THE FLORIDA STATE UNIVERSITY

COLLEGE OF SOCIAL SCIENCES AND PUBLIC POLICY

PREDICTING THE INTENSITY AND LOCATION OF VIOLENCE IN WAR

By

ANDREAS BEGER

A Dissertation submitted to the Department of Political Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy

> Degree Awarded: Summer Semester, 2012

Andreas Beger defended this dissertation on May 29th, 2012. The members of the supervisory committee were:

> Will H. Moore Professor Directing Dissertation

Victor Mesev University Representative

Jason Barabas Committee Member

Megan Shannon Committee Member

Mark Souva Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.

To my family, Vesna and Alexander, who kept me alive in 2010 and 2011.

ACKNOWLEDGMENTS

I would like to thank my committee, Will H. Moore, Mark Souva, Jason Barabas, Megan Shannon, and Victor Mesev, for their advice and feedback throughout this process. After my time in Iraq in 2009 and while in Washington, D.C. in 2010, Will Moore encouraged and prodded me to "get back in the saddle". I am very grateful for that.

Dale Smith, Charles Barrilleaux, Cherie Maestas, and Jerry Fisher in the Department of Political Science have been very flexible and generous in their support. I have spent considerable time away due to outside obligations while studying here, and the Department has been consistently accommodating and understanding.

Sally Anderson, Courtenay Conrad, Jacqueline DeMeritt, Matt Golder, Daniel Hill, Daniel Milton, Chungshik Moon, Marius Radean, Meredith Whiteman and Joseph Young have been incredibly helpful with their feedback over the years. One of the projects in this dissertation began in Paul Hensel's international conflict class.

TABLE OF CONTENTS

Li	st of [Fables		vii
Li	st of l	Figures		viii
Ał	ostrac	ct		xii
1	Intr	oducti	on	1
2	Pre	dicting	the level of fatalities in interstate wars	3
	2.1	Intro	luction	3
	2.2	How l	ong will states bear the costs of war?	5
		2.2.1	Incomplete Information	5
		2.2.2	Commitment Problems	6
		2.2.3	Military decision	7
	2.3	Data a	and methods	8
		2.3.1	Estimation strategy	9
		2.3.2	Explanatory variables	10
	2.4	Empi	rical results	12
		2.4.1	Regarding the hypotheses	12
		2.4.2	Robustness checks	13
		2.4.3	Regarding overall model fit	14
		2.4.4	Out of sample predictions	15
	2.5	Discu	ssion and conclusion	16
3	Usiı	ng fron	at lines to predict deaths in the Bosnian war	26
	3.1	Introc	luction	26
	3.2	Disag	gregated studies of civil conflict	28
	3.3	Ethni	c violence and point data on deaths	29
		3.3.1	How the data are used	29
		3.3.2	The special case of Bosnia	31
		3.3.3	Yugoslav identity	32
		3.3.4	Frontlines	33
	3.4	Data a	and model	35
		3.4.1	Model setup	35
		3.4.2	Model specifications and variables	37
		3.4.3	Estimation	38

	3.5	Results	9
		3.5.1 Parameter estimates 3	9
		3.5.2 Model fit	0
	3.6	Conclusion	2
4	Moo	deling the evolution of civilian deaths in the Iraq War 5	8
	4.1	Introduction	8
	4.2	Background on Iraq and IR forecasting	0
		4.2.1 Counterinsurgency and violence against civilians	0
		4.2.2 Research on the Iraq War	1
		4.2.3 Forecasting in International Relations	2
		4.2.4 Why we need well-fitting models for policy	3
	4.3	Violence against civilians in Iraq	5
		4.3.1 Data and machine-coding	5
		4.3.2 The dynamics of violence during the Iraq War 6	7
	4.4	Modeling civilian deaths	8
		4.4.1 LN-EWMA: modeling effects	9
		4.4.2 LN-logistic: From peak oil to peak conflict	1
		4.4.3 Model fit	3
		4.4.4 Peak violence and the troop surge 7	4
	4.5	Conclusion	5
5	Con	nclusion 8	7
۸.		diase	0
Ą	ppen	arces	9
A	Rep	blication code for Bosnia chapter 8	9
	A.1	JAGS code: Constant-only negative binomial model 8	9
	A.2	JAGS code: Constant-only spatial count model8	9
B	Rep	plication code for Iraq chapter 9	1
	B.1	Coding incident location	1
	B.2	JAGS code: LN-EWMA	3
	B.3	JAGS code: LN-logistic	4
Re	efere	nces	5
Bi	ograj	phical Sketch)6

LIST OF TABLES

2.1	Variable descriptions and sources	21
2.2	Log-normal regression of war duration	22
2.3	Truncated normal regression of war fatalities	23
2.4	Variable significance	24
2.5	Model predictions for selected wars	25
3.1	Comparison of model fit.	43
4.1	Civilian deaths from March 2003 to February 2012 by province	79

LIST OF FIGURES

2.1	Predicted and observed war deaths for the best-fitting model. The <i>x</i> -axis lists the wars in the sample, with the observed number of war deaths indicated by the red dot. The list is ordered by the total number of deaths, ranging from the Falklands war to the eastern front of World War 2. The number of war deaths is on a logarithmic scale. The blue bars show the interval containing 95% of the predicted war deaths in each bootstrap iteration. The intervals correctly bracket observed deaths for 69% of all wars	19
2.2	Predicted and observed war deaths for selected wars. The predictions for the Eritrean-Ethiopian and Kargil wars are out of sample forecasts. Bars show low and high estimates for observed fatalities and 95% range for model predictions. Points show best or median estimate for observed and predicted fatalities respectively. The scale is logarithmic with base 10, so for example $10^3 = 1,000$ deaths.	20
3.1	Illustration of conflict incident point data and aggregation to grid cells or administrative units.	43
3.2	Comparison of Armed Conflict Location and Events Dataset (ACLED) and Research and Documentation Center (RDC) incident data for the Bosnian War.	44
3.3	Proportion of people who self-identified as being Yugoslav, rather than Serb, Muslim, Croat, etc. during the 1991 census. The rates are higher in cities, especially Tuzla and Sarajevo, where 15 to 20 percent of all respondents la- belled themselves as Yugoslav, and lower in eastern Bosnia, Herzegovina in the south, and parts of western Bosnia adjacent to Dalmatia	45
3.4	Ethnic diversity by municipality. Based on 1991 census.	46

3.5	Bubbleplot of Yugoslav self-identification, ethnic diversity, and civil war death rates by municipality. Each circle represents a municipality, with its proportion of self-identified Yugoslavs on the x-axis and ethnolinguistic fractionalization index (ELF) value on the y-axis. The area of the circle is proportional to deaths per capita during the civil war. There are fewer Yugoslavs in ethnically less diverse municipalities, as one would expect, and those also tend to see relatively little violence. In municipalities with a high level of diversity, the proportion of Yugoslavs is related to deaths per capita, with diverse but nationalist municipalities.	47
3.6	Deaths per 1,000 by municipality during the Bosnian War, 1992-1995. The rate of deaths is highest in eastern Bosnia in the municipalities that con- tained Bosniak enclaves during the war, and in areas of western Bosnia that were ethnically cleansed early in the war. From the IDC Bosnian Atlas of War Crimes.	48
3.7	Histograms of deaths by municipality during the Bosnian War, 1992-1995. The untransformed count of deaths fits a negative binomial distribution with mean 712 and dispersion parameter 0.99. Both the log of deaths and population-adjusted deaths are skewed.	49
3.8	Approximate 1993 front lines. Digitized using information on land mines from BiH Mine Action Center (posted at www.mine.ba)	50
3.9	Predicted versus observed killings during the Bosnian War. Each panel shows a plot of predicted counts compared to observed counts for the base, ethnic, frontline, and combined models respectively. The axes are on a <i>log</i> ₁₀ scale. The black line shows observed = predicted. Each plot also gives the deviance information criterion for each model (constant-only has 1676). The base model includes a spatial lag (neighborhood average deaths) and popula- tion offset, but no other covariates. The ethnic model, which also includes measures of ethnic diversity and dispersion, as well as per capita income, has by far the best fit, but also draws on census data unlikely to be available for other countries. The frontline model, which uses distance to the 1993 front lines and other structural variables without census data, significantly improves fit over the base model as well.	51
3.10	Prediction error of the frontline and ethnic models. Blue shades indicate underprediction, red shades overprediction.	52
3.11	Posterior densities for the base model: r = negative binomial shape, rhos = spatial lag, b = theta constant.	53

3.12	Posterior densities for the ethnic conflict model: $r = negative binomial shape$, rhos = spatial lag, b[1] = theta constant, b[2] = income per capita, b[3] = ELF index, b[4] = local polarization, b[5] = strategic significance.	54
3.13	Posterior densities for the frontline model: $r = negative binomial shape,$ rhos = spatial lag, b[1] = theta constant, b[2] = centered urbanization rate, b[3] = percentage rough terrain, b[4] natural log of centroid distance to front- line	55
3.14	Posterior densities for the combined frontline and ethnic model: $r = nega-tive binomial shape, rhos = spatial lag, b[1] = theta constant, b[2] = centeredurbanization rate, b[3] = percentage rough terrain, b[4] natural log of cen-troid distance to frontline, b[5] = income per capita, b[6] = ELF index, b[7] =local polarization, b[8] = strategic significance.$	56
3.15	Posterior densities for the constant-only model: r = negative binomial shape, b = theta constant.	57
4.1	Civilian deaths in Iraq, 2003 to 2012. The black line shows total deaths, red are deaths in Baghad, and blue are deaths in the rest of Iraq. Source: Iraq Body Count.	78
4.2	Histograms of monthly, Iraq-wide civilians deaths and log-transformed month counts.	nly 78
4.3	Time series of civilian deaths, 2003 to 2012, by province. The plots do not share a common scale, allowing the dynamics of civilians deaths to be compared across provinces.	80
4.4	Time series of civilian deaths, 2003 to 2012, by province. The plots share a common scale, and show that relative levels of violence across Iraq were minor compared to the scale of deaths in Baghdad	81
4.5	Autocorrelation functions for Iraq-wide deaths and the first difference in logged deaths.	82
4.6	Parameter estimates for LN-EWMA and LN-logistic models. The parameter d[1] is for national constitutional or parliamentary elections, and d[2] is for calendar months that include part of the Arab month of Ramadan	83
4.7	LN-EWMA in-sample fit (top) and forecast (bottom). While the model matches observed data well, its forecasts are essentially a continuation of the last ob- served value, depending on the estimated value of ω and accounting for a slight upward drift due to the non-symmetrical log-normal outcome equa-	S 0 4
	uom	ŏ4

- 4.8 LN-logistic model in-sample fit (top) and forecast (bottom). The in-sample fit is decent for observed data but worse than the LN-EWMAs. However, the forecast from January 2007 on is over the long-run more useful. Although it does not match realized outcomes closely, it is more accurate than LN-EWMA forecasts and does accurately predict a downward trend in violence.
 85
- 4.9 Comparison of model forecasts based on forecast inaccuracy $(\hat{y}/y) 1$... 86

ABSTRACT

Forecasting in international relations is becoming increasingly common, accurate, and relevant for decision support to policymakers. One of the major subjects of forecasting has been inter- and intrastate war, through various observable outcomes like onset, duration, and casualties. The three projects in this dissertation focus on the latter, i.e. the human cost of war. The projects respectively: (1) develop a model to predict battledeaths in interstate wars and assess out-of-sample forecast accuracy, (2) use a Bayesian spatial count model to examine the relationship between front lines and civilian deaths during the Bosnian civil war, and (3) develop Bayesian time-series count models for long-range forecasting of deaths in civil conflicts using Iraq Body Count project data on civilian deaths during the Iraq War.

CHAPTER 1

INTRODUCTION

Real-time forecasting of political violence using quantitative techniques is becoming increasingly common. Relevant academic work over the past decades has proceeded in roughly three streams: substantive forecasting efforts, data discovery, and methodological development.

Substantive examples of forecasting in the academic international relations literature rarely, but with some exceptions (Bennett and Stam, 2006; Cioffi-Revilla, 1991), provide true *ex ante* forecasts of future, unobserved outcomes. Partly this may be because publishing cycles take a long time. But *ex post* forecasting, where outcomes of interest have already been observed, but are excluded from the training data used to fit a model, is necessary to validate the predictive ability of statistical models, and can be interesting from a theoretical perspective insofar as it reveals the substantive relevance of a theory. In the end, if we have good theories, they should allow us to anticipate future outcomes, whereas the obverse, that an accurate forecasting model implies a good theory, is not necessarily true. As an example consider the range of possible ARIMA time-series models (Box, Jenkins and Reinsel, 1994), which can produce accurate forecasts but require no substantive theory.

Yet in applied research, *ex ante* forecasting must be the goal if a statistical model is to be useful for decision support. Considering that forecasting far into the future is very difficult, relevant data must be available in near real-time in order to create forecasts that are credibly accurate enough to be used by policymakers. A significant amount of research then has focused on processes for finding and generating relevant data efficiently and fast. A major effort has been in the development of machine-coded event data of interactions between actors, sentiment, etc. (Schrodt and Gerner, 1994; Shellman, 2008; Schrodt, 2010*b*) using news services like Reuters. Similarly, the Armed Conflict Location and Events Dataset (ACLED)(Raleigh et al., 2010) and Georeferenced Event Dataset (GED) (Melander and Sundberg, 2011; Sundberg, Lindgren and Padskocimaite, 2010) use human coding of news reports to create conflict event data. A third major source of data, which is only starting to enter the realm of academic research, are government databases like the various U.S. military significant activities databases for Iraq and Afghanistan. Finally, social media, to the extent that it is publicly available, can provide another basis for event data (Zeitzoff, 2011).

With the availability of increasingly complex data exhibiting spatial, temporal, and other forms of dependence have come efforts to develop appropriate statistical models. The range of models with potential applications to forecasting include spatial, timeseries, vector time-series or panel, and count models, with outcomes that include continuous or near-continuous, binary, and count variables. Particular examples of models with application to forecasting include time-series count models (Brandt et al., 2000; Brandt and Freeman, 2006), spatial regression (Ward and Gleditsch, 2008), and vector regression models (Schneider, Gleditsch and Carey, 2011).

A third stream in the academic literature consists of substantive forecasting applications, driven in the past decade by growing interest from the defense community (Weinberger, 2011). Some of this research is the academic side of government-funded projects like the Political Instability Task Force (PITF) (Goldstone et al., 2010) and the Integrated Crisis Early Warning System (ICEWS) (O'Brien, 2010; Schrodt, 2010*b*), which aim to predict events such as regime instability, rebellions, and civil war in countries across the world. There are however examples of purely academic forecasting efforts, e.g. the duration of the Iraq War (Bennett and Stam, 2006) or fatalities during the Persian Gulf War (Cioffi-Revilla, 1991), and retrospective forecasts of civilian deaths during the Bosnian War (Weidmann, 2011).

The three projects in this dissertation are about developing approaches to predict deaths, civilian or military, in inter- and intra-state wars: (1) predicting battle deaths in interstate wars, (2) comparing model fit for spatial Bayesian models of civilian deaths in Bosnia, and (3) forecasting civilian deaths over time in the Iraq War. They share two common themes: (1) a focus on model fit and prediction, and (2) theory-driven model development and specification. All three projects use count data of deaths, disaggregated spatially in the Bosnia paper, and disaggregated over time in the Iraq paper. Although count models are commonly used in contexts where observations can be treated as independent, they are not well developed for spatially or temporally dependent data, and the last two papers develop new Bayesian models to deal with these challenges.

CHAPTER 2

PREDICTING THE LEVEL OF FATALITIES IN INTERSTATE WARS

Bombardment, barrage, curtain-fire, mines, gas, tanks, machine-guns, handgrenades–words, words, but they hold the horror of the world.

- Erich Maria Remarque, All Quiet on the Western Front

It is forbidden to kill; therefore all murderers are punished unless they kill in large numbers and to the sound of trumpets.

- Voltaire, War

2.1 Introduction

Between 1816 and 1997, there were 79 interstate wars involving 281 states as combatants that led to more than 31 million military deaths (Sarkees, Wayman and Singer, 2003). The causes and conduct of these wars have been extensively studied, addressing questions like why and when states fight, how they fight, and how one can explain the outcomes of a war. What is sometimes obscured is the sheer human toll these conflicts cause. Even when it comes to "legitimate" victims of these wars—uniformed soldiers fighting on behalf of organized states—the human cost is large. More than 31 million soldiers lost their lives in battle in this time period. There is however large variation in the wars that contribute to this figure. The Falklands War in 1982 led to 964 battle deaths, while near the other extreme, World War 1 led to more than 8.5 million military fatalities. What explains this variation in the human cost of war? And, given that wars occur, why are some so much deadlier than others?

The answers to these questions are interesting for several reasons. First, presumably an important part of the reason that we study war in the first place is their destructive nature, yet there is large variation in the cost of individual wars that has received comparatively little attention in the quantitative study of wars. If the study of war is interesting because they destroy things and people and cost money, then the question of fatalities should also be intrinsically interesting. The focus in this project is on battle-deaths due to limits in the data available, and these are admittedly only part of the true cost of war. These same limitations, along with the more fundamental problem of quantifying the cost of human death, make it difficult to ascertain how much battle deaths contribute to the overall cost of a war. It seems safe to presume that battle-deaths are a significant part of the true cost of war, and maybe even to presume that they are correlated with the overall cost of a war. And regardless, it provides at least a practically useful starting point for distinguishing wars by their severity.

Second, examining the determinants of war fatalities can inform existing research on war by (1) empirically evaluating the implications such theories have for fatalities, and (2) due to the particular way in which war onset, termination, and duration are operationalized. Bargaining theories of war have implications for the level of fatalities in war in the sense that these constitute part of the cost of war that states bear while fighting. Although the primary motivation of this paper is *not* to evaluate bargaining theories of war (rather, the focus is on systematic prediction of fatalities), the results can be informative for the original theories.

