Estimating State Preferences in International Crises: Promise and Limitations in Fully Structural Estimation[1]

PRELIMINARY AND INCOMPLETE,
DO NOT CITE WITHOUT PERMISSION.

Jeffrey B. Lewis[2]
Department of Political Science
UCLA

Kenneth A. Schultz[3]
Department of Political Science
Stanford University

February 11, 2005

[1]Earlier versions of this paper was presented the Midwest Political Science Association Meetings, Chicago IL, April 2004, and at the American Political Science Association Annual Meeting, Chicago IL, Sept. 2004. We gratefully acknowledge the support of the National Science Foundation (NSF-SES-0241647). Thanks to Cesar Zucco for research assistance.

[2]email: `jblewis@ucla.edu`
[3]email: `kschultz@stanford.edu`

1 Introduction

A central problem in the study of international conflict is to understand the factors that influence states' incentives and behavior in crisis situations involving the threat or use of force. What political, military, and economic variables influence how states (or, perhaps more accurately, their leaders) assess the attractiveness of waging war? Under what conditions will the targets of threats choose to yield or resist? What factors make it likely that leaders will implement the threats they make? There is no shortage of hypotheses regarding these questions in the IR literature. Recent interest, for example, has centered around such questions as

- whether regime type affects the propensity to win wars (e.g., Lake 1992; Reiter and Stam 2002), the costs of war (e.g., Russett 1993; Bueno de Mesquita and Siverson 1995), and the costs of backing down from a threat (e.g., Fearon 1994; Schultz 1999, 2001b);

- how territorial interests influence the stakes and outcomes of international crises (e.g., Huth 1996; Vasquez 1993; Diehl 1991);

- how alliances and strategic interests influence the the willingness to engage in crises or wars (e.g., Gowa 1999);

- whether economic interdependence and/or development makes war less attractive (e.g., Russett and Oneal 2001; Mousseau 2000);

- and how historical rivalries influence states' willingness to wage war or make concessions (e.g., Crescenzi and Enterline 2001; Crescenzi, Kathman, and Long 2004).

A challenge that arises in testing such hypotheses is that the outcomes of international crises are the product of interdependent choices: what a state decides to do depends not only on its own preferences but also on its beliefs and expectations about the preferences of the opponent. Moreover, these decisions are typically made under a great deal of uncertainty. In such settings, it is generally not straightforward to infer actors' underlying preferences from their observable choices (Frieden 1999). Thus, understanding how factors like regime type, power, and alliances affect states' preferences over war and peace calls for an empirical technique that is sensitive to dynamics of strategic interaction (Signorino 1999, 2003; Smith 1999).

In this paper, we provide some preliminary results using such a technique. The method that we employ involves making inferences from revealed preferences through the use of a fully structural strategic model. That is, we use observational data on state choices in crises to infer the distribution of preferences over the possible outcomes and the effects of covariates on those preferences. The empirical estimator is derived directly from the equilibrium of an extensive form crisis bargaining game with incomplete information. Many of the details and properties of this estimator were discussed in an earlier paper (Lewis and Schultz 2003). Here, we apply this technique to a new data set on international crises in the period 1919-1939. The data combine events from two widely used data sets—the International Crisis Behavior (ICB) and Militarized Interstate Dispute (MID) projects. Using the cases identified in these sources, we coded the outcome of each incident according to the four possible outcomes of the theoretical model.

While the small size of the current data set prevents an extensive test of all the relevant hypotheses in the IR literature, here we provide some preliminary results on the effects of several variables of interest. Most of the results we obtain are sensible, suggesting that there is promise in pursuing this line of inquiry. Of particular note are some interesting effects associated with regime type; in particular, democracy seems to both increase the costs associated with backing down from a threat and to decrease the attractiveness of war. Such a multi-faceted effect is not only important in its own right, but it demonstrates the ability of our method to uncover complex relationships that might be hard to tease out using standard methods.

That said, we pursue this research with a certain amount of caution and skepticism. As we hope to make clear later, fully structural empirical models have limitations that researchers will have to appreciate before embracing them. In particular, this approach requires the analyst to put enormous faith in the underlying theoretical model, a leap that many may be uncomfortable taking. Nonetheless, the field has shown recent interest in this kind of

A Crisis Bargaining Game

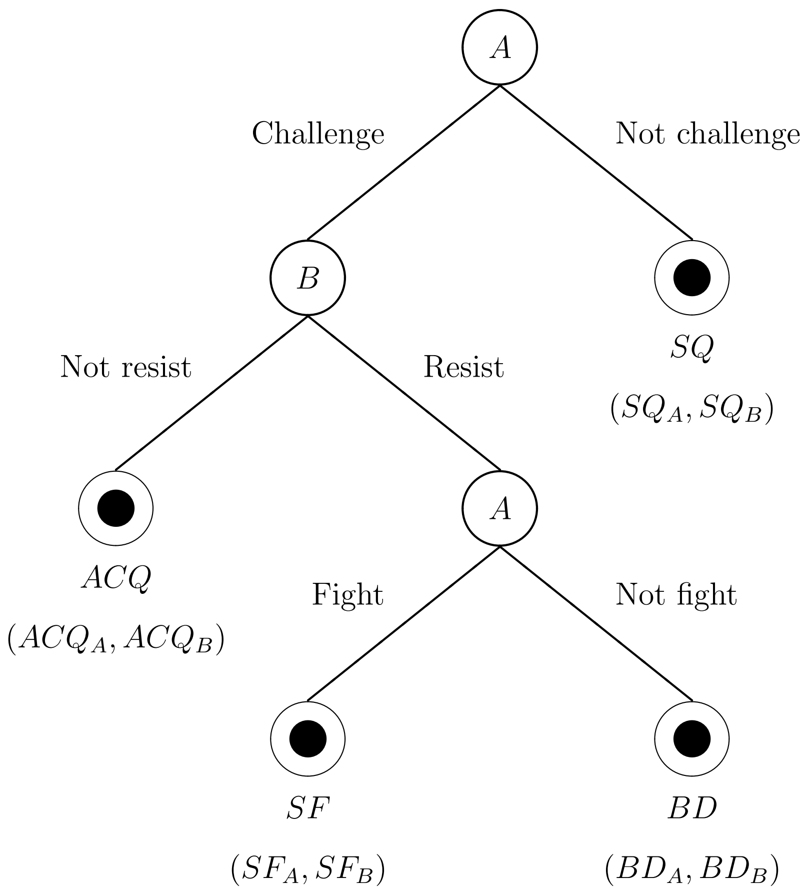

Figure 1: Simple crisis bargaining game.

analysis (esp. Signorino 1999, 2003; Smith 1999), and it is worthwhile to push this frontier forward to see where it leads us.

This paper proceeds as follows. The first section sketches the method we employ by showing how we derived an empirical estimator from an extensive form crisis bargaining game. Section 2 then discusses some the of the pros and cons of using this approach to estimating state preferences in international crises. Section 3 discusses the data used in this analysis. Section 4 then presents our preliminary estimation results.

2 From Theoretical Model to Empirical Estimator

2.1 Theoretical Model

The theoretical model underlying this analysis is a simple crisis bargaining game with incomplete information. Figure 1 depicts the extensive form.

Two states, A and B, are assumed to have some dispute. The game begins with a decision by A whether or not to challenge B by issuing a demand to alter the status quo in A's favor. A challenge is assumed to involve an explicit threat to use military force in the event that B does not acquiesce to the demand. If A chooses not to make a challenge, the status quo (SQ) prevails.[1] If A does make a challenge, then B must decide whether or not to resist the demand. If B does not resist, it is said to have acquiesced (ACQ), and the crisis ends peacefully. If B resists, then A must decide whether or not to carry out its threat. If A chooses not to fight, then it is said to have backed down (BD). Otherwise, A stands firm (SF) and implements its threat. We will sometimes refer to this last outcome as "war," but nothing requires that the action that ensues meet the usual requirements for full-scale war.

We label the payoffs generically and make only a few ex ante assumptions about their ordering. In particular, it makes sense that A should value acquiesence more than the status quo, or $ACQ_A > SQ_A$. Likewise, B should generally prefer the status quo and a back down by A to making concessions, or $SQ_B,\ BD_B > ACQ_B$. It is also common to assume, following Fearon (1994), that A pays a cost for backing down from its challenge, or $BD_A < SQ_A$. We will not, however, impose this restriction; rather, we will leave the characteristics of this payoff, including its likely sign and magnitude, as an empirical matter.

