174)	(1.311)
Asymmetry	-1.287	-1.376	-1.866*	-1.617*
	(1.034)	(0.803)	(0.937)	(0.756)
Non-military Cooperation	-2.790*	-2.798*	-3.372*	-2.347
	(1.112)	(1.239)	(1.085)	(2.940)
Ratification	0.851	2.358	1.662	0.063
	(3.319)	(1.682)	(1.944)	(2.517)
Military Cooperation	5.010	3.479*	4.853*	6.792*
	(3.556)	(1.104)	(1.524)	(2.462)
Time	-0.174	-0.197	-0.105	-0.085
	(0.435)	(0.253)	(0.267)	(0.390)
Time Squared	-0.002	0.001	-0.005	-0.001
	(0.011)	(0.011)	(0.011)	(0.015)
Time Cubed	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-3.199	-3.722*	-2.616	-3.082
	(1.938)	(1.869)	(1.371)	(1.979)
Relevance Equation				
Territorial Threat	3.946*	4.292*	4.973*	6.045*
	(1.091)	(0.986)	(1.057)	(2.057)
Number of Borders	0.145		0.044	-0.005
	(0.100)		(0.050)	(0.137)
Proportion of Democratic Borders	-0.329*			0.010
	(0.121)			(0.252)
Democracy		-1.129*		
		(0.561)		
Major Power		0.512	0.215	-0.010
		(0.413)	(0.409)	(0.842)
Rivalry		0.030	0.036	-0.492
		(0.376)	(0.374)	(0.493)
Cold War	0.416	0.449	0.473	1.892
	(0.520)	(0.271)	(0.290)	(1.092)
Economic Development		-0.125	-1.598*	-0.528*
		(0.090)	(0.075)	(0.202)
Oil Producer		0.485	0.642	1.598*
		(0.415)	(0.477)	(0.553)
Constant	-4.773*	-4.376*	-5.051*	-2.441
	(0.617)	(0.512)	(0.468)	(1.804)
Log-likelihood	-316.297	-314.076	-311.835	-151.948
Observations (Alliances)	6582 (223)	6424 (223)	6395 (223)	4896 (137)

Note: $*p < 0.05$, two-tailed. Standard errors in parentheses. Point estimates and standard errors were calculated from 10 draws using Rubin's (1987) formula for multiple imputation to account for uncertainty in the *territorial threat* instrumental variable.

Alliance Violators

Table A.5 provides information about individual cases of alliance violations. In addition to the violator (state A), the partner state violated (state B), and the year, we also report state A's degree of *territorial threat*, whether there was a *change in leader's societal coalition*, and whether it is a *democracy*.

Stata Code to Estimate Split-population Logit

We include *Stata* code of the program we wrote to estimate the split-population logit. For shorthand, *DV* represents the binary outcome variable, *IVO* represents the regressors in the *outcome* equation, and *IVR* represents the regressors in the *relevance* equation. The first line defines that the following code is a program; the second and third line specifies the arguments (coefficients to be specified), while the fourth and fifth creates two temporary variables (the two equations). Finally, the sixth and seventh lines specify the likelihoods to be summed for $Y = 0$ and $Y = 1$, while line eight ends the program. Note that line six, specifying the likelihood when $Y = 0$, treats the outcomes as coming from two distinct processes, i.e. a mixture model, as $Y = 0$ can occur because either (a) the observation is not relevant ("1-'rel"') or (b) the observation is relevant but there is no violation (" 'rel'*(1-'violate')"). $Y = 1$ occurs only if the observation is relevant and a violation occurred.

The variables are specified in line nine, with the equation from the first set of parentheses providing the independent variables (on the right hand side of the equal sign) for the 'beta' argument and the second set of parentheses providing the independent variables for the 'gamma' equation. Any desired options, such as estimating clustered standard errors or selecting a maximization technique, can also be specified after the parentheses. Finally, the likelihood is maximized in line ten.[3]

[3] In the case of a mixture model, it may be helpful to specify initial conditions to help identify the global maximum.

Stata code:

```
program define spl_lf, rclass
    args lnf beta gamma
    tempvar rel violate
        quietly gen double `rel' = 1/(1+exp(-`gamma'))
        quietly gen double `violate' = 1/(1+exp(-`beta'))
        quietly replace `lnf' = `lnf' = ln((1-`rel')+(`rel'*(1-`violate'))) if $ML_y1==0
        quietly replace `lnf' = `lnf' = ln((`rel')*(`violate')) if $ML_y1==1
end
ml model lf spl_lf (DV = IVO) ( = IVR)
ml maximize
```

Table A.5: List of Abrogated Alliances.