More abstractly, research on war fatalities is related to research on other aspects of war like onset as well due to the way in which the operational definitions of these concepts depend on fatality levels. In any given conflict, war onset is conventionally only coded if fatalities surpass a certain threshold of fatalities, usually 1,000, war termination presupposes that fatalities levels are zero or very low, etc. Research that examines these aspects of war thus also indirectly makes claims about fatality levels, and hence research like this can be useful for it. For example, one way to interpret questions of conflict escalation is that they attempt to determine, given that there is some form of conflict between two states, whether total fatalities will surpass 1,000 or not.¹

And third, the answers to the questions above also hold potential significance for policy-making. Presumably most state leaders confronted with situation that could lead to war will care about how costly and deadly such a war would be if it were to occur. The problem is that predicting how costly wars will be is to my knowledge not straightforward. Before the Persian Gulf War, estimates for the number of fatalities were far higher, up to an order of magnitude, than the approximately 30,000 that did die. Lacina and Gleditsch (2005) give a best estimate of 29,171 for the conflict, whereas Cioffi-Revilla (1991) predicted between 100,000 and 1,000,000 fatalities. Conversely, one may wonder whether the prospect of a Second Gulf War would have encountered more resistance in 2003 had it been obvious that the aftermath would drag out for more than 6 years and cost the lives of more than 4,200 U.S. soldiers and a large but unknown number of Iraqi soldiers, police, militia, and civilians.

The focus in this project is to answer the questions posed in the first paragraph, and

¹The empirical analysis below lacks a temporal dimension, mainly due to the questions it is meant to address, but also due to lack of data. Thus it is not directly comparable to empirical work that examines war onset in cross-sectional time-series data. But it could be used to study the issue of conflict escalation, for example, and more generally the point that there is a link between the study of fatality levels in general and other aspects of war remains.

doing so ultimately is interesting to the extent that one can actually predict fatality levels with some measure of accuracy. If the goal is to explain variation in the levels of fatalities across wars, raw coefficients and significance levels are of secondary interest to substantive effects and how well statistical models can actually predict fatalities. What conditions lead to particularly deadly wars? Thus one of the goals is to identify theoretically reasonable statistical models that perform well in terms of prediction.

2.2 How long will states bear the costs of war?

On a fundamental level, the level of casualties and other costs in a war is a function of the extent to which states will tolerate further loss, since hypothetically they could end a war at any moment by surrendering or making large concessions. In that sense, the question of how many fatalities a war will produce is theoretically similar to the question of how long a war will last(Bennett and Stam III, 1996; Slantchev, 2004). Bargaining theories of war and previous work on war duration suggest that states will bear the costs of war either until the underlying cause of a war, e.g. information asymmetries about relative capabilities, or commitment problems, have been sufficiently addressed, or until a military resolution to a conflict is reached (one side defeats the other). Since fatalities are presumably a significant component of war costs, one should also expect that these factors influence war fatalities as well (Wagner, 2000).

2.2.1 Incomplete Information

The incomplete information explanation of war focuses on the role that uncertainty among state leaders about relative capabilities and resolve plays in their expectations. As long as war is costly, there should always be a bargaining outcome that both states in a dispute should prefer(Fearon, 1995). Of course in practice obstacles such as difficulty in observing material and less tangible capabilities and incentives for state leaders to misrepresent information can lead to unrealistic expectations among states about the proper share due to them, which in turn can lead to costly war (Slantchev, 2010). In this sense, war serves as a mechanism to reveal information about the combatants and ends when it loses its informational value (Filson and Werner, 2002; Powell, 2004; Slantchev, 2003*b*; Smith, 1998).

Wars can convey information through two mechanisms, the battlefield outcomes they entail, and the strategic behavior of the combatants during the war. Because battlefield outcomes are in practice determined independent of negotiations (i.e. they are not easily subject to strategic manipulation), state behavior and negotiation offers are usually more informative in regard to the establishment of a set of reasonable expectations (Slantchev, 2003*b*). They are so however exactly because war is costly, which helps to separate weak from strong states. The fatalities states sustain in a conflict, as a major component of costs, are a key part in allowing war to convey useful and credible information to uncertain combatants.

This leaves open the question of what influences uncertainty and information asymmetries in the first place. Formal models have picked up on a long line of reasoning about balance of power and preponderance of power systems to examine the potential role of different power distributions on information asymmetries. Such a relationship is by no means straightforward, but there is some evidence to suggest that parity is linked to uncertainty (Reed, 2003). Although observable capabilities may be taken into account in pre-war bargaining by fully informed actors, given that two states already are at war, observable capabilities should be related to uncertainty because they influence how important unobservable capabilities are (Gartzke, 1999). Reed gives the example that, regardless of how large it is, Denmark's resolve will probably not influence perceptions of the likely course of a war with the USA, whereas resolve might change things quite a bit if two states are more or less equally matched in observable capabilities. Furthermore, battlefield outcomes themselves are less informative when observed capabilities are near parity and as a result it will take more battles and more deaths for credible information to be conveyed (Slantchev, 2004, 816). Thus when observed capabilities are at or near parity, states will face a relatively high level of uncertainty and will tolerate more costs:

H 1. Parity in observable capabilities increases war deaths.

Abstract discussion of war tend to focus on conflict between two states, but a significant number of wars involve more than two states. This adds additional sources of uncertainty to power calculations. There will be strategic interaction among states fighting on the same side which might influence how observable capabilities correlate with the unobservable true capability of one side. At the very least they will face common collective action problems. Thus wars fought between multiple states should increase uncertainty due to the potential strategic behavior involved, which in turn should make the main combatants willing to sustain higher fatalities than they otherwise might. **H 2.** As the number of states involved in a war increases, so will war deaths.

2.2.2 Commitment Problems

The second major rationalist explanation of war focuses on commitment problems which arise when states cannot reach a peaceful settlement because it is known or believed that one of the states involved will have incentives to deviate from the settlement in the future. War in this case arises because although a peaceful settlement is obvious to both sides, at least one of them for some reason is known or thought to have an incentive to renege on the settlement in the future because it will think that it can gain a more favorable bargain by doing so. As a result, neither side can credibly commit to upholding the current settlement under negotiation.

The archetypical reason underlying states' belief that they face a commitment problem is shifts in bargaining power over time, e.g. changes in observed capabilities or resolve (Powell, 1996, 2004). Other situations that may lead to commitment problems include significant surprise or first-strike advantages, or bargaining over issues which themselves can change the balance of power between adversaries (e.g. the Golan Heights between Israel and Syria) (Powell, 2006). The key in all three of these situations is that they hold the potential for rapid shifts of power that underlie commitment problems. When the stakes are high enough and large enough changes in power are looming, war arises because states will just not be able to resist the temptation of fighting to get all they want (Leventoğlu and Slantchev, 2007). Knowing this, states choose to engage in war instead.

Credible information provided through battlefield outcomes and state's behavior in response to them plays less of a role in resolving wars caused primarily through commitment problems. Rather, war incidentally resolves the problem by destroying enough of what states are fighting over to eventually make peace possible. (Leventoğlu and Slantchev, 2007, 767). Instead of dividing the pie, states fight war long enough to destroy so much of it that both can credibly commit to a peaceful settlement because there just is not enough left to make war an attractive gamble. This suggest that in situations where there is a credible commitment problem, states will be willing to endure fatalities and other costs of war in proportion to what is at stake. The more that is at stake the longer it will take to destroy enough of it to "sour" the prospect of continuing war. This is is consistent with the argument that the more important the issues underlying the dispute, the more states will be willing to sacrifice in order to get their share (e.g. Hensel et al., 2008): **H 3.** *High stakes in war increase deaths.*

2.2.3 Military decision

War can resolve obstacles to peaceful bargaining, but ultimately it is also a military contest that exposes states to the risk of collapse or military defeat (Reiter, 2003, 30). Regardless of what caused World War 2 in Europe, the occupation of almost all of Germany before the surrender ended it. Although most wars terminate in negotiated settlements, a significant number are decided militarily (Pillar, 1983). This suggests that military resolutions to a war are important for models of war termination or the costs of war—some wars simply stop because one side looses the ability to fight. The fatalities in a war should thus be constrained by factors that influence how long a state can keep fighting.

The most straightforward way to eliminate another state's ability to continue a war is to destroy or sufficiently weaken its armed forces. Rough terrain generally should make this harder to achieve since it makes large, conventional battles more difficult (Bennett and Stam III, 1996). This can allow combatants more time to capitalize on their reserves or to fight prolonged wars using non-conventional means (Fearon and Laitin, 2003). Wars fought in some types of rough terrain such as urban areas are also associated with higher levels of fatalities.

H 4. As the proportion of rough terrain increases, so will deaths.

Undermining the willingness of an enemy to continue fighting is an alternative to this brute force approach, and a very common argument relating to the ability of some states to bear the costs of war concerns the aversion of democracies to casualties (Bueno de Mesquita and Siverson, 1995; Filson and Werner, 2004; Gartner, 2008). If casualties be-

come too high, continuing a war becomes politically infeasible under certain regime types. There is empirical evidence to support this argument in the case of the United States and it seems to also be the case in other democracies (Carson et al., 2001; Gartner, 2008). As a result of this loss aversion, democracies will not only be unable to fight costly wars, but they might also self-select into fighting only short, easily winnable wars: **H 5.** *Deaths will be lower if one of the combatants in a war is a democracy.*

Another potential indication of a state's ability to bear or inflict costs lies in its selection of strategy to fight the war. Strategies aimed at producing quick military victories often indicate the unwillingness or inability of a state to wage prolonged warfare (e.g. Germany in World War 2), while strategies based on prolonged insurgency warfare can only be used by states with a high capacity for absorbing costs (e.g. North Vietnam during the Vietnam War):

H 6a. Wars in which at least one of the combatants uses a blitzkrieg-style strategy will produce less fatalities.

H 6b. *Wars in which at least one of the combatants uses a guerilla-style strategy will produce more fatalities.*

2.3 Data and methods

The data consist of 90 interstate wars from 1815 to 1991 with 910 or more battle deaths. The list of wars is based on the Correlates of War list of interstate wars (79 wars), with two changes (Sarkees, Wayman and Singer, 2003). It includes two additional wars that are consistent with COW state system membership (Slantchev, 2004), the Pastry War and Uruguayan Dispute. Furthermore, World War 2, the Vietnam War, and the Persian Gulf War are disaggregated into multiple smaller wars totaling 15 in number. The reasoning underlying this change is twofold. First, it makes this work empirically more consistent with previous work on war duration that uses similar or the same set of cases. Second, Bennett and Stam III (1996) argue that the cases which were split really consisted of distinct wars in the sense that actors in each distinct episode probably did not take into account the events that would later occur in the larger war. For example, when Germany invaded Poland in 1939, Hitler almost certainly did not expect the UK and France to respond forcefully in an effective manner. From the resulting 93 wars, 3 drop out of the final sample because fatalities were well below the traditional threshold of 1,000.

In addition to listing one major combatant for each of the two sides in a war, the data also include other combatants on each side from the list of states in the COW interstate war participants data and other historical references (Sarkees and Schafer, 2000; Dupuy and Dupuy, 1986; Holsti, 1991). The list of combatants is largely consistent with the COW interstate war participants list, unless Slantchev (2004) or Bennett and Stam III (1996) listed a state that was not included in the COW participants list as one of the major combatants. Historical sources were used to identify combatants for the added or disaggregated wars (e.g. World War 2) and to determine on which side of a conflict a

state fought.

Out of the total 93 wars, 35% (33 wars) were fought by 3 or more combatants, with the Korean War having the largest number of combatants (16–2 on the side of North Korea, and 14 on the UN side). The disaggregated World War 2 conflicts account for a large number of the remaining multilateral wars.

There are two rationales underlying this coding. First, as historical sources and the COW participants list indicate, a substantial number of wars genuinely were multilateral in the sense that more than 2 states contributed significant manpower and matériel to the conflict—e.g. consider the First World War or any of the Arab-Israeli wars. Thus only counting the material capabilities, population, etc. of a single state could be a misrepresentation of the true military situation prior to the outbreak of the war. Second, the addition of more combatants to a conflict likely influences the uncertainty of information for decision makers on either side. As a result, the explanatory variables dealing with material capabilities are calculated not only for the two major combatants, but for all states involved in a conflict.

The dependent variable consists of fatality data from three sources. Lacina and Gleditsch (2005) provide the most recent fatality data. These particular data only reach back to 1900, which covers 49 of the 93 wars. To obtain fatality estimates for all wars in the sample I used those from Slantchev and the COW project for the remaining 44 wars (Sarkees and Schafer, 2000; Slantchev, 2004). The correlation between fatality estimates from the two sources is 0.98 for the 49 wars in which the data overlap, suggesting that this is reasonable to do. The resulting fatality data range from a low of 12 (the German invasion of Denmark in WW2) to slightly above 10 million (the Great Patriotic War between Germany and the USSR in World War 2). Usually a conflict is coded as a war when battle deaths surpass 1,000. The Falklands War is coded as having 910 fatalities by COW, and using this as the threshold for annual battle deaths, the number of wars in the sample is reduced from 93 to 90.

2.3.1 Estimation strategy

Fatality figures are unlikely to be the result of a normal data generating process, and instead seem to follow a power law or similar type of distribution, where more serious wars will be less frequent and *vice versa* (Cederman, 2003; Levy and Morgan, 1984; Clauset, Shalizi and Newman, 2007). Existing studies of (civil) war fatalities use OLS regression of the log of fatalities to take this into account (Cioffi-Revilla, 1991; Lacina, 2006). This allows the explanatory variables to have what effectively is an exponential effect on the number of fatalities in a war: if $\ln y = x_i\beta$, then $y = e^{x_i\beta}$.

However, the fatality data used here are also truncated in the sense that only conflicts in which fatalities exceed 910 battle deaths are included. Regular statistical models like OLS that do not take this into account can produce biased and misleading results. A truncated regression model corrects this issue by rescaling the normal distribution to account for the truncation (where τ is the truncation point, ϕ is the standard normal pdf, and Φ is the standard normal cdf) (DeMaris, 2004; Greene, 2008, 756-761):

$$f(y|y > \tau) = \frac{\frac{1}{\sigma}\phi(\frac{y - x_i\beta}{\sigma})}{1 - \Phi(\frac{\tau - x_i\beta}{\sigma})}$$
(2.1)

This in turn leads to the log-likelihood function for truncated normal regression:

$$lnL = \sum_{i=1}^{N} \left(\ln\left[\frac{1}{\sigma}\phi\left(\frac{y - x_i\beta}{\sigma}\right)\right] - \ln\left[1 - \Phi\left(\frac{\tau - x_i\beta}{\sigma}\right)\right] \right)$$
(2.2)

Because the sample size of 90 is small, I use bootstrapped estimates to evaluate statistical significance. Bootstrapping does not rely on the asymptotic assumption of normality for errors and instead uses repeated re-sampling with replacement from the observed data to estimate errors. Significance levels for coefficients are determined by using 95% and 90% bias-corrected accelerated (BC_a) confidence intervals, which correct for bias and skewness in the bootstrap samples (DiCiccio and Efron, 1996).

Finally, war duration is a confounding factor that is related to both fatalities (longer wars are deadlier on average) and some of the independent variables. Since war duration is observed only after a war has concluded, it cannot figure in states' decision to continue fighting (Slantchev, 2004). Instead, it is possible to model and estimate war duration itself using observed conditions at the outset of a war. These include parity and total military personnel, parity and total population, number of combatants, rough terrain, and whether a combatant was democratic (table 2.2). Predicted war duration from this model in turn is used as a further independent variable in the fatality models (Slantchev, 2004).

The overall estimation strategy thus consists of three distinct steps: (1) estimate a duration model and generate estimates of war duration, (2) estimate a truncated normal regression model of war deaths and generate estimates of war deaths, and (3) repeat the first two steps 1,000 times to obtain bootstrap estimates for all parameters and model predictions.

2.3.2 Explanatory variables

Explanatory variables are measured during the beginning year of a war, except the number of states involved in a war and the resulting capabilities measures.

Parity in observed capabilities. Parity in observed capabilities is measured with two variables: active military personnel (in thousands) and COW's composite index of national capability (CINC). Both come from the COW National Material Capabilities data (Singer, 1987; Singer, Bremer and Stuckey, 1972). Military personnel are one of the easiest observable indicators of a state's power prior to a war, although they do not take quality into account. The composite index of national capability is more sensitive to this issue since it is based on military personnel, military expenditures, industrial capacity, and several other relevant variables. Each variable is used to construct a separate measure of parity. In the case of wars that have multiple combatants on a side, I add

the active military personnel or CINC figures for all states fighting on the same side in a war. Parity is calculated as the difference between observed capabilities of the stronger state and the weaker state relative to the sum of their capabilities. A 0 indicates complete superiority by one side, and a 1 indicates perfect parity.² The average war was initiated by a state or coalition with 430,000 military personnel (median) against a target state or coalition with 153,000 military personnel at the starting year of the war. The mean values for parity in military personnel and parity in CINC scores are 0.44 and 0.43 respectively.

Number of states. The total number of states involved in a war, as coded by the COW interstate war participants list (Sarkees and Schafer, 2000). A few wars have additional combatants if it appears that those states made a significant contribution in manpower or logistics (Dupuy and Dupuy, 1986). The variable ranges from 2, i.e. wars fought solely between two states, to 16 combatants for the Korean War.

Issue salience. Issue salience is identified for each side in a war with the following coding scheme: high salience if the issue is regime or state survival, national liberation, or autonomy; medium salience if the issue involves territory, integrity of state, or honor/ideology; and low salience if the issue is maintaining an empire, commercial disputes, or policy (Slantchev, 2004). Each of the two variables measures the salience of the issue underlying the conflict for one of the two major combatants.

Rough terrain. This variable identifies the extent to which a war was fought over rough terrain, e.g. heavy woods, jungles, swamps, and mountains (Slantchev, 2003*a*; Stam III, 1999). Higher values for the variable indicate that a war featured more combat in such types of terrain.