The game is played with two-sided incomplete information. In particular, we assume that B has private information about its payoff from war, SF_B, and A has private information about both its payoff from war, SF_A, and its payoff from backing down, BD_A. In assuming that A has private information about both of these payoffs, we depart from standard practice, in which it is generally assumed that the costs of backing down from a threat are common

[1]The term status quo is something of a misnomer here, since states may change the status quo without threats of force. Hence, this outcome reflects the absence of a militarized challenge, not necessarily preservation of the status quo.

knowledge (e.g., Fearon 1994; Smith 1998; Schultz 2001a). Introducing uncertainty about the audience costs makes the informational assumptions more realistic and also turns out to simplify estimation. Formally, we assume that Nature selects SF_A, BD_A, and SF_B according to some probability distribution. The realized values are observed only by the appropriate state. The probability distributions from which these values are drawn are assumed to be common knowledge. All other payoffs in the game are also assumed to be common knowledge.

In Lewis and Schultz (2003), we showed how to derive this game's unique perfect Bayesian equilibrium (PBE), in which all the strategies are sequentially rational given the actors' beliefs and beliefs are calculated from the equilibrium strategies according to Bayes' rule.[2] The equilibrium solution determines the probability of each outcome as a function of the payoff distributions.

2.2 Empirical Estimation

The objective of the estimation is to recover the effects of predetermined covariates on the payoffs of the crisis bargaining game from observations on those covariates and the corresponding outcomes of plays of the game. Hence, we now assume that the crisis game is played many times by different actors, and that, for each play of the game i, we can observe the outcome, $y_i \in \{SQ, ACQ, BD, SF\}$ and corresponding covariates, $\mathbf{x}_i$. We then rewrite the payoffs as a linear function of the covariates and, in some cases, stochastic shocks, as shown in Figure 2.

The shocks are assumed to be independently and normally distributed with mean zero and unit variance. Moreover, we assume that the shocks are observed only by the actor whose payoff is being perturbed. Thus, A observes ϵ_{SF_A} and ϵ_{BD_A}, while B observes ϵ_{SF_B}. All of the covariates, on the other hand, are common knowledge. Notice that these assumptions perfectly replicate the information structure in the game-theoretic model. Indeed, in moving from Figure 1 to Figure 2, we have simply rewritten the payoffs as index functions of some

[2]While in theory there can be multiple equilibria in this game, we assume below that the distribution of types is unbounded. Given this assumption, every decision node is reached with nonzero probability in equilibrium; therefore, Bayes' rule can be used to update beliefs at every decision node.

From Theoretical to Empirical Model

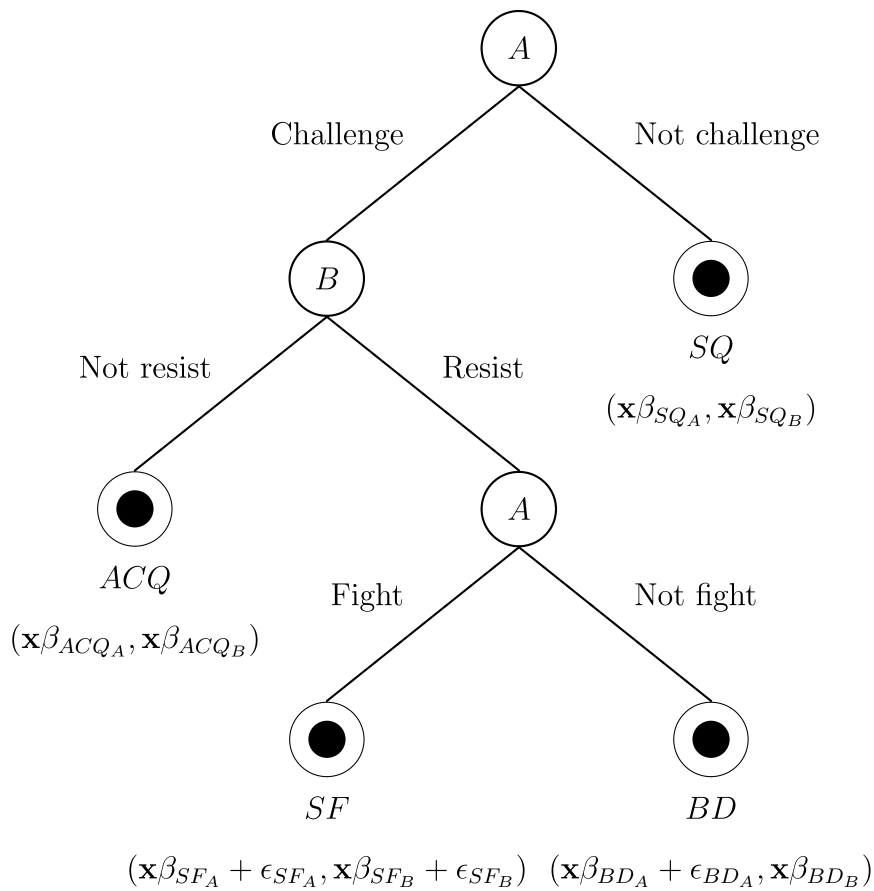

Figure 2: Crisis bargaining with indexed payoffs.

covariates and imposed a normal distribution on the unobserved components of SF_A, BD_A, and SF_B.

Once rendered this way, the connection between the theoretical and empirical exercise becomes clear. Game theory tells us how the payoffs map into equilibrium outcome probabilities. We can then take data on observed outcomes and covariates and use maximum likelihood estimation to invert the game theoretic solution: that is, to make inferences about the effects of the covariates on the payoffs (i.e., the β's).

Formally, the perfect Bayesian equilibrium provides a mapping from a vector of expected payoffs to the probability with which each outcome occurs,

$$T : \mathcal{U} \longrightarrow \mathcal{P}$$

where

$$\mathcal{U} \equiv (SQ_A, SQ_B, ACQ_A, ACQ_B, BD_A, BD_B, SF_A, SF_B)$$

and

$$\mathcal{P} = (\Pr(SQ), \Pr(ACQ), \Pr(BD), \Pr(SF)) .$$

Note that each element of $\mathcal{U}$ can be written as $U_j = \mathbf{x}\beta_j$ where $\mathbf{x}$ is a row vector of covariate values and β_j is a K-dimensional column vector of parameters associated with the jth utility (element of $\mathcal{U}$). Let $\boldsymbol{\beta}$ be the $K \times 8$ matrix of parameters formed by column joining each of the eight parameter vectors together. Given T and the mean payoffs, the probability of each outcome can be written as

$$\Pr(y|\mathbf{x}, \boldsymbol{\beta})$$

for $y = \{SQ, CD, SF, BD\}$. A multinominal likelihood function for a set of N observations generated by the model can be written as

$$L(\boldsymbol{y}|\boldsymbol{\beta}, \mathbf{X}) = \prod_{i=1}^{N} \Pr(y_i|\mathbf{x_i}, \boldsymbol{\beta})$$

where $\mathbf{y}$ is a vector of observed outcomes $(y_1, \ldots, y_N)$ and $\mathbf{X}$ is the $N \times K$ matrix formed by row joining the vectors $\mathbf{x}_i$ for $i = 1, \ldots, N$.

As is generally the case in random utility/discrete choice models, not all of the elements of $\boldsymbol{\beta}$ can be estimated. A certain number of identifying restrictions must be made. To begin, nothing can be estimated with respect to SQ_B.[3] This is because the strategic calculus of the game does not in any way depend on this payoff and, thus, nothing about this payoff can be revealed through the observation of game outcomes. As we will discuss below, other covariates and constants must be omitted from at least one payoff for each player. And, as a practical matter, even more restrictions are required to overcome what are sometimes referred to as empirical identification problems and more generally to avoid near colinearity in the parameter estimates.

Estimation of the model is not straightforward. Unlike in familiar models such as logit, in this model the mapping between the payoffs and the probabilities of the various outcomes has no closed-form solution. Rather, a set of implicit functions that define the PBE equilibrium must be solved by iterative numerical techniques. Thus, not only the likelihood itself, but also the link function relating $\mathbf{X}\boldsymbol{\beta}$ to the value of the dependent variable are computationally intensive. The efficiency and stability of the estimation routine was greatly enhanced through the inclusion of the analytically-derived gradient. Further improvements might result from the inclusion of the analytical Hessian or through the use of more sophisticated maximizing routines. We will be investigating these possibilities in future work.

3 Promise and Pitfalls

Why do this? What is the value-added of estimating a fully structural strategic model, rather than using more conventional techniques? In this section, we discuss some of the advantages of using this methodology rather than off-the-shelf models that are usually used in quantitative IR. At the same time, we are aware that this method is not without limitations, and we go through some of these as well.

[3] A general discussion of issues of identification in estimating models of strategic interaction are given in Lewis & Schultz (2001).

