

Presidential Uses of Force During the Cold War: Aggregation, Truncation, and Temporal Dynamics

Sara McLaughlin Mitchell Florida State University
Will H. Moore Florida State University

This study identifies three methodological issues that affect inferences drawn in studies of presidential decisions to use force: aggregation, truncation, and dynamics. We suggest that a dichotomous measure of uses of force introduces aggregation bias, while the decision to examine only major uses of force introduces truncation bias. In addition, we argue that the presence of rivalry creates temporal dependence or dynamics in the use of force series. We reexamine the empirical findings reported in a seminal study of U.S. presidents' use of force during the Cold War (Ostrom and Job 1986). Our findings demonstrate the importance of these three methodological issues. Results of a Poisson Autoregressive (PAR) model show dynamics in the use of force series. Contrary to Ostrom and Job, we find that international variables have a larger substantive effect on the president's decision to use force than political variables like approval and domestic variables like economic performance.

In their path breaking study, Ostrom and Job (1986) develop a cybernetic decision theory of Cold War U.S. presidential decision making to explain when presidents use force against other countries. It stands apart from other theories because it argues that presidents make foreign policy decisions through the lens of domestic politics. More specifically, Ostrom and Job argue that presidents are motivated by a desire to retain power and that they make policy decisions with an eye toward that goal. Sixteen years later, that argument is not nearly as radical as it was then: IR scholars commonly assume that an executive's primary goal is to retain power. The study was perhaps most controversial for concluding that domestic political factors such as the president's standing in the polls and economic performance have a larger substantive effect on presidential decisions to use force than do international factors such as U.S.-Soviet interaction and the U.S.-Soviet arms race.

We identify three primary methodological problems that affect the inferences drawn in Ostrom and Job (1986) and the uses of force literature more generally. At the broadest level we argue that the form of the data should not determine the substantive inferences drawn. This argument has two major components. First, we want to ascertain whether the coding of the dependent variable (binary, event count, or duration) and the subsequent choice of a statistical model (logit/probit, Poisson/negative binomial, hazard) influences the substantive inferences drawn. Second, we want to determine whether truncation of the dependent variable (such as looking at only major uses of force) affects the inferences drawn. Third, like Ostrom and Job, we submit that decisions to use force are not independent, as many scholars implicitly assume; rather they occur in a larger context of strategic rivalry. Thus it is important to take temporal dynamics into account when modeling presidents' decisions to use force.

Sara McLaughlin Mitchell is Assistant Professor of Political Science, The Florida State University, Tallahassee, FL 32306-2230 (smclaugh@garnet.acns.fsu.edu). Will H. Moore is Associate Professor of Political Science, The Florida State University, Tallahassee, FL 32306-2230 (whmoore@garnet.acns.fsu.edu).

A previous version of this article was presented at the 1999 North American Meeting of the Peace Science Society International, Ann Arbor, MI. A lengthier version of the article is available at http://www.fsu.edu/~polisci/wksp_papers/workingpapers/wp2000.htm. We would like to thank Matt Baum, Patrick Brandt, Ashley Leeds, and several anonymous reviewers for their helpful comments on previous drafts. A replication data set will be deposited with the ICPSR's Publication-Related Archive (<http://www.icpsr.umich.edu>).

American Journal of Political Science, Vol. 46, No. 2, April 2002, Pp. 438–452

©2002 by the Midwest Political Science Association

ISSN 0092-5853

We demonstrate that these methodological issues affect inferences drawn about the relative impact of domestic and international factors on decisions to use force. To address the issue of aggregation bias, we compare probit and event-count models. We compare models that analyze major, minor, and all uses of force to determine the extent of truncation bias. To model temporal dynamics, we employ a new statistical model developed by Brandt and Williams (2001), the Poisson Autoregressive (PAR) model. We replicate Ostrom and Job's (1986) seminal study on U.S. uses of force to determine the extent to which aggregation bias, truncation, and dynamics affect their results.

Our empirical results show that the inferences are fragile across the three methodological issues. First, we find evidence of dynamics in presidential uses of force. Increases in the number of uses of force in previous quarters leads to a greater number of uses of force in the current quarter. This supports our expectation that presidential decisions to use force are influenced by the overall U.S.-Soviet rivalry. Second, the PAR model of Ostrom and Job's data produces the inference that international factors play a much larger role in the president's decision to use force than domestic and political factors. Our analysis suggests that Ostrom and Job's finding that domestic factors have a strong influence on the president's decision to use *major* force hinges on coding decisions that may produce both aggregation bias (the decision to dichotomize rather than count) and truncation bias (the decision to analyze only major uses of force). We eliminate these potential biases by counting a broader set of foreign policy decisions to use force, and we find evidence of temporal dynamics in the data. The key point, then, is that scholars need to pay close attention to aggregation, truncation, and dynamics when planning their studies, as the inferences one draws from such studies can vary substantially.

The Ostrom and Job Study

The Ostrom and Job study is motivated by an interest in explaining political uses of force taken by U.S. presidents. They argue that presidents are charged with protecting U.S. national security interests and that the public will hold them accountable at the polls should they fail to do so. Ostrom and Job identify three functional responsibilities that inform presidential decision making: commander-in-chief of the U.S. military; chief executive of the U.S. polity; and leader of his political party. Each of these functional roles defines a relevant

information arena that the president must monitor in an effort to retain power.

The first arena is global politics, where the president closely monitors the U.S.-Soviet rivalry. The second arena is the domestic polity, especially as it relates to the international arena. Finally, Ostrom and Job suggest that the president must also monitor his—and thus his party's—political standing among the citizenry. The theory suggests that the president will monitor information in all three arenas and use that information to determine whether or not to use force.

Ostrom and Job test their theory using quarterly data from 1949 to 1976; the dependent variable is coded one if at least one major use of force occurred during a quarter and zero otherwise. They find that variables representing the international, domestic, and political environments have a statistically significant impact on presidential decisions to use force. However, they find that political factors, most notably presidential approval, have the largest substantive impact.

Several scholars have taken this finding for granted and focused their attention primarily on these political factors (e.g., DeRouen 1995, 2000; Fordham 1998a, 1998b; Hess and Orphanides 1995), whereas others have taken issue with the finding, arguing that international factors are paramount (e.g., Gowa 1998; James and Oneal 1991; Meernik 1994; Meernik and Waterman 1996; Wang 1996). We believe that some of the divergent findings in the use of force literature can be attributed to one of three methodological problems: aggregation bias, truncation bias, and dynamic misspecification. We discuss each of these problems in detail in the following sections. We then reexamine Ostrom and Job's (1986) study to determine the extent to which their empirical findings are altered when one accounts for each of these methodological issues.

