

Modeling “Effectiveness” in International Relations

Journal of Conflict Resolution
00(0) 1-32
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022002713499715
jcr.sagepub.com



Jonathan Renshon¹ and Arthur Spirling²

Abstract

Are democracies better at winning wars and militarized disputes? Is there an advantage associated with initiating a war or dispute? Noting that pairwise contest data are the norm in applied research, we motivate a straightforward Bradley–Terry statistical model for these problems from first principles, which will allow for a closer integration of theoretical and statistical practice for scholars of international relations. The essence of this approach is that we learn about the latent abilities of states from observing conflict outcomes between them. We demonstrate the novelty and appeal of this setup with reference to previous attempts to capture estimands of interest and show that for many questions of concern—especially regarding “democratic effectiveness” and “initiation effects”—our approach may be preferred on theoretical and statistical grounds. The evidence we find only partially supports the ideas of “democratic triumphalists”: democracy aids effectiveness, but only in certain contexts (while in others it actually impairs fighting ability). We also provide estimates of possible “initiation effects,” and show that moving first seems to carry little advantage in interstate wars, but a substantial one in lower-level disputes.

¹ Department of Political Science, University of Wisconsin–Madison, Madison, WI, USA

² Department of Government, Harvard University, Cambridge, MA, USA

Corresponding Author:

Jonathan Renshon, Department of Political Science, University of Wisconsin–Madison, 110 North Hall, Madison, WI 53706, USA.

Email: renshon@wisc.edu

Arthur Spirling, Department of Government, Harvard University, 1730 Cambridge Street, Cambridge, MA 02138, USA.

Email: aspirling@gov.harvard.edu

Keywords

conflict, contest data, Bradley-Terry, democratic effectiveness, power, initiation

Scholars of international relations (IR) have devoted considerable effort to translating theoretical understandings of behavior into statistical models of outcomes (Signorino 1999; Smith 1999; Lewis and Schultz 2003; Whang 2010; Carter 2010). Though the particular procedures used can be quite complicated, the motivation is very simple: the empirical approach must return estimates that correspond in some meaningful way to the parameters of our theoretical understanding of interstate conflict. If we fail to do this, we run the danger of being unable to interpret our results in terms of our original conjectures; put otherwise, we cannot test our theories (see, especially, Signorino 2003; cf. Carrubba, Yuen, and Zorn 2007). The flip side of this idea is that we need to understand the theoretical commitments we make—albeit implicitly—when we make choices among statistical models.

The current article takes this logic to the problem of estimating the “effectiveness” of states in IR, in particular their ability to prevail in conflict. Although much interest has focussed on “democratic effectiveness” (e.g., Reiter and Stam 1998b), the central question of “what makes states more likely to win conflicts?” is very general and long standing (e.g., Wright 1965). Here, we begin from the “ground up,” by giving plausible foundations for a statistical model of state effectiveness in interstate conflict. In contrast to all previous approaches of which we are aware, we concentrate on the estimation of latent state abilities as a function of covariates, and we discuss a logit linear form for that relationship that connects the two in a way that is new conceptually, but easily understood by political scientists who work with binary dependent variable observational data (see, e.g., Berry, DeMeritt, and Esarey [2010] for an overview). In so doing, we provide statistical foundations that might serve as a fruitful base for future theoretical and empirical efforts in this area.

The statistical model that subsequently emerges—a variant of the Bradley and Terry (1952) approach to pairwise contest data (see also Turner and Firth 2010)—is thus well motivated. We describe the model’s properties as they pertain to estimation, and we show how this technique may be taken to the type of data political scientists and IR scholars commonly use. Therein, we demonstrate the need for both states’ covariates to enter the linear predictor such that the effect of a given covariate can be understood as the difference in *relative* capabilities, and not one state’s (absolute) level of ability. Our model also allows for a satisfactory way to deal with a common problem endemic to quantitative IR research: that it is difficult to model (with the standard setup) outcomes of interactions between states if they are determined by heterogenous and unobserved factors specific to those countries involved (such as culture). We demonstrate how dependence across contests (in the sense of Green, Kim, and Yoon 2001; King 2001) may be taken into account with random effects via a mixed model arrangement for such pairwise data. We contrast the

approach here to that taken by other scholars, and show that, at the most basic level, our approach allows a more general question to be asked of data: in particular, we ask, “What factors make states more or less likely to win conflicts?” rather than “What factors make democracies more likely to win conflicts?” or “What factors make initiators more likely to win conflicts?” These latter enquiries emerge as special cases of our approach in the sense that we are not constrained to estimate the effectiveness of states with respect to a particular covariate: the model here is more theoretically agnostic. We make no attempts to pour scorn on previous efforts in the field, but rather seek to demonstrate what we believe is a sensible model that will hopefully give scholars pause when they reach for more conventional logit (and probit) arrangements.

We apply the model we discuss to “standard” (*Polity/COW [Correlates of War] / MID*) IR data, with particular focus on the impact of relative levels of “democratic-ness” and the open puzzle of whether there is an advantage associated with initiating conflicts in both wars and nonwars, a topic of some concern to scholars (see, e.g., Jervis 1978; Fearon 1997; Van Evera 1998). We do find (partial) support for a democratic advantage in wars, but to the extent that increasing “democratic-ness” increases the likelihood of victory, it only does so at certain (absolute) levels of democracy and this effect is partly dependent on an opponent’s characteristics. That is, in some cases, increasing democraticness seems to impair fighting ability. Similarly, we find that initiating those wars does not seem to be associated with an increased probability of victory. By contrast, for nonwar conflicts, there is something of an initiation “effect,” with first movers doing better, on average. In the following, we speculate on the likely mechanism/mechanisms behind such an observation.

Why We Care about Effectiveness in International Relations

Beginning at least with Wright (1965), Cannizzo (1980), and then Wayman, Singer, and Goertz (1983), there is a voluminous IR literature on what factors make states “powerful.”¹ Much attention has been paid specifically to “democratic effectiveness”—that is, the degree to which democracies are able to achieve more successful conflict outcomes than nondemocracies.

The modern incarnation of this debate grew out of the democratic peace research paradigm and was influenced by the work of Lake (1992, 31), who found that of the twenty-six wars fought between democracies and nondemocracies from 1816 to 1990, democracies won twenty-one (81 percent) of them. Lake’s explanation (that for a variety of reasons democracies are able to bring to bear overwhelming coalitions against nondemocracies) would later be classified into the *material capabilities* school, in opposition to works that argued that democracies were more likely to choose fights that they could win because they faced harsher political consequences for bluffing or losing (this is known as the *selection effects* argument, see Fearon 1994; Schultz 1999). A bevy of related work soon followed up on these initial findings.

Reiter and Stam (1998b) found that democratic initiators and targets have both been more likely to win wars than other types of states. They also argued for a

nonlinear relationship between regime type and effectiveness, while democratic initiators were the most likely to win, the least democratic countries (autocracies) were more likely to win wars than so-called *mixed regimes*. Gelpi and Griesdorf (2001, 641) examined international crises and found that democratic challengers (i.e., initiators) were more likely to prevail than other types of challengers, but that democratic defenders were no more likely to prevail than autocratic defenders, even when crucial values were at stake. Other work examined the temporal nature of the democratic advantage (Bennett and Stam 1998) or investigated how democratic effectiveness might explain micro-level outcomes at the level of the individual battles (Reiter and Stam 1998a; Biddle and Long 2004).

The main thrust of this research program may be summed up by Reiter and Stam (1998b), who write that there is “something about democratic regimes that makes it easier for them to generate military power and achieve victory in the arena of war.” Whether that “something” is primarily the ability to generate more wealth and mobilize more resources (a view advanced by, among others, Lake 1992; Bueno de Mesquita et al. 1999), the use of different strategies (Reiter and Meek 1999) or the tendency to select wars that are winnable (the “selection effects” argument, exemplified by Reiter and Stam 2002) is not yet settled. In recent work, the very existence of the so-called *democratic advantage* has been called into question. To wit, there has been lively debate over case selection (see Desch 2002, 2003, 2008); that is, “What conflicts should count?” (see also, Lake 2003; Reiter and Stam 2003). Scholars have also suggested differentiating states as “initiators” versus “joiners” of wars (see Downes 2009), with commensurately new data definitions.

The point here is that estimating the “effectiveness” of states matters. Rather than ponderously critiquing the various approaches of previous scholars, we start from “first principles” and suggest a statistical model that gets to the core of the issue and that provides a platform for future formal models as well as empirical research. The next section begins that process.

Modeling Effectiveness

What We Want

First, it is helpful to consider in some detail the quantity we *want* to estimate. At its heart, the “democratic effectiveness” debate revolves around the following query: are democracies more likely to win disputes or confrontations than nondemocracies? Put more generally, scholars are interested in what factors make countries more likely to win wars. Though this is easily stated, the implied statistical model is perhaps not so obvious.

To fix ideas, suppose that we observe a series of $1, \dots, t, \dots, T$ conflicts between a set of countries. We assume that this data set is “connected” insofar as there is no subset of states that never meets the others in disputes, an empirical requirement that is met for the great majority of observations in real investigations.² Any given dispute involves two of a total of N states, and we index the countries concerned as i and

j . Since i either wins or loses the dispute against j , the response (the “dependent variable”) to be predicted for observation t is binary; that is, $y_t \in \{0, 1\}$, where $y_t = 0$ implies state j is victorious over i , whereas $y_t = 1$ implies state i beats j . The goal is to model the probability, $\Pr(y_t = 1)$, that is, the probability that i wins. When we have two states in this scenario this probability is, in fact, $\Pr(i \text{ beats } j)$.

Keeping with just i and j for now, suppose that both countries have latent “abilities”³— α_i and α_j , respectively—and that the probability of any particular conflict outcome, for example, a win for state i , is some function of these unobserved quantities. More specifically, we suppose that

$$\frac{\Pr(i \text{ beats } j)}{\Pr(j \text{ beats } i)} = \frac{\Pr(i \text{ beats } j)}{1 - \Pr(i \text{ beats } j)} = \frac{\alpha_i}{\alpha_j}. \quad (1)$$

That is, the odds that i beats j is just the ratio of their abilities. The fact that $1 - \Pr(i \text{ beats } j)$ is equivalent to $\Pr(j \text{ beats } i)$ in this context simply means we are (for now, and in general accordance with the literature) ruling out “draws” in international conflicts.