Democratic combatants. I measure democracy using data from the Polity IV project (Marshall and Jaggers, 2002). First I created a polity score that equals the democracy score minus the autocracy score for each of the two major combatants in a war in the year that the war started. The resulting index ranges from -10 to 10. In cases where there was a regime transition or other turmoil, I use the democracy or autocracy score from the last year available. The final measure is dichotomous and indicates a democracy if a state had a polity score of 6 or higher, otherwise an autocracy. Thirty-two wars involved a democratic state as one of the two main combatants.

Strategy. Bennett and Stam measure strategy using a fairly complex variable that captures which side pursued and offensive/defensive doctrine and whether the particular strategy consisted of maneuver warfare, attrition, or punishment (e.g. insurgency) (Bennett and Stam III, 1996). I used their measure as well as historical references to construct two dummy variables indicating whether any side used either a maneuver strategy (19 wars) or a guerrilla strategy (5 wars) (Dupuy and Dupuy, 1986). The majority of wars were fought between opponents who both used attrition strategies (69 out of 93 wars), and there were no instances were at least one of the combatants did not employ an attrition strategy.

Control Variables. All regression models include three variables to control for exposure effects ((e.g. deaths cannot exceed the total population of all combatants) and

²Specifically, $1 - \frac{p_{stronger} - p_{weaker}}{p_{stronger} + p_{weaker}}$, where *p* is either military personnel or CINC.

relative population size: *total observed capabilities*, i.e. the sum of observed capabilities, either military personnel or CINC score, for all combatants involved in a war; *parity in total population*, i.e. how close to parity in population each side in a war is; and the natural log of *total population*, the sum of all combatant states' prewar population.

Duration model. The dependent variable is the duration of a war in days (Slantchev, 2004). The covariates used are similar to those in Slantchev's full model. They include parity in military personnel, parity in total population, total military personnel, total population, the number of actors, rough terrain, contiguity, and whether a democratic state was involved as principal combatant. I use the natural log of total military personnel and total population as those produced better fitting models. The purpose of this duration model is to generate predicted duration values for each war to use in the main regression models.

2.4 Empirical results

The duration model which was used to generate predictions that are used as an independent variable in the war deaths models is shown in table 2.2. Based on AIC scores, a log-normal form provided the best-fitting model. As with the war deaths models, the estimates are bootstrapped due to the small sample size. The results, in terms of specific associations and significance, are roughly consistent with previous models of war duration (Bennett and Stam III, 1996; Slantchev, 2004), suggesting this is a theoretically reasonable model. Parity in military personnel, the number of states involved, and terrain are associated with significant increases in war duration.

Moving on to war deaths, table 2.3 shows estimates for the two main truncated regression models, using either military personnel or CINC to measure parity. Table 2.4 shows coefficient signs and significance for several further models to check for robustness of results.

Raw coefficients in a truncated regression model provide the marginal effect of a variable in the underlying, unobserved population of cases, which here would mean all conflicts including those short of war (Greene, 2008, 867–869). Calculating marginal effects for just the truncated sample (i.e. wars) requires adjustment by a parameter that is bound by 0 and 1, depending on the extent of truncation. Thus the magnitude, but not the direction of effects from these results would be somewhat closer to zero in regard to traditionally-defined wars. Exponentiated coefficients for those that are significant show the factor change in observed fatalities associated with a one unit change in the independent variable.

2.4.1 Regarding the hypotheses

Two variables stand out for having highly significant and consistent effects across both models (and other specifications; table 2.4)—democracy and issue salience for the initiating state. Wars that involve at least one democracy appear to be consistently and significantly less deadly, with an average reduction in fatalities by a factor of about 0.2. Whether it is because of self-selection or internal politics, democracies tolerate far less deaths in wars than autocratic regimes.

Higher issue salience for the initiator increases deaths on average by a factor of between 5 and 7 times for a one unit increase in salience (it ranges from 0 to 2). However, the coefficient for issue salience for the target state is always insignificant. So it appears that issue salience for the initiating state matters, while issue salience for the target state does not. Maybe states that have a lot at stakes also tend to be the ones that initiate wars, to exploit the element of surprise or some other first-strike advantage. The data suggest otherwise. In most wars (82%), the target has more or roughly as much at stake as the initiator. War deaths thus seem to be driven by the extent to which the initiator values the issue over which it is fought, regardless of the target states stakes.

The two hypotheses relating uncertainty and fatalities overall receive only mixed support. The coefficients for both parity in observable capabilities and the number of states in a conflict are positive and significant when using military personnel, but not with the CINC score. One reason for this may be related to the fact that the variables are all measured in the first year of the conflict, so that if their values and hence state's expectation change over the course of a conflict, the models here are unable to account for any effect these changes have. Another alternative has to do with the measures themselves: while military personnel is a fairly straightforward measure of military capabilities, the CINC score is more broad and less observable in real time.

What about military resolutions to war? Support for a relationship between factors that influence the military conduct of a war and consequent fatality levels are also mixed. There is a very clear indication that wars involving a democracy as major combatant, all else being equal, have lower fatalities than those fought purely between nondemocratic regimes. On the other hand, terrain has a significantly positive impact in the first model, but not in the second and strategy does not seem to have a significant impact when controlling for other factors.

2.4.2 Robustness checks

The main results—significant effects for democracy and issue salience, a possible effect for parity in military personnel and the number of states—remain consistent across several different model specifications.

The first set of robustness checks both change the estimation strategy by either using a negative binomial regression instead of truncated regression in the second step of modeling war deaths, or does away with the first step, generating war duration predictions, away altogether by using actual duration. Both approaches simplify model estimation. Using a negative binomial regression (models 3 and 4 in table 2.4) to model counts of war dead leads to a significant effect for rough terrain, but otherwise leaves the results largely unchanged. The in-sample predictions become notably worse however. Using actual duration renders military personnel parity insignificant, and leads to a positive and significant effect for maneuver strategies and observed war duration. The variable used to measure issue salience in the base models is ordinal with 3 possible values. Using it as is assumes that the effect is linear, i.e. going from a low value to moderate value (0 to 1) is equivalent to going from a moderate to high value (1 to 2). Using dummy variables for moderate and high value issues removes this assumption. Both of the dummy variables for initiator issue salience have a significant, positive effect on war deaths, while neither of the two dummy variables for the target state has a significant effect. However, parity using either military personnel or CINC, the number of combatants, and terrain now have significant effects and are associated with deadlier wars.

The last four models look at how sensitive the results are to the inclusion of the World War 2 conflicts and World War 1, since these represent the major changes to the data and some of the deadliest wars in the data as well. The main change is that the number of combatants in the new subsample of wars has a consistent positive effect on casualties. Out of the wars excluded from this subsample are 4 conflicts with more than 3 combatants, of which 3 have a million or more casualties. Both parity variables, using military personnel or CINC, also show positive and significant effects.

2.4.3 Regarding overall model fit

Theoretical contributions aside, the actual fit of statistical models is of interest here for pragmatic reasons. Being able to accurately forecast war fatalities is practically use-ful to know. To assess how well the models do in predicting actual data, I estimated (1) in-sample predictions (which are *ex post*), as well as (2) out of sample forecasts for the Eritrean-Ethiopian War from 1998–2000 and the Kargil War in 1999. Fatality estimates are calculated during the bootstrapping process, which deals with the problem of calculating accurate fitted values in log-normal models using point estimates (Duan, 1983).

The in-sample predictions consist of point estimates for fatalities as well as the 95% interval for those point estimates.³ Point predictions for these data are not the only information that is interesting since the chances that a prediction will absolutely match actual fatalities is near zero. Intervals, if accurate, on the other hand can provide a better sense of the likely range fatalities span through their size. For example, two wars might have identical point fatality estimates, but if one has a much smaller interval than the other we can be much more confident in the general range that war is going to produce. The 95% interval for estimated fatalities from model 1, in relation to observed fatalities, are shown in figure 2.1. The y-axis shows fatality figures (on a logarithmic scale) for each of the 90 wars in the estimation sample. The red dot corresponds to the observed fatalities for that war, while the blue lines indicate the 95% interval for the predicted number of fatalities. The list of wars is sorted by the observed number of deaths, with less deadly wars on the left and deadlier wars on the right.

Statistical models of war, e.g. war onset or termination, tend to be limited in their ability to accurately predict observed values. Here, the 95% intervals of predicted values tends to, on average, do fairly well in capturing observed fatalities within it. The

³Estimates are linear predictions, i.e. $\ln \hat{y} = X\beta$.

predicted deaths intervals bracket observed fatalities 69.89 percent and 65.59 percent of the time for models 1 and 2 respectively. The point estimates for fatalities, using median value, are correlated with observed fatalities at 0.63 and 0.60 respectively.

On the other hand, the intervals are fairly wide. In three-quarters of the wars they span between 1 and 2 orders of magnitude (log_{10}) , which can translate to a difference between 100 and 10,000 deaths at the low end of the scale. One of the better predictions is for the Yom Kippur War 1973. Here the median prediction is 5,800 with an interval from 900 to 22,700—observed fatalities were 11,084. But at the other extreme the interval for predictions for the Korean War spans 5.4 orders of magnitude with a nonsensical upper estimate of deaths. So there certainly is room for much improvement in the quality of these estimates.

Compared to earlier efforts to predict the magnitude of war, the estimates seem reasonable in accuracy and variance. Similar efforts to predict war deaths in the Persian Gulf War in 1991 for example produced a best estimate of 100,000 and a few million (Cioffi-Revilla, 1991). Actual, observed fatalities were far lower, around 29,171 (plausible range of 28,945 to 44,271). Table 2.5 shows the fatality estimates produced by the two base models. Both models bracket observed fatalities, although the first model has a very high upper limit close to 1.2 million. Neither model, looking at the upper limit, would have entirely ruled out some of the very large fatality figures that were heard in the lead up to the war. The median, or guest guess, estimates of 46 and 13 thousand dead are close to the observed 29 thousand fatalities however.

Part of the reason for the over-prediction from the two base models lies with the large number of states that participated in this war as members of the US-led coalition. Fatality estimates for the Korean War, which also had a very large number of combatants, are similarly large relative to observed fatalities, and in fact they are by far the highest estimates generated by the models in absolute terms. The Korean War and Gulf War have respectively the largest and third largest number of combatants out the wars in the sample with 16 and 14 respectively. But compared to other wars with many combatants, both wars stand out as being fought primarily on one side by the U.S. at the forefront of coalitions sanctioned by the United Nations, and so the apparent large number of combatants distorts actual contribution of effort.

2.4.4 Out of sample predictions

How well would these models actually do in forecasting fatalities in a future war? Clearly, we would have a little bit of a problem in evaluating forecast accuracy for wars that have not occurred yet. But since the data used so far end in 1997, we can still do out of sample forecasts but with the benefit of having observed actual events since then. In other words, we can pretend it is roughly 1998/99 instead and create out of sample forecasts for hypothetical wars between Eritrea and Ethiopia (Eritrean-Ethiopian War, 1998–2000) and India and Pakistan (Kargil War, 1999). Using the statistical models above and data collected in the beginning years of these two wars, how accurately would we have predicted war deaths? The first step is to code some missing data for our independent variable. The Eritrean-Ethiopian War was fought over disputed territories (issue salience = 1, moderate, for both sides) on the countries' arid, mountainous border (rough terrain = 1) and both countries had conventional militaries (strategy = attrition). Ethiopia had both a larger military, with 200,000 against 100,000 troops (parity = 0.67), larger capabilities, and larger population with 50 million compared to Eritrea's 3 million. Neither side had a democratic government. With these inputs, the two base models provide median estimates of 21 and 15 thousand deaths, with a range from the low thousands to 80 and 96 thousand deaths, as shown in Figure 2.2 (see table 2.5 for the exact numbers). With the benefit of hindsight, actual deaths in the war were 50 thousand, and both models correctly bracket that number.

The Kargil War between India and Pakistan in 1999 was fought over a Pakistani incursion into mountainous terrain near Kashmir (issue salience = 1, rough terrain = 1). Both states used an attrition strategy. India had a democratic regime and was superior to Pakistan on all measures of capabilities and size (parity = 0.62). The resulting median estimates are 15 and 7 thousand respectively. The observed number of fatalities is between 884 and 4,527+ deaths, which provides a mean estimate of 2705 fatalities.⁴ Both models bracket the number of 2705 fatalities, but generally over-predict fatalities in the war.

Notably however, the Kargil War is the only war to date that has been directly fought between nuclear powers. Fighting in the war was restricted to a relatively small geographic area and it did not escalate into a general conflict between India and Pakistan. One can speculate that fatalities were low because both sides were reluctant to escalate and risk fighting a nuclear war. Since this is the only war so far fought directly between two nuclear powers, there is no way to take something like mutually assured destruction into account. In an interesting bit of speculation, maybe the predicted fatalities listed in the table are close to the level of fatalities we would have seen in a 1999 Kargil War between Pakistan and India had they not been nuclear-armed.

2.5 Discussion and conclusion

The statistical models employed here were derived from theoretical bargaining models about conflict. As I argued in the introduction, although they do not directly make arguments about war dead, they do include the cost of conflict. If one can accept that deaths are a significant part of war costs, then by extension these models should have implications for the level of war dead as well. Furthermore, those that deal with concepts like war onset and duration, some of which have been empirically tested in these contexts, also implicitly assume arguments about war dead due to the fact that onset, termination, and duration are typically operationalized using fatality counts.

⁴There are no clear fatality figures available for the Kargil War, mainly because Pakistani casualties are unclear. India claims 527 fatalities of its own, but Pakistan claims anywhere from 357 to 4,000 plus.

The positive, significant, and substantively interesting relationship between democracy and deaths is consistent with colloquial and academic arguments about the inability of democracies to tolerate war deaths. The causal mechanism is open to question but two possible explanations are that democracies either end wars that prove to be costly in human life (at least that of their own citizens/soldiers), or that they self-select into only fighting certain wars that are less likely to produce large fatalities. Indeed, democracies appear to fight shorter wars, although they are not more likely to win.⁵ Another factor that may explain this result is that democracy, aside from influencing the ability of a state to sustain high fatalities and continue fighting a war, may also help resolve information asymmetries surrounding war through the mechanism of audience costs and the transparency inherent in democratic procedures.⁶

Issue salience for the initiating side is also strongly associated with fatalities. Although it is probably difficult to distinguish which side in a conflict will initiate a war beforehand, the finding does suggest that particular attention should be paid to evolving conflicts between states that involve highly salient issues on either side. Further, the role that issue salience plays in prolonging wars fought over commitment problems suggests that third-party mechanisms to address these problems, like monitoring or enforcement, may be particularly apt for avoiding the escalation of such conflicts into fullscale war.

Variables related to information asymmetry, i.e. parity in observable capabilities, were less clearly related to war deaths. There is some indication that parity in military personnel increases deaths, but using CINC scores there is no relationship. Empirically, this might be because CINC scores are a more complex measure of military strength. Theoretically however the absence of a clear statistical relationship could be because the effect of parity is conditional on other factors like relative war costs.⁷ and the extent of commitment problems.⁸

Overall, the coefficients and statistical significance proceeded by these bargaining derived statistical models overall thus show limited evidence for relationships between bargaining conflict theories and fatalities. But looking at how these models match reality through in sample fit and out of sample forecasts does show that in the aggregate bargaining theories provide a useful starting point for specifying statistical models of war dead. Underlying this point is the fact that statistical significance can be as much a result of sample size as of a true association, and that by themselves, significant coefficients do not mean that a model matches the real world well.⁹ The fact that the model predictions here do seem to be accurate and useful for forecasts, compared to the standards of conflict research on war onset for example, speaks to their usefullness to both theory and applied research.

From both an academic and future policy perspective, the most important shortfalls

⁵Bennett and Stam III (1996); Slantchev (2004)

⁶Fearon (1994); Schultz (1999); Slantchev (2010)

⁷Wittman (2009)

⁸Wolford, Reiter and Carrubba (2011)

⁹Schrodt (2010*a*)

of this effort are that war dead are aggregated and do not distinguish among combatants, and that the model predictions are broad, typically spanning 1 or 2 orders of magnitude. Both of these issues are driven by lack of finer-grained data. Including lowerscale conflicts and disputes would do much to reduce uncertainty, assuming the same mechanisms were at work in producing deaths as in large-scale wars. Practical issues with missing and inaccurate data on fatalities makes this a difficult, but possible, future effort.

Despite these shortcomings, from an applied perspective, being able to forecast war dead even if only roughly is useful. Faced with a potential conflict, where would one even begin in predicting the likely human toll? This effort provides a reasonably accurate starting point that can be used to anchor more educated guesses specific to any given case.



Figure 2.1: Predicted and observed war deaths for the best-fitting model. The *x*-axis lists the wars in the sample, with the observed number of war deaths indicated by the red dot. The list is ordered by the total number of deaths, ranging from the Falklands war to the eastern front of World War 2. The number of war deaths is on a logarithmic scale. The blue bars show the interval containing 95% of the predicted war deaths in each bootstrap iteration. The intervals correctly bracket observed deaths for 69% of all wars.



Figure 2.2: Predicted and observed war deaths for selected wars. The predictions for the Eritrean-Ethiopian and Kargil wars are out of sample forecasts. Bars show low and high estimates for observed fatalities and 95% range for model predictions. Points show best or median estimate for observed and predicted fatalities respectively. The scale is logarithmic with base 10, so for example $10^3 = 1,000$ deaths.