Data Collection: Aggregation and Truncation

The use of force by a U.S. president is a political event that can be initiated on any day while the president is in office. Given that most of the factors posited to influence the president's decision are not changing on a daily basis (such as presidential approval), it is common for scholars to aggregate uses of force over a standard time unit. As we explain below, such aggregation produces a variety of measures of uses of force. It is also common for scholars to examine only the most severe and visible uses of force, usually major or nuclear capable uses of force (DeRouen

1995, 2000; Fordham 1998a, 1998b; Ostrom and Job 1986). The researcher makes both of these choices; the former can produce aggregation bias in events data, while the latter is arguably truncation of the dependent variable. These decisions may seem trivial, but in fact both aggregation and truncation can have a serious impact on the inferences drawn in an empirical model.

Aggregation Bias

The way in which events are measured is an aggregation choice by the researcher, but little attention has been paid to the bias that can be introduced by this choice (Alt, King, and Signorino 2001). Ostrom and Job (1986), for example, analyze quarterly uses of force from 1949 to 1976 with a dummy variable that equals one if *one or more* uses of force occur in a quarter and zero otherwise. A possible problem with this approach is that it removes potentially important information that could be relevant to the study of uses of force. Uses of force could be measured as a dummy variable (Ostrom and Job 1986), as an event count (the number of times force was used per time period), or as an event history (the number of days between uses of force). Given the multiple options available for representing the information, one might ask whether the inferences drawn are affected by the aggregation decision. To the extent that they are, our inferences suffer from aggregation bias. As Alt, King, and Signorino note, "We do not want the form in which the data happen to be collected to determine the substantive ideas which we can explore" (2001, 22).

The variety in how uses of force are measured is quite evident in a glance through the recent literature. We find dichotomous (James and Oneal 1991; Meernik and Waterman 1996; Ostrom and Job 1986), ordinal (DeRouen 1995), continuous (Wang 1996), and event count (Fordham 1998a) representations of U.S. presidents uses of force.

Ideally, we would identify the underlying data generating process for our dependent variable and then use a statistical model most appropriate for that process. In practice, however, individual researchers often set up their statistical models based upon the way in which their data are collected. Those who utilize dichotomous dependent variables often turn to logit or probit, while others who collect event counts estimate Poisson or negative binomial models. The problem is that "the statistical literature does not generally provide ways of comparing results across these different models. This should be quite frustrating to scholars, since binary, count, and duration data are all coded from precisely the same underlying events" (Alt, King, and Signorino 2001, 22). These scholars argue that we need to develop statistical models that

will produce the same inferences regardless of the aggregation decision of the researcher.

Unlike Alt, King, and Signorino (2001), we do not explore solutions to the aggregation problem by developing new statistical models. Instead, we hope to increase awareness of this problem and advocate that scholars make a case for why they feel that their aggregation decision best reflects their conceptualization of the phenomenon in question and is most appropriate given their theory.

The question, then, is whether Ostrom and Job's conceptualization and/or theory provide us with guidance regarding the most appropriate aggregation rule for measuring U.S. presidents' use of force during the Cold War. Ostrom and Job adopt Blechman and Kaplan's definition of a political use of force: "physical actions...taken by one or more components of the uniformed armed military services as part of a deliberate attempt by national authorities to influence or be prepared to influence, specific behavior of individuals in another nation without engaging in a continuing contest of violence" (1978, 12). Thus, they are military actions that are intended to influence the behavior of other states.

As noted above, the variables scholars have used to measure such behavior include a dichotomous measure that counts whether at least one use of force took place; a count of the uses of force that occurred; and an ordinal (or continuous) measure of the intensity of the use of force. Ostrom and Job explain that because they are "following the logic of the cybernetic perspective, and thereby seeking to ease the burden of calculation, we assume the president's choice set to be limited to two alternatives: (a) do not use major force and (b) use major force" (550). One interpretation of this quote is to make a case in favor of using a dichotomous variable, which is what Ostrom and Job did. However, we submit that doing so essentially ignores the problem of aggregation bias. We do not wish to argue that Ostrom and Job were "wrong." Rather, we wish to make the case that an event-count measure better represents the theory and then determine whether the use of a different aggregation rule produces different findings.

Our case begins with the observation that on any given day the president has the opportunity to make a decision to use force. A decision to use a dichotomous measure rests on the implicit assumption that the actor makes only one decision in the given unit of time. In this case, the implicit assumption in Ostrom and Job's aggregation rule is that the president makes a quarterly decision whether or not to use force. Put differently, it treats quarters where the president makes a single decision to use force and quarters where the president uses force on multiple occasions as equivalent. Yet Ostrom and Job do not make any assumptions about the frequency with

which the president makes foreign policy decisions. We contend that a reasonable reading of their theory is that during quarters where the president uses force more than once, he uses the cybernetic decision-making rule in each instance. Such a reading suggests that an event-count representation of the data, where the number of uses of force is counted per quarter, better represents Ostrom and Job's theory.

If we are correct that the conclusions drawn by Ostrom and Job (1986) could be influenced by aggregation bias (i.e., their decision to use a binary count of uses of force), then it is quite possible that we will draw different inferences when we use a different measure. Below we compare the results from analyses using both the dichotomous measure of uses of force and an event-count representation of the same data. This is especially important for the use of force literature because of the great variation in uses of force measures. Some of the disparate findings in the literature could be an artifact of the use of event counts in some studies and dichotomous measures in others. Yet, before turning our attention to these empirical models, we must first consider another measurement issue: truncation bias.

Truncation Bias

In addition to exploring the potential for aggregation bias, it is also important to consider what behavior is appropriate for inclusion in a study of presidential decisions to use force. Ostrom and Job (1986) argue that only major or nuclear capable levels of force should be analyzed. Further, they note that this decision contradicts the measurement decision made by the authors who originally collected the data (Blechman and Kaplan 1978). Noting that their findings about the importance of political and domestic variables relative to international variables is at odds with Blechman and Kaplan's (1978) findings, Ostrom and Job explain that "these authors were analyzing data including both minor and major uses of force—the former constituting about 150 incidents, many of them merely 'sail bys'" (1986, 556). Thus, Ostrom and Job justify their decision to examine a limited set of uses of force on the apparent grounds that the other events are not important.

We suggest that the decision to eliminate minor uses of force¹ from the analysis introduces truncation bias. Our contention is based on the argument that uses of force are best conceptualized as a single class containing

what Blechman and Kaplan code as "minor" and "major" uses of force. The argument that minor and major uses of force belong to distinct classes of behavior implicitly assumes that presidents know in advance what the highest level of force will be. We believe that any use of force has the potential to escalate into a more serious crisis, and thus that minor and major uses of force differ primarily in the ultimate intensity level the use of force reaches. Similarly, Blechman and Kaplan (1978) conceive of political uses of force as any use of the military to achieve a political objective. Minor and major uses of force are utilized for similar purposes, but simply differ in the branches and extent of the military employed in a given crisis.