How to Get It

To take this approach to contests to real data, we need to specify a way to estimate α_i and α_j , presumably as a function of covariates. One approach is to take the logarithm of “both” sides of equation (1) such that we have

$$\log \left[\frac{\Pr(i \text{ beats } j)}{1 - \Pr(i \text{ beats } j)} \right] = \log \left[\frac{\alpha_i}{\alpha_j} \right] \quad (2)$$

$$\text{logit } p_{ij} = \lambda_i - \lambda_j,$$

where $\lambda_i = \log(\alpha_i)$ for all i , and we substitute p_{ij} for $\Pr(i \text{ beats } j)$ and thus $\text{logit } p_{ij}$ for $\log \left[\frac{\Pr(i \text{ beats } j)}{1 - \Pr(i \text{ beats } j)} \right]$. We can then make λ_i a function of covariates of interest (i.e., “ X ”s) in a way that accords with the usual setup of a generalized linear model (GLM), in the sense that we have a linear predictor (“ βX ”) on the “right-hand side.” Thus, we have $\lambda_i = \beta x_i$ and $\lambda_j = \delta x_j$ where x_i and x_j are one variable that takes (possibly) different values for the countries, with (possibly) different coefficients β and δ . It is no problem to have a long vector of covariates in practice (say, “democracy,” “capability,” “nuclear power,” “gdp per capita,” etc.), but we assume a single x for exposition in this subsection. Moreover, though we specify more elaborate models later, for now we suppose that the relevant covariate profile for i and j do not change over the course of the data. That is, we assume that x_i is the same for any observation t or, more plainly, that the relevant attributes that contribute to i ’s or j ’s power are constant over time.

What This Implies

From this simple but logical arrangement, four remarks immediately follow that, as we will see, denote this model as different from others in IR. In part, these remarks are summaries of properties that are “built-in” to the approach, and we present them here to clarify matters to readers.

Remark 1: An observation consists of one conflict, which is used exactly once. Thus,

$$\Pr(\widehat{i \text{ beats } j}) + \Pr(\widehat{j \text{ beats } i}) = 1.$$

A particular conflict between i and j contributes one observation to the data: that is, we do not count the conflict as a case of, for example, i attacking j and subsequently count the case again as j defending itself from i . Indeed, any notion of attacking or defending roles may be incorporated via the covariate profile (as discussed in Remark 2). Note that we are agnostic as to what, precisely, a particular “conflict” consists of it might be a battle, or war, or event defined some other way. The point is that it is a “contest” between states.

In any case, a consequence of this arrangement is that the sum of the predicted probabilities of i beating j and j beating i is equal to one. This follows directly from the equivalence $\Pr(i \text{ beats } j) = 1 - \Pr(j \text{ beats } i)$, but it is distinctive insofar as a situation cannot obtain in which we predict a “victory” for both i and j ; that is, those predicted probabilities cannot *both* be $\geq \frac{1}{2}$. In a model which uses each conflict as two observations, such “non-sensible” predicted probabilities can occur.

Remark 2: logit p_{ij} is modeled as a function of features specific to the states and their roles, not features of the conflict they are involved in.

Note that p_{ij} is defined in terms of the abilities of the states: $\lambda_i = \beta x_i$. Here x_i is properly indexed with a subscript since it pertains to characteristics of i and i alone. In a given contest, this might include a characteristic like “initiator,” but it cannot include a characteristic of the conflict itself such as “fought at sea.” This is because conflict characteristics cannot per se make it more less likely that state i wins over state j . To see the distinction, suppose, for example, that state i is “better” at fighting at sea and thus wins more sea conflicts. Then this “better-ness” must enter x_i in some way (e.g., “number of fighting ships”⁴), else it cannot affect the relative estimated abilities of the states: absent any information entering x_i , fighting at sea affects both parties equally.

To clarify matters here, notice that the model makes the assumption that we wish to estimate the effectiveness of states in some “overall” sense as a latent trait of a state, wherein the conflicts from which we generate our estimates are fundamentally comparable.⁵ If the analyst deems that the contests are not comparable—for example, if he or she wishes for some a priori theoretical reason to estimate effectiveness in sea conflicts as a quantity separate to effectiveness in land battles, and that some x

should have a differential effect therein—then this implies that the model ought to be fit to the different data sets (“sea conflicts,” “land conflicts”).

Remark 3: The expression for $\text{logit } p_{ij}$ contains no (nonzero) intercept term.

To see this, suppose that the intercept term is in fact nonzero. We have stated in equation (2) that $\text{logit } p_{ij} = \lambda_i - \lambda_j$. Adding an intercept, γ , we have that

$$\text{logit } p_{ij} = \gamma + \lambda_i - \lambda_j. \quad (3)$$

Of course, by symmetry (just flipping the i and j index over), we must also have

$$\text{logit } p_{ji} = \gamma + \lambda_j - \lambda_i. \quad (4)$$

Since i either wins or loses against j —there are no “ties” here—we must have $p_{ij} = 1 - p_{ji}$. This implies⁶ that

$$\text{logit } p_{ij} = -\text{logit } p_{ji},$$

and thus

$$\gamma + \lambda_i - \lambda_j = -\gamma - \lambda_j + \lambda_i. \quad (5)$$

But then, $\gamma = -\gamma$, which is only true for $\gamma = 0$ which contradicts the premise that $\gamma \neq 0$.

Substantively, the foregoing remark says that there is no “bump” (up or down) in effectiveness, simply because a country is listed as being i (first) or j (second) in a conflict. These indices are arbitrary. Note that it may well be that initiation-status affects the probability of winning, but this status is already included in λ via the linear predictor for the earlier model.

Remark 4: Only relative covariate levels matter for determining $\text{logit } p_{ij}$

From Remark 1 we have that

$$\gamma + \beta x_i + \delta x_j = -\gamma - (\beta x_j - \delta x_i), \quad (6)$$

and that $\gamma = 0$. Hence

$$\beta(x_i + x_j) = \delta(x_i + x_j),$$

implying that $\delta = \beta$. Using these identities, the original regression equation (2) is rewritten as

$$\text{logit } p_{ij} = \beta(x_i - x_j). \quad (7)$$

And thus only the relative values of x_i and x_j (not their absolute values) determine the log odds of a win for i (or j).

Table 1. Stylized Monadic Data Example.

War	Country	Year	Y_{VICT}	X_{INIT}	X_{DEM}	...
Russo-Hungarian	RUS	1956	1	1	0	...
Russo-Hungarian	HUN	1956	0	0	0	...
Yom Kippur	ISR	1973	1	0	1	...
Yom Kippur	EGY	1973	0	1	0	...
Iran-Iraq	IRQ	1980	1	1	0	...
Iran-Iraq	IRN	1980	0	0	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

This assumption of the model reiterates the point that indices (i or j) are arbitrary: it does not matter “who,” in state terms, has more or less of a particular covariate. All that enters the model is how much one state has relative to the state they face in a contest.

How It Differs from Previous Approaches

Our four remarks are not, of themselves, remarkable: they follow directly from the way we conceptualized “effectiveness” as akin to an ability, and then the functional form assumptions we chose to make. Nonetheless, they do mark our approach as distinctive with respect to the way that other scholars have thought about the modeling problem. We make no particular claims that our model is “better,” but we do think it answers the right question—in the sense that we posed it—and in a logically consistent way.

One obvious difference is that our model uses each conflict (or contest) only once. One previous approach to the question of democratic effectiveness has been to estimate models based on monadic data (e.g., Reiter and Stam 1998b; Downes 2009). In this setup, each observation (or row) contains information on one state in a conflict. Hence, though state i and state j might fight a particular war against each other, they and their covariate profiles will take two separate entries. In Table 1, we give a segment from a stylized version of such data. In the table, the six detailed observations concern three conflicts: between Russia (RUS) and Hungary (HUN) in 1956, between Israel (ISR) and Egypt (EGY) in 1973, and then between Iraq (IRQ) and Iran (IRN) in 1980. Each row then refers to one party to a given dispute. Thus, Israel was the victor in 1973 ($Y_{VICT} = 1$), but was not the initiator ($X_{INIT} = 0$), though it was democratic ($X_{DEM} = 1$). By contrast, for the same conflict, Egypt was not a democracy ($X_{DEM} = 0$), initiated the conflict ($X_{INIT} = 1$), but ultimately lost ($Y_{VICT} = 0$).

From the perspective of simply estimating a regression equation, monadic data present few issues. The models will generally be identified in the econometric sense, and the properties of the estimators—unbiasedness, efficiency, and so on—will apply. Given available technology at the time, these models were perfectly reasonable

choices. However, there are serious logical and inferential problems with this setup. More concretely, there are at least three concerns. First, there is the potential for double counting in the sense that half the observations are not independent of the other half (in fact, they are entirely dependent). For example, given that $Y_{\text{INIT}} = 1$ for Egypt in 1973, Y_{INIT} for Israel the same year *must* take the value 0. Our approach obviously differs in that each contest is used exactly once in the data.

Several scholars have noted that IR are characterized by heavy dependency between observations. One extreme version of this problem results in the “double counting” we noted earlier; a slightly different problem emerges when the outcomes of interactions between states are determined by heterogenous and unobserved factors specific to those countries involved (see, e.g., Green, Kim, and Yoon 2001; King 2001). Thus, the second issue is that the standard setup shown previously does not allow for variability between states that have the same covariate profiles, a serious problem if there is a chance that unobserved factors (e.g., resolve or culture) contribute to a state’s power and effectiveness. The statistical model we discuss below allows for random effects (in essence, player-specific residuals) to alleviate such concerns, and is to our knowledge, the first contest-data model to do so.

Third, and most importantly in the standard setup, only one state’s characteristics enter the regression equation for a given observation. To the extent that we believe the model we derived earlier, this is clearly problematic. We acknowledge that there are examples of scholars taking “hybrid approaches” in which some characteristics of both sides are included, for instance, Sullivan (2007) includes a number of variables related to the characteristics of the target state, while Reiter and Stam (2002) include covariates for the Polity score of the target state (in addition to the initiating state). We see these approaches as further evidence that most researchers would grant the importance of including opponent characteristics if possible. Crucially, if opponent characteristics are not included in the *way* we suggest—that is, for every contest, it is the relative magnitude of the covariates that matters—then the “first principles” justification for the contest model *must be different* to the account we gave. And, to the extent that our foundations are reasonable, as they are to the statistics literature, our model has a sensible interpretation in terms of abilities that is missing in other approaches. Just as importantly, we believe that our framework and assumptions (particularly our focus on relative capability levels) will make intuitive sense for IR scholars.