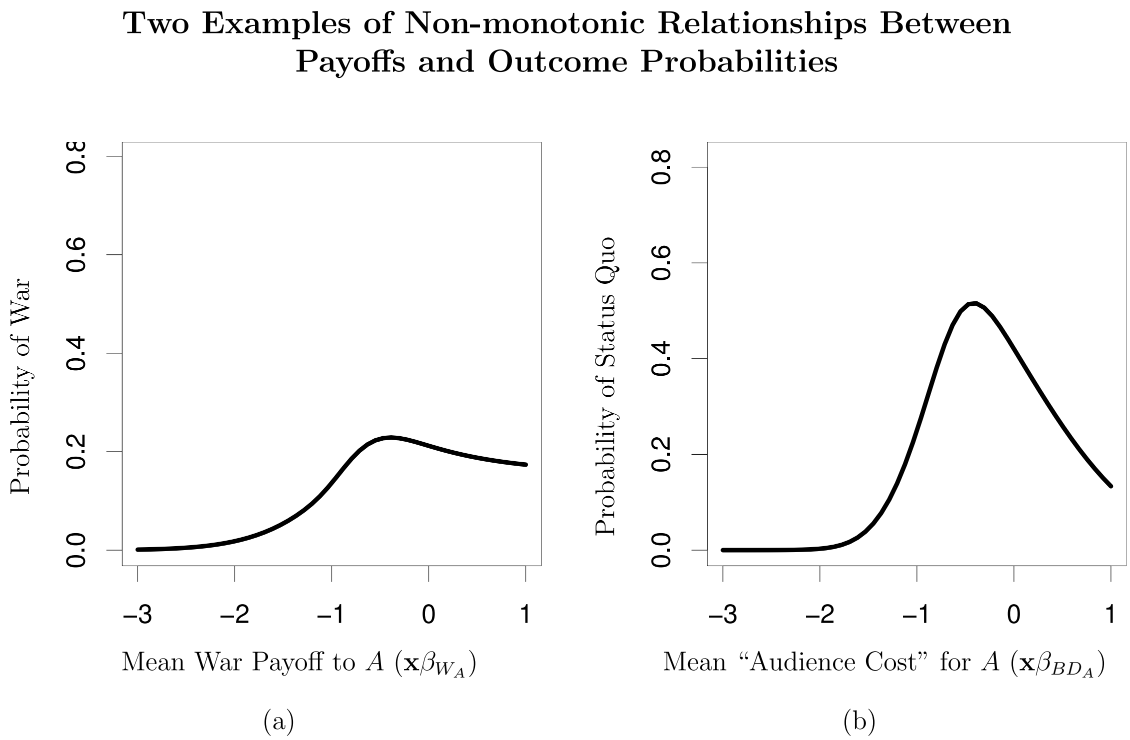

Figure 3: *As described in the text, increasing the expected value of war for state A at first increases the probability of a war outcome, but ultimately reduces that probability as state B increasingly acquiesces to A's demands. Similarly, as A's audience costs fall (utility from backing down increases), the probability of a status quo outcome first rises as A is decreasingly able to coax B into acquiesce through a credible commitment to fight and, consequently, A is decreasingly likely to challenge. As this utility further increases, A does not experience a loss from backing down and increasingly challenges resulting in fewer status quo outcomes.*

3.1 Advantages to Fully Structural Models

The main argument for employing this methodology is that, if we believe that crises take the strategic form shown in Figure 1, then our method fully internalizes the mechanism that is believed to generate the data. This is important because strategic settings can cause the relationship between variables of interest and observable outcomes to be more complicated than standard techniques force us to assume (Signorino 1999; Smith 1999). At a minimum, the functional forms of these relationships may look very different from those imposed by logit or probit regressions. At worst, the implied relationships may be non-monotonic, making them awkward to capture in off-the-shelf linear models.

Figure 3 shows two examples of non-monotonic relationships that can arise in our model. Panel (a) shows how the equilibrium probability of war varies with the mean war payoff to A, $\overline{SF}_A \equiv \mathbf{x}\beta_{SF_A}$. In this example, the probability of war rises with $\mathbf{x}\beta_{SF_A}$ until it reaches a peak at around -0.5, at which point the probability of war starts to fall and then level off. The reason for this is straightforward. When the systematic, observable component of A's payoff from war is very low, then it is very unlikely that A will be in a position to make a challenge or, if it does, to carry out its threat. At the other extreme, when $\mathbf{x}\beta_{SF_A}$ is very high, it is very likely that A will want to fight; however, the probability of war is low because B, knowing that A is resolved, is unlikely to resist a challenge. The highest probability of war falls somewhere in between.

Similarly, panel (b) shows the equilibrium probability of a status quo outcome as a function of A's mean payoff from backing down, $\overline{BD}_A \equiv \mathbf{x}\beta_{BD_A}$. When the observable component of this payoff is very low—in which case A is expected to suffer very high costs from backing down—then B must expect that a threat will be carried out. Because it is easy to get B to acquiesce under these circumstances, A finds challenges very attractive. At the other extreme, high positive values for backing down also make challenges attractive, since A must make a challenge to enjoy the benefits of backing down. The highest probability of an SQ outcome once again occurs somewhere in between the two extremes. The possibility of such non-monotonic relationships underscore the need for an estimator that is faithful to the underlying strategic interaction.

A final virtue of this approach is that it allows us to test the full model rather than just isolated pieces. In particular, we can ask not only whether a given factor influences, say, the probability of war, but we can in principle determine through which payoff or payoffs that factor acts.

To see how this might matter, consider the potential pitfalls of the current best practice in the literature, which involves testing isolated comparative statics from a model. For example, Schultz (1999) uses a game theoretic model similar to the one here to test some competing theories for how democratic institutions influence crisis outcomes. The model showed that, if democracy in A decreases W_A—as suggested by "institutional constraints" arguments (e.g., Bueno de Mesquita and Lalman 1992; Morgan and Campbell 1991)—then

B should be more likely to resist challenges by democratic states. On the other hand, if democracy in A decreases BD_A—as suggested by Fearon (1994)—then B should be less likely to resist democratic challenges. Schultz then presented empirical evidence that threats by democratic initiators are less likely to be reciprocated by their targets than are threats by nondemocratic initiators—a result that is consistent with the claim that democracies face higher "audience costs" for backing down.

While this is an interesting and important result, an increase in the probability of the ACQ outcome can be caused in more than one way. Comparative statics from the model show that the same empirical pattern would arise if democracy in A increased W_A or decreased W_B, both of which are consistent with recent theories about the superior war-fighting ability of democracies (e.g., Bueno de Mesquita et al. 1999; Reiter and Stam 2001). Hence, if democratic challengers enjoy lower rates of resistance, ceteris paribus, it would be good to know the exact mechanism: that is, through which payoff(s)—BD_A, SF_A, and/or SF_B—democracy in A operates. As we will see, the method we employ can make it possible to tease out these different pathways (though the actual effects of regime type in our data set are somewhat different from those described in Schultz [1999]).

3.2 Caveat Emptor

The potential benefits of fully structural empirical models are thus quite impressive. Nonetheless, we do not argue here that all empirical work on international conflict should go this route. Our hesitation is driven by an awareness of the costs and limitations of this approach, some of which we mention here.

The first concern is that this method asks us to take the theoretical model very literally, both in the collection of data and in the construction of the estimator. The data need to identify, for each event, who is state A, who is state B, and what outcome node of the game was reached. While we have made a valiant effort to do this, the exercise makes it evident that, not surprisingly, the world is a more complicated place than the model suggests. Choices that are modelled as dichotomous are actually continuous in the real world. Interactions that are modelled as having two players may often, in reality, involve many more. This is not a criticism of the theoretical model: the whole point of a model is not to capture the world in all of its complexity, but rather to strip away the complexity and boil down a class of interactions to their essential features. Once we build our empirical test directly out of the model, however, we are making a strong assumption that this model actually generated the data. This is an assumption that many modelers might be unwilling to make. While this is a concern, we suggest that this method can still be an advance over current best practice—in which the connection between theory and empirics is often indirect or obscured.

A second caveat to mention is that not all interesting hypotheses regarding crisis behavior deal with how covariates influence outcome utilities. For example, Sartori (2002) argues that past behavior influences the credibility of current threats through the use of history-contingent strategies, rather than through payoffs. And Schultz (1998) argues that democracy in the challenger leads to a fundamentally different game structure, in which there is an observable move by a strategic opposition party. Finally, various arguments that have been made about the transparency of democratic polities (Schultz 1999; Gaubatz 1998) suggest that regime type influences the variance of the unobserved shocks.