To the extent that uses of force are best conceptualized as a single class of events, exclusion of the minor uses can produce a serious methodological problem called truncation bias, which will result in causal effects being estimated closer to zero than they actually are unless the independent variables account for the selection rule (King, Keohane, and Verba 1994, 130). The effects of truncation (or what Achen refers to as censoring) are even more complex in the multivariate case. "For moderate amounts of censoring the coefficients may resemble neither their value in the misspecified regression nor their true values. That is, individual coefficients need not converge to values between those they would take on with no censoring or complete censoring" (Achen 1986, 81).

In the case of truncation, parameter estimates will be consistent as long as the explanatory variables can account for the selection rule. However, if minor and major uses of force are members of a single class and a study uses a truncated measure of uses of force without including independent variables that can account for the selection of major versus minor uses of force, it will produce inconsistent parameter estimates (Long 1997, 204). The question then is whether the ten independent variables employed by Ostrom and Job can adequately account for the difference in the president's decision to use major force as opposed to minor force. We believe several factors are omitted that can help account for the level of force employed by the president in a crisis. Ostrom and Job do not consider any characteristics of the crisis such as the issue at stake or the strategic importance of the crisis area, nor do they consider the potential for reaction to the actions of the adversary (e.g., major components of the military may be employed if the adversary does something very provocative). It is also possible that the history of relations between the disputing parties may influence the level of force employed. All of these possibilities lead us to suspect that Ostrom and Job's focus on only major uses of force may well have introduced truncation bias into their study. We have argued that decisions about data collection can

¹Blechman and Kaplan (1978, 50) create an ordinal scale to reflect the military level of effort for each use of force. They distinguish between major and minor force components, defining the former as two or more aircraft carrier task groups (naval), more than one ground battalion, or one or more combat wings (land-based air).

introduce aggregation and truncation bias. In the following section we identify an additional source of bias, the failure to consider dynamic specification.

Dynamic Specification Issues

That certain political time series should exhibit persistence (e.g., an autoregressive process) is not surprising to political scientists. Should uses of force exhibit persistence? Ostrom and Job's emphasis on the U.S.-Soviet rivalry suggests that it should.² Ostrom and Job conceive of foreign policy decisions in the 1946–1976 time period as being driven in large part by the ongoing Cold War rivalry, which is reflected in their assumption that presidents share goals of anti-communism and containment of the Soviet Union. Furthermore, they consider only interactions between the United States and the Soviet Union to be relevant for international tension, excluding tension that may exist between the United States and other states.

If the president's decisions to use force are made in the context of the U.S.-Soviet rivalry, then the probability of using force in one instance should be related to the chances of using force in another instance. Each situation where force could be used is framed in terms of the U.S.-Soviet rivalry, which means that success or failure in one interaction can affect behavior in a subsequent interaction. One finds a similar logic in the research on enduring rivalries: a repeated history of militarized confrontation makes future militarized disputes between adversaries more likely (Hensel 1998). Additionally, the outcome of one crisis also has a large impact on the outbreak and escalation of future disputes. For example, nations that adopt successful strategies in one crisis are likely to retain similar strategies in the future, whereas nations that lose in a crisis might become more belligerent in the future because of failed past strategies (Leng 1983).

Ostrom and Job's emphasis on the U.S.-Soviet competition implies a similar process. U.S. presidents should be more likely to view the use of force as imperative when they believe that the Soviets are leading the rivalry, either in terms of their exportation of communism or in terms of their military capabilities. Ostrom and Job (1986) make a similar argument about strategic balance; they expect the use of force to be more likely in situations where the United States is losing its strategic advantage.

Furthermore, the president may view some regional conflicts as much more important in the context of the Cold War rivalry than others. This could help to account for differences between opportunities to use force and when force is actually used (Meernik 1994). The broader issue is that if Ostrom and Job are correct about the importance of the U.S.-Soviet rivalry, then presidents' uses of force should exhibit temporal dynamics. Decisions to use force are not independent because they are part of a larger competitive interaction between the U.S. and the Soviet Union. A model that assumes such independence, like probit, may be underspecified and omit interesting dynamic properties of the uses of force over time. This is more than a "methods" point; different theories imply different positions on the assumption of independence.

One can usefully distinguish three approaches to modeling time-series dynamics. The first is to treat them as a nuisance; that is, to note that the error structure in a regression is disturbed when time-series dynamics are present in data, such that the residuals are not white noise. The solution, from this view, is to transform the data in an effort to produce white noise residuals. The second approach is to model the dynamics directly; that is, to try and specify the data generation process that produces the observed dynamics in the data.³ The third approach is to assume (often implicitly) that residuals are white noise.

The first and second approaches have come to dominate the statistical analysis of time-series data that are normally distributed, but the third approach is still commonly used by scholars who model either event-count or binary dependent variables that they have measured over time. For example, many dyadic analyses of the Militarized Interstate Dispute data (Jones, Bremer, and Singer 1996) employ logit or probit and ignore the potential dynamics across disputes over time. One simple solution for the violation of autocorrelated residuals is to include lagged dependent variables as regressors in these models. Another solution has recently been posed by Beck, Katz, and Tucker (1998) where they treat binary cross sectional time-series data as grouped duration data and include temporal dummy (or spline) variables in their logit model. While this approach corrects any potential inefficiency posed by autocorrelation, it does not attempt to model the dynamics of the data-generating process explicitly (the second approach).

²Many processes could create an autoregressive process in uses of force. While the presence of an AR process is consistent with our rivalry-based explanation, we do not deny that other factors (such as domestic politics) could create dynamics as well.

³The first approach is frequently called generalized least squares, while the second motivates most other time series analysis including ARIMA, regression models (such as the London School of Economics approach), Vector Autoregression, Error Correction Models, etc.

Several recent methodological papers (e.g., Alt, King, and Signorino 2001; Brandt and Williams 2001; Brandt, Williams, Fordham and Pollins 2000; Jackman 2000) develop models that encompass both the distributional and dynamic properties of the data-generating process. In this study we are particularly interested in Brandt and Williams' two general models that capture dynamics between event counts across time periods. For persistent data, the Poisson exponentially weighted moving average (PEWMA) is appropriate (Brandt et al. 2000). For data that have cyclical properties and memoried processes that are mean reverting, the Poisson autoregressive model (PAR) of order (p) is suitable. As we show below, the PAR is the appropriate model for our reexamination of the Ostrom and Job study.