To see an example of this logic in action, consider a very simple model of international conflict where only one independent variable, x , is used to predict whether or not a state wins a conflict. For clarity, suppose x is such that a one-unit change in x has some intuitive meaning; for example, x might be gross domestic product (GDP) per capita (a one-unit increase is a one-dollar increase) or x might be major power status (a one-unit increase means a state was not a major power and now has that status). Suppose also, that a researcher decides to use a “hybrid” approach such that both states’ values on this covariate— x_i and x_j —are included on the “right-hand side” of the (logistic) regression equation. To make our example especially stark,

we will further suppose that $x_i = x_j$ (e.g., they have identical GDP figures, or are both nonmajor powers, or whatever). With respect to the relevant logit, this model would be set up as follows:

$$\text{logit Pr}(Y = 1) = \beta_0 + \beta_1 x_i + \beta_2 x_j, \quad (8)$$

where the “left-hand side,” as usual, is the log of the probability of state i winning divided by the probability of state i losing. Of course, when that log odds ratio increases, the probability of i winning is increasing relative to the probability of i losing. Suppose that we fit the model, and obtain estimates of β_0 , β_1 , and β_2 ; we then attempt to predict the outcome of an out-of-sample contest with its own value for the i th state, written x_m and for the j th state, written x_n . To make life simple, suppose that this out-of-sample x_m is $x_i + 1$ and that x_n is $x_j + 1$. What is our prediction for the outcome? By definition, a one-unit increase to x_i (giving us x_m) increases the log odds by β_1 . And a one-unit increase to x_j (to give us x_n) increases the log odds by β_2 (bear in mind that β_1 and/or β_2 could be negative). What is the overall effect on the log odds? It is not zero, unless β_1 and β_2 happen to sum to zero; that is, the log odds must increase or decrease unless $\beta_1 = -\beta_2$. But this is strange: previously, the states were exactly evenly matched. Now, having increased the x by one unit each, they are once again exactly evenly matched. And yet the probability of state i winning relative to the probability of state i losing has *changed*. This seems theoretically wrongheaded and logically undesirable. What about in the Bradley–Terry case? There, increasing x_i and x_j by one each makes no difference to the log odds: the changes “cancel out” since the expression $(x_i - x_j)$ has not changed in value. So, to the extent that we want to include covariates from both sides, the standard logit set up is not the way to proceed; instead, we want to deal with differences rather than sums.

The logic of our approach has consequences for the question asked of the data. Recall that we ask, “What factors make *states* more or less likely to win conflicts?” Other approaches typically have to be a priori more specific. In terms of the “standard” logit specification described above, those approaches have to pick (possibly arbitrarily) a state i for each observation to which the log odds applies. Thus, they ask a question more akin to “What factors make *democracies* more likely to win conflicts?” or “what factors make *initiators* more likely to win conflicts?” This is not a semantic or trivial difference. In particular, other work assumes that the interest is in estimating a version of equation (3); that is, $\text{logit } p_{ij} = \gamma + \beta x_i + \delta x_j$, where i and j have specific characteristics, such as being democratic or being initiators. As a result, the relevant x with respect to which the equation is estimated (e.g., initiator status in much of the “democratic effectiveness research”) cannot usually appear with its coefficient as an “effect.” Our model makes no such restrictions: i and j are simply states with whatever characteristics they happen to exhibit. An obvious consequence of this is that any given x_i or x_j can pertain to initiator status, without having to rearrange the data or our interpretation of what the estimated $\hat{\beta}$ s represent in terms of the underlying

quantity of interest. In a very real sense then, other approaches are special cases—both theoretically and statistically—of our model. This also comes with significant benefits, not the least of which is that it allows us to generate (what we believe to be the first) estimates of the effect of initiating conflicts.

Statistical Choices

While the basic Bradley–Terry model seems a logical choice for the task at hand, in practice the restrictive nature of the covariate vectors make it unacceptable for most research in IR. The extended version we consider below does not suffer from these shortcomings. More technically, we now discuss two types of quantities that we (as well as most IR scholars, we presume) would like to estimate: fixed effects and random effects. The former refer to the impact of independent variables that (may) change between contests (such as Polity level or military capabilities). The latter are essentially player-specific residuals (i.e., abilities of the state unaccounted for by the covariates we include in our model).

Fixed Effects

By “fixed effects,” we mean the modeling of parameters as a function of independent variables (X s), where the parameters are considered nonrandom. In the current context, the parameters of interest are the state abilities, denoted λ . Of course, equation (7) does just this, but in that form, the model forces us to assume that covariates are constant over sequential contests. This seems unrealistic in practice: Britain’s level of democracy in 1900 is very different to that in 2000. To make equation (7) more general, we can allow the x_{ri} to vary across contests. Indexing a contest as k we are interested in

$$\lambda_i - \lambda_j = \sum_{r=1}^p \beta_r (x_{ikr} - x_{jkr}). \quad (9)$$

where x_{ikr} is indeed varying across bouts, we have a “contest-specific” effect⁷ represented by our coefficient on x_r , $\hat{\beta}_r$. The special case—such as the “region” of a country that is unchanging over time—is then constant for all k .

Random Effects

In the typical case of logit (or Bradley–Terry [BT] model), as can be readily seen in equation (3), there can be no variability between states that have identical covariate profiles: a state’s ability is “all” fixed effect: $\lambda_i = \beta x_i$. This is problematic if we believe there may be unobserved heterogeneity in factors that contribute to state abilities. One solution is to model “random effects” for each player. Analogous to panel data, the idea here is to include an error term composed of two parts. The first is the extent to which that player’s ability differs from the overall one estimated for that state’s covariate profile

and the second is simply a random deviation specific to that contest. If the variance of the random effect component is statistically significant, it must be that the random deviations from the explanatory variable modeled abilities are in fact different. Thus, in this context, after controlling for covariates, the players *still* differ in ability.

Mixed Bradley–Terry Models

We would like to make the model in equation (9) more general, by allowing for correlation between the same players in different contests. But we still wish to make statements about the influence of various independent variables which may vary over contests. Fitting the following general Bradley–Terry model facilitates these aims:

$$\text{logit } p_{ijk} = \underbrace{\sum_{r=1}^p \beta_r (x_{ikr} - x_{jkr})}_{\text{fixed effects}} + \underbrace{(U_i - U_j)}_{\text{random effects}}. \quad (10)$$

In this way, we have combined the player-specific (possibly contest specific) covariates of equation (9) with random effects represented as the difference between the error terms. It is this latter term that will facilitate modeling correlation in unobserved influences over contests. In matrix notation, we have

$$\text{logit } p_{ij} = \mathbf{X}\beta + \mathbf{Z}e, \quad (11)$$

where $e \sim N(0, D)$ and D is a diagonal matrix whose diagonal elements are equal to σ^2 . The assumption here is that the sampled states are representative of a population of nations and thus the variance term reflects the underlying heterogeneity of the dispute parties.

When we have the basic Bradley–Terry form of equation (7) or, in fact, equation (9), estimation can take place within a GLM framework (see Firth [2005] for an implementation). But GLMs will not do once we add the error term: we must now move to generalized linear mixed models (GLMMs; see McCulloch and Searle [2001] for detailed discussion). Parameter estimation is usually via some variant of maximum likelihood, with computational complexities arising from the inclusion of the random effects component. The approach we take here relies on penalized quasi-likelihood (PQL; in the sense of Breslow and Clayton 1993) which approximates the full-maximum likelihood solution. We need an approximation because the original likelihood equation involves a complex integration for which there is no closed form. The specific suggestion of Breslow and Clayton (1993) involves a quadratic expansion of the Laplace approximation of the likelihood, but the essence of the technique is that the random effects, the Z matrix earlier, are treated as if they are part of a linear mixed model. Fitting algorithms obtain estimates of the fixed effects, the β s, and then use them to get *best linear unbiased predictions* (BLUPs) of the random effects, \hat{e} (see, e.g., Robinson [1991], for discussion).

Those BLUPs are then used to improve the estimated $\hat{\beta}$ s and so on in an iterative scheme. The variance of those random effects in our analysis will be of interest substantively, since it will give a sense as to whether, in fact, states' performances in conflicts are correlated over contests.

There are two immediate consequences of our approach here. First, PQL is relatively straightforward in computational terms, but because it does not make use of a log likelihood, measures of fit that rely on that likelihood (like Akaike Information Criterion [AIC] and Bayesian Information Criterion [BIC]) will not be available. This will make it difficult to compare the fit of a mixed Bradley–Terry model in some “overall” sense (assuming that AIC is a sensible choice for the comparison under consideration). Second, our model predictions—which in our case will be predicted probabilities of winning or losing—are derived from “plugging in” the estimated fixed and random effects along with the relevant X and Z matrices. We note, in passing, that despite the fact that the assumption of normality for the random effects is a requirement for PQL, this presents no theoretical bar to utilizing the logit form of the equation. Turner and Firth (2010) devote considerable effort to designing software for the fitting of such models, and we use their R package for what follows.

An Application

To show the model in action, we now embark on an application using “standard” data and assumptions simply to highlight how (fitted) model interpretation differs with our approach. In what follows, we concentrate on the estimation and interpretation of “democratic” effectiveness, since that has dominated the literature in this area. Readers may wonder why we do not consider reestimating an example from the literature; the short answer is that the models previously used are sufficiently different from ours—in particular, they use location variables common to both players, which are simply not sensible in our setup (for reasons given in Remark 2)—such that a straight replication is not meaningful or helpful.

What We Used: Data

We used the EUGene software program (Bennett and Stam 2000, 2007) to generate a directed dispute data set of Militarized Interstate Disputes (MID; Jones, Bremer, and Singer 1996), both wars and nonwars, from 1816 to 2001.⁸ The first COW/MID data sets were not dyadic in nature, and thus any analysis of dyadic data prior to that (which is a great deal of the published work in IR and security studies) must rely on a conversion scheme in order to take pre-1993 data and convert it to dyad (or directed-dyad) form. There are two options for doing so. In the default option, the EUGene software performs the conversion (our discussion focuses on this data set). The second option uses, for pre-1993 data, the Maoz Dyadic MID data (Maoz 2005), which has a different (and purportedly stricter)

set of criteria for generating pre-1993 dyads (both options use the same data for post-1993 dyads). Though we focus the bulk of our discussion on the analysis of the EUGene dyads, we note that the different conversion schemes do not change much of the substantive interpretation, and the results are nearly identical for our findings on the relationship between democracy and effectiveness.