State A	State B	Year	Territorial Threat	Δ in Leader's Societal Coalition	Democracy
Germany	Russia	1933	.0134181	1	0
France	Italy	1935	.0194196	1	1
Turkey	Italy	1935	.0438603	0	0
Greece	Italy	1935	.0658665	0	1
Belgium	France	1936	.0074095	0	1
Italy	Spain	1936	.0549720	0	0
Russia	Czechoslovakia	1938	.2272077	0	0
France	Czechoslovakia	1938	.1936811	0	1
Germany	Austria	1938	.1865337	0	0
Russia	Poland	1939	.279217	0	0
France	Germany	1939	.2953013	0	1
Russia	Finland	1939	.279217	0	0
Italy	Albania	1939	.1462826	0	0
Russia	France	1939	.279217	0	0
Russia	Lithuania	1940	.5524079	0	0
Yugoslavia	Romania	1940	.2755803	0	0
Russia	Estonia	1940	.5524079	0	0
Germany	Denmark	1940	.6677183	0	0
Thailand	United Kingdom	1940	.0000000	0	0
Russia	Latvia	1940	.5524079	0	0
Italy	Russia	1941	.5609003	0	0
Germany	Russia	1941	.5609003	0	0
Russia	Iran	1941	.5609003	0	0
Italy	Germany	1943	.259393	0	0
Russia	Japan	1945	.5433966	0	0
Russia	Turkey	1945	.5433966	0	0
Russia	Yugoslavia	1949	.4135729	0	0
Albania	Yugoslavia	1949	.0933776	0	0
Hungary	Yugoslavia	1949	.0779969	0	0
Bulgaria	Yugoslavia	1949	.0941740	0	0
Poland	Yugoslavia	1949	.015158	0	0
Czechoslovakia	Yugoslavia	1949	.0170141	0	0
Romania	Yugoslavia	1949	.0794406	0	0
Pakistan	Turkey	1950	.1850240	0	0
Egypt	United Kingdom	1951	.2974145	0	0
Russia	France	1955	.2527492	0	0
Russia	United Kingdom	1955	.2527492	0	0
Egypt	United Kingdom	1956	.3046337	0	0
Jordan	United Kingdom	1957	.2971661	0	0
Iraq	Jordan	1958	.176082	1	0
Iraq	United Kingdom	1959	.1818257	0	0
Mali	France	1960	.0829011	0	0
Egypt	Yemen Arab Republic	1961	.2875846	0	0
Saudi Arabia	Yemen Arab Republic	1962	.1485351	0	0
Nigeria	United Kingdom	1962	.0385331	0	1
Saudi Arabia	Egypt	1962	.1284466	0	0
France	Morocco	1966	.0029804	0	1
Egypt	Yemen Arab Republic	1967	.3355091	0	0
Libya	United Kingdom	1970	.0069660	0	0
Madagascar	France	1973	.0000000	1	0
Tunisia	Libya	1974	.0015907	0	0
United Kingdom	South Africa	1975	.0149923	0	1
Egypt	Russia	1976	.1617086	0	0
Iraq	Egypt	1977	.1838835	0	0
Somalia	Russia	1977	.1971040	0	0
Uganda	Sudan	1979	.1853800	1	0
Russia	Pakistan	1979	.4478039	0	0
Iran	United States	1979	.3403858	1	0
United States	Taiwan	1980	.0682471	0	1
Syria	Libya	1980	.1067835	0	0
Niger	Libya	1981	.0030048	0	0
Chad	Libya	1982	.1290733	1	0
Algeria	Libya	1984	.0460337	0	0
Sudan	Egypt	1985	.1514078	0	0
Morocco	Libya	1986	.0980619	0	0
Malta	Russia	1987	.0000000	1	1
Senegal	Gambia	1989	.0008766	0	0
Jordan	Saudi Arabia	1990	.1075147	0	0
Russia	Iraq	1990	.2336497	0	0
Poland	Russia	1991	.1297963	1	1

References

Bennett, D. Scott and Allan C. Stam. 2004. *The Behavioral Origins of War.* Ann Arbor, MI: The University of Michigan Press.

Clarke, Kevin A. 2001. "Testing Nonnested Models of International Relations: Reevaluating Realism." *American Journal of Political Science* 45(3):724–744.

Clarke, Kevin A. 2003. "Nonparametric Model Discrimination in International Relations." *Journal of Conflict Resolution* 47(1):72–93.

Clarke, Kevin A. 2007. "A Simple Distribution-free Test for Nonnested Model Selection." *Political Analysis* 15(3):347–363.

Gibler, Douglas M. and Jaroslav Tir. 2014. "Territorial Peace and Democratic Clustering." *Journal of Politics* 76(1):27–40.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses." *Econometrica* 57(2):307–333.