Operational Measures

There are a total of ten independent variables in the Ostrom and Job (1986) study, reflecting the three information environments the president must monitor: the international arena, the domestic environment, and the political environment. The first international variable, the level of international tension (I1) is taken from Azar's (1982) COPDAB event data base, and is measured as the "total directed conflict and cooperation between the U.S. and the U.S.S.R for each quarter divided by the total directed behavior of either a conflictual or a cooperative nature" (Ostrom and Job 1986, 552). The second international variable is the strategic balance (I2), which ranges from -1 (total U.S.S.R. capability dominance) to $+1$ (total U.S. capability dominance). They also code a war variable (I3), which takes on the logged values of the sum of battle deaths during the Korean and Vietnam wars.

The domestic environment is captured through four variables. The first, (D1), is an interaction term between the level of international tension (I1) and the propensity for the public to identify foreign policy issues as most salient. The second domestic variable, (D2), is a similar interaction measure for strategic balance. Aversion to war (D3) is the mirror image of the war measure (I3), while the misery index (D4) multiplies the sum of unemployment and inflation and the percentage of the U.S. public identifying the economy as the most important problem. Finally, Ostrom and Job code three political variables. The first, presidential approval (P1), is the percentage of people who approve of the job the president is doing. The second, (P2), is the difference between the president's approval upon taking office and his level of approval in the current quarter. Finally, the electoral vari-

able (P3) is a dichotomous variable that is coded one during the third quarter of even number years (i.e., during midterm and general elections).

In the next section, we turn to empirics and report the parameter estimates found when one uses different aggregation decisions to code the dependent variable (i.e., dichotomous versus event count), different truncation decisions to code the dependent variable (i.e., "major" versus "minor" versus "all" uses of force), and different statistical models (i.e., probit versus negative binomial versus Poisson autoregressive).

Empirical Findings

Probit Model Results

We begin with a discussion of the findings produced using a probit model. Table 1 reports parameter estimates, standard errors, and predicted probabilities for major uses of force, minor uses of force, and all uses of force, respectively. Model 1 (major uses of force) replicates Ostrom and Job's (1986, 555) Table 1. All of the parameter signs in the replication model are in the same direction as the original Ostrom and Job model, and the variables that were statistically significant in the original model are significant in the replicated model.

The primary difference between our replicated findings and Ostrom and Job's findings is that elections have a significant and positive impact on the major use of force in the replicated model. This is consistent with Ostrom and Job's (1986, 558) claim that while their election parameter was not statistically significant, its substantive impact was quite large. We find a similar result; the probability of using force increases by 29.02 percent in an election period. Other than the economic misery index (D4), the election variable has the largest substantive impact on the propensity for the president to use major force.

Ostrom and Job's decision to study only major uses of force reduces the total number of Blechman and Kaplan uses of force from 202 to 71 in the 1949–1976 time period. To determine the extent to which the selection of major or nuclear capable uses (truncation) influences the results, we created two additional dichotomous variables: the first is coded one when at least one minor use of force is taken, and the second is coded one when at least one use of force (either minor or major) is taken. The results we obtained when using these dependent variables are reported in Table 1, Model 2 (minor uses), and Model 3 (all uses).

These results give us some insight into what is driving Ostrom and Job's findings as we see that the only

TABLE 1 Probit Analysis of Major, Minor, and All Uses of Force

Variable	Model 1		Model 2		Model 3	
	Replicated Parameter (s.e.)	Percentage Change ^a	Parameter (s.e.)	Percentage Change ^a	Parameter (s.e.)	Percentage Change ^a
I1: International tension	0.951* (0.550)	-0.047/ 16.67	0.655 (0.568)		1.121* (0.650)	-0.073/ 11.09
I2: Strategic balance	-1.121 (0.913)		-1.813* (1.029)	-27.12/ -27.17	-2.436* (0.998)	-26.43/ -34.66
I3: Cumulative war dead	-0.194* (0.111)	-14.71	-0.446* (0.127)	-35.30	-0.404* (0.119)	-24.64
D1: International tension when foreign policy is primary public concern	-1.854* (0.811)		-0.789 (0.825)		-1.700* (0.913)	
D2: Strategic balance when foreign policy is primary public concern	1.211* (0.648)		0.078 (0.761)		1.457* (0.769)	
D3: Aversion to war	-0.205* (0.097)	-14.49	-0.273* (0.095)	-19.77	-0.215* (0.111)	-10.29
D4: Weighted economic misery index	0.166* (0.092)	21.19	-0.052 (0.093)		0.028 (0.099)	
P1: Presidential approval	0.083* (0.030)	3.45	0.005 (0.034)		0.051 (0.036)	
P2: Overall presidential success	0.054* (0.026)	2.12	0.023 (0.031)		0.060* (0.035)	1.16
P3: National elections	0.747* (0.441)	29.02	0.053 (0.427)		0.988* (0.595)	12.55
Constant	-5.039* (2.108)		2.072 (2.521)		2.634	
Percent cases correctly predicted	74%		74%		75%	

^aThe percentage change in predicted probability is based on a one standard deviation or one-unit increase from the mean holding all other variables at their mean or mode. The first percentage values for I1 and I2 are the effects of those variables when foreign policy is the public's primary concern (D1 = I1, D2 = I2); the second percentage values for I1 and I2 are the effects of those variables when foreign policy is not the public's primary concern (D1 = 0, D2 = 0).

* $p = .05$ for a one-tailed test.

N = 112

domestic variable that has a statistically significant impact on minor uses of force is aversion to war (D3), which is the mirror image of cumulative war dead: the major domestic and political variables no longer have a statistically significant impact. Perhaps most interesting, the impact of presidential approval is no longer significant in Model 2 (minor uses of force). This implies that the president may perceive the need for an approval "buffer" before using major or nuclear capable force, but

that approval is not important for the decision to use minor force.⁴

⁴Baum's (1999) formal model of presidential decisions to use force may help to explain these empirical results. He argues that public attentiveness creates greater domestic audience costs. For those events where the public is paying attention, the president is strongly influenced by public opinion (approval is significant), whereas for those events that exhibit lower levels of public attentiveness (all uses of force), the president is not influenced either by poor economic conditions or by his approval rating.

We can also see the influence of the public when comparing the substantive effects of international tension when foreign policy is or is not salient. When foreign policy is the public's primary concern, increases in international tension actually make the president less likely to use major force (−0.047 percent). On other hand, when foreign policy is not the public's primary concern, a one standard deviation increase in international tension increases the president's propensity to use major force by 16.67 percent. We see a similar difference in the substantive results for I1 in Model 3 (all uses of force).