In the data, there is one observation per dispute,⁹ side A is always the initiator, and for each observation there is a categorical outcome variable coded one through nine to denote different possible settlements.¹⁰ For regime types/Polity scores, we use the Polity2 variable from the 2010 version of the Polity IV data set (Jagers and Gurr 1995).¹¹ Polity2 scores were introduced in 2002 in order to standardize an approach to dealing with the “special codes” (–66, –77, and –88) that had previously been dealt with in an ad hoc manner.¹² Military capabilities data are taken from Singer, Bremer, and Stuckey (1972). Diplomatic data were taken from the diplomatic exchange data set (Singer and Small 1966; Bayer 2006).

For our current application, and keeping in line with common practice, we do not use “draws” as outcomes. We note in passing that several authors have discussed the nature and frequency of draws, and what leaving them out might mean for estimation (e.g., Bennett and Stam 1998; Reiter and Stam 2002; Fortna 2009; Downes 2009). We refer readers there for further information.¹³

This leaves us with 929 MID dyads of all levels and 340 war dyads. COW coding procedures distinguish between “originators” (those countries involved in the war on day 1) and “joiners” (those states that joined the conflict after the first day). Again, to make matters simple we treat both groups the same way: our theoretical framework means that the distinction between joiners and originators is irrelevant. Our goal is to model the relative *power* or effectiveness of different states, and not to model their decisions concerning whether to go to war. Whether the decision to join an ongoing war is different from a decision to initiate one (we believe it probably is) does not matter so much as the fact that in each case there is a contest between two players, and that contest tells us something about each player’s relative abilities. We refer readers to extensive debates on this subject elsewhere rather than offer a redux in this article (see, e.g., Downes 2009).¹⁴

By including MIDs, we are not implying that interstate wars are in any way similar to the types of disputes that are picked up by the MIDs database. MIDs themselves represent a broad spectrum of conflict, from fishing and trade disputes to crises, to violent conflict that falls just below the threshold of war. Thus, even among the MIDs, there is likely to be great variation in the dynamics at hand. Which states, and which types of states, win these disputes may not tell us much about the ability to prevail in interstate wars (we would not make that inferential leap), but—given our theoretical framework—each dispute should still be thought of as a contest between two actors which can tell us something about the relative abilities of the two players. With that in mind, we fit models to both wars and more limited disputes next.

What We Found: Fit

Before turning to the substantive findings, we first note the relative advantage of the Bradley–Terry method in terms of model fit over “monadic” designs.¹⁵ A seemingly natural method for such a comparison is via some metric that considers the value of the (log) likelihood, and penalizes by parameter number. Unfortunately, however the relevant regressions are constructed, monadic and dyadic approaches use observations (the Y) that are fundamentally different in nature and number. In any case, with random effects in place, the Bradley–Terry GLMM has no available (log) likelihood and thus model fit procedures relying on the deviance cannot be used. For both these reasons then, we cannot rely on measures such as the AIC or BIC, at least for the models that uses random effects. Subject the caveats noted, we *can* report the AIC for the model without random effects relative to the usual logit, and the Bradley–Terry approach clearly does better: its AIC is less than half the value of the logit for the wars and nonwars specification.

For a more complete picture of model fit, we consider the proportion of cases correctly predicted by the relevant fitted models. The dyadic case is as described in the previous sections in the sense that, for each country party to a conflict, we model $\text{Pr}(\text{victory}) = X\beta$, where the X s entering the likelihood for a given observation refer only to one country. Our specific interest is the fraction of observations for which the models predict that $\text{Pr}(i \text{ wins}) > 0.5$ where i is the true winner in the data. Note that the independent variables are the same in every case and correspond with those in the coefficient tables we introduce shortly. We fit a Bradley–Terry model without random effects (77 percent correctly predicted for the war model) as given in equation (9); a Bradley–Terry model with random effects (88 percent) as given in equation (11); a monadic logit model (73 percent); and finally, a monadic logit model where the predictions are subject to a straightforward “sensibleness” restriction (28 percent). In the latter case, we only count as a “correct” forecast those cases where the relevant i is predicted to have won, and the relevant j is predicted to have lost. Without this restriction, the logit is able to predict a win (or loss) for *both* states involved in a particular dispute, which is impossible given the actual data. Perhaps unsurprisingly, the Bradley–Terry approach with random effects (our choice in the following) outperforms the other models. For the nonwars, our results are similar with the percentage correctly predicted, respectively, being 76 percent, 83 percent, 75 percent, and 33 percent. The full table is given in Table 2.

To clarify our procedure here, and to recap our methodology section above, recall that predictions for the mixed model are based on “plugging in” the BLUPs of the random effects and the $\hat{\beta}$ fixed effects along with their relevant covariate matrices. This is not the case for our figures that follow, however: here, we simply use the fixed effects, taking the random effects as zero. This means that we are plotting predictions for an “average” initiating player with the relevant Polity score versus an “average” noninitiating player (and its Polity score). So, while the predicted

Table 2. Proportion Correctly Predicted, BT Model versus Monadic Logit for Different War and Nonwar Data Sets.

	BT model, no random effect	BT, random effect	Logit	Logit (logic restriction)
Wars	0.774 AIC= 275.717	0.878	0.726 AIC = 649.090	0.278
Nonwar (MID < 5)	0.760 AIC = 551.208	0.829	0.749 AIC = 1191.767	0.329

Note: AIC = Akaike Information Criterion; MID = Militarized Interstate Disputes.

Table 3. Results: Bradley–Terry Models.

	EUGene dyads		Maoz dyads	
	Wars (Model 1)	Nonwars (Model 2)	Wars (Model 3)	Nonwars (Model 4)
Polity (continuous)	−0.173 (0.119)	0.322* (0.07)	−0.108 (0.138)	0.351*** (0.084)
Polity × Initiation	0.036 (0.044)	−0.055* (0.025)	0.113* (0.057)	−0.051 (0.033)
Polity ²	0.026* (0.012)	−0.023* (0.008)	0.009 (0.015)	−0.027** (0.010)
Major power	1.165* (0.524)	−0.035 (0.373)	0.483 (0.603)	−0.117 (0.434)
Diplomacy score	−2.201 (1.22)	−0.825 (0.848)	−2.557 (1.418)	−1.728 (1.012)
Dyadic capability ratio	−0.207 (0.553)	0.895* (0.361)	0.736 (0.622)	−0.229 (0.454)
Region 2	−1.752 (0.964)	−1.055* (0.614)	0.741 (0.766)	1.331 (0.698)
Region 3	2.423 (2.423)	0.507 (0.948)	31.37 (>100)	−0.692 (1.252)
Region 4	−0.529 (0.815)	−1.027* (0.616)	0.279 (0.956)	1.359 (0.782)
Region 5	1.816* (0.901)	0.009 (0.486)	−1.178 (1.014)	−1.114 (0.652)
Initiation	−0.235 (0.234)	0.725* (0.135)	1.354*** (0.289)	2.314*** (0.196)
Random effect SD	1.112* (0.222)	0.824* (0.151)	0.718 (0.252)	1.2256*** (0.189)
<i>n</i>	279	538	147	605

Note: $\hat{\beta}^* \Rightarrow p < .05$, $\hat{\beta}^{**} \Rightarrow p < .01$, $\hat{\beta}^{***} \Rightarrow p < .001$.

probabilities for the model fit are based on *specific* cases in the data, the graphics are based on *general* (predicted) patterns of conflict.

What We Found: Overview

Here we focus on control variables that have been used in other recent quantitative analyses of international conflict (see, e.g., Danilovic 2001; Reiter 2001; Hegre 2008). Table 3 reports our estimates for two subsets, interstate wars¹⁶ and conflicts

that fall below the level of war (non wars), as well as for the two sources of data (EUGene dyads and Maoz dyads).¹⁷ All variables represent the difference between the relevant covariate for i and j , as implied by equation (10). Thus, Polity for i is simply its Polity2 score as taken from the Polity Data set (where it ranges from -10 to $+10$), while Polity in the results table should be interpreted as the effect of a one-unit increase in i 's score relative to j , with the caveat—explained later—that we need to be careful both with interactions and with the sign of the resulting difference between the states.¹⁸ The quadratic term, $Polity^2$, is the state's absolute polity value multiplied by its signed polity value such that the original direction is preserved: thus, a Polity of -3 corresponds to a $Polity^2$ of -9 . This variable gives us a sense of possible increasing or decreasing “returns to scale” for democraticness: for example, it may be that being democratic is helpful for effectiveness and that more democratic countries are *disproportionately* better off. *Major Power* is a binary variable indicating whether or not a state is a major power as defined by the COW data set. Its “effect” should be interpreted as the effect of increasing i 's score by one unit (or in other words, turning it into a major power) relative to j . Diplomacy scores are the proportion of states in the system that have sent diplomatic representatives to a given country in a given year (e.g., 1 if all countries have sent representatives, 0.5 if half have sent representatives). Here, it is the difference between i and j 's diplomacy scores. Dyadic capability ratio is the difference between i and j 's capability ratio (where i 's share of the total dyadic capabilities is $CINC_i / (CINC_i + CINC_j)$). *Region* variables are dummy variables that denote where a country is located (as per COW: Europe = 1, Middle East = 2, Africa = 3, Asia = 4, North and South America = 5). Region 1 is the “base category” in this model (so, regions are “factors” from the point of view of taking a difference between covariates). *Initiation* indicates whether or not i initiated the conflict or dispute.

Recall that the justification for using the difference between i 's and j 's covariate values (for *Polity*, *Major Power*, *Diplomacy*, etc.) is as given in *How It Differs from Previous Approaches* section: that this is the only way that the logit equation can make sense as written. That is, to the extent that states have latent abilities and to the extent that the functional form we have given is the one implied when scholars typically think about interstate interactions—or is a reasonable approximation to it—we need to use differences rather than sums here.

Interpreting the coefficients is an operation similar in nature and complexity to that in the conventional logit or probit case. As in those scenarios, coefficient signs imply direction, as regard the relationship between that variable and the probability of a state winning an interaction it is involved in (all else equal). More specifically, the $\hat{\beta}$ s tell us how the log odds of victory for country i change as we increase the relevant X by one unit. This is exactly as suggested by equation (10), where the $\log \frac{Pr(i \text{ wins})}{Pr(j \text{ wins})}$ is a (linear) function of the difference between i 's and j 's covariate values.