Source	(Lacina and Gleditsch, 2005)	(Singer, 1987; Singer, Bremer and Stuckey, 1972)			(Singer, 1987; Singer, Bremer and Stuckey, 1972)					(Sarkees and Schafer, 2000; Dupuy and Dupuy,	1986)	(Slantchev, 2003 <i>a</i> ; Stam III, 1999)	(Marshall and Jaggers, 2002)		(Slantchev, 2004)		(Bennett and Stam III, 1996; Dupuy and Dupuy,	1986)	(Singer, 1987; Singer, Bremer and Stuckey, 1972)		(Singer, 1987; Singer, Bremer and Stuckey, 1972)	(Singer, 1987; Singer, Bremer and Stuckey, 1972)	(Singer, 1987; Singer, Bremer and Stuckey, 1972)	(Singer, 1987; Singer, Bremer and Stuckey, 1972)		
Indicator	Number of military battle deaths.	Ranges from 0 (one side dominant) to 1 (equally	strong). Calculated using active duty military	strength for all combatants at beginning of the war.	Ranges from 0 (one side dominant) to 1 (equally	strong). Calculated with composite index of na-	tional capabilities (CINC), which itself uses mili-	tary personnel, military expenditures, and indus-	trial capacity as inputs.	Number of states that fought in the war.		Ranges from 0 (open country) to 1 (rough).	0 or 1. For one of the major combatants, was Polity	Score (democracy-autocracy) >6 at start of war?	2 if state/regime survival, 1 if territory or ideology, 0	if maintaining empire, commercial, policy disputes	1 if at least one side used maneuver or punishment	strategy, 0 otherwise; reference is pure attrition war	Sum of military personnel for all combatants.		Sum of all CINC scores.	Parity in total population.	natural log of sum of each combatant's population.	0 or 1	War duration in days from duration model (table	2.2).
Variable	War deaths	Parity military per-	sonnel		Parity CINC					Number of states		Rough terrain	Democratic com-	batant	Issue salience		Strategy		Total military per-	sonnel	Total CINC	Parity population	In Total population	Contiguity	Predicted duration	

Table 2.1: Variable descriptions and sources

	β	95% CI
Parity military personnel	1.570**	(0.483, 2.741)
Total military personnel ^a	-1.138	(-0.437, 0.179)
Parity total population	-0.947	(-2.085, 0.182)
Total population ^a	0.134	(-0.218, 0.492)
Number of States	0.118^{*}	(-0.008, 0.266)
Rough terrain	3.186**	(2.009, 4.433)
Democratic combatant	-0.201	(-0.813, 0.412)
Constant	2.305^{*}	(-0.423, 4.791)
σ	1.247^{**}	(1.168, 1.406)
No. observations	90	

 Table 2.2: Log-normal regression of war duration

Wald χ^2 72.30^{**} Notes: *a* natural log used. Log-normal duration model of war duration in days as dependent variable. Bootstrapped bias-corrected accelerated (BC_A) confidence intervals. Significance levels (two-tailed): * $p \le 0.10$, ** $p \le 0.05$.

		Mod	del 1			Mode	el 2	
	β	$\exp(\beta)$	959	6 CI	β	$\exp(\beta)$	959	6 CI
Parity military personnel	2.494^{**}	12.11	(0.690)	6.142)				
Parity CINC					2.165		(-0.560,	5.455)
Number of States	0.290^{**}	1.34	(0.025,	0.663)	0.190		(-0.116,	0.389)
Rough terrain	2.762^{*}	15.83	(-0.007 ,	8.404)	1.934		(-1.465,	5.454)
Democratic combatant	-1.570**	0.21	(-3.081)	-0.135)	-1.727**	0.18	(-3.112,	-0.556)
Issue salience, initiator	1.639^{**}	5.15	(0.723,	2.787)	1.975^{**}	7.21	(0.963,	3.048)
Issue salience, target	-0.039		(-1.207 ,	0.974)	0.102		(-0.894,	1.132)
Maneuver strategy	0.787		(-0.787 ,	3.010)	1.114		(-0.328,	3.056)
Punishment strategy	1.589		(-1.427 ,	5.468)	1.614		(-1.659,	4.775)
Total military personnel	0.0002^{*}	1.0002	(-0.0001,	0.0005)				
Total CINC					6.493^{**}	660.50	(2.506,	12.733)
Parity total population	-1.644^{*}	0.19	(-4.823 ,	0.249)	-2.404^{*}	0.09	(-5.398,	0.185)
ln Total population	0.464^{*}	1.59	(-0.028 ,	1.048)	0.345		(-0.120,	0.851)
Contiguity	-0.336		(-2.532 ,	0.988)	0.402		(-1.368,	1.901)
Predicted duration	-0.001		(-0.007 ,	0.001)	-0.0004		(-0.005,	0.002)
Constant	-0.773		(-9.631,	4.928)	0.265		(-6.602,	6.045)
σ	1.826^{**}		(1.642,	2.503)	1.785^{**}		(1.594,	2.324)
Me chemitan		C	9			00		
INO. ODSELVAUOIIS		כנ	2			90	_	
Wald χ^2		63.7	**62			75.22	**	
Notes: Dependent variabl	e is <i>ln</i> of fa	atalities.	Bootstrap	ped BC _a c	onfidence i	ntervals.	Signifi-	

cance levels (two-tailed): * $p \le 0.10$, ** $p \le 0.05$.

Table 2.3: Truncated normal regression of war fatalities

Coefficient	-	5	с С	4	5 L	9	2	8	6	10	11	12
	Ba	se	Neg	ative	Act	ual	Dum	mies	Exclu	Iding	Exclu	lding
			Bino	mial	Dura	ation	for I	ssue	M	N2	/IWW	'WW2
Milper parity	++		+				+++		+		+	
Total milper	+						+++++++++++++++++++++++++++++++++++++++					
CINC parity				+				+		+		
Total CINC		+ +		+ +				+++				
Pop parity	I	I		 				 	 	 	I	I
Total pop	+		‡ +	Ι	+++	+++++			+++		‡ +	+ +
Number of states	+ +		+++++++++++++++++++++++++++++++++++++++				+ +	+ +	+ +	+	+ +	+++++
Terrain	+		‡	‡ +			+ +	+				
Dem	 	 	 	 	 	 	 	 	 	 	 	
Salient 1	+ +	‡ +	‡	‡ +	+ +	+ +	+ +	+ +	+ +	+ +	+	‡
Salient 2												
Maneuver					+	+ +						
Punishment									+		+	
Contiguous												
Duration				I	+ +	+ +						‡
Observations	06	06	90	06	06	90	90	06	80	80	79	29
Fit	69.89	65.59	54.84	58.06	66.67	63.44	72.04	63.44	64.52	67.74	62.37	68.82
Notes: Signs denot	e wheth	ier a coe	efficient	had a si	gnificar	nt positir	ve or ne	gative el	ffect, wi	th one s	ign for	
90% significance a	nd two	for 95%	signific	ance, e.g	g. ++ dei	notes a p	positive	effect si	gnifican	it at 95%	ó level.	
Significance levels	are deri	ived fro	m boots	trapped	$ BC_a co$	nfidenc	e interv	als.				

Table 2.4: Variable significance
	Lower bound ^a	Median estimate	Upper bound ^a		
Persian Gulf War, 1991; observed: 29,171 (28,945 to 44,271)					
1	2,219	45,685	1,198,118		
2	450	13,435	162,274		
Eritrean-Ethiopian War, 1998–2000; observed: 50,000 (no interval)					
1	3,450	20,617	96,032		
2	1,797	14,543	80,202		
Kargil War, 1999; observed: 2,705 (884 to 4,527+)					
1	2,131	15,146	119,597		
2	630	7,382	70,669		

Table 2.5: Model predictions for selected wars

Notes: ^{*a*} 95% confidence interval using percentiles.

CHAPTER 3

USING FRONT LINES TO PREDICT DEATHS IN THE BOSNIAN WAR

3.1 Introduction

What readily observable indicators can we use to predict killings in civil war? This paper compares an "ethnic conflict" model for reported killings during the Bosnian War from 1992 to 1995 that uses data likely to be unavailable for ongoing civil conflicts (e.g. Weidmann, 2011) to a "front lines" model based on readily observable or static indicators like areas of active fighting (e.g. front lines) and terrain characteristics. The goal is to evaluate whether such an alternative model can predict killings in civil war as well as an ethnic conflict model.

Research on the Bosnian War from 1992 to 1995 has developed sophisticated models of ethnic conflict that provide a detailed examination of ethnic violence (Weidmann, 2011), and similarly sophisticated models have been developed for other cases in which disaggregated data is available (e.g. Berman, Shapiro and Felter, 2011; Hegre, Østby and Raleigh, 2009).

As more detailed data on violence in civil wars, with exact geographic locations and dates giving days and even the hour of a violent event, have become available, models of civil war violence are becoming increasingly complex. Previous research for example has examined the relationship between complex measures based on ethnic group data and ethnic violence in Bosnia (Weidmann, 2011), poverty and war in Liberia (Hegre, Østby and Raleigh, 2009), and reconstruction spending and violence in Iraq (Berman, Shapiro and Felter, 2011).

With the focus provided by disaggregated data, statistical models are now working at the level of small geographic units the size of U.S. counties or smaller, and with time periods measured in weeks or months. The complementary utility of such models for real-world forecasting of violence in ongoing conflicts has accordingly increased. We now have data to, at least in theory, construct statistical models that can forecast levels of violence in specific parts of a country.

One of the obstacles to implementing such models for forecasting in practice is that much of the existing research uses high-quality data, like detailed census data on ethnicity and income figures, unemployment levels, etc. These are desirable from a theoretical standpoint for developing models designed primarily to test hypotheses about causal inferences, but are likely not available in countries experiencing large-scale insurgency or civil war in which the government is to weak or it is too dangerous to attempt to collect similar statistical or census data.

This study develops an "ethnic conflict" model of killings during the Bosnian War from 1992 to 1995 similar to Weidmann (2011). It takes into account detailed census information on ethnicity and income. To address several technical challenges, I also develop a negative binomial model appropriate for the count nature of the data, capable of dealing with overdispersion and exposure effects, and which includes a spatial lag to allow for spatial diffusion of violence.

In contrast to models focused on hypothesis testing, the requirements for a model that is useful for forecasting are different: data needs to be practical and either available in real-time or static in nature (Ulfelder, 2012), while also producing good fit with observed violence. An intuitive source of such data are information on where fighting occurs, e.g. front lines in a relatively conventional conflict, and time-invariant (at least in the short-term) features such as terrain.

Thus I also estimate an alternative "front lines" model based on distance to the frontline, population, urbanization, and rough terrain. All of the covariates are either timeinvariant in the short term, like terrain and urbanization, or could be estimated in real time using some form of remote sensing. For example, aerial imagery has been used to track fighting by military forces for a century, and there have been many efforts to estimate population sizes, etc. using remote sensing with satellites (Wu, Qiu and Wang, 2005).

Using front lines to estimate civilian deaths sounds almost tautological. That certainly is true, although the extent to which it is can be disputed. For example, Kalyvas (2006) argues that selective violence against civilians will not be as high in directly contested areas as in those in which a combatant is trying to consolidate control, which suggests that selective killing of civilians will occur at some distance from the frontline. Using a theoretical model and data on the Soviet counterinsurgency in the Ukraine from 1943 to 1950, Zhukov (2012) on the other hand argues that this logic only holds under certain conditions and that violence against civilians in contested areas can be a rational strategy. In the end, and more importantly, the issue is somewhat irrelevant: endogeneity is a problem for accurately determining causal effects, whereas the goal here is to assess if front lines can be used to accurately predict civilian deaths. The more endogeneity, the better.

A comparison of the "ethnic conflict" and "front lines" models in Bosnia shows that both improve fit over a baseline model, and that both fit the data roughly equally well. This suggests that models using data on front lines and other static variables are appropriate alternatives for forecasting efforts.

The next sections review current studies using disaggregated spatial data on violence and how they are used. Focusing on the case of Bosnia, an example of the relationship between Yugoslav self-identification, ethnic diversity, and violence shows how unusual these data are for researchers interested in ethnic conflict. An intuitive alternative model based on front lines is developed, and a statistical model appropriate for the spatial count data on killings is introduced.

3.2 Disaggregated studies of civil conflict

An increasing number of studies examine civil war and other forms of political violence at a sub-national level using disaggregated data and GIS (geographic information systems), rather than cross-nationally with samples that span continents or the entire globe. One of the advantages of this approach is that it allows for a closer examination of arguments that might otherwise be subject to ecological fallacy (Buhaug and Lujala, 2005). Rather than relying on cross-national variation of measures of interest in different civil wars to evaluate arguments, sub-national data allows researchers to evaluate hypotheses in a single country, and thus to establish with more confidence whether the data fit the proposed explanation.

Several common arguments in the civil war literature have implications for the location of violence within countries. For example, if civil wars are fought primarily because the opportunity for predation is high (e.g. weak state and structural conditions that favor control over primary commodity export), then we should expect to see violence in areas within a country where such conditions are met (Collier and Hoeffler, 2004). Rebel groups in Colombia should be based in rural jungles and mountains which allow for drug production and where government control is weak, not in urban areas, where government control is typically strong. Using disaggregated data to develop smaller units of observation thus is useful to ensure that the lower-level implications of an argument are empirically reasonable.

Because data on civil violence was unavailable below the country level, early work in this literature focused on identifying the location and extent of conflict zones within countries that experienced civil war (Buhaug and Gates, 2002; Buhaug and Lujala, 2005). This presents new problems in using geographically disaggregated data on various explanatory variables in a consistent manner, and limits the ability of statistical models to predict specific conflict locations within a country, but is also more easily adapted to existing, cross-national conflict research.

More recently, point data that associate events such as killings or press reports of violence with a specific geographic location have become available. The three major types of such data are distinguished by the source for the coding of events: public media reports, e.g. ACLED and Uppsala GED (Raleigh et al., 2010; Melander and Sundberg, 2011; Sundberg, Lindgren and Padskocimaite, 2010), state collected statistics, e.g. crime statistics or U.S. military significant activities (SIGACT) databases (Felter, 2005; Berman, Shapiro and Felter, 2011; Berman et al., 2011), and (post-war) fact-finding efforts or historical research (Toft and Zhukov, 2012; Zhukov, 2012). None of these sources are likely to be error or bias-free, and this is an important caveat to research that uses them (Davenport and Ball, 2002). Event-coded data for media reports are at minimum biased by the extent to which public media have access to certain areas of a country, and typically do not include measures of event intensity (a demonstration and massacre might both be coded as single events each). State collected statistics reflect state control and bureaucratic efficiency, assuming they are not outright manipulated. And even factfinding efforts are unlikely to catch, accurately, all violence.

3.3 Ethnic violence and point data on deaths

Point data consist of incidents, e.g. the murder of a civilian, and geographic coordinates at which the incident occurred, with other information such as the time, information source, and so on. Assuming that the coordinates associated with an incident are precise, there is conceptually no further level of spatial disaggregation. Panel (a) in Figure 3.1 shows a fictitious example of point data for Bosnia, plotted as the red dots and shown in the table below. Each of the four data points depicts a civilian murder or similar incident, and the associated coordinates allow us to plot this information.

With other data that describe a geographic area, for example an administrative unit such as districts or an arbitrary grid cell roughly analogous to a square piece of ground, one can however aggregate these point data to that larger geographic area. In this context, the end result might be a count of incidents by district, or some other summary measure for variables contained in the point data, like the average time of day during which incidents occurred. Panel (b) in Figure 3.1 shows how the data could be aggregated up to grid cells by superimposing an arbitrary grid of latitude and longitude lines, and coating the number of incidents that fall within each grid square. In this case we have one grid with three events, one grid with one event, and the remaining grids have no event occurrences. An alternative is to aggregate to administrative divisions, and panel (c) shows the four data points as well as Bosnia's municipalities. Again only counting the number of incidents, the resulting data shows that 3 events occurred in Mostar municipality, 1 in Konjic, and none in other municipalities.

Because point data, in theory, provide the exact location of an event or death, there thus is flexibility in how they are aggregated for empirical analysis. One of the major constraints is the availability of other spatial data for covariates of interest, and how to translate concepts and measures to a spatial context. An additional choice when aggregating is how to summarize information at the higher level, e.g. counts of incidents, counts of casualty information if it is available, etc. The next subsection provides an overview of spatially disaggregated conflict research. Bosnia has been studied in the context of ethnic conflict, but in many ways is a quite unique case for empirical research, and further below I provide an example that illustrates this point.

3.3.1 How the data are used

Specific point data on violence or violent events are now available and for empirical research they are typically aggregated to areas to produced count data of the number

of events, number killed, per capita killed, etc. Cross-national studies tend to use grid cells as the basic unit of observation to which all other data are merged or aggregated (Buhaug and Rød, 2006; Raleigh and Urdal, 2007), while studies focusing on a single or small number of countries tend to use local administrative units (Weidmann and Ward, 2010).

The grid cells used in such research are approximately square cells created using a coordinate system. Nearly square at the equator, they become more distorted at higher latitudes. Although arbitrary, they are easier to use in cross-national studies that cover multiple countries with potentially very different sub-national administrative units. Global data on population, income, etc. estimates are also commonly available for similar arbitrary grids.

Buhaug and Rød (2006), using 100×100 km cells covering Africa, found several relationships between the incidence of territorial and governmental conflict within states and factors such as the distance to the border or capital city, proximity of diamonds, and open terrain. Raleigh and Hegre (2009) uses smaller 8.6 km cells, also covering Africa, to examine complex relationships between distance, population, and other factors. Using population density as a proxy for the value of an area, their results imply that although distance has an effect on conflict, the most significant factor is population density. And, in an example of a single country study using grid cells, Hegre, Østby and Raleigh (2009) examine conflict and poverty in Liberia and find that events were more frequent in wealthy areas. They interpret this finding as offering support for arguments in which opportunities for profit drive civil violence.

The source of conflict data is in all three of these studies is ACLED, which codes media reports of violence. An important issue to keep in mind therefore is also how various factors that are hypothesized to impact violence relate to the probability that an event is reported in Western media.

The alternative, sub-national administrative areas, are more intuitive, but spatial data for them is often more difficult to obtain. Research using districts and provinces in Iraq, Afghanistan, and the Philippines has found that unemployment does not increase insurgent attacks against either the military or civilians (Berman et al., 2011), but that improvements in the provision of government services are related to reduced violence, at least in Iraq (Berman, Shapiro and Felter, 2011).