It is also interesting to note that the substantive impact of the independent variables is quite different in the three models. Two of the three international variables (I2 and I3) have a larger impact on the decision to use minor force, while the substantive effect of most other domestic and political variables declines. The results in Table 1 suggest that the impact of domestic and political factors on a president's decision to use force may depend on how attentive the public is to such events. In the general uses of force data set where attentiveness is lower on average, the president is influenced more by international factors, and less so by domestic ones.

Yet, we do not want to make too much of these results: while the results using all uses of force address truncation bias, both aggregation bias and temporal dynamics remain unaccounted for. We thus turn our attention to eliminating aggregation bias. The solution is to use a count, rather than a dichotomous, representation of the uses of force in a quarter.

Negative Binomial Model Results

Event counts are simply variables that record the frequency with which a type of event occurs over a defined unit of time. In our case, we are interested in the frequency of uses of force over the quarter year. King (1988, 1989) explains that OLS regression is inefficient and biased when used to estimate parameters of a model that has an event-count dependent variable. He proposes several alternatives: the Poisson regression model, the negative binomial regression model (or his generalized event count model), and a hurdle Poisson regression model. The distinctions among these models hinge on an independence assumption. More specifically, the Poisson regression model assumes that events are independent of one another, both across and within the unit of observation. That is, we should use the Poisson regression model if we are willing to assume that the decision to use force in any given quarter is made without regard to (1) whether force was used last quarter and (2) whether

force has already been used this quarter. If one is willing to make the first assumption, but not the second, then the negative binomial regression is appropriate.

In this subsection we accept the first assumption, relax the second assumption, and estimate negative binomial-regression models. Put differently, we estimate models that correct for both aggregation bias and truncation bias. The results do not yet account for the possibility of temporal dynamics in the use of force series.

Table 2 reports the results obtained when we estimate the Ostrom and Job model using an event count operationalization of uses of force with a negative binomial regression model.⁵ As in Table 1, we report results using three operational measures of the dependent variable: a count of all major uses of force, a count of all minor uses of force, and a count of all uses of force. If one accepts the arguments we have made in the previous section, then the first two sets of results suffer from truncation bias, and all three sets suffer from assuming that the events are independent.

The negative binomial results in Table 2 are similar to the probit results, especially for Model 2 (minor uses of force). Again, we see that while domestic and political factors have a strong influence on the president's decision to employ major force, of these variables, only D3—the mirror image of cumulative war dead—has a statistically significant impact on decisions to use minor force. The substantive differences are quite striking when we compare Model 1 (major uses of force) to Model 2 (minor uses of force). National election periods increase the expected number of major uses of force by 89.46 percent, while the effect of elections on decisions to use minor force is negative and insignificant. Likewise, we see a strong substantive effect for international tension when foreign policy is not the primary public concern. The expected number of major uses of force increases by 56.44 percent when this variable (I1) increases one standard deviation above its mean; the variable has an insignificant impact in Model 2. Model 3 demonstrates that international and political factors influence the president's decision to use force in general. All three international variables are significant (statistically and substantively).

⁵The parameters were estimated using the nbreg command in Stata, version 6.0. For the negative binomial models in Table 2, the substantive values reported are the percentage changes in the expected count of uses of force for a one standard deviation increase in each independent variable, holding all others at their mean or mode. We should note that these percentage changes are not directly comparable to the probit percentage change values. We will be able, however, to compare the substantive effects of each variable in the negative binomial versus the Poisson Autoregressive (PAR) model, as both are event-count models.

TABLE 2 Negative Binomial Analysis of Major, Minor, and All Uses of Force

Variable	Model 1		Model 2		Model 3	
	Replicated Parameter (s.e.)	Percentage Change ^a	Parameter (s.e.)	Percentage Change ^a	Parameter (s.e.)	Percentage Change ^a
I1: International tension	1.011* (0.592)	22.07/ 56.44	0.494 (0.372)		0.628* (0.313)	14.72/ 32.05
I2: Strategic balance	-0.811 (0.933)		-1.851* (0.667)	-54.82/ -51.58	-1.378* (0.530)	-41.68/ -41.74
I3: Cumulative war dead	-0.197* (0.118)	-33.43	-0.386* (0.088)	-54.95	-0.305* (0.069)	-46.74
D1: International tension when foreign policy is primary public concern	-1.301* (0.769)		-0.170 (0.570)		-0.588 (0.449)	
D2: Strategic Balance when foreign policy is primary public concern	0.522 (0.625)		-0.308 (0.459)		0.006 (0.362)	
D3: Aversion to war	-0.033 (0.087)		-0.274* (0.091)	-40.94	-0.155* (0.061)	-25.77
D4: Weighted economic misery index	0.082 (0.094)		-0.032 (0.067)		-0.003 (0.054)	
P1: Presidential approval	0.080* (0.031)	8.33	0.029 (0.020)		0.045* (0.017)	4.60
P2: Overall presidential success	0.053* (0.029)	5.44	0.030 (0.019)		0.036* (0.016)	3.67
P3: National elections	0.639* (0.310)	89.46	-0.061 (0.273)		0.208 (0.203)	
Constant	-5.430* (2.057)		2.072 (2.521)		-1.066 (1.265)	
α	.000 (.000)		.000 (.000)		.000 (.000)	

^aThe percentage change in the expected count is based on a one standard deviation or one-unit increase from the mean holding all other variables at their mean or mode. The first percentage values for I1 and I2 are the effects of those variables when foreign policy is the public's primary concern (D1=I1, D2=I2); the second percentage values for I1 and I2 are the effects of those variables when foreign policy is not the public's primary concern (D1=0, D2=0).

* $p = .05$ for a one-tailed test.

N = 112

Presidential approval and overall presidential success are also positive and significant.

What these results demonstrate is that aggregation and truncation decisions can have a large impact on one's inferences. The impact of aggregation is clear when comparing the models employing a dichotomous measure of uses of force and an event-count measure. For example, compare Model 1 in Table 1 to Model 1 in Table 2. We can see that several variables that appear significant in the probit model drop out in the negative binomial (D2, D3, D4). The inference that poor economic conditions influence presidential uses of force finds no support when one

adopts an event-count operationalization. We can also see the strong impact of truncation bias in Tables 1 and 2. Examining only a portion of the full range of uses of force (Models 1 and 2) produces very different results than those obtained in the nontruncated model (3).