Democratic “Effectiveness”?

In line with the democratic effectiveness literature we cited earlier, we first turn our attention to the potential link between a country’s Polity score and its overall military effectiveness. For our sample of wars, the estimate for Polity is negative (-0.173), but not statistically significant. Moreover, using the Maoz data, with its stricter criteria for dyads, generated almost identical results (-0.108), suggesting that this is not simply a matter of “what observations count.” In the following paragraphs, we explain how this estimate should be interpreted.

Holding all other variables constant, a one-unit increase in Polity corresponds to a one-unit increase in the difference between two states. This follows from our earlier equations, in the sense that the interest is in $\beta(x_i - x_j)$. Given this difference formula, an increase of one unit can occur via an decrease in x_j (fixing x_i) or an increase in x_i (fixing x_j) or some combination thereof. An extra subtlety is added by the fact that the original values of x_i and x_j affect our interpretation of the effect on the log odds. Suppose, for example, that $x_i > 0$ (i is at least slightly democratic), and that $x_j > 0$ too, but that $x_i < x_j$. In this case, a positive coefficient will be multiplied by a positive difference and the effect of a one-unit increase is to raise the log odds of a victory for i . This is not necessarily true, however, if $x_j > x_i$. Now a one-unit increase in the difference may still be a part of a negative difference, in which case a positive coefficient will act on that negative covariate and we will decrease the log odds of a victory for i . This could happen, for example, in the case where i ’s Polity score is -5 , but $x_j = 6$. Now the difference is $-5 - 6 = -11$, meaning that a one-unit increase still leaves a negative difference with commensurate effects on the log odds of victory.

Furthermore, an increase in Polity will affect the log odds of a victory via several coefficients: Polity itself, of course, but also the (absolute) quadratic version of the same term, plus the interaction with initiation, and initiation status per se. To make matters stark, consider a country which initiates and for which $x_i > x_j$ after the one-unit increase; for a one-unit change in Polity, the effect on the log odds is $\beta_{\text{Polity}} + \beta_{\text{Polity} \times \text{Initiate}} + \beta_{\text{Polity}^2} + \beta_{\text{Initiate}} = -0.173 + 0.036 + 0.026 - 0.235 = -0.346$. In words, the negative sign implies the state is more likely to lose as it become more democratic. If the state is a noninitiator, the effect of a one-unit increase in its Polity score is -0.147 : still negative, and implying that more democratic states are not better off (i.e., the effect is not simply driven by initiation status).

What of nonwars? Here, our conclusions are somewhat different. For nonwars, we find strong evidence for a democratic advantage: Polity has a positive and statistically significant effect on effectiveness using both the EUGene (0.322^*) and Maoz (0.351^{***}) data. Suppose, again, that we increase the Polity difference by one for initiators, and consider the effect on the log odds of a victory for a situation where $x_i > x_j$. We have $\beta_{\text{Polity}} + \beta_{\text{Polity} \times \text{Initiate}} + \beta_{\text{Polity}^2} + \beta_{\text{Initiate}} = 0.322 + -0.055 + -0.026 + 0.725 = 0.966$: a positive result, suggesting more democratic countries

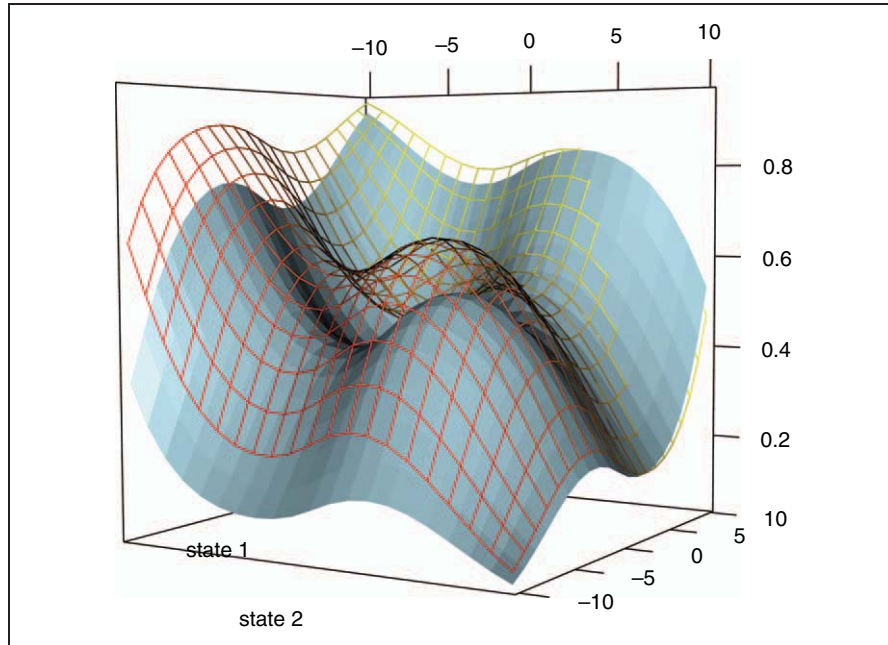


Figure 1. Wars.

Note: Effect of Polity score on the probability that state i wins the conflict. Unfilled (red) surface has i as target (and j as initiator), filled (blue) surface has i as initiator. Note that the surfaces are not monotonic in increasing democraticness for either i or j . Vertical axis is the relevant probability of victory for i , the axis emerging from the page refers to i 's Polity score. Horizontal axis moving from -10 to 10 left to right is j 's Polity score.

are generally more likely to win. Note that when we remove the initiator status, this log odds effect shrinks back to 0.296, though it is still positive.

As suggested, these results are not “cut-and-dried,” because they depend on differences. In Figures 1 and 2, we display the consequences of this point for wars and nonwars, respectively, in graphical form. There, the (red) unfilled surfaces refer to the probability of winning contingent on not initiating, while the (blue) filled surface correspond to the probability when initiating. Importantly, the non-vertical axes refer to Polity scores, and it can clearly be seen that the relationship between democraticness and effectiveness is nonmonotonic. For *some* ranges, more democratic states are more likely to win, but for others they are *less* likely. For example, in the wars case, Figure 1 shows that if we fix state 2 (i.e., state j) at a Polity of around 5, and move state 1 (i.e., state i) from -5 to zero on the Polity scale, state i 's probability of winning is actually slightly decreasing: implying a decreasing return to democraticness. Elsewhere, for example, fixing j in the same place, but increasing i 's score from 5 to 10, we see a steep increase in the

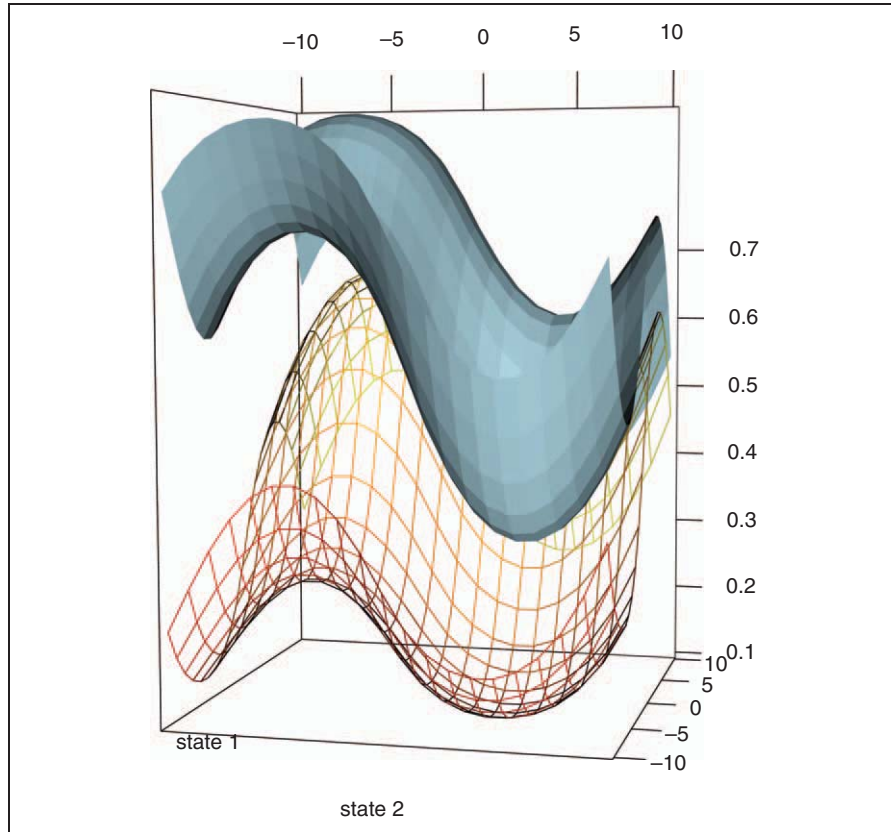


Figure 2. Nonwars.

Note: Effect of Polity score on the probability that state i wins the conflict. Unfilled (red) surface has i as target (and j as initiator) and filled (blue) surface has i as initiator. Note that the surfaces are not monotonic in increasing democraticness for either i or j . Vertical axis is the relevant probability of victory for i , the axis emerging from the page refers to i 's Polity score. Horizontal axis moving from -10 to 10 left to right is j 's Polity score.

probability that i wins. The story from Figure 2 is similarly subtle, though not quite as complicated. Here, for a wide central swath of democraticness for i , increasing j 's Polity score makes i less likely to win. But this is not true everywhere: for very undemocratic and very democratic j values, increasing j 's democraticness actually makes it less likely to win against i .

As detailed in earlier sections, our model and our data are set up in a manner different enough from previous work (on democratic effectiveness) that a direct comparison would be more misleading than illuminating. However, it is reasonable to speculate for a moment on how our findings fit with the conventional wisdom that

link democracy and military effectiveness. We noted above that research on this question had been divided into two broad schools: democratic triumphalists (who believe that, either through skillful selection or material capabilities, democracy boosts military effectiveness) and skeptics (who are unconvinced by the evidence for a democratic advantage). Our results are mixed: for some ranges, for some conflict types, democraticness is associated with advantages, but for some it is not. This (partially) corroborates the “swoosh” pattern found by Reiter and Stam (1998b, 387) in their related analysis of the Pr(initiator wins) at different levels of Polity. We see this as providing further evidence of the nonlinear relationship between Polity and effectiveness in war (and counter to the claim of Downes in Reiter, Stam, and Downes 2009). Similarly, the curvilinear relationship between democraticness and victory underscore the heterogeneity among autocratic regimes (see Weeks 2008).