A final concern revolves around the structure of the data needed for this approach. We assume that we can identify discrete "plays" of the game, which form our unit of observation. If a threat is issued and a clear outcome reached, it is not hard to aggregate those events into a single play of the game. Cases in which the outcome is SQ, however, are harder to observe. If nothing happens in a given year between two states, should we consider that one play with an SQ outcome? Two? Or none—on the grounds that there was no opportunity for interaction? Is it even sensible to use a unit of time to proxy for a play of the game, and, if so, why is a year better than, say, a decade or a month? The basic problem is that we cannot observe opportunities to play the game in those cases in which the opportunities did not lead to a challenge. This means that the frequency of SQ outcome is hard to determine. Note that this problem is not unique to the methodology we employ here, and others have grappled with the same issue (see, e.g., Lemke and Reed 2001; Huth and Allee 2002, 23-24; Clarke and Reed 2003). Nonetheless, the nature of our method, focusing as it does on discrete plays of a repeated game, makes the issue transparent. We will revisit this issue when discussing construction of the data set.

For all of these reasons, we approach this project gingerly, aware of its potential benefits but also wary of its limitations. We are not suggesting that everyone in IR should stop doing probits and embrace fully structural models; rather, we suggest that this method is sufficiently intriguing that it is worth exploring whether it can be done and how it can be done right.

4 Description of Data

4.1 Crisis outcomes

The primary challenge in applying the estimator to real data is to gather systematic information about the empirical frequency of the different outcomes to the game. While several data sets on international crises already exist, none provides the codings required for this test, which made it necessary to perform additional data collection.

To this end, we drew on two data sets already widely used in the study of international conflict: the Correlates of War project's Militarized Interstate Dispute (MID) data (Gochman and Maoz 1984; Jones, Bremer, and Singer 1996) and the International Conflict Behavior (ICB) data (Brecher and Wilkenfeld 1997). Both data sets have appealing features, as well as limitations. The MID data set has unparalleled spatial and temporal coverage (over 150 countries from 1816-2001) and a very low threshold for inclusion. Any militarized action between two states that entails at the least a threat to use force qualifies as a MID. Unfortunately, for our purposes, the MID codings are static: they report the levels of hostility that the actors eventually reached, but say nothing about the path of play. There is no reliable way to map the outcome codings provided onto our game. In addition, the over two thousand cases in the MID data set are not well documented, making any effort to augment the information about these cases quite difficult. The ICB data set is well documented, provides richer information about the course and outcome of the event, and contains only 434 cases. This smaller size is only partly due to a more modest temporal domain (1919–2001) and mainly reflects a more restrictive inclusion criterion. An ICB crisis requires that decision makers perceive a threat to basic values, a compression of time, and heightened danger of violence.

Our data set was built by first going through all of 434 ICB cases and coding them according to rules which will be discussed shortly. To determine whether there was additional important information in the MIDs, we examined all MID cases in the interwar period, 1919-1939. Any MIDs that were not already covered by an ICB case were then coded according to the same rules. As we will see, the distribution of outcomes varied quite dramatically depending on whether the case came from ICB or MID. For this reason, we only feel confident in applying the estimator to the combined data from the interwar period.

Three variables were coded for each case: (1) the identity of state A, (2) the identity of state B, and (3) the outcome of the interaction. The most difficult coding decision is generally the first one. Given that status quo is generally disputed, and the question of "who started it" can be quite subjective, we have tried to develop an objective coding rule that best captures the assumptions of the model. In particular, we define a challenge as:

> Any act that is made deliberately by a central state authority with the intent of altering the pre-crisis relationship between itself and at least one other state and that is backed by the threat of military force. The threat of force may take the form of a diplomatic ultimatum, a show of force, or a limited use of force.

Given this coding rule, incidents of violence caused by non-state actors (such as rebel groups or assassins) and spontaneous, low-level troop clashes are not considered challenges, though they may provide an opportunity for a challenge to arise.

We have found that this coding instruction permits a reasonably straightforward identification of challenging acts. In addition, it deals with the problem that not all crises or militarized disputes are cases of coercive diplomacy as captured in the model. For example, a large number of cases in both the MID and ICB data sets take the form of a clash between troops patrolling a disputed border. Such an event constitutes a MID because force has been used and an ICB crisis if the fear of war has increased. For our purposes, however, this event alone generally does not constitute a challenge, though it may trigger one. If one state seizes on the clash as the pretext to make an ultimatum designed to settle the border dispute, then this ultimatum is the challenging act, and a case of coercive diplomacy follows from there.

If, on the other hand, the clash simply peters out, and the states do not pursue the matter through any additional threats of violence, then no coercive diplomacy has taken place. We have labelled such cases SQ outcomes in terms of the model. This coding reflects the fact that, in some inter-state relations, the status quo involves occasional clashes as troops patrol poorly marked or contested borders.

This coding rule also avoids the subjective issue of which state was the aggressor or instigator in the crisis. A provocative act that does not entail a threat of force is not considered a challenge—though, again, such an act may trigger a challenge if others seek to reverse it.[4] This decision is appropriate given that the theory's substantive focus is on threats of force used to change other states' behavior, and not on the broader class of provocative actions.

Once the challenge has been identified, the identities of the challenger and the target follow immediately. The outcome is then coded according to whether the target satisfied the challenger by making some material concession to its demands (ACQ), the target made no concession but the challenger failed to carry out the threat (BD), or the challenger implemented its threat (SF). In general, coding the outcome once the challenge was identified was relatively straightforward. Some ambiguity can arise if the target makes only a partial concession to the challenger's demands. In those cases, the outcome was coded as ACQ on the grounds that that challenger was materially better off as a result of having made its threat.

After applying these coding rules to ICB and MID cases in the period 1919-1939, we identified 71 cases in which a challenge was issue; as some crises involve more than one dyad, these cases encompass 101 crisis dyads. The list of these cases, including our coding of the challenger, the target, and the outcome, is given in the appendix [XX APPENDIX FORTHCOMING]. Note that there are fewer crises than there are distinct MIDs and ICB

[4]For example, ICB crisis no. 203 was triggered by Israel's decision to open a water pipeline to the Negev from Lake Tiberias. Arab states considered, but then rejected, threatening force to end the project (see Brecher and Wilkenfeld 1997, 279-80). Though Israel's action was provocative, the case involved no coercive diplomacy. Had the Arab states made a threat of force to close the pipeline, they would have been coded as the challengers.

Distribution of Outcomes in Crisis Dyads, 1919–1939

Coded outcome	ICB	MID	Total
Acquiescence(ACQ)	28	15	43
Back Down (BD)	5	9	9
Stand Firm (SF)	40	4	44

Table 1: *Distribution of outcomes among dyads in which a challenge took place.*

crises in this period, since a number of those events were coded not entailing a militarized challenge by our definition. [5]

The observed frequency of outcomes in these cases is shown in Table 1. The table also breaks down the outcomes according to whether the case came from ICB (with or without an associated MID) or from MID. The difference in the distribution of outcomes is striking. In particular, the cases that came only from MID account for a large portion of the ACQ and BD outcomes in our data set. This is consistent with Hewitt's (2003) finding that the ICB criteria tend to select those interactions that are most severe, so that cases in which the target quickly gave in or in which the challenger backed away from the threat can fall below its radar screen. Because of this observation, it would be misleading to apply our estimator to a data set based only on the ICB cases, since there is good reason to believe that they do not capture the true distribution of outcomes. Thus, we must restrict our attention to the interwar period until the coding of MIDs for the remaining years can be completed.

4.2 Dealing with SQ Outcomes

As mentioned earlier, identifying the population of SQ cases is not a trivial matter, especially when these make up the vast majority of outcomes: most countries are at peace most of the time. How should determine which observations belong in the estimation sample? If one uses a loose criterion for inclusion (e.g., every dyad year or even every "politically relevant" dyad year), then the crisis cases make up an exceedingly small proportion of the data set. As a result, all of the action is in the extreme tails of the probability distributions,

[5]There were also a small number of MIDs about which we could not find sufficient evidence to determine whether or not a crisis, by our definition, took place.

where numerical approximations of the cumulative normal become less reliable. In addition, admitting too many observations into this population risks introducing unmeasured heterogeneity, as observations with very little interaction or grounds for conflict are pooled with those that are more conflict prone (Green, Kim, and Yoon 2000). On the other hand, of course, overly restrictive criteria for inclusion risks deleting valuable information from the data set.

For the purposes of this paper, we use a modified version of the "politically relevant" dyad restriction. Recall that two states are politically relevant, as defined by Russett (1993, XX), if they are territorially contiguous or if at least one of them is a major power. This second restriction assumes that a major power is politically relevant to all other states in the system. Our more limited definition includes the first criterion but imposes a tighter version of the second. In particular, we assume that major powers are politically relevant to (1) other major powers and (2) to all other states in their geographic region. For the purposes of this coding, the United States' region is the Americas (COW country codes 2-199), the region of Britain, France, Italy, and Germany is Europe (COW country codes 200-399), Japan's region is Asia (COW country codes 700-899), and Russia sits in both Europe and Asia.[6]. Every crisis dyad in our data set meets this definition of political relevance.