A seemingly interesting result is that the α parameter is not statistically significant in any of the three models.⁶ This finding appears to suggest that the prob-

⁶ α is known as the dispersion parameter. The Poisson model contains an assumption that the mean is equal to the variance. For cases where the variance exceeds the mean (overdispersion), α will

ability of a use of force (whether major, minor or all) is independent of other uses of force in any given quarter. Yet, as Brandt and Williams (2001) explain, when temporal dynamics are present in event-count data, the negative binomial model's α parameter can be misleading. The alpha parameter estimated in a negative binomial model may be insignificant, even in the presence of temporal dynamics. We will show in the following subsection that this is very similar to the situation obtained in our analysis. There we present the results we obtain when we estimate a model developed by Brandt and Williams (the Poisson Autoregressive model) that is designed for event-count data that exhibit temporal dynamics.

The Poisson Autoregressive (PAR) Model

While selecting the all uses of force count variable addresses both aggregation and truncation bias, the negative binomial regression model can neither account for, nor diagnose, temporal dynamics. The first task, then, is to determine whether it is necessary to use the PAR model, as we cannot rely on the α parameter in the negative binomial model to help us evaluate the presence or absence of dynamics. Brandt et al. (2000) recommend the autocorrelation function as an appropriate diagnostic tool.⁷ If the number of significant lags is extremely large (the time-series process is persistent), then the Poisson exponentially weighted moving average (PEWMA) model is appropriate. Series that are mean reverting, on the other hand, are characterized by an ACF with a smaller number of significant lags.

Figure 1 plots the all uses of force count series over time, and displays the ACF (with Bartlett corrected standard errors). The total uses of force in one quarter are dependent on the uses of force in previous quarters: the ACF shows strong, though not particularly long-memoried, persistence in the data (i.e., the first four lags of the series are significantly correlated). This indicates that the Poisson Autoregressive (PAR) model is suitable.

be greater than zero. The null hypothesis for the test reported in Table 2 is that α is equal to zero (which means the Poisson model is adequate if the parameter is insignificant). If α is insignificant, the parameter estimates for a negative binomial model are identical to the parameter estimates of a Poisson model.

⁷The autocorrelation function (ACF) of a time series, y_t , is the covariance of y_t and subsequent values of y_{t-k} divided by the variance of y_t . Each autocorrelation value falls between -1 and $+1$. Each spike in the partial autocorrelation function (PACF) represents the partial derivative for each y_{t-k} with respect to y_t .

The two models developed by Brandt and Williams (PEWMA and PAR) are derived in state-space form and estimated via a Kalman filter. State-space models consist of a measurement equation, a transition equation, and assumptions about the mean and variance of the initial state vector. The transition equation describes the evolution of a set of state variables (the dynamic process), while the measurement equation describes how the data are generated from the state variables (the distributional aspect). The Poisson Autoregressive (PAR) model assumes that the event-count series is generated by a Poisson distribution (the measurement equation). The state (transition) equation is assumed to be generated by a gamma distribution. These assumptions produce an event-count model that follows a simple linear autoregressive (AR) process (Brandt and Williams 2001).⁸

The PAR differs from the Poisson and negative binomial event-count models by specifying an explicit time-series process. This alters the interpretation of the parameters:

For the PAR(p) model, the instant effect of the change in the independent variable or impact multiplier for the number of counts at time t depends on the estimated value of the regression parameters and on the estimated values of ρ . . . The implication is that standard factor change and percentage change calculations that are used to interpret instantaneous changes in event count models with exponential link functions are not valid if the true data generation process is a PAR(p). (Brandt and Williams 2001, 169–170)

In other words, the calculated substantive effects of a negative binomial or Poisson event count model will be biased if the events are characterized by a dynamic, temporal process.⁹

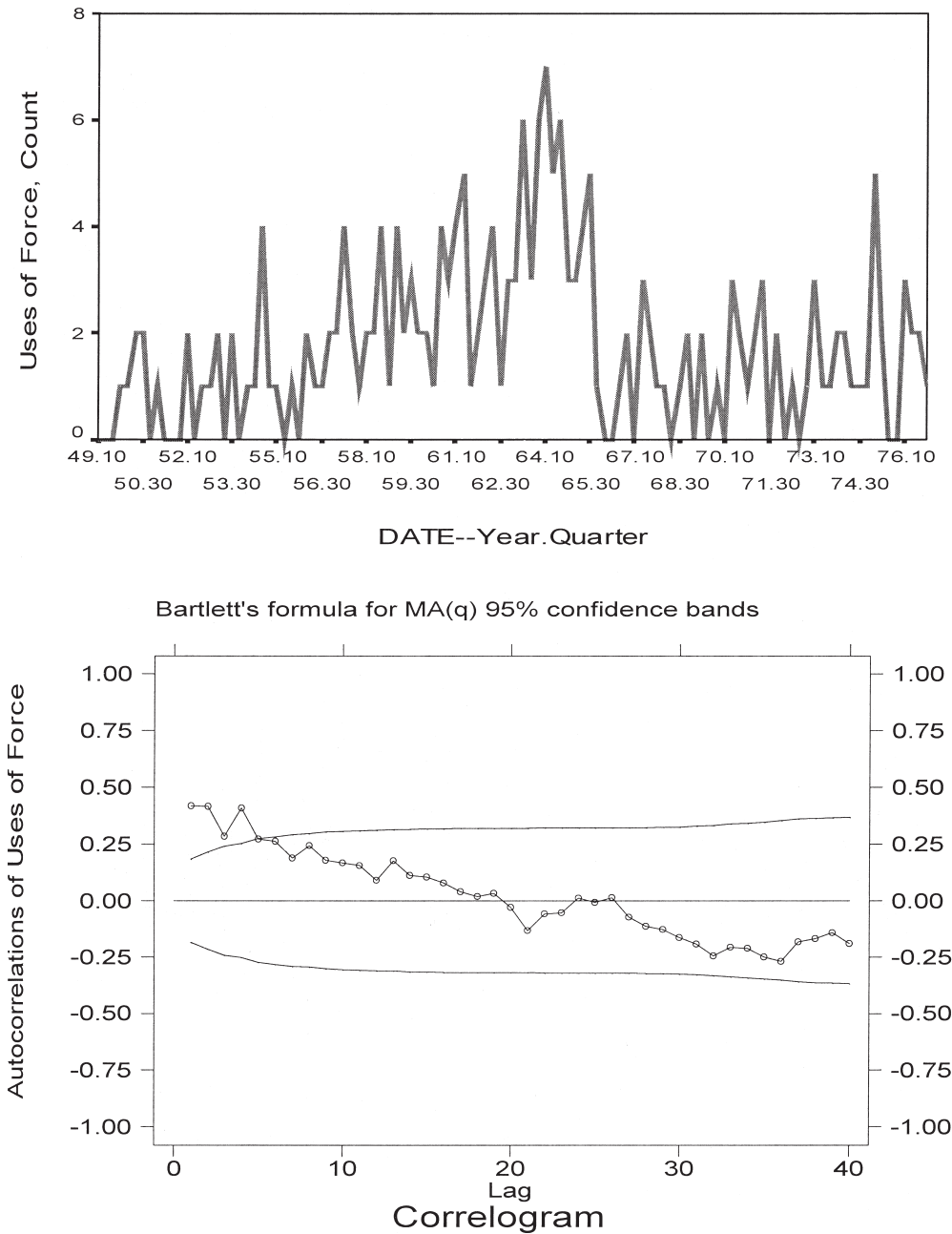
Table 3 presents the results for the estimated PAR model for all uses of force from 1949 to 1976 using Ostrom and Job's independent variables.¹⁰ We estimated a series of PAR(p) models, such as PAR(1) and PAR(2).