With this in mind, we cannot unequivocally support the arguments of democratic triumphalists: it is not the case that increasing democraticness is always associated with higher probabilities of winning wars. However, we do find more consistent (though by no means completely consistent) evidence for a democratic advantage in nonwars, which corroborates the one other effort (of which we are aware) to investigate democratic effectiveness in a context other than major interstate wars (Gelpi and Griesdorf 2001).

Thus, on the one large question that drives much of this research—*are democracies better at fighting?*—our answer is yes, in some cases. On the many other methodological issues that separate triumphalists and skeptics (such whether a particular country *really* initiated a conflict), we remain agnostic. The passage of time is likely to give us more, and more accurate data to work with, but that will not change our argument that the method detailed in this article is the correct way to estimate effectiveness based on contest data.

We return to the specific issue of initiation next. The rest of our variables were essentially controls, and we can deal with them relatively quickly here: major powers are (significantly) more effective in wars, and less so (not significantly) in nonwars. States with greater diplomatic status do worse in both scenarios, though not significantly. States with a greater share of the capability ratio do significantly better in nonwars, but this seems not to help in wars. Finally, countries in region 5, North and South America (relative to those from Europe) are significantly better off in wars, while countries from Asia and the Middle East (relative to those from Europe) are worse off. To round out this section, we note that unsurprisingly, both sets of random effects were significant. This implies that there are, indeed, correlations in performance across contests involving the same states.

Initiation: Findings

Notice that this setup allows us to properly estimate the effect of initiating a war. This is, we believe, the first “clean” estimate of the value of moving first in war and helps to shed light on a question of substantial interest to IR scholars (see, e.g., Jervis

1978; Fearon 1997; Van Evera 1998): has offense been dominant in modern political history? Previous works have debated the answer to this question for different periods, but substantial difficulties remain since it is often difficult to come to an agreement on whether a specific technological advance favors offense or defense.

Figure 1 illustrates more clearly the pattern we observe in the regression results in Table 3, namely that there is no evidence of an “initiation” effect in wars. That is, moving first in wars provides no advantage to initiators in our model. If anything, the pattern depicted in Figure 1—specifically that the unfilled (red) lattice surface is above the filled (blue) surface for almost every value of Polity for both i and j —is suggestive of the reverse: that initiating wars carries with it a disadvantage. However, we caution against overinterpretation here, as the same model applied to the Maoz dyadic data yielded conflicting findings. In that model, initiation was positively associated with victory in wars. When divergent findings such as this emerge, one positive benefit is in spurring future research designed to further sharpen our understanding.

With that in mind, we note (speculatively) that the findings in model 1 related to initiation do fit well with extant theories of initiation dynamics. For example, Nevin (1996) finds that war outcomes influence future decisions to initiate such that winning makes states more likely to initiate in the future, while losing makes them less likely to do so. This seems likely to lead to overconfidence, such that an easy victory in the past may embolden leaders to make more risk-seeking decisions. This finding also dovetails with that of Geller (2000), who found that initiators were equally likely to be superior or inferior in capabilities relative to their opponents.

Intriguingly, the results for nonwars are substantially different from those for wars. These, shown for models 2 and 4 in Table 3 and then depicted in Figure 2 show clear evidence of an “initiation effect.” In contrast to the results for war, in nonwars it is obvious that the filled (blue) surface is substantially higher on the y -axis than is the unfilled (red) surface. The gap between these two lattice surfaces represents the difference between initiating and not initiating. The higher y values for initiating suggest that doing so is associated with a higher probability of victory. Critically, the same result emerges using the Maoz data as well, suggesting once again that this result is not sensitive to different specification of dyads.

Why does initiation increase the likelihood of victories for disputes short of war, but provide no discernible advantage for initiators in wars? One possible explanation lies with the different dynamics and incentives that states face in the two contexts. For example, a quick glance at the data reveals that many disputes that fall short of war occur between allies (the United States and Canada, e.g., have “fought” in six disputes since 1974). The lack of any true threat between the disputants in such cases lends itself to a higher likelihood of conciliatory bargaining since better relations between the states both makes issue linkage easier (“if you give in on issue x , we will give you what you want on issue y ”) and allows the states to trust that their concessions will be reciprocated over time. Similarly, the stakes involved in disputes that fall short of war are almost certain to be lower than those that are involved in wars, so states might find it easier to concede a dispute than they would be to concede the issue that is the cause of an interstate war.

Table 4. Table Comparing Some Selected Predicted Probabilities for “Vanilla” Logit and BT Model.

Year	Conflict	State 1	State 2	Logit probability (state 1)	BT probability (state 1)
1843	Mexican American War	USA	MEX	0.236	0.515
1853	Crimean War	TUR	RUS	0.160	0.157
1876	First Central American War	GUA	SAL	0.178	0.315
1896	Greco-Turkish War	TUR	GRC	0.184	0.030
1912	First Balkan War	BUL	TUR	0.138	0.517
1939	World War II	GMY	FRN	0.184	0.113
1940	World War II	GRC	FRN	0.129	0.244
1940	World War II	GRC	ITA	0.132	0.210
1940	World War II	FRN	ITA	0.183	0.243
1944	World War II	FRN	GMY	0.145	0.148
1944	World War II	FRN	JPN	0.151	0.100
1944	World War II	BRA	FIN	0.240	0.530
1964	Vietnam War	DRV	USA	0.070	0.018
1964	Vietnam War	CHN	USA	0.082	0.062
1964	Vietnam War	RUS	USA	0.086	0.056
1974	Cyprus War	TUR	CYP	0.212	0.150
1977	Ethiopian–Somalian Ogaden War	CUB	SOM	0.232	0.466
1977	Ethiopian–Somalian Ogaden War	ETH	SOM	0.234	0.575
1990	Persian Gulf War	SAU	IRQ	0.137	0.636
1990	Persian Gulf War	SAU	JOR	0.167	0.322

Note: In every case, state 1 “won” the conflict in practice, and we present the predicted probability of this outcome from the two models. Boldface values indicate the more accurate prediction.

Comparing Models

Given these distinctive findings, we wanted more information on what—that is, which conflicts—drive it relative to the “vanilla” logit approach. To that end, we looked at the twenty “worst-case” scenarios for the usual logit: in particular, ten cases where the logit predicted a loss, but the state actually won, and ten cases where the model predicted a win, but the state actually lost. We then compared the Bradley–Terry predictions (without random effects, to keep the models relatively similar) to those from the usual procedure. Table 4 contains the results of that analysis, and we present everything in “probability that actual winner won” form to make things easily comparable (recall, these are two-sided contests, but the logit treats them as “one sided”). One interesting pattern emerges from this. The BT model does better more often (twelve times of the twenty) than does the “vanilla” logit model, but simply counting correct predictions understates the higher accuracy of the BT model. For the twelve predicted better by the BT model, they were better by an average of 0.21, while for the eight cases for which the logit model made better predictions, its predictions were better by only .06 on average. In other words, for cases

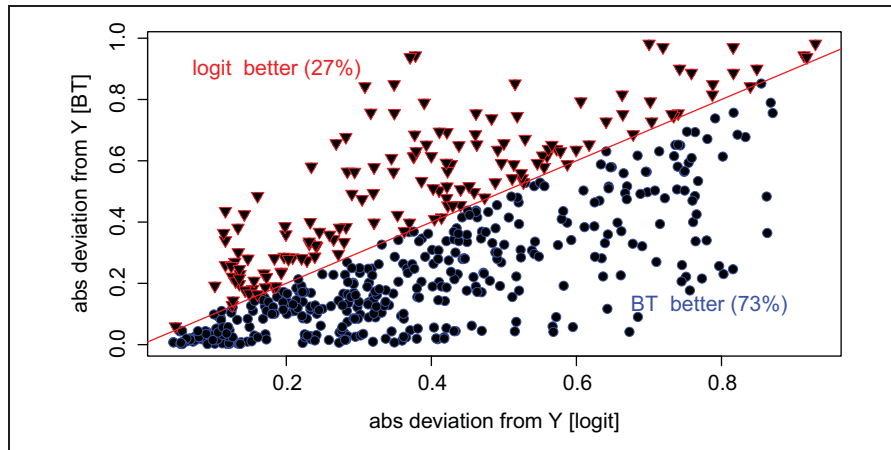


Figure 3. Relative performance of the logit and Bradley–Terry models across full (wars) data set. Note: Points below the lines are where the Bradley–Terry model does better in terms of its predicted probability relative to y . Points above the line are areas where the logit does better.

where the logit model was “better,” it was not better by much, whereas when the BT model outperforms the logit model, its predictions are much better.

So, where does the BT model do better than logit? One obvious pattern is the overrepresentation of large, multilateral wars in this table: Crimean War, First Balkan War, World War II, Vietnam War, and Persian Gulf dyads account for fourteen of the twenty worst predictions of the logit model. This bears out the warnings of Poast (2010) and Valeriano and Vasquez (2010) that more flexible methods may be better suited to analyzing this very specific subset of wars (though we note that even under these very difficult conditions, the BT model generally performs significantly better than a comparable logit model). The BT model also does significantly better in the Persian Gulf War, which involves several nondemocracies. Here, it is likely that the ability of the BT model to take differences in democraticness into account aids in its predictive power.

Although we may be able to draw substantive insight on model performance for particular conflicts from Table 4, it is helpful to see the larger picture of relative estimation accuracy. In Figure 3, we present a plot that does just this. There, every point represents one side of a conflict. The x -axis represents the predictive accuracy of the logit model, in terms of the deviation from the “truth” of the predicted probability produced by that model for that side of the conflict. The “truth” is whatever value the dependent variable took in practice. Thus, if a state won a conflict, and the logit predicts they won it with probability .63, the deviation is calculated as $1 - .63 = .37$. The y -axis is the same deviation, but from the point of view of the Bradley–Terry model. So, any (blue, circle) points falling *below* the 45° line are cases where the Bradley–Terry model is doing better than the logit; any (red triangle) points falling *above* the 45° line are cases where the logit model does better.

As can be readily seen, there are many more points below the line: indeed, some 73 percent of the observations are better predicted by the Bradley–Terry approach; moreover, where it predicts well, the Bradley–Terry model does much better than the logit—this can be seen by the “piling up” of the points at the bottom left of the plot, implying very little prediction error for that model.