In one of the most comprehensive examinations of the role of ethnic conflict in civil war violence, Weidmann (2011) looks at two complementary processes of ethnic violence: top-level pressures to consolidate ethnic territories, and violence due to local interethnic resentment. The two alternative processes are evaluated using data for Bosnia, including killings documented after the war through fact-finding efforts. Top-level territorial contestation is measured using the relative value of a municipality to each ethnic group's territory, given the municipalities and it's neighbors' ethnic populations. For example, a municipality with significant populations of 2 ethnic groups bordered on one side by areas in which one of the groups dominates and on the other side by areas dominated by the other group would have a high score for contestation, whereas a municipality with a clear majority borders by areas in which coethnics also have a majority would have a low score. Local ethnic resentment is measured using municipality-averaged village level ethnic polarization (Montalvo and Reynal-Querol, 2005).

Weidmann finds that both measures are associated with significant increases in per capita violence in Bosnia. The potential for ethnic violence is usually measured using an ethnolinguistic fractionalization (ELF) index constructed as $1 - \sum e_i^2$, where e_i is ethnic group *i*'s share of the population. The index can be interpreted as an overall measure of ethnic diversity. Although the two measures of territorial contestation and local polarization constructed by Weidmann are based on specific theoretical processes while the ELF is not, the latter has been used much more widely and both are useful starting points for an ethnic conflict model of violence.

3.3.2 The special case of Bosnia

The Bosnian civil war is becoming a canonical case in the study of civil war and ethnic conflict, and for several reasons it is well suited for sub-national research. Several sources of event data are available for the conflict, a census was conducted several months before war broke out by an efficient bureaucracy, and there is significant variation in the distribution of several ethnic groups within the country.

Bosnia and Herzegovina was the ethnically most diverse republic in the former Yugoslavia, with significant minorities of Bosnian Muslims, Serbs, and Croats. After secession and war in Croatia, Bosnia was by early 1992 left in a Serb-dominated rump Yugoslavia. A successful vote for secession and months of preparation by nationalist leaders lead to open war in April 1992.

During the Cold War, Yugoslavia's defense strategy was based on guerrilla warfare in the mountainous center, and there were extensive, republic-controlled territorial defense units and local militias (Nation, 2003, 150-151). As in the other republics, these territorial defense units and local police formed the core of the new sectarian militaries. Bosnian Serbs benefited from material, personnel, and command support from the remnants of the JNA (Yugoslavia's federal army) in Bosnia as well. The resulting Serb army was dominant in the beginning of the war and able to secure half of Bosnia's territory.

Despite the initial gains in the conflict by the Bosnian Serb army, in the end all three ethnic armies were fairly makeshift and not strong enough to decisively defeat any other side. By the end of 1992, the war had militarily become a stalemate with static front lines. Behind the lines, in areas controlled by each factional army, ethnic cleansing perpetrated by semi-professional armies and paramilitary forces became a strategy to consolidate control (Nation, 2003, 151-168).

For empirical studies of civil war, Bosnia is particularly well suited. A census was conducted in the republic in 1991, right in the prologue to the civil war which was already starting to engulf other republics in Yugoslavia. The communist federal state had long recognized the fact that Yugoslavia was a multi-ethnic state and distinct categories for Serbs, Croats, Muslims, and other minorities had been well-established for decades (Glenny, 2001). Considering the set of countries which have seen civil war, Yugoslavia was well-developed and thus likely had an efficient bureaucracy. It's 1991 GDP per capita puts it in the top 20 percent of countries experiencing civil war onset, while it's much higher 1990 GDP per capita puts it in the top 5 percent (replication data for Fearon and Laitin, 2003). Comparison to the 1981 census shows that there are no gross inaccuracies and that it is probably accurate (Stevanovic and Breznik, 1991). The results show that three ethnic groups, Bosniaks (or Bosnian Muslims), Serbs, and Croats together constituted the vast majority of Bosnia's population, with 43, 31, and 17 percent of the population each, and there is significant variation in the ethnic diversity of municipalities.

The other side of the equation, (civilian) deaths, has also been fairly well documented. A post war fact finding organization, the Research and Documentation Center (RDC; www.idc.org.ba/) in Sarajevo, has collected extensive data on dead and missing, and made that data publicly available. The records include names and the location of documented deaths are are available for use with Google Earth. In total there are 95,000 casualties, with over 77,000 confirmed deaths and 17,000 missing.

Media-based event data for Bosnia are also available from ACLED (Raleigh et al., 2010). The maps in figure 3.2 show ACLED events and RDC documented killings. In terms of their spatial distribution one can see noticeable discrepancies between the two sets of data. They also have quite different frequency distributions–many municipalities have no ACLED events (shaded light grey on the map)–while at least some deaths are recorded in every single municipality by the RDC.

3.3.3 Yugoslav identity

In addition to the historical ethnic groups one could identify with in Bosnia, e.g. Muslim, Serbs, Croats, etc., there was an additional category for those who saw themselves as Yugoslav (Malcolm, 1996; Glenny, 2001; Judah, 2000; Bringa, 2010). The word "Yugoslav", which literally means "South Slav", started to enter use in the late 19th and early 20th centuries and eventually was adopted as the term of choice for a pan-south slavic movement among slavs in the Austro-Hungarian and Ottoman Empires and their successor states, as opposed to the traditional ethnic movements in Croatia, Serbia, and Bulgaria. Maybe the best way to describe it is as nascent, supra-national identity, similar to the way Americans at some point started to identify as "Americans", not as residents of their particular state, or similar to German nationalism that developed in the 18th century from distinct regional (Bavarian, Prussian, Austrian, etc.) identities.

The south slavic Kingdom of Serbs, Croats and Slovenes became Yugoslavia after a coup which led to royal dictatorship in 1929. The communist Yugoslavia after World War 2 continued to use the term as a means to foster federalist sentiments and in opposition to ethnic nationalism (Glenny, 2001).

Although few people ever identified as Yugoslavs, no more than 2 percent nationally in Yugoslavia and around 5 percent in Bosnia, it was explicitly a pro-federal and anti-nationalist concept/ethnicity. Common reasons for identifying as a Yugoslav were probably intermarriage or lack of clear membership in one of the defined nationalities, communism, and idealism or some other belief in federalism (Bringa, 2010).¹ Regardless, it was distinctly opposed to nationalism, and thus can be used as a measure of anti-nationalist sentiment.

The distribution of people who self-identified as Yugoslav in the 1991 census in Bosnia is shown in map 3.3. It was highest in several urban areas, reaching past 15 percent in Tuzla and parts of Sarajevo. Compare this to the distribution of ethnic diversity in figure 3.4. The fractionalization index of ethnic diversity (ELF) is calculated as $1 - \sum e_i^2$, where e_i is ethnic group *i*'s share of the population. Higher numbers indicate diversity, and it can be interpreted as the chance that two randomly picked individuals will be from different ethnic groups. Bosnia in general was a very heterogenous place and the diversity scores for most municipalities reflect this. The national-level diversity score is 0.68.

The plot in figure 3.5 shows the relationship between self-identification as Yugoslav and ethnic diversity, measured using the fractionalization index. The x-axis shows the proportion of Yugoslavs, and the y-axis shows ethnic diversity. Each circle represents one of the 109 pre-war municipalities. As one would expect, few people identified as Yugoslav in ethnically homogenous areas. In ethnically diverse areas, near the top of the plot, there is more variance in the proportion of self-identified Yugoslavs.

The size of the circles shows the per capita rate of killed. Looking at the top right quarter of the plot, one can see that in municipalities that were ethnically diverse but also anti-nationalist (high Yugoslav identification) there were relatively few per capita deaths during the civil war. Municipalities that were ethnically diverse but also nationalized, in the top left quarter of the plot, had some of the highest per capita deaths during the civil war. In terms of ethnic diversity by itself, violence was worst in municipalities that are moderately high in diversity, and actually lower in those with very high diversity. The extent of nationalism, as measured by Yugoslav self-identification, is what seems to distinguish those highly diverse municipalities that see a lot of deaths from those with few.

Higher ELF values seem to be associated with more variance in death rates-they are uniformly low with low diversity, but vary quite a bit in the top half of the plot. This is consistent with previous, cross-national research on civil war incidence (Blimes, 2006). Theoretically this makes sense. Deaths will not be high in ethnically homogenous areas unless people are killing members of their own group, which probably makes little sense in the logic of ethnic violence. In very diverse areas the potential for unrestrained killing is high since there are many people who are not members of the same ethnic group, and thus many potential victims. But clearly, the potential for violence does not seem to uniformly lead to actual violence, depending on nationalist sentiment and other factors.

3.3.4 Frontlines

As much as it is interesting, this example shows how unique the case of Bosnia is in terms of data availability. Although media-coded event data is widely available now, detailed ethnic data at a subnational level is not available for many conflict regions.

¹Also, (Makul and McRobie, 2011).

There are two potential sources for information on ethnography: population censuses, and surveys. As mentioned above, Bosnia is unique in that a census was carried out a few months before violence broke out, by a government that explicitly recognized ethnic minorities. While accurate census data is easily available for much of the world, those places in which it is not are probably also the most relevant cases for ethnic conflict studies, either because conflict is currently ongoing or because conflict is likely. For example, and unsurprisingly, the last population census in Afghanistan was a partially completed effort in 1979, before the Soviet invasion (Gupta et al., 2011). In countries not experiencing open conflict population census can be a sensitive and even politically impossible issue, as is currently the case in Bosnia, but also in other countries like Nigeria (Bamgbose, 2009) and Lebanon (Ghosn and Khoury, 2011) where censuses have not been conducted at all or are subject to manipulation. Thus in countries at risk for ethnic conflict, census data is likely unavailable, outdated, or inaccurate.

For demographic and ethnic conflict research an alternative is to estimate data using surveys. For example, current population estimates in places like Afghanistan or the Democratic Republic of Congo are survey-based estimates (Coghlan et al., 2006). Selway (2011) uses various surveys like the World Values Survey and regional barometer surveys to create various aggregate ethnic diversity measures for the world. Surveys to some extent however still suffer the same problem as censuses: they are more likely to be unavailable for exactly those places in which they would be most useful for conflict prediction. For example, in Selway's data effort, most of the missing coverage is in African and the Middle East.

For a model to be useful in prediction violence in other conflicts, it has to rely on other information that is easier to collect. An intuitive alternative is to look at where the fighting is–frontlines–and assume that is where people tend to die. Collecting data on civil war front lines using remote surveillance and some combination of local reports is possible for governments and non-governmental groups. The Satellite Sentinel Project (www.satsentinel.org/about) for example uses satellite images to monitor violence in southern Sudan.

Is the relationship between front lines and civilian deaths tautological? Front lines depict boundaries between opposing military positions and in a conventional conflict delineate areas controlling by opposing sides. In a conventional war it is where we should reasonably expect most of the fighting to occur, and if so then there certainly will be civilian fatalities as a consequence as well. Furthermore, in an ethnically-driven conflict like Bosnia, the front lines also reflected underlying ethnic settlement patterns. But there are at least two reasons for why one might expect significant civilian deaths somewhere other than near the front lines. The first is simply a matter of logistics. If the armed groups are busy fighting each other, they will have less time to engage in systematic violence against civilians. And second, an armed group might even have an incentive to be more lenient in contested areas than areas further back in which it has a more secure presence (Kalyvas, 2006).

In any case, the aim of this paper is not causal inference, but to assess whether front lines can be used to predict civilian deaths as accurately as a state of the art ethnic conflict model. This turns the argument about endogeneity on its head: more endogeneity is desirable, because it will increase predictive accuracy.

The long military fronts in the war and relatively static nature, with sieges and limited offensives, encouraged the use of land mines, and Bosnia remains heavily mined (Nation, 2003, 158-159).² Data collected by post-war demining efforts shows clear dividing lines and matches other depictions of the wartime front lines. The static nature of the front lines simplifies the empirical analysis of their relationship with deaths since changes over time are less important, but it is also possible to work with front lines that change over time, provided data on violence also includes the time of the event.

In addition, measures of rough terrain and urbanization are easily collected or available, and can reasonably be expected to relate to violence. Rough terrain has been extensively examined in the context of civil war, with various expectations and results. Fearon and Laitin (2003) find that rough terrain and poverty, as a consequence of weak states, increase the probability of civil war. Areas with rough terrain and weak state control allow rural militant groups to challenge conventional government forces that usually would be dominant in a conventional conflict. Therefore violence should be more intense in inaccessible areas. Similarly, one should expect more violence in less developed rural areas with low urbanization rates.

Thus a reasonable alternative to the ethnic conflict model is one that uses front lines, urbanization, general (non-ethnic) population data, and measures of rough terrain. How well does such a model perform compared to a ethnic conflict model taking advantage of the detailed census data?

3.4 Data and model

Prior to the war, Bosnia had 109 municipalities. These constitute the basic unit of observation. Data on deaths comes from the Research and Documentation Center (RDC), a non-governmental, non-profit institution based in Sarajevo. It has documented and published extensive and detailed records of killed and missing persons. The records for confirmed killed total 78,000. Their distribution across Bosnia is depicted in figure 3.6, adjusted for area and population, and shows that fighting was most intense in central Bosnia around Sarajevo and in eastern Bosnia, along the border with Serbia.

3.4.1 Model setup

There are several issues in modeling the death counts, including spatial autocorrelation and choosing an appropriate statistical model for the data. Death counts are by definition positive integers, which suggests a standard linear normal model is inappropriate. One possibility is to transform the counts by taking the natural log or dividing by total population (thus creating a rate) so as to create something closer to a normal

²See BiH Mine Action Center and State Department BiH information .

distribution. As the histograms in figure 3.7 show, neither of those transformations produces a normal distribution. In both transformations, the data are skewed. Using a linear model in practice can produce several problems, among them worse predictions (Cameron and Trivedi, 1998).

More appropriate is a Poisson or negative binomial distribution. The observed deaths data by municipality have a mean of around 700, but the variance far exceeds this and overdispersion is potentially a problem. A standard Poisson model is inappropriate if there is overdispersion in the conditional mean, and as an alternative I use a negative binomial distribution to model deaths, with mean λ and a dispersion parameter, *r*. Another parameter, θ , includes the exponentiated covariates, $x_i\beta$, with two exceptions.

The first is to account for varying exposure. Deaths are theoretically dependent on a municipality's total population, which varies significantly. To correct for this, standard MLE negative binomial regression models would include an offset of logged population, with coefficient constrained to 1, among the exponentiated covariates. Equivalently, θ is multiplied by municipality population in the model here (i.e. $e^{x_i\beta + \ln p} = e^{x_i\beta} \times p$).

Since the deaths counts in each municipality likely are related to other, near municipalities, spatial dependency or correlation is also a potential issue. The second exception therefore is a spatial lag to model this. Spatial autocorrelation in general occurs when the value of a variable is related to values in some set of neighbors, and can be positive or negative (Ward and Gleditsch, 2008). Negative correlation occurs when high values in one area are associated with low values in the neighborhood, and would show on a map colored according to that variable as a checkerboard pattern. Positive spatial autocorrelation shows on a map as clustering and occurs when neighboring units have similar values. Previous research on Bosnia has identified it as an issue with different data, and it is a problem here as well (Moran's I, z= 3.38, p<0.0004) (Schutte and Weidmann, 2011).

One way to deal with spatial autocorrelation is to include a spatial lag that captures a measure of dependent variable values in the neighborhood. A straightforward version of this is a neighborhood average that provides, for any municipality, the average number of deaths in bordering municipalities. It can be constructed by multiplying a row-standardized matrix indicating contiguity for each pair of municipalities (W) with the list of death counts (y). Although other ways of accounting for spatial diffusion, e.g. spatially-lagged error terms, using distance or some other weighting scheme instead of contiguity, and so on, are possible, a spatially-lagged dependent variable (or neighborhood average) is more intuitive and appropriate in this context given that we would expect deaths to be positively related (Ward and Gleditsch, 2008).

The spatial lag, $y_s = Wy$, enters the model as a multiplicative term in λ , with an exponentiated coefficient ρ that measures its impact (Lambert, Brown and Florax, 2010). Exponentiating the coefficient allows for both positive and negative spatial diffusion, while constraining the overall spatial term $(y_s)^{\rho}$ to be a positive number. Holding the remainder of λ constant, a positive ρ will lead to a spatial term greater than 1, thus increasing mean deaths, while a negative ρ will lead to a spatial term between 0 and 1, thus reducing mean deaths.

The final model is:

$$y_i \sim NegBin(mean = \lambda_i, dispersion = r)$$
 (3.1)

$$\lambda_i = (y_{s,i})^{\rho} \times \theta_i p_i \tag{3.2}$$

$$\ln \theta_i = x_i \beta \tag{3.3}$$

where y_s is the neighborhood average or spatial lag, and p_i is the municipality's population. The parameter ρ can be interpreted as the strength of spatial diffusion of violence, with values greater than 0 indicating positive diffusion. The other term, $\theta_i \times p_i$, can be seen as the municipality-specific factor by which observed deaths differ from what one might expect due to spatial diffusion alone.

3.4.2 Model specifications and variables

I examine four versions of this model: a base, ethnic conflict, frontline, and combined model. They differ in the covariates included in theta, but otherwise match the form above. The base model only has a spatial lag and local population offset. The ethnic conflict model furthermore includes ELF index scores, territorial contestation and local polarization, as well as income per capita. The front line model includes distance to the front line, urbanization rate, and a measure of rough terrain. The combined model includes the variables from both the frontline and ethnic models. The variables are described in more detail below.

Population. Population for each municipality is taken from the 1991 census.

ELF index. Scores for each municipality were computed as $1-\sum e_i^2$, where e_i is ethnic group *i*'s share of the population, using municipal data from the 1991 census (Petrovic, 1992; Weidmann, 2011).

Strategic importance. An indicator that measures the extent to which a particular municipality is contested by two ethnic groups vying for territorial consolidation. It comes from the replication data for Weidmann (2011). The indicator is fairly complex and the original source has a detailed explanation for which there is no room here. But in essence, it depends on the population shares the the two largest ethnic groups in a municipality and it's neighboring municipalities.