⁸We do not present the mathematical details of this model due to space limitations. Interested readers should see Brandt and Williams (2001).

⁹In fact, for an AR(p) process with positive coefficients, the estimated effects in a Poisson regression are biased downward from their true values (Brandt and Williams 2001, 173).

¹⁰The PAR model cannot begin with a zero count, therefore we omitted the first three quarters of 1949 (which had no uses of force) and began estimation in the fourth quarter. We estimated the parameters using PESTS, a program developed by Patrick Brandt and John Williams. Patrick Brandt and Mitch Sanders kindly helped us implement the program.

FIGURE 1 Plot and ACF of Total Uses of Force



We compared these models based on the significance of the AR coefficients and the overall fit of the model. We selected the PAR(2) model because it produced the lowest Akaike's information criterion, even though the first autoregressive lag was not statistically significant. We compare this model with the results obtained in the negative binomial event count model (Table 2, Model 3).

The signs of the parameters in the PAR(2) model are similar to the signs in the negative binomial model (two are different), yet only four independent variables are

statistically significant: International tension (I1), Cumulative War Dead (I3), the interaction between I1 and foreign policy salience (D1), and war aversion (D3). All of these variables were significant in the negative binomial model except D1. On the other hand, the negative binomial model produced a significant relationship between Strategic Balance (I2), Presidential Approval (P1), Presidential Success (P2), and the propensity for the president to use force. These factors were not significant in the PAR(2) model that takes into account the dynamic

TABLE 3 Poisson Autoregressive (PAR) Model of All Uses of Force

Variable	Parameter (s.e.)	Percentage Change ^a
I1: International tension	1.291* (0.586)	32.96/76.58
I2: Strategic balance	-1.025 (0.883)	
I3: Cumulative war dead	-0.533* (0.107)	-4.91
D1: International tension when foreign policy is primary public concern	-1.802* (0.738)	
D2: Strategic balance when foreign policy is primary public concern	0.250 (0.661)	
D3: Aversion to war	-0.281* (0.097)	-42.55
D4: Weighted economic misery index	-0.102 (0.010)	
P1: Presidential approval	-0.041 (0.034)	
P2: Overall presidential success	-0.019 (0.030)	
P3: National elections	0.217 (0.370)	
AR parameter 1 (lag one)	0.119 (0.124)	
AR parameter 2 (lag two)	0.246* (0.126)	
Constant	5.097* (2.521)	

^aThe percentage change in the expected count is based on a one standard deviation or one-unit increase from the mean holding all other variables at their mean or mode, taking into account the autoregressive (AR) parameters.

* $p = .05$ for a one-tailed test.

N = 109

process that characterizes presidential uses of force. This demonstrates that to the extent that our arguments about aggregation bias, truncation bias, and temporal dynamics are reasonable, domestic and political factors have a smaller impact on the president's decision to use force than we would have thought estimating an event count model that ignores the dynamics in uses of force across time.

Furthermore, the substantive significance of each independent variable is quite different in the PAR(2) model as compared to the negative binomial model. In the PAR(2) model, a one standard deviation increase in international tension increases the expected number of presidential uses of force by 76.58 percent (when foreign

policy is not the public's primary concern). This effect is more than twice as large as the effect estimated in the negative binomial model in Table 2 (32.05 percent). The impact of cumulative war dead (I3) appears much larger in the negative binomial model than in the PAR model. A one standard deviation increase in war dead decreases the expected number of uses of force by only 4.91 percent, which is almost ten times smaller than the substantive effect estimated by the negative binomial model. Also, strategic balance (I2) has little effect in the PAR model (the parameter is insignificant and the percentage change is under 5 percent). When the public views foreign policy issues as the most salient, the impact of international tension is diminished. However, this substantive

effect is quite large as well; when foreign policy is the public's primary concern, the expected number of presidential uses of force increases by 32.96 percent (half the effect when foreign policy is not a primary concern). This effect is quite a bit larger than the substantive effect estimated in Table 2 (14.72 percent).¹¹

Finally, recent war deaths (D3) diminish the president's proclivity to use force (42.55 percent), and this effect is also larger than the percentage change for D3 in the negative binomial model (−25.77 percent). The three political factors have no impact on the president's decision to use force in general. This is interesting because the substantive impact of variables such as elections may have been overestimated by ignoring the dynamic relationship between uses of force over time.

The first and second autoregressive parameters are positive, suggesting that increases in the number of uses of force in previous quarters leads to a greater number of uses of force in the current quarter. This supports our expectation that presidential decisions to use force are influenced by the overall U.S.-Soviet rivalry. As the rivalry heats up (international tension increases), the president is much more likely to use force. This in turn creates a dynamic process whereby increased uses of force in one time period create even more uses of force in subsequent periods (the dynamic process observed in the AR component of the model). Regardless of what process is driving dynamics in the use of force (rivalry, presidential administrations, and so on), the results clearly show that the statistical and substantive conclusions one would draw in a study such as Ostrom and Job are changed when one models such dynamics directly.

To conclude, our replication of Ostrom and Job's study produces substantively different inferences, especially with respect to their primary finding that political factors have a larger impact on uses of force than do international factors. Researchers who find our arguments about aggregation bias, truncation bias, and temporal dynamics convincing will want to reconsider what has become a widely held belief about the importance of presidential approval, economic performance, and elections on presidential uses of force during the Cold War. In addition, regardless of whether one accepts our arguments with respect to the Ostrom and Job study, we have demonstrated that the decisions we make about measurement and statistical modeling can have rather dramatic effects

on the inferences we draw. Thus, all researchers in international relations would do well to pay close attention to these issues when designing their studies.