Discussion

The foregoing article set out a statistical model of “effectiveness” in IR, and did so from first principles of states with latent abilities that are revealed from the contests they win or lose. We allowed for random effects (thus avoiding the error of overattributing effects to variables of interest) and showed how researchers might naturally account for the relevant characteristics of both players in a dispute. We also showed that, in model fit terms, the BT model outperforms more traditional options in predicting war outcomes.

The key innovation, though, is that this model asks the “right” question of conflict data—in particular, “what factors make states more likely to win wars”—and does so in a way that has a sensible interpretation in terms of underlying (but unobserved) “abilities.” We make no claims that we have radically overturned previous findings in this area: democracy still matters, though perhaps in more subtle ways than previously thought. But what is new is that our estimates can be related back to a sensible model of the data generating process and interpreted in that light. What is also new is the flexibility and generality of the model, and we hope that researchers will find its introduction here helpful.

Substantive findings are of course intertwined with empirical strategies, and because our approach freed us from conditioning the results on initiation status, we were able to provide a “clean” estimate of an initiation effect. We found strong evidence that moving first was associated with a significantly higher likelihood of victory in nonwars. However, we found that initiation effects for wars were highly sensitive to how one constructed the data set, whether using EUGene or Maoz codings. This has relevance both for traditional IR debates on whether or not offense or defense is “dominant,” but for contemporary foreign policy debates as well. However, as is often the case, answering one question (Does “moving first” provide an advantage in lower-level disputes?) generates numerous additional research topics. In this case, we provided evidence that initiation matters for nonwars, but how it affects war outcomes remains an open question. Additionally, we did not fully explore *why* this is the case. Some previous work (e.g., Betts 1982) has suggested that the surprise itself is what contributes to a high probability of victory. After all, an unexpected first strike that devastates the opposition should logically contribute to a higher likelihood of winning. However, it might also be the case that smart leaders simply select lower-level conflicts that they are likely to win (with or without the advantage of surprise). Notice that these theories are not mutually exclusive, and more evidence is needed (on everything from leaders’ beliefs about relative capabilities to more granular data on surprise attacks) in order to assess the relative merit of

the two explanations as well as the empirical observation that we do not find evidence for a strong “initiation effect” in interstate wars.

What is next for this approach? Earlier, we had little to say about a possible “selection stage” in conflicts. That is, it may well be the case that certain characteristics of states see them (strategically) initiate wars with one another, while a (possibly different) set of covariates determines the outcomes of those conflicts. It has been the latter stage that this note has sought to model. Future work might consider the ramifications of selection effects—especially with respect to bias in the sense of Heckman (1979)—and how scholars might deal with such concerns. In the Appendix we give some speculative comments specifically on these issues. All in all, there is much still to do, but we think we have provided a fruitful new avenue for this line of research.

Appendix

Bradley–Terry Models, Selection, and Selection Bias

One way to conceive of the data generating process is that states are randomly paired with others (i.e., the pairings are not dependent on the various characteristics the states possess) and that a contest takes place as soon as one of the states initiates. In this scenario, there is no selection bias in the estimation of the abilities.

An obvious concern is that the contests are, in fact, not exogenous. Instead, there is a *selection* stage in which states choose (in the sense that they attack) another state with whom to fight and then an *outcome* stage in which the relevant winner or loser is determined. To keep matters simple, suppose that a given state can choose at any time to attack any other: what, if any, bias might we expect? In this case, we are in a situation akin to that of Heckman (1979), albeit with binary outcomes at both stages (see Dubin and Rivers [1990], for discussion of fitting such models: we use their notation shown below).¹⁹ For the selection equation, we might represent matters as

$$y_{1i}^* = \beta_1(\mathbf{x}_i - \mathbf{x}_j) + \varepsilon_{1i}.$$

Here, the decision by i to attack (initiate) against j is partly due to (the difference between) i 's and j 's observed covariates, and partly due to some other unmeasured variables and residing in the error term ε_{1i} .²⁰ As written, y_{1i}^* is latent, and we would let $y_{1i} = 1$ if $y_{1i}^* > 0$ and be zero otherwise. One interpretation of this scenario is that i attacks if he expects to win. The outcome equation may be specified in a similar way.

$$y_{2i}^* = \beta_2(\mathbf{x}_i - \mathbf{x}_j) + \varepsilon_{2i},$$

where, this time, the outcome of victory or defeat for i is a function of the covariate differences between him and j and some error term. In the preceding article, it is from a variant of this second equation that we have sought to estimate the abilities of the states in our data. To make clear the potential for bias, suppose that one

component of ε_{1i} is having a “clever leader”—someone who is good at picking relatively weak states to prey on—whether that leader be democratic or a dictator. Suppose, in addition, that the interest is the effectiveness of democracies in the outcome equation and that we expect more democratic countries to be better at fighting wars. The problem is that states with “big” errors (clever leaders) end up winning conflicts along with more democratic states who are also good at fighting wars. As a result, if democracy does aid fighting effectiveness, our estimate of its contribution to ability is biased down. Generally speaking, the larger the effect that the unmeasured variable (clever leader) has on selection and the higher the correlation between the error terms, the worse this selection bias becomes.

Under what circumstances would this potential bias *not* be a problem? First, if the unmeasured variables in the selection equation—for us, “clever leader”—are uncorrelated with the unmeasured factors influencing victory in the outcome equation. This seems admittedly unlikely: surely a leader good at picking fights is good at winning them too. Second, if all factors that determine selection (here, initiation) are controlled for in the outcome equation. Again, this seems unlikely for the simple reason that some things are simply impossible to measure or observe, especially for international relations data sets.

Where does this logic leave the merits of the Bradley–Terry model presented in this article? First, one can choose to buy in to the somewhat heroic assumptions that mean potential bias may be ignored. These are the assumptions generally maintained in the effectiveness literature that was our jumping off point for this article.²¹ Second, and more importantly, our claim here is that the Bradley–Terry model is the right choice for the modeling of the outcome equation: it estimates the right thing (abilities) in the right way (as a function of differenced covariate values). We accept that estimates may be biased, but we are now at least focusing on the correct estimand.

Acknowledgments

We gratefully acknowledge comments from Iain Johnston, Gary King, Vipin Narang, Jane Vaynman, and the audiences at both the IQSS Applied Statistics and Security Studies workshops. Two anonymous referees and the editor provided us with helpful guidance on both content and structure.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. In economics as well, see the (theoretical) literature on “contest success functions” (Hirshleifer 2000).

2. This does not mean that every state meets every other in a conflict, simply that no two states are in conflict of which neither party meets any other. So, if there are three states *A*, *B*, and *C*, and *A* fights *B* and *B* fights *C*, we do not need *A* to also fight *C*. However, if there were four states and *A* fights *B* only, while *C* fights *D* only, we do not have a “connected” design: we have no way to compare the two sets of conflicts since they do not have a common player.
3. Note that while IR scholars typically use the term “latent capabilities” to refer to industrial capacity (and thus power that has not yet been brought to bear in a military conflict, but could be; see Walt 1985; Mearsheimer 2001), we use “latent abilities” in its more narrow sense: as an unobserved quality inferred through the observation of data that are measurable (here, contest outcomes).
4. This “better-ness” might well be something difficult to measure, such as “more advanced strategy” or “more experienced Admirals,” but in principle, that something is still a characteristic of the state (and not the conflict itself).
5. We are assuming we wish to know how effective the United States is “overall” (and what that depends on, in terms of predictors) rather than the effectiveness of the “US at sea” or the “US on land.”
6. Recall that $\text{logit } p_{ij} = \log(p_{ij}) - \log(p_{ji})$ and $\text{logit } p_{ji} = \log(p_{ji}) - \log(p_{ij})$.
7. We mean “contest-specific” in the sense that some state covariate varies by contest, not that the contest itself is exerting a particular effect.
8. Thus, one significant difference between our data and that of previous works (over and above the already discussed difference in the structure of the data) is the temporal scope. Reiter and Stam (1998b) contains 197 observations (remember, they count each country in each conflict as one observation), beginning in 1823 with the Franco-Spanish War, and ending in 1982 with the Lebanon War between Israel and Syria. Reiter, Stam, and Downes (2009) and Downes (2009) both contain 234 observations of wars, beginning in 1823 with the Franco-Spanish war, and ending with the Second Sino-Vietnamese War, which took place from 1985 to 1987.
9. Thus, multilateral wars (such as World War I and II [WWI and WWII, respectively]) are decomposed into several dyads each. Although we agree with Valeriano and Vasquez (2010) that these “complex” wars are in many ways different than wars involving only one dyad, we note only that they can still be thought of as contests in which the outcome can provide information about relative abilities. Poast (2010) similarly notes several potential problems that arise from using dyadic data to analyze “*k*-adic” events. A multilateral Bradley–Terry model is a possible avenue for future research.
10. After generating the data set from EUGene, we modified our data set to match it as closely as possible to that used by Reiter and Stam (1998b), and the rest of the literature on democratic effectiveness. In practice, this simply meant correcting some obvious coding errors in COW, such as changing Germany to the initiator in the Germany–Poland dyad in 1939 and changing the United States to the winner of the United States–Japan dyad in WWII.
11. Polity scores vary from -10 to $+10$. Generally, scores above $+7$ are considered democracies, below -7 are considered autocracies, and those in the middle are considered “anocracies” or mixed regimes.

12. More specifically, the Polity2 variable converts: –66 (foreign interruption) to “system missing”; –77 (interregnum or anarchy) to a neutral score of 0 and –88 (transition) codes are prorated across the spans of the transition. For more information on this, see the *Polity IV Project Codebook*.
13. Dealing with draws in the Bradley–Terry model is not impossible, though, it involves an extension of the method; Rao and Kupper (1967) and Davidson (1970) consider such an approach, though, interestingly, Turner and Firth (2010) find that simply treating a tie as “half a win” works well in practice.
14. However, excluding joiners from the nonwars or wars + nonwars models does not change the results presented below relating to *Polity* and *initiation*. Restricting joiners from the wars-only model yields too few observations to make any meaningful inferences.
15. Again, our specific discussion here applies to models 1 and 2, the EUGene-created dyadic data, but results are similar for the Maoz data as well.
16. Following the traditional definition from the Correlates of War project: conflicts in which there were at least 1,000 battle-related fatalities; that is, conflicts in which the *cwhostd* or *mzhostd* variable is equal to 5.
17. In general, our analysis distinguishes between wars and nonwars. However, including MID level 4 (use of force) observations along with MIDs at level 5 (wars) to create a sample of “violent conflicts” does not affect estimates for our main variable of interest, Polity.
18. We acknowledge that many previous accounts focus on a binary Democracy variable. However, much recent research uses a continuous measure and we follow that lead here (see, e.g., Oneal and Russett 1997; Elkins 2000; Bennett 2006).
19. See also Freedman and Sekhon (2010) for discussion of options for fitting specifically for the probit case.
20. As in the material, this might be a difference in error terms, but we write it as one term here for notational clarity.
21. We fully acknowledge the increasingly large literature that takes selection (e.g., Reed 2000) and strategy (e.g., Signorino 1999) seriously in IR, but note that this has not, to our knowledge been applied to abilities estimation per se.