The second step we take to reduce the overwhelming preponderance *SQ* outcomes is collapse the 21-year time period covered by the data into seven 3-year periods. If a dyad experiences no challenges in a given 3-year period, this is coded as one observation of an *SQ* outcome. In such cases, we cannot not know who the potential challenger was. To deal with this problem, we assume that, in every period, each state in the dyad had an equal chance of being the challenger. Hence, in each period with no challenge, one state in the dyad is randomly selected to be state *A*. If the dyad experiences one or more challenges in the given

[6]We recognize most major powers have active interests outside their geographic region, generally due to colonial possessions. To capture the potential conflicts involving major powers outside of their home region, our definition of contiguity takes into account contiguity arising through colonies or other dependencies. Thus, for example, Britain is politically relevant to Brazil through its possession of British Guyana, and it politically relevant to China through its possession of India. Direct and colonial/dependency contiguity data were provided by the COW 2 project.

period, then the ensuing crises are all included. For SQ outcomes, covariates were converted into three-year averages.

After implementing these inclusion rules, the resulting data set has 1873 SQ outcomes, in addition to the 101 crisis dyads mentioned above. Hence, crises are still quite rare, accounting for only 5 percent of the observations in the data set.

5 Covariates and Specification

The next step is to specify which covariates should be included in the analysis. With this approach, we need to consider not only which independent variables matter but also which payoff(s) they should influence: model specification requires a list of covariates for each outcome utility for each player. These decisions are by no means trivial, since identification problems are endemic to efforts to estimate preferences from observed choices.

5.1 Identification Issues

The main identification issues arise from the nature of the von Neumann-Morgenstern utilities which underly the game theoretic model. It is well-known, for example, that adding the same quantity to every payoff for a given player will have no impact on the outcome of the game. Similarly, changing the scale of the payoffs such as by multiplying them all by the same amount has no impact on the outcome probabilities. In In Lewis and Schultz (2003),we considered the implications of these considerations for identification in our estimator and developed three results that must guide specification decisions:

- no covariate can appear in every payoff for a given state;
- a constant must be dropped from at least one payoff for each state; and
- nothing can be inferred about SQ_B since it has no effect on the outcome probabilities.

Hence, in deciding which covariates belong in the specification for each payoff, we have to exclude each covariate from at least one of the payoff from each player (not including SQ_B). Furthermore, at least one payoff for each player cannot include a constant term.

To implement these restrictions, we fix SQ_A and BD_B to zero by including neither constants nor covariates in the specification of these payoffs. This normalizes the utility scale for each player such that the utility associated with the B's possession of the good is zero. This assumption entails no loss of generality, and it means that all other payoffs are estimated relative to the base category for each actor. [7] In practice, then, covariates that are thought to influence the challenger's assessment of the status quo should generally appear only in the specification of the ACQ_A and SF_A payoffs. Since backing down implies a return to the status quo, any such covariates can reasonably be excluded from BD_A. The BD_A payoff should only include those covariates that are thought to influence the costs or benefits of backing down from a challenge, holding aside the value of the status quo that has been retained.

One additional restriction is warranted: we constrain the constant associated with ACQ_A to be greater than zero and the constant associated with ACQ_B to be less than zero.[8] We then demean all of the covariates so that the constants reflect the average payoffs across the sample. Substantively, then, the constraints ensure that, on average, the good that is being contested is in a fact a good, in the sense that A would like to have it, and B would rather not give it up. Hence, A prefers B's acquiesence over the status quo ($ACQ_A > SQ_A$) and B prefers A's backing down to acquiesence ($BD_B > ACQ_B$). While this restriction entails a loss of generality, it is sensible given the assumptions of the game.

Even after dealing with these basic identification issues, there remains a danger of empirical under-identification. That is, even though parameters are theoretically identified, it may in practice be difficult to estimate them because of insufficient data. This problem is particularly troublesome in strategic settings, where a change in the probability of a given outcome can be the result of a change in any number of payoffs. For example, when we find that being involved in a civil war makes a target more likely to resist a challenge, it can be difficult to know whether this is because being in a civil war makes acquiescing worse

[7]The units of the utility scales for each player can also be fixed without loss of generality by fixing variance of the payoff shocks (in this case to unity as noted above).

[8]These constraints were imposed by including penalties in the likelihood function for parameters that are out of the specified bounds.

or fighting easier, or both. In principle, one can estimate the effects of a civil war on both ACQ_B and SF_B, as long as this variable is excluded from BD_B. In practice, however, these parameters will be highly correlated, which makes identification difficult, particularly in small data sets. We will provide some examples of this problem below.

5.2 Covariates

The existing literature suggests a number of variables that should appear in the specification of the payoffs. Given the small numbers of observations of non-SQ outcomes, we found it necessary to keep the specifications relatively simple. It is hoped that, with a more complete data set, richer specifications will be possible. What follows is a discussion of the covariates we used and the payoffs in which they were included.

Relative Capabilities. Using the Correlates of War's measure of national material capabilities (Singer and Small 1993), we constructed a variable indicating state A's share of capabilities in the dyad. The resulting variable, $Capshare_A$, goes from zero to one. Initially, this variable was included in the specification of four payoffs: ACQ_A, SF_A, and SF_B. The inclusion of relative capabilities in the payoffs for the "war" outcome needs little justification, since threat implementation by A leads to a clash of military forces. The inclusion of $Capshare_A$ in A's payoff for acquiescence stems from the assumption that, as the share of A's power increases, so too does its ability to obtain changes in the status quo without resorting to threats of force. This is consistent with a longstanding argument that a preponderance of power can diminish military conflict by leading peaceful changes in the status quo to the benefit of the stronger state (e.g., Organski 1968). Put another way, our expectation is that the distribution of goods is generally correlated with the distribution of power (Gilpin 1981). If this is true, then $Capshare_A$ is positively correlated with the status quo evaluation and hence negatively correlated with the value of acquiescence.

Regime Type. Using the Polity III data set (Jaggers and Gurr 1996), dummy variables were created indicating whether or not state A and B were democratic at the time of the crisis. This was done by subtracting the 10-point autocracy score from the 10-point democracy scale, leading to a combine scale running from -10 to 10. The variables Dem_A and Dem_B

were set to one if the country's score on this scale was greater than 5.

We include Dem_A in the specification of BD_A to test Fearon's (1994) argument that democratic governments incur larger audience costs, on average, from backing down. We also include the regime type variables in each state's payoff from war to test various theories about how regime type influences governments' assessments of war. Thus, Dem_A appears in the specification for SF_A, and Dem_B appears in the specification for SF_B. The theoretical literature is inconclusive regarding the expected effects of these variables. On the one hand, a long tradition suggests that democratic leaders consider war to be more costly, ceteris paribus, than do their nondemocratic counterparts, due to institutions of accountability (e.g., Kant 1983 [1795]; Morgan and Campbell 1991; Bueno de Mesquita and Lalman 1992). On the other hand, some have argued that democratic states are better at fighting wars, due to advantages in resource mobilization and troop effectiveness (e.g., Reiter and Stam 2002; Lake 1992). As these arguments are not mutually exclusive, the expected impact of regime type on the value for war is not only contested but also ambiguous.

The vast literature on the democratic peace (e.g., Doyle 1986; Russett 1993; Russett and Oneal 2001) suggests that we should also include a dummy variable indicating whether or not both states in the dyad are democratic. In our data set, however, there is only one instance in which democratic states were engaged in a crisis: the 1919 Teschen dispute between Poland and Czechoslovakia (which, incidentally, is coded as a SF outcome even though the fighting did not reach the requirements for a full-scale war). Consequently, we are pessimistic about the ability of our estimator to determine the effects this variable on the payoffs.

Territorial Contiguity. Using data from the Correlates of War, a dummy variable, $Contiguous$, was created indicating whether or not the states in the dyad were territorially contiguous. Following convention, contiguous states either share a land border or are separated by less than 150 miles of water. Both direct and colonial contiguity are captured by this variable. Contiguity is known to have several effects on crisis behavior (e.g., Diehl 1991). The distance between states influences their ability to mobilize military force against one another. A shared border may also provide a source of conflict, as a large proportion of militarized disputes take place over the control of territory (e.g., Huth 1996). Because we have already restricted the sample to a subset of politically relevant dyads, we are already focusing on states that can presumably project power against one another. Hence, we are primarily interested in the effect of contiguity on the potential stakes of a dispute, ACQ_A and ACQ_B.

Civil War in Target. Civil wars provide opportunities for foreign states to press demands on a government, either for significant pieces of territory or for a change of government. Empirically, we observe that a number of interstate disputes are associated with civil wars in the target country: e.g., Allied intervention in the Russian Civil War, German and Italian intervention in the Spanish Civil War, and Japanese predations against China during that country's spells of civil conflict. Hence, using the COW data on civil wars, we create a variable $CivilWar_B$ indicating whether or not the target state in the dyad was involved in a civil war.