Local polarization. Ethnic polarization scores are calculated on the village level and averaged to the municipality using 1991 census data, also from Weidmann (2011). The score measures how far a the distribution of ethnic group population shares deviates from an ideal in which two groups share half the population each, and is calculated as $4 \times \sum p_i^2 (1 - p_i)$ (Montalvo and Reynal-Querol, 2005).

Per capita income. Measures local wealth using 1991 per capita income. The original source is also the 1991 census via the replication data for Weidmann (2011). It is included as a control variable in the ethnic conflict model, but not the frontline model.

Distance to front lines. Since the front lines in the Bosnian War were fairly static (Nation, 2003), I digitized on set of front lines using period maps of front lines and data

on minefield location from the BiH Mine Action Center (posted at www.mine.ba). The latter has includes a map of minefield locations and rough front lines, which I traced to recreate front lines. The result is shown in the map in figure 3.8. They reflect the approximate location of the 1993 front lines, and in any case the major changes in front lines were in sparsely populated areas. Only with the joint Croat-Muslim offensive in 1995 following the collapse of the Serb state in Croatia did the front lines change drastically–and the war ended as a result of this offensive. The distance to the front lines was calculated as the straight line distance from centroids for each municipality to the closest frontline, not counting borders of Bosnia with other republics. The natural log of this measure is used as a covariate.

Urbanization. Urbanization is measured as population concentration in each municipality (or the chance that two randomly selected individuals will be from the same sub-municipality village) using 1991 census data (Weidmann, 2011).

Rough terrain. Rough terrain is measured as the proportion of a municipality's terrain that has a slope greater than 15 degrees. Sloped is calculated using SRTM (Shuttle Radar Topography Mission) 90 meter raster elevation data from the U.S. Geological Survey. I reduced the resulting slope data were in resolution by a factor of 10 to make the data more manageable and then aggregated to the proportion of cells in each municipality with a slope greater than 15 percent. In other words, this indicator measures the extent to which a municipality has mountainous terrain.

3.4.3 Estimation

The four models were estimated using Markov-chain Monte Carlo simulation with 5 separate chains in which every 5th observation after the first 4,000 of 20,000 draws was recorded. Using MCMC estimation and model coding in JAGS provides a more flexible approach that does not require deriving likelihood functions and their partial derivatives. The large number of simulation iterations was necessary due to autocorrelation in the Markov chains and to ensure convergence (Gelman and Hill, 2007, 356-358).

Without going too much into the differences between Bayesian and maximum-likelihood estimation (MLE), the basic intuition is that, given assumptions about the prior probability densities for our parameters, we use the data to update and calculate posterior densities for the parameters in the model. The parameters for the models here include the coefficients, but also the fitted values for each municipality. The end result is a distribution of estimates, i.e. posterior densities, for each parameter in the model that can be summarized using means and percentile intervals in a way similar to MLE coefficient estimates. Although Bayesian and frequentist (MLE) approaches have fundamentally different interpretations of the concept of probability (Gill, 2002, 65–79), the results presented below can be interpreted just like those from any other regression model.

Unlike for regression modeling in Stata or well-established models like linear regression in R, the workflow for estimating these models in JAGS and R is more complicated and time-intensive, and as a result I do not conduct robustness checks for different covariate specifications. Each version of the model requires a separate text file describing the model and its covariates for JAGS. Within R, the model call and related commands each take about 15 lines of code to call on JAGS, process the results, and create the 3 plots presented for each model below. This includes calls to 2 subroutines I wrote to process the objects returned by JAGS, which are large sets of numbers of posterior density samples for each of the 100+ parameters (including fitted values) per model. These total an additional 35 lines of R code. Furthermore, the JAGS call for each model itself takes several minutes to complete.³ This complexity makes it less feasible to estimate a large number of models for robustness checks, even if there are only minor specification differences.

3.5 Results

3.5.1 Parameter estimates

The posterior densities for the models are shown in figures, 3.14, 3.11, 3.12, and 3.13 for reference. I also show density estimates from a model in which theta has a constantonly in figure 3.15, although the model fit is otherwise not discussed. One way to look at these is as densities similar to the ones from which we derive coefficient and standard error estimates using maximum likelihood estimation, with the mean matching the coefficient estimate, and variance of the posterior density akin to a standard error.

The densities for ρ show the direction and strength of spatial diffusion. Values below 0 indicate negative spatial diffusion in which high counts of deaths in one area reduce counts of deaths in neighboring areas, while values above 0 indicate positive diffusion in which high deaths are associated with high deaths in the neighborhood. All three models show positive diffusion averaging between 0.4 and 0.5. The dispersion parameter *r* for the negative binomial distribution shows the extent to which variance diverges from that in a Poisson distribution–as *r* reaches infinity a negative binomial distribution approaches a regular Poisson distribution in which mean and variance equal.

The other coefficients can be interpreted similar to the way one would interpret maximum likelihood regression coefficients, with distributions centered on 0 indicative of no relationship, either positive or negative. A density excluding zero for most simulation draws could be interpreted as significant evidence for a positive or negative relationship. For example, there seems to be a certain positive association between local polarization (b[4]) and deaths in the ethnic conflict model in figure 3.12, but not strategic significance (b[5]). This is in contrast to Weidmann (2011), where both measures, independently or in a combined model, have a positive relationship to civilian deaths in a linear spatial model.

Three of the variables in the front line and combined models seem to have non-zero associations with the number of civilians deaths: more urbanized areas were less violent (b[2]), as were areas far away from the front lines (b[4]), but mountainous areas were

³The JAGS call from R for the combined model takes 332 seconds, or 5.5 minutes, using R 2.13 and JAGS 3.1 on a 2.7 GHz Intel Core i5 running Mac OS X 10.7 in 2012.

more violent (b[3]). Although the fighting in Sarajevo and Mostar received much media attention during the war, these findings seem to suggest that civilian killings were most intense in rural, mountainous areas like eastern Bosnia. It is also important to consider that these results are offset for population, and while Sarajevo and other cities may have seen a lot of fighting, there also was a large pool of potential civilian victims.

3.5.2 Model fit

There does not seem to be a uniformly accepted measure of goodness of fit for count models analogous to the R^2 measure for linear models. Using a statistic like R^2 or mean-squared error provides a measure of fit that is relatively easy to interpret, but using it with count models is problematic (Cameron and Windmeijer, 1996). Count data are by definition zero or positive and likely (as in this case) asymmetrically distributed, placing lower bounds on the theoretically possible residuals. For example, with an observed outcome of 100, the error at most can be -100 but has no positive bound. This makes using a sum of squared residuals inappropriate, and can lead to R^2 values outside the 0 to 1 interval.

The deviance information criterion (DIC) is a general model fit statistic developed in the context of Bayesian modeling that can be used to compare non-nested models (Gelman and Hill, 2007, 524-526). It is a generalization of the Akaika and Bayesian information criteria (AIC and BIC) for hierarchical models, and is calculated as: $DIC = \overline{D} + p_D$, i.e. the mean deviance plus the effective number of parameters p_D . Deviance is defined as $-2 \times log(p(y|\theta))$, where $p(y|\theta)$) is the likelihood of the data given the model parameters. The effective number of parameters is estimated as the variance of the deviance, and penalizes for model complexity (Spiegelhalter et al., 2002). The DIC can be interpreted as the ability of the model to predict new cases, with lower values indicating higher accuracy (Gelman and Hill, 2007, 525).

Figure 3.9 shows plots of predicted over observed values for all 3 models as well as DIC (deviance information criterion) values. The plots are on a logarithmic scale. The black line indicates points on which observed equal predicted values, and the plots give a rough visual indication of model fit. The DIC allows comparison across the models, with lower values indicating better fit. A constant-only model (not shown) has a DIC of 1676. The base model with spatial lag and population offset improves this by more than 120 to a DIC of 1545. Both the ethnic conflict and front lines models slightly improve fit over that, although not as drastic. Compared to each other they have similar fit to the data with DIC values of 1529 and 1530 respectively. Finally, a combined model with coefficients from both the ethnic conflict and front lines models improves fit a little further with a DIC of 1522, although the improvement is not as large as between the base model and the front and ethnic models. This suggest that much of the prediction accuracy is driven by difference in the pre-war population of an area and diffusion of violence from neighboring areas, i.e. the population offset and spatial lag.

While the DIC allows for comparison across models, it is somewhat difficult to interpret substantively and does not provide much information on how well the model predictions fit particular cases. The deviance-based R^2 is an alternative summary statistic based on the ratio between model deviance and deviance for a saturated model based on mean observed outcomes (Cameron and Windmeijer, 1996). For a poisson model, it is calculated as:

$$R_{dev}^{2} = 1 - \frac{D(y;\hat{\mu})}{D(y;\bar{y})} = 1 - \frac{\sum \left[y \times \log(y/\hat{y}) - (y - mean(\hat{y})) \right]}{\sum \left[y \times \log(y/\bar{y}) - (y - \bar{y}) \right]}$$

where $\hat{\mu}$ are the mean fitted values. Using this form here is somewhat inappropriate since the model here is closer to a negative-binomial regression, and since it is not estimated using maximum likelihood techniques. It does however give a rough impression of fit and the calculated values are similar to measures one could derive from the simple correlation between observed and fitted values.

The deviance- R^2 ranges from 0 to 1 and can be interpreted as the proportion of variance in outcomes explained by the model. Table 3.1 summarizes the deviance- R^2 values for the four models as well as the DIC for the four models and a constant-only model. The third column also shows the bivariate correlation between observed and mean predicted deaths. In comparison to each other, the model R^2 values are similar to the DIC, with both the ethnic and front lines models improving fit over the base model. A combined model further improves fit, although not as drastically as the ethnic and frontline models do over the base. The values indicate that the models generally explain roughly 59 to 68 percent of the variance in observed civilian deaths.

Aside from the summary measures above, the spatial distribution of prediction errors across Bosnia can also be interesting as it shows areas in which the models overor under-predict. Figure 3.10 plots the prediction error from the ethnic and front lines models on a map of Bosnia. The prediction error is calculated as the difference between observed and mean predicted deaths. The map is shaded so that blue indicates a negative difference, i.e. under prediction, and red indicates a positive difference, i.e. over prediction. The range of errors is from -2,900 to 2,500 and from -2,800 to 1,600 for the ethnic and front models respectively. Comparing the blue areas of the map, it is apparent that both models under predict civilian killings in western Bosnia, and over predict killings in central Bosnia and Banja Luka in northwest Bosnia. During the war there were several Bosniak enclaves surrounded by Serb forces in western Bosnia, including Srebrenica. The similarity in the error patterns also suggests that most of the error (and hence prediction) is driven by factors common to the two models, i.e. the spatial lag, population offset, and general model structure.

Overall, a combined model is preferable if the goal is to maximize predictive accuracy. If there is no detailed ethnic data but there is information on the position of front lines, then such a model is not any less accurate, and vice versa. To some extent the two models, at least in the case of the Bosnian War, are exchangeable.

3.6 Conclusion

This paper introduced an example that showed that Yugoslav self-identification in municipalities in pre-war Bosnia was associated with lower levels of wartime killing in order to demonstrate the unique quality of ethnographic information available for Bosnia. Data with such measures and comparable quality are going to be difficult to obtain for other countries at risk for, or experiencing, conflict.

Statistical models of the Bosnian conflict that include detailed, high-quality information as covariates can, unsurprisingly, improve fit to observed data and therefore predictive ability. At the same time, comparably accurate and effective ethnographic data will be difficult to obtain for other countries at risk for, or experiencing, ethnic and civil conflict. Thus while models that take advantage of Bosnia's ethnographic data may be highly interesting and useful in a theoretical context, they will be less appropriate for forecasting in general because they can only be applied where similar high-quality information is available. The cross-sectional analysis of killings in the Bosnian war in this paper shows that a model based on front lines and other structural factors fits observed killings as well as an alternative model based on detailed ethnic information. This provides some confidence that using such a frontline/structural factor model for other conflicts will produce a model fit comparable to theoretical models of ethnic conflict.

The front lines during the Bosnian War were fairly static, allowing a simpler, crossnational analysis to be used. But to be practically useful for forecasting future violence in a conflict, a feasible-data model will also have to include a time component. Disaggregating data and modeling not just in space but also over time is a logical extension of this project that is feasible given currently available event data, such as for Iraq and Afghanistan.

Another problem, in applying this approach to other conflicts, is that not all civil wars will have well-defined front lines. If one side is largely predominant, as in Iraq or Afghanistan, there will be no equivalent to front lines. And in those conflicts where the sides are more balanced, whatever front lines there are will probably not be as static as those during the Bosnian War.

Despite these shortcomings, the contributions this project makes are threefold: (1) the example illustrating the relationship between ethnic diversity, Yugoslav self-identification, and deaths suggests and interesting case for further research, (2) it develops and uses a spatial count model that is more appropriate for spatial data on violence-related counts of deaths or events, and (3) it suggests that a model based on static factors and front line distance is appropriate for the purpose of prediction if detailed ethnic and income data are not available.



Figure 3.1: Illustration of conflict incident point data and aggregation to grid cells or administrative units.

 Table 3.1: Comparison of model fit.

Model	DIC	deviance- <i>R</i> ²	$\operatorname{cor}(y, \hat{y})$
Constant	1676	-0.07	-
Base	1545	0.59	0.70
Ethnic	1530	0.64	0.72
Frontline	1529	0.67	0.74
Combined	1522	0.68	0.74





western Bosnia adjacent to Dalmatia.





Figure 3.5: Bubbleplot of Yugoslav self-identification, ethnic diversity, and civil war death rates by municipality. Each circle represents a municipality, with its proportion of self-identified Yugoslavs on the x-axis and ethnolinguistic fractionalization index (ELF) value on the y-axis. The area of the circle is proportional to deaths per capita during the civil war. There are fewer Yugoslavs in ethnically less diverse municipalities, as one would expect, and those also tend to see relatively little violence. In municipalities with a high level of diversity, the proportion of Yugoslavs is related to deaths per capita, with diverse but nationalist municipalities experiencing more violence than diverse but federalist municipalities.





Figure 3.7: Histograms of deaths by municipality during the Bosnian War, 1992-1995. The untransformed count of deaths fits a negative binomial distribution with mean 712 and dispersion parameter 0.99. Both the log of deaths and population-adjusted deaths are skewed.





Figure 3.9: Predicted versus observed killings during the Bosnian War. Each panel shows a plot of predicted counts compared to observed counts for the base, ethnic, frontline, and combined models respectively. The axes are on a log_{10} scale. The black line shows observed = predicted. Each plot also gives the deviance information criterion for each model (constant-only has 1676). The base model includes a spatial lag (neighborhood average deaths) and population offset, but no other covariates. The ethnic model, which also includes measures of ethnic diversity and dispersion, as well as per capita income, has by far the best fit, but also draws on census data unlikely to be available for other countries. The frontline model, which uses distance to the 1993 front lines and other structural variables without census data, significantly improves fit over the base model as well.







Figure 3.12: Posterior densities for the ethnic conflict model: r = negative binomial shape, rhos = spatial lag, b[1] = theta constant, b[2] = income per capita, b[3] = ELF index, b[4] = local polarization, b[5] = strategic significance.



rhos = spatial lag, b[1] = theta constant, b[2] = centered urbanization rate, b[3] = percentage rough terrain, b[4] natural log of centroid distance to frontline.



Figure 3.14: Posterior densities for the combined frontline and ethnic model: r = neg-ative binomial shape, rhos = spatial lag, b[1] = theta constant, b[2] = centered urbanization rate, b[3] = percentage rough terrain, b[4] natural log of centroid distance to frontline, b[5] = income per capita, b[6] = ELF index, b[7] = local polarization, b[8] = strategic significance.



CHAPTER 4

MODELING THE EVOLUTION OF CIVILIAN DEATHS IN THE IRAQ WAR

4.1 Introduction

Data on specific incidents of violence in a civil conflict are becoming increasingly common: U.S. military significant activities reports (SIGACTs) for Iraq and Afghanistan, the Armed Conflict Location and Events Dataset (ACLED) (Raleigh et al., 2010), the Georeferenced Event Dataset (UCDP-GED) (Sundberg, Lindgren and Padskocimaite, 2010; Melander and Sundberg, 2011), Armed Forces of the Philippines (AFP) incident data (Felter, 2005), North Caucasus incident data (Toft and Zhukov, 2012), and the Iraq Body Count project (IBC; http://www.iraqbodycount.org/). These sources increasingly allow us to look at violence, through reports of civilian and battle deaths, over time and at a resolution of months, weeks, or even days, in near real time. Being able to make intelligent predictions about the future evolution of violence during a civil conflict is becoming feasible and has tremendous applications in policy and decision-making, but how do we model and forecast violence in civil conflicts?

Statistical models in political science are commonly used for hypothesis testing and estimating marginal effects using all available data. These quantitative models of civil war are usually meant to determine underlying causal effects in order to evaluate theoretical arguments related to civil war. Model fit, let alone forecast performance, typically is not evaluated or of interest, although this appears to be changing (Schneider, Gleditsch and Carey, 2011). Unsurprisingly then, model fit and predictive performance in well-known civil war studies is poor and driven by a few variables (Ward, Greenhill and Bakke, 2010). Yet at the same time, predictive performance is of great interest to decision makers who want to evaluate the impact of decisions on future events (Montgomery, Hollenbach and Ward, 2012; Ward, Greenhill and Bakke, 2010).

An additional challenge arises from the fact that the most commonly used statistical models in political science in the context of civil war and civil violence were developed in the context of hypothesis testing. Common models like pooled regression models, panel models with fixed effects, and count models are potentially inappropriate for incident data on violence (Brandt et al., 2000) and in any case the nature of time-dependence,

which is critical for forecasting into the future, is often treated in a static or ambiguous way, e.g. with dummy variables, lagged dependent variables, or splines.

The goal of this paper is to develop well-fitting statistical models of violence in a conflict over time, and to provide the ability to produce both a good fit in-sample but also the ability to provide long-range forecasts months into the future. I use machine-coded data on monthly civilian deaths in Iraq from the IBC project to develop and evaluate statistical models for in-sample fit and forecast usefulness, but the resulting models should be useful across a range of conflict applications with time-dependent count data, e.g. incident data for other countries.