References

- Bayer, R. 2006. “Diplomatic Exchange Data Set, v2006.” <http://www.correlatesofwar.org/COW2%20Data/Diplomatic/Diplomatic.html>.
- Bennett, D. 2006. “Toward a Continuous Specification of the Democracy–Autocracy Connection.” *International Studies Quarterly* 50 (2): 313–38.
- Bennett, D., and A. Stam. 1998. “The Declining Advantages of Democracy: A Combined Model of War Outcomes and Duration.” *Journal of Conflict Resolution* 42 (3): 344–66.
- Bennett, D. S., and A. C. Stam. 2000. “EUGene: A Conceptual Manual.” *International Interactions* 26:179–204. <http://www.eugenesoftware.org/documentation.asp>
- Bennett, D. S., and A. C. Stam. 2007. “EUGene Documentation.”
- Berry, W., J. DeMeritt, and J. Esaarey. 2010. “Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?” *American Journal of Political Science* 54 (1): 248–66.

- Betts, R. 1982. *Surprise Attack: Lessons for Defense Planning*. Washington, DC: Brookings Institution Press.
- Biddle, S., and S. Long. 2004. "Democracy and Military Effectiveness: A Deeper Look." *Journal of Conflict Resolution* 48 (4): 525-46.
- Bradley, R., and M. Terry. 1952. "Rank Analysis of Incomplete Block Designs, I. The Method of Paired Comparisons." *Biometrika* 39:324-45.
- Breslow, N., and D. Clayton. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88:9-25.
- Bueno de Mesquita, B., J. Morrow, R. Siverson, and A. Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review* 93 (4): 791-807.
- Cannizzo, C. 1980. "The Costs of Combat: Death, Duration, and Defeat." In *The Correlates of War II: Testing Some Realpolitik Models*, edited by J. David Singer, 233-57. New York: Free Press.
- Carrubba, C., A. Yuen, and C. Zorn. 2007. "In Defense of Comparative Statics: Specifying Empirical Tests of Models of Strategic Interaction." *Political Analysis* 15 (4): 465-82.
- Carter, D. 2010. "The Strategy of Territorial Conflict." *American Journal of Political Science* 54 (4): 969-87.
- Danilovic, V. 2001. "The Sources of Threat Credibility in Extended Deterrence." *Journal of Conflict Resolution* 45 (3): 341-69.
- Davidson, R. 1970. "On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments." *Journal of the American Statistical Association* 65 (329): 317-28.
- Desch, M. 2002. "Democracy and Victory: Why Regime Type Hardly Matters." *International Security* 27 (2): 5-47.
- Desch, M. 2003. "Democracy and Victory: Fair Fights or Food Fights?" *International Security* 28 (1): 180-94.
- Desch, M. 2008. *Power and Military Effectiveness: The Fallacy of Democratic Triumphalism*. Baltimore: Johns Hopkins University Press.
- Downes, A. 2009. "How Smart and Tough are Democracies? Reassessing Theories of Democratic Victory in War." *International Security* 33 (4): 9-51.
- Dubin, J., and D. Rivers. 1990. "Modern Methods of Data Analysis." In *Selection Bias in Linear Regression, Logit and Probit Models*, edited by J. Fox and S. Long, 410-42. Thousand Oaks, CA: Sage.
- Elkins, Z. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 44 (2): 293-300.
- Fearon, J. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577-92.
- Fearon, J. 1997. "The Offense-Defense Balance and War Since 1648." <http://www.stanford.edu/~jfearon/papers/OFFDEF1.pdf>.
- Firth, D. 2005. "Bradley-Terry Models in R." *Journal of Statistical Software* 12 (1): 1-12.
- Fortna, P. 2009. "Where have all the Victories Gone? Peacekeeping and War Outcomes" <http://www.columbia.edu/~vpf4/victories%20Sept%202009.pdf>.
- Freedman, D., and J. Sekhon. 2010. "Endogeneity in Probit Response Models." *Political Analysis* 18 (2): 138-50.

- Geller, D. 2000. "Status Quo Orientation, Capabilities, and Patterns of War Initiation in Dyadic Rivalries." *Conflict Management and Peace Science* 18 (1): 73-96.
- Gelpi, C., and M. Griesdorf. 2001. "Winners or Losers? Democracies in International Crisis, 1918-94." *American Political Science Review* 95 (03): 633-47.
- Green, D., S. Kim, and D. Yoon. 2001. "Dirty Pool." *International Organization* 55 (2): 441-68.
- Heckman, J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153-61.
- Hegre, H. 2008. "Gravitating Toward War: Preponderance May Pacify, but Power Kills." *Journal of Conflict Resolution* 52 (4): 566-89.
- Hirshleifer, J. 2000. "The Macrotechnology of Conflict." *Journal of Conflict Resolution* 44 (6): 773-92.
- Jagers, K., and T. Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research* 32 (4): 469-82.
- Jervis, R. 1978. "Cooperation under the Security Dilemma." *World Politics: A Quarterly Journal of International Relations* 30 (2): 167-214.
- Jones, D. M., S. A. Bremer, and J. D. Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15 (2): 163-213.
- King, G. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Dyad." *International Organization* 55 (2): 497-507.
- Lake, D. 1992. "Powerful Pacifists: Democratic States and War." *The American Political Science Review* 86 (1): 24-37.
- Lake, D. 2003. "Fair Fights? Evaluating Theories of Democracy and Victory." *International Security* 28 (1): 154-67.
- Lewis, J., and K. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11 (4): 345-67.
- Maoz, Z. 2005. "Dyadic MID Dataset (version 2.0)." <http://psfaculty.ucdavis.edu/zmaoz/dyadmid.html>
- McCulloch, C., and S. Searle. 2001. *Generalized, Linear and Mixed Models*. New York: John Wiley.
- Mearsheimer, J. 2001. *The Tragedy of Great Power Politics*. New York: W. W. Norton.
- Nevin, J. 1996. "War Initiation and Selection by Consequences." *Journal of Peace Research* 33 (1): 99-108.
- Oneal, J., and B. Russett. 1997. "The Classical Liberals were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly* 41 (2): 267-94.
- Poast, P. 2010. "(Mis) Using Dyadic Data to Analyze Multilateral Events." *Political Analysis* 18 (4): 403-25.
- Rao, P. V., and L. L. Kupper. 1967. "Ties in Paired-comparison Experiments: A Generalization of the Bradley-Terry Model." *Journal of the American Statistical Association* 62 (317): 194-204.
- Reed, W. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44 (1): 84-93.
- Reiter, D. 2001. "Does Peace Nurture Democracy?" *The Journal of Politics* 63 (3): 935-48.
- Reiter, D., and C. Meek. 1999. "Determinants of Military Strategy, 1903-1994: A Quantitative Empirical Test." *International Studies Quarterly* 43 (2): 363-87.

- Reiter, D., and A. Stam. 1998a. "Democracy and Battlefield Military Effectiveness." *Journal of Conflict Resolution* 42 (3): 259-77.
- Reiter, D., and A. Stam. 1998b. "Democracy, War Initiation, and Victory." *American Political Science Review* 92 (2): 377-89.
- Reiter, D., and A. Stam. 2002. *Democracies at War*. Princeton, NJ: Princeton University Press.
- Reiter, D., and A. Stam. 2003. "Understanding Victory: Why Political Institutions Matter." *International Security* 28 (1): 168-79.
- Reiter, D., A. Stam, and A. Downes. 2009. "Another Skirmish in the Battle over Democracies and War." *International Security* 34 (2): 194-204.
- Robinson, G. K. 1991. "That BLUP is a Good Thing: The Estimation of Random Effects." *Statistical Science* 6 (1): 15-32.
- Schultz, K. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53 (2): 233-66.
- Signorino, C. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93 (2): 279-97.
- Signorino, C. 2003. "Strategic Misspecification in Regression Models." *American Journal of Political Science* 47 (3): 551-66.
- Singer, J., S. Bremer, and J. Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War and Numbers*, edited by B. Russett, 19-48. Thousand Oaks, CA: Sage.
- Singer, J., and M. Small. 1966. "The Composition and Status Ordering of the International System: 1815-1940." *World Politics: A Quarterly Journal of International Relations* 18 (2): 236-82.
- Smith, A. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43 (4): 1254-83.
- Sullivan, P. 2007. "War Aims and War Outcomes: Why Powerful States Lose Limited Wars." *Journal of Conflict Resolution* 51 (3): 496-524.
- Turner, H., and D. Firth. 2012. "Bradley-Terry Models in R: The BradleyTerry2 Package." *Journal of Statistical Software* 48 (9): 1-21.
- Valeriano, B., and J. Vasquez. 2010. "Identifying and Classifying Complex Interstate Wars." *International Studies Quarterly* 54 (2): 561-82.
- Van Evera, S. 1998. "Offense, Defense, and the Causes of War." *International Security* 22 (4): 5-43.
- Walt, S. 1985. "Alliance Formation and the Balance of World Power." *International Security* 9 (4): 3-43.
- Wayman, F., J. Singer, and G. Goertz. 1983. "Capabilities, Allocations, and Success in Militarized Disputes and Wars, 1816-1976." *International Studies Quarterly* 27 (4): 497-515.
- Weeks, J. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35-64.
- Whang, T. 2010. "Empirical Implication of Signaling Models: Estimation of Belief Updating in International Crisis Bargaining." *Political Analysis* 18 (3): 381-402.
- Wright, Q. 1965. "The Escalation of International Conflicts." *Journal of Conflict Resolution* 9 (4): 434-49.