Historical Rivalry. Dyads that have a history of conflict are very likely to experience conflict in the future (Crescenzi and Enterline 2001; Goertz and Diehl 2000; Beck, Katz, and Tucker 1998). A history of repeated hostile interactions presumably indicates that the states have a conflict over a relatively high value good. A condition of ongoing rivalry might also make concessions to the opponent both materially and politically costly.

The literature on rivalries has produced several different attempts to code interstate rivalries (e.g., Diehl and Goertz 2000; Bennett 1998), and we have no desire to step into this controversy. It is important for our purposes, though, that we have an ex ante indicator of rivalry: that is, we want an indicator of rivalry that can be measured independently of whether or not there is a crisis in the current period. This is a nontrivial issue because most efforts to code interstate rivalries are ex post: once an entire history of conflict is observed, the rivalry and its start date are identified. For this reason, Bennett's (1998) treatment of rivalries is best suited for our purposes. Bennett notes that, to be classified as rivals according to one common coding rule, a pair of states must experience at least six MIDs within at least 20 years, with a gap of no more than 15 years between disputes. Hence, an ex ante indicator should code states as rivals only after they have experienced both 20 years of conflict and 6 MIDs. Bennett (1998) provides a coding to this effect, determining for each rivalry the year in which this threshold was met. Based on this, we create a dummy variable, $Rivalry$, indicating whether a given dyad was in an identifiable rivalry at the start of a given period.[9]

Similarity of Strategic Interests. The final covariate is a measure of the similarity of the states' strategic interests, as captured in the similarity of their alliance portfolios (Bueno de Mesquita 1975). The indicator we employ here is the tau-b score using alliances in the dyad's region, a measure which goes from -1 to 1, with increasing scores reflecting more similar portfolios.[10]

6 Estimation and Results

6.1 Initial Results

A large number of models with different specifications were estimated. In some cases, the estimator failed to converge, quite possibly because of issues of empirical under-identification. Table 2 presents the estimates from two specifications. The first column shows the results from a "full" specification, in which the covariates were included liberally in a variety of payoffs. In the second column, some of variables with insignificant coefficients in the first model were dropped. In particular, *Contiguity* and *Alliance* were taken out of the specification of the SF_A and SF_B payoffs.[11] The χ^2 statistic reported at the bottom compares the full model including covariates with a constant-only model. In both cases, the null hypothesis that the estimated coefficients are all zero is clearly rejected.

Since the constant in ACQ_A and ACQ_B were constrained to have the signs reported, there are only three constants to consider, those in BD_A, SF_A, and SF_B. These constants are all negative and, except for the last of these, statistically different from zero. This

[9]In only a small number of cases did a dyad met the rivalry criteria during the window of observation, 1919-1939, In these cases, *Rivalry* is coded as one only in years after teh criteria was met. If a rivalry was coded as ending during the interwar period, we continue to code the dyad as having a history of rivalry. This is because the coding of rivalry termination is always ex post: we only know that a rivalry has ended if there is a sufficiently long period without conflict.

[10]We note that Signorino and Ritter (1999) propose an alternative measure of alliance portfolio similarity based on the S-score. We conducted tests using this alternative measure and found no effects.

[11]A likelihood restriction test cannot reject the null hypothesis that the four parameters eliminated were jointly equal to zero.

Basic Results from the Crisis Bargaining Model

		(1) Est.	(1) (SE)	(2) Est.	(2) (SE)
SQ_A					
	Constant	0.00		0.00	
ACQ_A					
	Constant	2.15	(1.42)	1.62	(1.26)
	Alliance	-1.27	(0.96)	-0.08	(0.81)
	Capshare A	-1.73	(1.06)	-0.96	(0.66)
	Civilwar B	2.24	(1.30)	2.52	(1.30)
	Contig	1.62	(1.05)	1.15	(0.82)
	Rivalry	0.78	(0.72)	1.73	(0.96)
ACQ_B					
	Constant	-1.81	(1.51)	-2.25	(1.50)
	Alliance	-0.64	(1.25)	-0.21	(0.19)
	$Civilwar_B$	0.01	(0.14)	-0.01	(0.15)
	Contig	-0.32	(0.99)	-0.28	(0.13)
	Rivalry	0.22	(0.19)	0.11	(0.17)
BD_A					
	Constant	-4.06	(1.26)	-3.80	(1.16)
	Dem_A	-0.66	(0.43)	-0.78	(0.41)
BD_B					
	Constant	0.00		0.00	
SF_A					
	Constant	-3.65	(1.24)	-3.40	(1.15)
	Alliance	-0.07	(0.24)		
	$Capshare_A$	0.78	(0.25)	0.75	(0.23)
	Contig	0.08	(0.17)		
	Dem_A	-0.67	(0.39)	-0.78	(0.38)
SF_B					
	Constant	-2.51	(2.12)	-3.22	(2.15)
	Alliance	-1.14	(1.58)		
	$Capshare_A$	0.86	(0.82)	1.40	(0.83)
	Contig	0.19	(1.36)		
	Dem_B	0.01	(0.07)	0.01	(0.08)
N		1,974		1,974	
Log likelihood		-397.60		-398.30	

Table 2: *Estimates from crisis bargaining model applied to interwar conflict data. Parameters without standard errors are constrained to equal the value shown.*

suggests that, on average, both war and backing down are less desirable to A than the status quo. War also appears to be less desirable to B than having the status quo restored, but this effect is not statistically different from zero at conventional levels. The negative constant on BD_A confirms that there are, on average at least, costs from backing down from a challenge. This is not particularly surprising given the scarcity of challenges in the data set: if it were generally costless to issue challenges, we would see many more. This result does not, however, tell us whether the costs from domestic audiences (Fearon 1994; Smith 1998) or international audiences (Sartori 2002).

There is some evidence that democratic leaders face higher costs from backing down than do their nondemocratic counterparts, as suggested by Fearon (1994). The effect of Dem_A on BD_A is negative, and significant at the 10 percent level. However, any benefit democratic leaders might have in generating "audience costs" is offset by the fact that democratic challenges are estimated to have lower values for war on average. The coefficient on Dem_A in the equation for SF_A is negative and statistically significant, providing evidence in favor of the view that democratic leaders face higher costs for war. By contrast, the coefficient on Dem_B in the equation for SF_B is essentially zero. While this result might be the product of insufficient data, it is also plausible that the domestic constraints on democratic leaders bite harder when they are the challenger in the dispute than when they are they are the target.

State A's share of dyadic capabilities has the expected positive effects on its payoff from war (SF_A). $Capshare_A$ is also negatively associated with A's payoff from acquiescence, consistent with the conjecture that more powerful states do well in the status quo and so have less to gain by making militarized demands. This effect is not, however, statistically significant at conventional levels. The coefficient on $Capshare_A$ in B's expected payoff from war (SF_B) is unexpectedly positive, though its significance is borderline. This is a an observation which we will explore further below.

The variable indicating territorial contiguity has a positive effect on ACQ_A and a negative effect on ACQ_B, though only the latter coefficient is statistically significant. All other things equal, a shared border makes acquiescence less attractive to B. These results are consistent with the hypothesis that contiguity increases the stakes of a potential dispute. Two other indicators for the potential stakes of a dispute–participation in an ongoing rivalry and the presence of a civil war in the target–have positive and significant impact on ACQ_A, as expected. Neither variables, however, has a measurable affect on ACQ_B. The coefficients on the alliance portfolio similarity have no statistically significant effect.

6.2 Additional Explorations

Of the findings reported above, two call out for further exploration. First, the estimated effects of democracy in the challenger are striking, not only because they are in line with several existing theories, but also because, if true, they demonstrate the value of the empirical technique. Without a fully structural model, it would be quite difficult to estimate the effects of democracy on two payoffs separately. Second, the positive association between $Capshare_A$ and B's payoff from "war," is unexpected.

Though intriguing, the estimated effects on democracy on BD_A and SF_A are also suspicious, precisely because they are almost exactly equal. This means that, compared to nondemocracies, democratic challengers are no more likely to stand firm or back down given a challenge, nor are the targets of democratic challenges less likely to resist. Rather, this effect must be driven by the fact that democracies were less likely to issue challenges in the first place.[12] Thus, before accepting these results at face value, we should consider two other possible explanations for the lower rate of democratic challenges. The first is that democracies face higher costs for making a challenge, regardless of the ensuing outcome. In this case, democracy in A would also diminish the value of B's acquiescence.