Do policymakers care specifically about the ability to forecast civilians deaths? Then-Secretary of Defense Rumself and high-ranking military officers have frequently disputed estimates of civilian deaths in both Iraq and Afghanistan, and they certainly are not presented openly. Yet on the other hand, winning the support of the population is one of the centerpieces of the U.S. military's revised counterinsurgency doctrine (U.S. Army and Marine Corps, 2006; Hultman, 2012). According to a memorandum issued by the International Security Assistance Force (ISAF) in Afghanistan in 2009, "gaining and maintaining [the support of the population] must be our overriding operational imperative" and that "we must respect and protect the population from coercion and violence" (2009). Similarly, the first goal set out by General Petraeus in a 2010 ISAF guidance is to "secure and serve the population" (2010). Among the congressional benchmarks for Iraq presented in a GAO report are reducing the level of sectarian violence (2007) and in a 2007 report Gen. Petraeus presented figures on civilian casualties as a measure of progress in Iraq (Petraeus, 2007). There is clear interest among high-level policymakers in civilian deaths in both the Iraq and Afghanistan counterinsurgency conflicts.

At the same time, the U.S. Army includes prediction as one of the purposes of military intelligence (U.S. Army, 2010, 1-4). Given the clear interest in civilian deaths as a counterinsurgency objective, this implies that being able to predict civilian deaths is relevant to both policy and military decision-makers.

The two requirements that a model should forecast well but also be appropriate for hypothesis testing create opposing demands on a statistical model: flexibility to match data in sample, but also some measure of rigidity in order to be able to forecast accurately more than a few steps into the future. This tradeoff becomes practically relevant in the context of conflict data with a non-monotonic trend, i.e. increasing up to some peak level of violence and decreasing subsequently. A model without assumptions about this trend will fit well in sample, whereas a model that assumes a specific shape for the underlying conflict forecasts better. Thus although the initial question motivating this paper is how to model and forecast violence in civil conflicts, the two models presented as answers present only partial solutions.

The next section reviews existing quantitative work on Iraq, using disaggregated data, as well as existing approaches to forecasting in IR. The remainder of the paper introduces the data on Iraq and their dynamics, and the two statistical models as well as their particular advantages and disadvantages.

4.2 Background on Iraq and IR forecasting

The basis for much of the existing research on civil war violence, in Iraq and elsewhere, lies with current counterinsurgency doctrines in the U.S. military. So, before discussing specific examples of work on Iraq, it's useful to review the military and academic origins of current thinking on civi war and insurgency.

4.2.1 Counterinsurgency and violence against civilians

The intellectual roots of the current, population-focused U.S. counterinsurgency doctrine lie in the 1950's and 1960's with the British and French counterinsurgency efforts in southeast Asia and Algeria respectively. Based on his experience as a military officer in French Algeria (Galula, 2006), Galula emphasizes the importance of the civilian population to both sides and the need for local security combined with political and economic development (Galula, 1964). Writing around the same time, Thompson (1966) recounts the British counterinsurgency campaign in Malaysia. The key lessons he draws are the need for civil-military cooperation, governance, and intelligence and police work, supplemented with direct military actions when appropriate. Significant portions of both accounts were distilled, along with a U.S. Marine Corps manual (U.S. Marine Corps, 1940), into the widely-publicized new Counterinsurgency Field Manual (U.S. Army and Marine Corps, 2006). It summarizes counterinsurgency warfare as a struggle for the support of the population, where its protection and welfare are paramount goals. Violence against civilians is undesirable and counterproductive.

While the focus in military doctrine is on how to conduct counterinsurgency, academic research on civil war has approached the question from the perspective of those who rebel, join the rebellion, and why. Competing strands of research have argued about whether or when rebels are motivated by "greed" rather than "grievances". The greed story posits that rebellion is a criminal form of organized violence with the goal of securing valuable resources (Collier and Hoeffler, 2004), while the grievances story is about rebellion to effect popular political or cultural changes, often related to ethnic minority politics (Gurr and Harff, 1994; Posen, 1993). Leaving motivation behind a confounding set of arguments examines "opportunity" for rebellion due to weak states, high unemployment, etc (Fearon and Laitin, 2003). The set of academic arguments is diverse and not all of them are consistent with a military strategy centered on winning hearts and minds through protection and welfare.

One of the key questions motivating much of the research on Iraq is how much popular support matters for insurgents and counterinsurgents. A strategy centered on population security and development ("winning hearts and minds") evolved as a result of the U.S. military's experience in Vietnam, earlier conflicts usually conducted by the U.S. Marine Corps, and as the result of peacekeeping in the 1990's. To some extent the existing research seeks to quantitatively validate (or evaluate) a doctrine that was developed in a more subjective manner.
4.2.2 Research on the Iraq War

Looking broadly at how research on counterinsurgency in Iraq, and other places, matches with this view of effective counterinsurgency shows that there is empirical support for a population security and development-centered strategy. Research examining development spending in Iraq and the Philippines confirms that local reconstruction spending increases the perceived well-being of the population and reduces violence (Beath, Christia and Enikolopov, 2011; Berman, Shapiro and Felter, 2011; Crost, Felter and Johnston, 2012), although only if there is an adequate level of security initially (Beath, Christia and Enikolopov, 2011) and when funding is done in a manner that promotes local participation and avoids the creation of high-profile targets (Crost and Johnston, 2010). Harming civilians reduces support for whatever faction is responsible, although this effect can be somewhat moderated by ethnic and political allegiances (Condra and Shapiro, 2010; Toft and Zhukov, 2012). Notably, this pattern may break down in areas where control is contested if the combatants start using violence to force support (Zhukov, 2012), although this is less likely to be a problem in a conflict like Iraq where the government side largely holds the upper hand. Consistent with the argument that information and intelligence are key to a counterinsurgent, Shapiro and Weidmann (2011) provide evidence that increased cell coverage in Iraq increases the information flow to the counterinsurgent. In general these findings are consistent with a counterinsurgency doctrine focused on population security and economic development.

Berman et al. (2011) reveal a pattern contrary to expectations, finding that unemployment does not increase violence and may in fact decrease it. They speculate that when unemployment is high it may be cheaper for the government to obtain information, or that unemployment is the result of increased security efforts like checkpoints and registration requirements.

Finally, Lyall and Wilson (2009) argue that increased mechanization of military forces over the last two centuries has made it more difficult for counterinsurgents to obtain the kind of information that is necessary to defeat insurgents. Looking at Iraq, they compare the level of violence across military units with varying levels of mechanization, and find that those that consist of light infantry perform better.

This research, regardless of whether it is consistent with current counterinsurgency doctrine or not, does have clear policy implications that are often noted by authors. To conduct a successful counterinsurgency, a counterinsurgent should place emphasis on widely dispersed infantry forces with a solid background knowledge of local culture and language (Lyall and Wilson, 2009), provide local development assistance (Beath, Christia and Enikolopov, 2011; Berman, Shapiro and Felter, 2011; Crost, Felter and Johnston, 2012), improve public communication access (Shapiro and Weidmann, 2011) and take care to not harm civilians (Condra and Shapiro, 2010). There is little need to worry about unemployment however (Berman et al., 2011). Should decision-makers be interested in this? I will come back to this question after briefly discussing existing forecasting efforts in IR.

4.2.3 Forecasting in International Relations

Schneider, Gleditsch and Carey (2011) group forecasting efforts in international relations into three categories: structural models well-suited for cross-sectional or geographically disaggregated predictions, game-theoretic approaches using expert opinion as input, and time-series forecasting. Although time-series analysis for the purpose of forecasting is well-established in fields like finance and economics (Greene, 2008) or voting and public polling (Lewis-Beck, 2005), it is less common in international relations.

The relative paucity of such work in IR is partly because time-series of interest to international relations scholars are typically only available as yearly data, and even then the actual variation in a series often is relatively modest or the events modeled are rare occurrences like civil war outbreak or genocide. As a result, much of the existing IR time-series research is based on event data generated using automated coding software and depicting interactions between international actors or relevant subnational actors (Schrodt and Gerner, 1994; Goldstein and Pevehouse, 1997; Brandt, Freeman and Schrodt, 2011). In terms of statistical models, the data typically consists of series that are related to each other, like cooperative and hostile actions between two actors within the same country, and hence are modeled using vector analogues to single time series models like vector autoregression (VAR), bayesian VAR (BVAR), and Markov-switching BVAR (MS-BVAR) (Brandt and Freeman, 2006; Brandt, Freeman and Schrodt, 2011). Although these types of data provide information on the behavior of the actors involved in a conflict or international event, they tend to be somewhat abstract and modeling them often is difficult.

The growing availability of data on incidents of violence, like, Iraq Body Count (IBC), or U.S. military significant activities reports (SIGACTs) provides information on simply the level of violence during a conflict. This allows the simpler question of how a conflict will evolve over time to be addressed, and what factors will impact this. Existing research has adapted existing panel regression methods to work with these data, but this neglects one of the main features of such conflict data over time, i.e. how it reflect the escalation and resolution of conflicts over time. From a time-series perspective, the data show a particular kind of non-stationary trend over time.

Academic research, on Iraq and elsewhere, has already exploited the new kinds of event data that provide information on civil war incidents at a fine spatial and temporal resolution, largely in order to evaluate hypotheses and causal arguments. This research has potentially significant policy implications. At the same time, there is a growing stream of work on forecasting in international relations, but the focus has often been on describing the behavior of actors in a system towards each other, rather than outcomes like deaths directly. Would it be useful to take elements from both literatures, i.e. the focus on policy relevant questions in a current conflict on one hand, and the focus on model fit and forecast accuracy from the latter? I provide an example below that illustrates why the answer to this question is yes.

4.2.4 Why we need well-fitting models for policy

I raised several questions in the previous sections: to what extent should we heed policy recommendations from existing work on Iraq, and would it be useful to place more emphasis on combining policy relevant questions with a modeling approach that values model fit and predictive accuracy?

Before answering, let us take a step back: statical models can be useful in two ways (Brandt, Freeman and Schrodt, 2011, 41) from the perspective of someone who makes decisions that impact the conduct of a counterinsurgency. First, by providing information on the likely impact of a policy choice. For example, if I send another combat battalion to this province, how much will civilian violence decrease? Second, by providing information on the likely future state of the conflict, e.g. what will future levels of civilian deaths in Iraq be?

Statistical models for forecasting need to fit the data well in order to be able to forecast well. The statistical significance of coefficients is less or even no consideration and the goal is rather to maximize model fit (Ward, Greenhill and Bakke, 2010). A common approach to the general problem of building forecast models is to subset historical data into training and validation data that can be used to estimate a model and evaluate it's predictive accuracy out of sample (Brandt, Freeman and Schrodt, 2011; Weidmann and Ward, 2010). Although it is maybe not very common to examine forecast accuracy in international relations, quantitative political scientists have the skills needed to do this.

Instead of prediction, academic research has focused on hypothesis testing to determine causal effects, using statistical significance. One might be tempted to conclude that this also provides policy makers with an answer to the first concern, how policy choices will impact future conflict. But it is important to note that the relevant counterfactual from a policy standpoint is what impact a change will have *in the future*. Statistical models that show significant coefficients but do not fit the underlying data very well provide misleading answers to this question by neglecting predictive power (Ward, Greenhill and Bakke, 2010). A statistically significant coefficient may have a very small substantive effect, while simpler models with less statistically significant coefficients may provide better fit to the data. Thus even in the realm of hypothesis testing, in the context of policy implications, there are good reasons to develop well-fitting statistical models that predict well.

The empirical research discussed before (Condra and Shapiro, 2010; Beath, Christia and Enikolopov, 2011; Berman, Shapiro and Felter, 2011; Berman et al., 2011; Shapiro and Weidmann, 2011; Hultman, 2012) has a particular drawback when it comes to predicting into the future, aside from any question of their fit in sample. The models use subnational data of violence over time derived from two sources: U.S. military SIGACTs and IBC media-derived incidents. The data is in a panel format with observations across space (districts or provinces) and time (quarterly, monthly, weekly), and these articles employ a common strategy for addressing the statistical dependencies in this type of data: panel regression models with some combination of space and time fixed effects. Time fixed effects, as opposed to dynamic models of temporal dependence, make it difficult to provide predictions into the future since there is no basis for expected future time effects. Although one may get accurate estimates of effects, these models do not directly answer the question of how much policy recommendations will impact future levels of violence, even assuming that they fit well in sample.

Another important, but different, consideration for policy-relevant models has to also be how they are presented and how this relates to their credibility. Coefficients provide estimates of marginal effects, and in a complicated system in which many factors influence observed outcomes, it may well be that expectations fail to match reality. Imagine a scenario in which a scientist develops an accurate estimate of the marginal effect β_x of policy *x* on violence with the following model and an estimate of $\beta_x = -500$:

$$violence_{T+1} = violence_t + \beta_x \times x + e_1$$

 $violence_{T+1} = violence_t - 500 \times x + e_1$

which in the simplest form suggests that implementing *x* will reduce expected violence by 500 deaths. Now imagine an alternative model that includes another factor previously in the error term *e* which is not needed to accurately estimate the marginal effect, β_x of policy *x*, but does improve model fit, like an upcoming religious holiday, *z*. Furthermore, assume it will increase violence by 600 deaths:

$$violence_{T+1} = violence_t - 500 \times x + 600 \times z + e_2$$

Assuming that the random error is zero other than the effect of the religious holiday, that the policy is implemented, and that previous violence consisted of 1,000 deaths, there will 1,100 deaths in the current period.

The fact that observed violence is higher after implementing a policy expected to reduce violence might seem contradictory to a general audience (that has not taken statistics). Clearly the marginal effect of implementing x was to reduce violence, but only presenting a negative (or positive) marginal effect can also create the expectation that changes in overall violence will be negative (or positive). Well-fitting models that not only accurately estimate marginal effects, i.e. minimize bias, but also fit well, i.e. minimize deviance, reduce the chances of observed outcomes that contradict conventional, non-statistical expectations.¹

It might also be useful to consider the certainty with which one can evaluate the accuracy or truth value of research: predictions about the near future can easily be evaluated once outcomes are observed. For example, we predict that civilian deaths next month will be between 400 and 600. Next month there are 800 civilian deaths. The prediction was inaccurate, and false in a sense. Academic research tends to be more complex and abstract, which often can make it more difficult to evaluate it's accuracy.

¹For a current (2012) political example consider debates about the effectiveness of President Obama's 2009 stimulus package, e.g. "[the] weight of the evidence suggests that fiscal policy softened the impact of the recession", yet "polls show that a sizable majority of voters think that the stimulus either did nothing to help or actively hurt the economy" (The New Yorker, http://www.newyorker.com/talk/financial/2010/09/20/100920ta_talk_surowiecki).

Imagine I find a statistically significant relationship between ethnolinguistic fractionalization (ELF) and civilian deaths in a cross-sectional model that includes 10 other covariates. Past research has found mixed results for the relationship between ELF measures and various aspects of civil conflict (e.g. onset, duration). How do I evaluate the accuracy of this claim? What if the relationship becomes insignificant with a different set of covariates? The point is that it is much easier to show that a prediction was definitely inaccurate, while it is usually more difficult to definitely show that a statistically significant correlation is inaccurate, e.g. due to correlations within the model, omitted variable bias, endogeneity, etc. This places additional value on model fit in the context of policy-related research.

Therefore for a statistical model to be useful to support decision-making, and regardless of whether the main interest is in evaluating potential policies or pure forecasting, it has to match the data well in-sample and forecast well out of sample. Other than being accurate, this will help build confidence in the ability of those models to anticipate the future.

4.3 Violence against civilians in Iraq

How does one develop a model of violence in civil war over time that fits and predicts well? One important task is to find data to model, and I use data on Iraqi civilian deaths from the Iraq Body Count (IBC) project (Hicks et al., 2011).² IBC collects event data that consists of incidents depicting violence against civilians during the Iraq War, starting in January 2003 and updated on a regular basis to the present. Incidents are primarily identified from media sources, but have also incorporated incidents in the Wikileaks database of U.S. military SIGACTs. There are over 27,000 recorded incidents of violent events like bomb explosions, discovered bodies, etc. from 2003 to February 2012.

The IBC data, unlike the various U.S. military SIGACT databases, are publicly available and easily accessible online. They are regularly updated (at least through 2012) and since they incorporate Wikileaks incidents, are likely a more comprehensive source of civilian deaths in Iraq than SIGACTs alone. The majority of coding is done on the basis of English-language commercial media and translations of Middle Eastern and Iraqi media, which however themselves may also include primary local sources, non-political NGOs like the Iraqi Red Crescent, and official cumulative government figures (IBC collection methods).³

4.3.1 Data and machine-coding

For this project, I downloaded the IBC incident data from March 2003 to February 2012, with roughly 27,000 incidents depicting civilian deaths in Iraq. Each incident records a violent event such as bombings, assassinations, discovery of mass graves or

²http://www.iraqbodycount.org/

³http://www.iraqbodycount.org/about/methods/

bodies, etc., and includes the date(s), low and high reported civilian deaths, and location. The data has low and high estimates of civilian deaths to account for incidents in which the exact number of deaths is uncertain or where different media sources give different figures. These data are, in theory, sufficient information to construct data on civilian deaths over time and by location.

Unfortunately the location variable has a large number of idiosyncratic values, typically consisting of strings of multiple words describing a location. Some common patterns are a location relative to a major city or province, e.g. "Hamda street, west Mosul", a location that gives the province, e.g. "Baghdad province", a location which includes a known province or city but with additional characters, e.g. "Baghdad's ... neighborhood" or "Mosul?", strings that include a village name but do not occur very frequently, i.e. it is not a major city, and various misspellings and transliteration errors of all place names ("al-Baghdad", "Bagdad", "Qadisiya", "Qadisiyah", "al-Qadisiyyah").