Conclusion

A great deal of international relations research involves the analysis of international events, such as militarized disputes and the use of force by the U.S. president. Such events are typically measured dichotomously or as event counts, yet little attention is given to the tradeoffs among these choices or a defense of the choice made. Further, few students of international relations make an explicit effort to guard against truncation bias—indeed, many of the datasets produced in our subfield truncate their measurement at a specific threshold of one kind or another. Third, the dynamic properties that may characterize such data over time have only recently attracted attention. And yet many theories of international politics (such as rivalry) hypothesize that international events are related to each other over time.

In this study, we reexamine U.S. presidents' use of force during the Cold War, exploring aggregation bias, truncation bias, and temporal dynamics. We begin by replicating Ostrom and Job's (1986) probit analysis of quarterly uses of force. Our replication produces almost identical results to the original study, although we are concerned that the selection of major or nuclear capable uses of force creates an artificial truncation in the data. Thus we expand the data set to include all major and minor uses of force. We also wonder whether Ostrom and Job's operational choices influence the substantive conclusions one would draw about presidential uses of force. To address this potential aggregation bias, we create event counts of the uses of force per quarter, and estimate negative binomial regression models for those series using the same independent variables as Ostrom and Job (1986).

Analysis of probit and event-count models of these measures reveals that the impact of domestic and political factors (in Ostrom and Job's study) depends on the level of force examined. The economy, presidential approval, and elections have a much greater impact on the president's decision to employ major components of the U.S. armed forces (which are highly visible), but these factors do not influence the president's decision to use force in general. Furthermore the substantive impact of international factors is much larger for all uses of force than it is for the major use subset. This suggests that the influence of domestic and political factors on a president's decision to use force may vary based on the

¹¹The differences in these substantive effects stem both from the different magnitude of the coefficients and from the fact that in the dynamic PAR(2) model, the effect is seen in later quarters whereas in the static negative binomial model, the effects of a shock are restricted to a single quarter. We would like to thank Patrick Brandt for clarifying this for us.

level of force employed and the level of public attentiveness to the dispute.

Yet, while the negative binomial model accounts for the distributional aspect of the data, it cannot account for any dynamic properties of such data. When testing theories that imply that foreign policy behavior is related to its own past, we ignore time-series properties at our peril. To show that this is the case, we use Brandt and Williams' (2001) Poisson Autoregressive (PAR) model to estimate the parameters of the Ostrom and Job model. The parameter estimates are different from those found with either the probit or the negative binomial model. First, the autoregressive parameters are positive and significant, suggesting that increases in the number of uses of force in previous quarters leads to a greater number of uses of force in the current quarter. By contrast, the negative binomial model produced the inference that there were no dynamics present in the data. Second, with respect to substantive significance, the PAR parameters demonstrate that international factors play a larger role in decisions to use force than domestic and political factors. This finding contradicts the main finding of Ostrom and Job that domestic and political factors drive a president's decision to use force. Their finding may be sustained with the use of major force, but we have argued that a focus on major uses of force truncates the data, and that all uses of force should be included in analyses.

To conclude, we have shown that the substantive implications one draws can hinge on one's measurement decisions and one's choice of statistical model. International relations scholars are becoming increasingly aware of the importance of using the most appropriate measures and technique. We hope this study demonstrates the utility of not only using all of the information available in the data, but also in accounting for the temporal dynamics so prevalent in time-series data.

Manuscript submitted March 3, 2000.

Final manuscript received August 23, 2001.

References

- Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.
- Alt, James E., Gary King, and Curtis S. Signorino. 2001. "Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9:21–44.
- Azar, Edward E. 1982. *Conflict and Peace Databank (COPDAB), 1948–1978*. Second release, ICPSR study #7767. Ann Arbor: Inter-university Consortium for Political and Social Research.
- Baum, Matthew A. 1999. "The Ties that Bind: Public Opinion, Domestic Audience Costs, and the Decision to Use Force Abroad." Presented at the Annual Meeting of the American Political Science Association.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42:1260–1288.
- Blechman, Barry M., and Stephen S. Kaplan. 1978. *Force Without War*. Washington, D.C.: The Brookings Institution.
- Brandt, Patrick T., and John T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9:164–184.
- Brandt, Patrick T., John T. Williams, Benjamin A. Fordham, and Brian Pollins. 2000. "Dynamic Modeling for Persistent Event Count Times Series." *American Journal of Political Science* 44:823–843.
- DeRouen, Karl R. 1995. "The Indirect Link: Politics, the Economy, and the Use of Force." *Journal of Conflict Resolution* 39:671–695.
- DeRouen, Karl R. 2000. "Presidents and the Diversionary Use of Force: A Research Note." *International Studies Quarterly* 44:317–328.
- Fordham, Benjamin. 1998a. "Partisanship, Macroeconomic Policy, and U.S. Uses of Force." *Journal of Conflict Resolution* 42:418–439.
- Fordham, Benjamin. 1998b. "The Politics of Threat Perception and the Use of Force: A Political Economy Model of U.S. Uses of Force." *International Studies Quarterly* 42:567–590.
- Gowa, Joanne. 1998. "Politics at the Water's Edge: Parties, Voters, and the Use of Force Abroad." *International Organization* 52:307–324.
- Hensel, Paul R. 1998. "Interstate Rivalry and the Study of Militarized Conflict." In *Conflict in World Politics: Advances in the Study of Crisis, War and Peace*, ed. Frank P. Harvey and Ben D. Mor. London: Macmillan Press.
- Hess, Gregory D., and Athanasios Orphanides. 1995. "War Politics: An Economic, Rational Voter Framework." *American Economic Review* 85:828–847.
- Jackman, Simon. 2000. "In and Out of War and Peace: Transitional Models of International Conflict." unpublished manuscript, <http://jackman.stanford.edu/papers.shtml/>.
- James, Patrick, and John R. Oneal. 1991. "The Influence of Domestic and International Politics on the President's Use of Force." *Journal of Conflict Resolution* 35:307–322.
- 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* 15:163–213.
- King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32:838–863.
- King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

- Leng, Russell J. 1983. "When Will They Ever Learn? Coercive Bargaining in Recurrent Crises." *Journal of Conflict Resolution* 27:379-419.
- Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.
- Meernik, James D. 1994. "Presidential Decision Making and the Political Use of Force." *International Studies Quarterly* 38:121-138.
- Meernik, James, and Peter Waterman. 1996. "The Myth of the Diversionary Use of Force by American Presidents." *Political Research Quarterly* 49:573-590.
- Ostrom, Charles W., Jr., and Brian L. Job. 1986. "The President and the Political Use of Force." *American Political Science Review* 80:541-566.
- Wang, Kevin. 1996. "Presidential Responses to Foreign Policy Crises: Rational Choice and Domestic Politics." *Journal of Conflict Resolution* 40:68-97.