The second possibility is that democracy is associated with something else that would cause a lower rate of challenges. In the interwar period, it is quite plausible that the democratic states were more satisfied with the status quo, given that the most prominent democracies—Great Britain, France, and the United States—had a large hand in establishing the post-war order. To the extent that power explains satisfaction, we have already controlled for this by including $Capshare_A$ in A's payoffs. An additional control that we could include is suggested by Kim (1991), who follows Gilpin (1983) in arguing that satisfac-

[12] All of these implications can be confirmed through standard analyses, using probit models or cross tabulations.

tion increases with a state's ties to the leading state in the international system. Kim (1991) thus proposes that a state's evaluation of the status quo is indicated by the similarity of its alliance portfolio with that of the leading state, which is generally taken to be Great Britain in this period. Alliance portfolio similarity between state A and Britain is measured using the method discussed earlier, leading to a new variable, Tau_{UK}. Using this method, Great Britain is, by assumption, the most satisfied state in the system. The correlation between Tau_{UK} and Dem_A is 0.38, confirming that democracies in this period were relatively more satisfied.

Table 3 shows the estimates of several models which explore these alternatives. The specification in column (1) is the same is in Table 2, column (2), but Dem_A has been added to the equation for ACQ_A. The new parameter is negative, implying that democratic challengers value acquiescence less than do nondemocratic challengers, consistent with the idea that it is costlier for the former to make militarized threats. This coefficient, however, is not significantly different from zero, and its inclusion renders the coefficients on Dem_A in the other payoffs insignificant as well. As this exercise shows, including the same variable in more than one payoff can be similar to putting two variables that are highly correlated in the same regression model: each washes out the effect of the other. Here, this phenomenon is driven by the correlation in the parameters across equations. Because decreasing ACQ_A has a similar effect on the observable outcomes as decreasing SF_A and BD_A simultaneously, the estimator cannot easily figure out which payoffs are changing.

In column (2) of Table 3, the democracy indicator is only in the equation for ACQ_A, constraining its effect on BD_A and SF_A to be zero. Not surprisingly, democratic challengers now appear to have a significantly lower value for acquiescence. Notice that the log-likelihoods of the last three models are all virtually the same. This means we cannot reject the null that Dem_A has no effect on ACQ_A, nor can we reject the null that Dem_A has no effect on SF_A and BD_A jointly. This set of results casts doubt on the conjecture that democracy increases the sunk cost of making a challenge, thereby lowering the value of all three non-SQ payoffs. However, eliminating Dem_A entirely (estimates not reported), generates a log-likelihood of -407.6, so a likelihood ratio test easily rejects that null that Dem_A has no effect whatsoever.

Is this effect is spurious, a product of the fact that democratic states in this period were

The Effect of Democracy: Alternative Specifications

		(1)		(2)		(3)	
		Est.	(SE)	Est.	(SE)	Est.	(SE)
SQ_A							
	Constant	0.00		0.00		0.00	
ACQ_A							
	Constant	1.49	(1.26)	1.41	(1.24)	1.43	(1.16)
	Alliance	-0.09	(0.76)	-0.08	(0.73)	0.53	(1.00)
	$Capshare_A$	-0.81	(0.73)	-0.74	(0.71)	-0.80	(0.66)
	$Civilwar_B$	2.55	(1.22)	2.58	(1.15)	2.48	(1.37)
	Contig	1.07	(0.72)	1.02	(0.66)	1.02	(0.82)
	Dem_A	-0.40	(0.75)	-0.74	(0.32)		
	Rivalry	1.68	(0.87)	1.70	(0.75)	2.48	(1.30)
	Tau_{uk}					-0.41	(0.40)
ACQ_B							
	Constant	-1.79	(1.34)	-1.55	(1.07)	-3.27	(1.61)
	Alliance	-0.19	(0.17)	-0.18	(0.16)	-0.38	(0.23)
	$Civilwar_B$	-0.04	(0.13)	-0.07	(0.10)	0.05	(0.16)
	Contig	-0.25	(0.12)	-0.23	(0.11)	-0.30	(0.14)
	Rivalry	0.08	(0.15)	0.06	(0.14)	0.10	(0.18)
BD_A							
	Constant	-3.67	(1.16)	-3.60	(1.16)	-3.65	(1.16)
	Dem_A	-0.37	(0.88)			-0.54	(0.47)
BD_B							
	Constant	0.00		0.00		0.00	
SF_A							
	Constant	-3.27	(1.15)	-3.20	(1.14)	-3.27	(1.15)
	$Capshare_A$	0.81	(0.26)	0.85	(0.24)	0.67	(0.20)
	Dem_A	-0.36	(0.83)			-0.57	(0.43)
SF_B							
	Constant	-2.56	(1.93)	-2.22	(1.54)	-4.73	(2.32)
	$Capshare_A$	1.27	(0.80)	1.19	(0.73)	1.79	(0.86)
	Dem_B	0.00	(0.07)	-0.01	(0.06)	0.02	(0.09)
N		1,974		1,974		1,974	
Log likelihood		-398.20		-398.30		-397.40	

Table 3: *Estimates from crisis bargaining model applied to interwar conflict data. Parameters without standard errors are constrained to equal the value shown.*

simply more satisfied with the status quo? Column (3) of Table 3 reports the estimates of a model in which Tau_{UK} is added to the equation for ACQ_A. As expected, the coefficient on this variable is negative—meaning that more satisfied states had less to gain from making a demand—but it is not statistically significant. Including this variable attenuates the effects of Dem_A on SF_A and BD_A, so that they are no longer individually significant; however, a likelihood ratio test easily rejects the null hypothesis that these coefficients are jointly equal to zero.[13] Hence, the original effect holds up even with this additional control for the challenger's status quo evaluation. Of course, this control is by no means perfect, and further work is needed to identify good indicators of this concept. For now, though, the evidence is consistent with the hypotheses that democratic challengers experienced higher costs from backing down and higher costs from going to war.

We now turn to further consideration of the unexpected finding that A's share of military capabilities is positively associated with B's value from the SF outcome. This positive effect appears in all of the models reported so far, though its statistical significance is generally borderline by conventional standards. Since the result seems implausible at face value, it is worth considering what is going on. One possibility is that this result is a product of the fact that powerful states do not issue enough challenges against weak states given their military advantages. By including $Capshare_A$ in the equation for ACQ_A, we tried to account for the fact that relatively more powerful states do not need to make explicit threats to get their way, meaning that they have less to gain from acquiesence. Nonetheless, it is possible that the estimator is also attributing high values of war to weak targets in order to explain why strong challengers seem to be deterred from making more threats. If so, then the result could be an artifact of how the sample was created; in particular, by including every pairing between a major power and every state in its geographic region, we have probably included many highly asymmetrical dyads in which there is no grounds for conflict. Dealing with this requires better ex ante indicators of the potential for conflict between two states.

A second possibility is that the capability variable, by only measuring the dyadic balance of power, does not take into account alliances which effectively augment the power of weaker

[13] The log-likelihood of the restricted model is -403.7, generating an LR statistic of 12.6.

states. To test for this possibility, additional specifications were employed that included in the equation for SF_B a variable indicating whether or not the target had a major power ally. This variable had no impact on the results.

A final possibility to consider is that weaker targets actually do resist challengers more than would be expected, perhaps because the stakes are higher for them. We observe, for example, that the data set contains a number of wars in highly asymmetrical dyads: e.g., Italy-Ethiopia, Japan-Mongolia. It is possible that weakness invites larger demands, and the weaker states resist in large part because it is costly for them to acquiesce to stronger predators. To test this possibility, we put $Capshare_A$ in the equation for ACQ_B. Table 4 reports the estimates from two specifications, one in which $Capshare_A$ appears in both of B's payoffs (column 1), and one in which it appears only in ACQ_B (column 2).

Not surprisingly, when is included in both of B's payoffs, neither coefficient is individually significant. When it is included only in ACQ_B, the coefficient is significant at the 10 percent level. Nonetheless, a restricted model in which $Capshare_A$ is excluded from both of B's payoffs has a log-likelihood of -402.0. Hence, a likelihood ratio test easily rejects the null hypothesis that the effects are jointly zero. In both models, the negative coefficients in the equation for ACQ_B is as expected: the more powerful the challenger, the less attractive it is for the target to acquiesce to its demands. This is consistent with the conjecture that more powerful states make larger demands. It is also worth noting that the utilities that we are estimating are essentially continuation values; that is, they indicate not only the value of the current outcome, but also the expected value of the future conditional on having reached that outcome in the current play of the game. Hence, it is possible that acquiescence to a more powerful state is particularly unattractive because of its expected impact on future interactions as well.

While this is a plausible explanation for the unexpected effect of $Capshare_A$, it is by no means definitive. Moreover, even if this interpretation is true, it mostly points to the need for better ex ante indicators of the kinds of demands challengers are likely to make.

The Effect of Capabilities: Alternative Specifications

		(1) Est.	(1) (SE)	(2) Est.	(2) (SE)
SQ_A					
	Constant	0.00		0.00	
ACQ_A					
	Constant	1.54	(1.23)	1.62	(1.28)
	Alliance	0.13	(0.91)	0.30	(0.99)
	$Capshare_A$	-1.16	(0.77)	-1.37	(0.75)
	$Civilwar_B$	2.53	(1.34)	2.53	(1.36)
	Contig	1.01	(0.74)	0.99	(0.71)
	Rivalry	1.88	(1.06)	1.89	(1.02)
ACQ_B					
	Constant	-2.63	(1.56)	-2.69	(1.54)
	Alliance	-0.27	(0.20)	-0.29	(0.20)
	$Capshare_A$	-0.80	(1.10)	-1.21	(0.72)
	$Civilwar_B$	0.01	(0.16)	0.02	(0.15)
	Contig	-0.25	(0.13)	-0.24	(0.11)
	Rivalry	0.13	(0.18)	0.14	(0.17)
BD_A					
	Constant	-3.75	(1.16)	-3.83	(1.22)
	Dem_A	-0.83	(0.44)	-0.86	(0.46)
BD_B					
	Constant	0.00		0.00	
SF_A					
	Constant	-3.32	(1.15)	-3.37	(1.20)
	$Capshare_A$	0.87	(0.30)	0.92	(0.26)
	Dem_A	-0.82	(0.42)	-0.85	(0.44)
SF_B					
	Constant	-3.65	(2.15)	-3.65	(2.09)
	$Capshare_A$	0.56	(1.22)		
	Dem_B	0.00	(0.08)	0.00	(0.08)
N		1,974		1,974	
Log likelihood		-397.80		-398.00	

Table 4: *Estimates from crisis bargaining model applied to interwar conflict data. Parameters without standard errors are constrained to equal the value shown.*

7 Conclusion

The results in this paper are obviously preliminary. As more data become available, we expect to be able to test richer specifications with more covariates identified by the literature. Nonetheless, the exercise has been instructive in illuminating both the promise and limitations of this kind of technique.

Fully structural empirical estimators are not a cure-all for the problems that arise analyzing observational data from strategic situations. While there is an appeal to designing empirical tests that faithfully capture the underlying theoretical model, it must be recognized that this method comes with costs. Imposing so much structure on the empirical model is useful only to the degree that the structure is itself valid. Remember: the exercise here was not to test a game theoretical model, but rather to estimate the unobserved utilities *under the assumption that the model is correct*. Hence, a substantial leap of faith is involved. Beyond that, as we have seen, there are considerable issues of identification that have to be taken into account. While identification can be achieved through relatively innocuous restrictions in theory, empirical under-identification can remain a problem, particularly given the correlation in the parameters.

That said, the results here are promising in several ways. We have shown that it is possible to construct an estimator directly from a game theoretic model, to capture in that estimator the informational assumptions that are common in the literature on crisis bargaining, and to apply that model to real observational data. The fact that some of the results are not surprising—e.g., that more powerful states have higher expected value for war—is reassuring at this early stage. Moreover, the findings regarding the challenger's regime type are intriguing. They not only speak to a debate in the literature, but they also demonstrate the potential value of estimating the effect of covariates on payoffs rather than on outcome probabilities. With standard techniques, it would be difficult to tease out the countervailing effects of regime type that we found. While the analysis of these results was admittedly inconclusive in some respects, they clearly point to the potential of this kind of work.

Finally, the results demonstrate that the path to advancement in this area lies not only in methodological innovation, but also in data collection. Applying a fully structural estimator requires data to be collected with the underlying strategic model in mind. The fact that existing data sets on international conflict were not motivated by an explicit strategic structure means that considerable work remains to be done.

References

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Beyond Ordinary Logit: Taking Time Seriously in Binary Time-Series Cross-Section Models. *American Journal of Political Science* 42: 1260–88.

Bennett, D. Scott. 1998. "Integrating and Testing Models of Rivalry Termination." *American Journal of Political Science* 2 (October): 1200-1232.

Brecher, Michael, and Jonathan Wilkenfeld. 1997. *A Study of Crisis.* Ann Arbor, MI: University of Michigan Press.

Bueno de Mesquita, Bruce. 1975. "Measuring Systemic Polarity," textitJournal of Conflict Resolution 19: 187–216.

Bueno de Mesquita, Bruce. 1981. *The War Trap.* New Haven: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason: Domestic and International Imperatives.* New Haven: Yale University Press.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. "An Institutional Theory of the Democratic Peace." *American Political Science Review* 93: 791–808.

Bueno de Mesquita, Bruce, and Randolph Siverson. 1995. 1995. War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability. *American Political Science Review* 89: 841–55.

Clark, David H., and Patrick M. Regan. 2003. "Opportunities to Fight: A Statistical Technique for Modeling Unobservable Phenomena." *Journal of Conflict Resolution* 47 (February): 94–114.

Crescenzi, Mark J. C., and Andrew J. Enterline. 2001. "Time Remembered: A Dynamic Model of Interstate Interaction." *International Studies Quarterly* 45(4):409431.

Diehl, Paul F. 1991. "Geography and War: A Review and Assessment of the Empirical Literature." *International Interactions* 17: 16–23.

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

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (Summer): 379–414.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (September): 577–92.

Frieden, Jeffry A. 1999. Actors and Preferences in International Relations. In David A. Lake and Robert Powell, eds., *Strategic Choice in International Relations.* Princeton: Princeton University Press. pp. 39–76.

Gaubatz, Kurt Taylor. 1999. *Elections and War.* Stanford, CA: Stanford University Press.

Gilpin, Robert. 1981. *War and Change in World Politics.* Cambridge: Cambridge University Press.

Gochman, Charles S., and Zeev Maoz. 1984. Militarized Interstate Disputes, 1816–1976: Procedures, Patterns, and Insights. *Journal of Conflict Resolution* 28 (December): 585–615.

Hewitt, J. Joseph. 2003. Dyadic Processes and International Crises. *Journal of Conflict Resolution* 47 (October): 669-692.

Huth, Paul K. 1996. *Standing Your Ground.*Ann Arbor, MI: University of Michigan Press.

Huth, Paul K., and Todd L. Allee. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century.* Cambridge: Cambridge University Press.

Jaggers, Keith, and Ted Robert Gurr. 1996. Polity III: Regime Change and Political Authority, 1800–1994 [computer file] (Study #6695). 2nd ICPSR ed. Boulder, CO: Keith Jaggers/College Park, MD: Ted Robert Gurr [producers], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1996.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns. *Conflict Management and Peace Science* 15(2): 163–213.

Kant, Immanuel. 1983 [1795]. "To Perpetual Peace: A Philosophical Sketch." In *Perpetual Peace and Other Essays on Politics, History, and Morals.* Indianapolis, IN: Hackett Publishing Company.

Lake, David. 1992. Powerful Pacifists: Democratic States and War. *American Political Science Review* 86: 24–37.

Lemke, Douglas, and William Reed. 2001. "The Relevance of Politically Relevant Dyads." *Journal of Conflict Resolution* 45: 126–44.

Lewis, Jeffrey B., and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11 (Fall): 345–67.

Morgan, T. Clifton, and Sally H. Campbell. 1991. "Domestic Structure, Decisional Constraints and War: So Why Kant Democracies Fight?" *Journal of Conflict Resolution* 35: 187–211.

Morrow, James D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33 (November): 941–72.

Mousseau, Michael. 2000. Market Prosperity, Democratic Consolidation, and Democratic Peace. *Journal of Conflict Resolution* 44:472–507.

Partell, Peter J., and Glenn Palmer. 1999. "Audience Costs and Interstate Crises: An Empirical Assessment of Fearons Model of Dispute Outcomes." *International Studies Quarterly* 43: 389–405.

Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War.* Princeton, NJ: Princeton University Press.

Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World.* Princeton, NJ: Princeton University Press.

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

Sartori, Anne. 2002.The Might of the Pen: A Reputational Theory of Communication in International Disputes.*International Organization* 56 (Winter): 123–51.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy.* Cambridge: Cambridge University Press.

Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform?:Contrasting Two Institutional Perspectives on Democracy and War," *International Organization* 52 (Spring): 233–66.

Signorino, Curtis S. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11 (Fall): 316–344.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(June): 279–98.

Singer, J. David, and Melvin Small. 1993. National Material Capabilities Data, 1816-1985 [computer file] (Study #9903). Ann Arbor, MI: J. David Singer, University of Michigan, and Detroit, MI: Melvin Small, Wayne State University [producers], 1990. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1993.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43(October): 1254–83.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(September): 623–38.

Smith, Alastair. 1996. "To Intervene or Not to Intervene: A Biased Decision." *Journal of Conflict Resolution* 40 (March): 16–40.

Vasquez, John A. 1993. *The War Puzzle.* New York: Cambridge University Press.