I use an R function to parse these location strings to identify the province in which the event occurred. The full R code is included in the appendix and it is also available online (GitHub).⁴ It should be able to code provinces for any version of the IBC incident data, i.e. regardless of the time period included.⁵

The function works by identifying a candidate word in the location string and matching it against dictionary of roughly 200 known city province pairs. I coded this dictionary based on online map searches and the NGA GEOnet Names Server (http://earthinfo.nga.mil/gns/html/), starting with the most frequent city occurrences in the raw data. The candidate word is identified by starting with the last word in the location string (in more than half of cases this results in a match), looking for keywords like "province", and lastly by evaluating every remaining word in the string. Each candidate word is parsed to remove special phrases or characters like "al-", "'s" (possessive form), commas, and so on. If none of the candidate words are found in the dictionary, the function returns the original location string, providing a list of failed matches that can be used to improve the dictionary or identify common misspellings in the future.

At this point the function identifies the province for approximately 94 percent of the 27,000 records. The resulting ~26,000 records then have information on the day on which an event occurred and in which province, allowing the data to be aggregated to a province-day format or higher. The majority of missing values appear to be due to two factors: (1) misspelling of a location name that is in the dictionary, and (2) small locations like villages that are not in the dictionary. A few incidents do have completely missing location information.

I use two versions of this data here: panel data of province-months, and a single time series of monthly deaths for all of Iraq. The panel version of the data captures civilian deaths over a period of 108 months (March 2003 to February 2012) for each of Iraq's 18 provinces. The total number of observations for all of Iraq is 1,944 province months,

⁴https://github.com/andybega/Code_IBC_Province

⁵IBC does code the province and city in which an incident occurred as part of their original coding process, and has in the meantime provided me with this information. Their publicly available data does not include this information however and other users may find this script of interest therefore.

and the data encompass 100,839 to 110,107 recorded civilian deaths. The discussion below will include this province-level data, but note that for the statistical modeling I use country-wide totals, giving 108 months of civilian deaths over all of Iraq. Table 4.1 shows the total number of civilians deaths by province, and I will discuss it further below.

4.3.2 The dynamics of violence during the Iraq War

Figure 4.1 shows total monthly civilian deaths for all of Iraq from March 2003 to February 2012, as well as deaths in Baghdad and the rest of Iraq. Violence in Baghad accounts for around 55 percent of all total civilian deaths. Looking the level of monthly deaths in Baghad compared to the rest of Iraq, Baghdad represents the majority of monthly deaths up to 2007, when violence dropped below that in the rest of Iraq.

A brief history of the Iraq conflict is well-reflected in the evolution of civilian deaths over time. During the initial invasion in March and April, more than 200,000 coalition forces fought from stating areas in Saudi Arabia and Kuwait through the south of Iraq and to Baghdad, while lightly supported Kurdish forces fought in the north. Military resistance was weak and most of the Iraqi Army dispersed along with its equipment, thus forming the core of the insurgency. Although the fighting was less intense than expected, it did produce an initial spike of civilian deaths in the southern provinces and Baghdad (see Figure 4.3).

The time period from the end of major combat in May 2003 to late 2007/early 2008 saw a steady increase in violence, and especially in Baghdad. By late 2006 civilian deaths averaged more than 2,500 a month. During this period there are also notable spikes in civilian deaths, corresponding to the First Battle of Fallujah in April 2004, the Second Battle of Fallujah in November 2004, a bridge stampede in Baghdad that killed almost 1,000 in August 2005, and a string of car bombings in Baghdad in July 2006.

In light of a growing civil war, the U.S. announced a surge of troops and change in strategy, i.e. the Surge, in early 2007. Although overall U.S. troop levels increased, the relative change in levels was minor (Belasco, 2009) (CNN chart). Most of the additional troops were concentrated in Baghdad and its vicinity, and forces were dispersed rather than sequestered at large bases (Thiel and Hogan, 2011). Civilian deaths in Baghdad started decreasing in December 2006, before the surge was announced and before the total troop levels increased, suggesting an alternative explanation, e.g. more effective strategy (Thiel and Hogan, 2011), changes in the number or effectiveness of Iraqi security forces, or ethnic homogenization of previously mixed areas in Baghdad. Meanwhile civilian deaths continued to increase in the rest of Iraq until the middle of 2007, after which they started falling as well.

Violence decreased throughout Iraq from the beginning of 2008 on, except for a spike during the Iraqi government's offensive against Shi'a militias in the south in March 2008. They have remained relatively stable since 2009 despite the progressive withdrawal of U.S. forces.

Looking at the underlying trend in violence for all of Iraq, Baghdad, and the rest of Iraq, each series has a particular non-monotonic trend in which violence escalated to some peak before dropping again. Figures 4.4 and 4.3 show monthly deaths for all 18 provinces with and without a common y-scale. Table 4.1 lists deaths by province, showing the minimum and maximum numbers recorded in IBC as well as the percentage of total deaths. More than half of the total deaths occurred in Baghdad, and 3 other provinces, Nineveh in the north (Mosul), Diyala just north of Baghdad, and Anbar in the west, account for another quarter of civilian deaths. Looking at the panels in the figure without a common scale (Figure 4.3), Nineveh and Diyala also seem to have a underlying trend with a peak level of violence around 2007 and 2008. In Anbar the two battles of Fallujah show as the two spikes in 2003 and 2004, but without them there also seems to be a peak around 2007.

Looking at variance, it seems that in provinces with few total deaths, several intense episodes of violence account for most of the violence and show as irregular spikes in the time-series. Consider for example Arbil, with around 230 total deaths. The three spikes are three suicide bombings that together account for around 160 of the 230 total deaths. In general it seems like irregular spikes in violence account for more of the data in provinces with few reported deaths. This is probably for two reasons. First, if the province is remote only high-profile incidents might draw media attention, while other forms of violence like assassinations go unreported. Second, with less overall violence the data are bound to look less smooth, the same way a random draw of 10 values from a normal distribution will not produce a smooth histogram. In any case, this makes it difficult to estimate any underlying trend in violence and so I aggregate the deaths to monthly totals for all of Iraq.

4.4 Modeling civilian deaths

The next two subsections present two statistical model of civilian deaths over time. There is overdispersion in the counts, with a mean number of monthly deaths is 1019 and variance of 857,227. Figure 4.2 shows histograms of monthly deaths and logged monthly deaths. The first visually shows the extent of overdispersion. One approach to modeling death counts is to use normal linear models of the logged counts, but the histogram on the right indicates that the logged counts do not have an unconditional normal distribution. To address these issues, the first model below uses a log-normal exponentially weighted moving average (LN-EWMA) that closely matched observed data, while the second uses a log-normal logistic (LN-logistic) model that uses a fitted logistic density to provide long-range forecasts.

Although the main challenge this paper seeks to address is how to model future civilian deaths given past, observed deaths, the two models presented here do allow for the inclusion of exogenous covariates, and can hence also be used for more traditional hypothesis testing. I use two variables to demonstrate how this can be done: a dummy variable for national-level elections, and a dummy variable for Ramadan.

The national elections variable is coded as 1 if, during that month, Iraq conducted national constitutional referenda or elections. This occurred four times during the time

period covered here: in January 2005 to elect a national assembly to write a constitution, in October 2005 to approve the constitution, and in December 2005 and January 2010 for national parliamentary elections under the new constitution.⁶ Although the elections naturally drew media attention and threats of violence, they were also accompanied by increases in security in anticipation of election violence, and so it is not quite clear whether we should expect more or less violence against civilians during the elections.

The second dummy variables is coded as 1 if the Islamic month of Ramadan included any part of that western calendar month. Ramadan is a month in the Islamic calendar that lasts 29 or 30 days, depending on the moon, and has religious significance as a time of fasting and prayer. Since the Islamic calendar is lunar, the timing of Ramadan changes relative to the western calendar months, and typically overlaps two calendar months. Exceptions, where Ramadan fell within a western calendar month, where in 2008, when it lasted from 1 to 30 September, and 2011, when it lasted from 1 to 30 August.⁷ Given the religious significance of Ramadan, I would except there to be less violence.

4.4.1 LN-EWMA: modeling effects

Returning to the monthly total civilian deaths in all of Iraq shown in Figure 4.1, there is a non-monotonic trend over time, with death counts increasing up to 2007, and thereafter decreasing. An augmented Dickey Fuller test of non-stationarity for this series gives a value of -1.7398, with associated p-value of 0.68 (Said and Dickey, 1984). In a nonstationary time-series, the joint probability distribution changes when shifted in time (Greene, 2008, 636), and here the mean obviously changes over different time periods. Figure 4.5 shows the autocorrelation function for the untransformed time-series on the left, and the large number of significant lags reflect the strong trend in the data.

In the context of count data, Brandt develops two models (2000; 2001) for time-series of counts. The poisson autoregressive model (PAR) adapts an autoregressive latent process to poisson-distributed count processes that are mean-reverting. Since the series here is non-stationary, the alternative Poisson exponentially weighted moving average model (PEWMA) (Brandt et al., 2000), based on Harvey and Fernandes (1989), is more appropriate. It is written in state-space form, with separate equations to describe a latent process giving the state of a system, and a separate observation equation that determines how the state of the system at any given time is observed. The dynamic portions of the model are entirely in the state equations, i.e. this is where past information enters the model.

The model I use here is a simplified adaption with the following observation equation:

$$y_t \sim LogNormal(ln\mu_t, \tau)$$
 (4.1)

⁶From http://en.wikipedia.org/wiki/Politics_of_Iraq.

⁷Based on these conversion tables: http://www.staff.science.uu.nl/~gent0113/islam/ummalqura.htm.

where μ is a latent mean and τ is an estimate of the variance. The original PEWMA uses a poisson distribution to characterize the observation equation, while I use a log normal distribution with a separate variance parameter instead to allow for over dispersion.⁸ In practice, a poisson produces vastly tighter confidence bands that probably understate actual uncertainty given the data.

The system or state transition equation is:

$$\mu_t = \mu_{t-1}^* \times e^{x_t \delta} \tag{4.2}$$

$$\mu_t^* \sim Gamma(a_t, b_t) \tag{4.3}$$

$$a_t = y_t + \omega \times a_{t-1} \tag{4.4}$$

$$b_t = e^{x_t o} + \omega \times b_{t-1} \tag{4.5}$$

where μ_{t-1}^* is a gamma distributed prior that captures all systematic information in the model up to time t-1. Unlike Brandt et al. (2000), the transition equation here is simplified and does not include a period specific growth rate. The variables a_t and b_t are parameters for the prior gamma density and introduce the past history of observed outcomes y and predictors $x\beta$ into the transition equation. The parameter ω provides an estimate of how much the current state of the system depends on past states, i.e. time dependence. At the extremes, and assuming for now that there are no exogenous variables, i.e. $e^{x_t\delta} = 1$, then when $\omega = 1$ the term μ^* collapses into the mean of all past observations of y up to time t. In that case the overall model is equivalent to a standard poisson regression model. When $\omega = 0$, the term μ^* forms a gamma distribution with mean and variance y_{t-1} , akin to an autoregressive model. The model essentially provides an estimate of latent mean violence at any given time and which ranges anywhere from a local average of all past outcomes and an autoregressive term.

This model was estimated with MCMC methods using JAGS and R, and the JAGS model code is provided in the appendix. Because the initial state of the system is unobserved, I assume values for *a*, *b* and μ in the first time period, akin to an assumption about the initial values of a moving average. Subsequent states of the system are then updated according to the transition equations above. I use beta and normal prior densities for ω and δ respectively. The parameters are sampled using one chain with 100,000 iterations, with half discarded for burn-in.

Estimating a complete model requires writing a specific text file describing the model for JAGS and calling it from R. The model is estimated three separate times to obtain parameter estimates, fitted values, and forecast values respectively. Any change in covariates requires writing a new model file, and given that estimation through MCMC itself takes half a minute to a minute, I do not conduct any robustness check or alternative

⁸I use a log-normal rather than negative binomial density because a log-normal is easier to implement in JAGS. The negative-binomial density in JAGS is modeled with parameters for the number of trials and success in each trial, and requires transforming these two parameters if one wants to model the mean instead. This adds complexity, and in this case, an alternative with a negative binomial density has convergence problems.

model specifications.⁹ The covariates included in the model are mainly intended for demonstration in any case, not for their substantive effects themselves.

Coefficient estimates for a LN-EWMA model of all violence in Iraq are shown in the left panel in Figure 4.6. The model includes two exogenous covariates. Both are dummy indicators of national elections (d[1]) and Ramadan respectively (d[2]). The lines show 80% credible intervals, and neither excludes zero. Both do however lean in the direction of a negative effect. Notably, elections are not associated with increased violence. Figure 4.7 shows the in-sample fit of the LN-EWMA model. The red line shows the mean of the posterior density for each month, while the band shows the 80% interval. It nicely captures the general trend in violence except where events like the battles in Fallujah or the Baghdad bridge stampede caused a spike in civilian violence.

The next plot in Figure 4.7 shows forecasted civilian violence from January 2007 on. The forecasts were created by dropping observed deaths after Jan. 2007, refitting the model, and predicting out of sample for the forecast period. Except for changes in the exogenous variables, the forecast is a continuation of the last known value for the moving average. There is a slight drift upwards over time because the log-normal distribution is slightly asymmetric, and the mean forecasts and credible bands are slightly jagged because JAGS samples a posterior density for each individual month. More importantly, the forecasts are not very useful. A forecast from January 2008 or 2009 would show a better fit with observed levels of violence, but only because there are no large changes in the overall level violence from then on. Essentially, the flexibility that allows the model to fit well in sample also seems to make it less useful for forecasting.

4.4.2 LN-logistic: From peak oil to peak conflict

While the LN-EWMA model developed above matches the trend of violence in Iraq well, it is somewhat less useful for long-range forecasts because of the ad hoc nature in which the underlying trend is modeled. For any given conflict, we do have a prior belief about the evolution of violence over time: escalation up to a peak period of violence and subsequent deescalation. Rather than empirically tracing the trend with a moving average, an alternative option is to explicitly model observed deaths as consisting of an underlying trend that is a function of time, a random error component, and exogenous influences.

A relatively simple choice for modeling the underlying trend is to fit a logistic density function of time. A well-known example of this modeling choice is in estimates of oil production. In the 1950's, geologist M.K. Hubbert showed that oil production in a given oil field or country is well approximated with a symmetrical, bell-shaped curve, e.g. a logistic or gaussian density, and that with known total reserves and historical production figures, one can reasonably predict the timing of peak production and future production levels in general (Hubbert, 1956; Bartlett, 2000). Similar models have been used to model

⁹The three JAGS calls through R take 66, 67, and 31 seconds respectively using R 2.13 and JAGS 3.1 on a 2.7 GHz Intel Core i5 running Mac OS X 10.7 in 2012.

the population dynamics of agricultural pests (Matis et al., 2005) and to model market shares of products (Meyer, Yung and Ausubel, 1999).

A general logistic density model with peak at time *p* and growth rate of *r* might look like this:

$$y_t = K \times \frac{r \times e^{(t-p)r}}{(1+e^{(t-p)r})^2}$$

$$y_t = K \times Logistic(t, p, r)$$

The basic assumption underlying application of this type of curve is that the outcome at any given time, whether oil production or violence, is dependent on the cumulative outcomes so far, and specifically, that as some maximum level of total possible output, *K* is reached, growth will be attenuated and eventually reversed by a saturation effect. An intuitive example, although inappropriate as an analogy, of this is in the evolution of an aphid population over time (Matis et al., 2005). Aphids are small insects that feed on plant sap. From an initial population, they will grow exponentially until their number is large enough to exhaust the available food supply. The scarcer food supply becomes, the slower population growth will be in what is called a saturation effect. Eventually, as food supply is depleted, the population will apex and collapse. This decrease will be attenuated over time since less aphids compete for the remaining food supply, but asymptotically nears zero.

In the context of industrial production and population ecology the total output K has a natural interpretation, e.g. total proven oil reserves or carrying capacity of an ecosystem, and it is usually an exogenous, empirically derived parameter. In the context of violence, the timing of peak violence, p, is easy to understand, but interpreting the parameters K and r can be somewhat artificial. The growth parameter r controls how fast violence escalates and subsequently deescalates, while the parameter K is related to the intensity of a conflict. It might seems strange to make *ex ante* assumptions about the intensity of a conflict while it is still going on, but fortunately there is no need to as we can let the data inform our prior expectations about K.

Adopting a general logistic density function to model the latent trend in violence here produces this logistic model of violence:

$$y_t \sim LogNormal(ln\mu_t, \tau)$$

 $\mu_t = trend_t \times e^{x_t \delta}$
 $trend_t = C + K \times Logistic(t, p, r)$

where *K* is an estimate of the total deaths that will occur, *C* is a constant to allow for a background level of violence, *p* is an estimate of the peak conflict timing, and where *r* is an estimate of the rate of escalation and de-escalation. The exogenous factor $x\delta$ enters the model by modifying the logistic trend at time *t*.

Model estimation, as before, is with MCMC sampling using JAGS through R, and the full JAGS model code is shown in the appendix. I use a gamma prior for r and normal priors for the other parameters. Because this model is not dynamic there is no need to

separately model an initial system state, as before. The parameters are sampled using one chain with 100,000 iterations, with half discarded for burn-in. I estimate the model three times to obtain samples for the parameter estimates, fitted values, and forecast values respectively. As before, I do not conduct any robustness checks with alternative sets of covariates, because of the complex workflow involved and the time required to estimate any given model.¹⁰

Using the same exogenous variables as above, i.e. national elections and Ramadan, the coefficient estimates and in-sample fit for a LN-logistic model are shown in figures 4.6 and 4.8. The coefficient estimates are similar to the previous results in that both lean on the negative side but include zero in the 80% credible intervals. The in sample-fit is worse than with the LN-EWMA model but a logistic density is not an unreasonable fit for the observed levels of violence. The main discrepancy other than the previously noted high-casualty battles and incidents is after the peak of violence in early 2007. While the trend seems to be a good description of the escalation of violence up to that point, violence in late 2007 and 2008 dropped to low levels faster than estimated. This is the period that coincides with the surge.

The forecasts were created using the same procedure as before, by dropping observed deaths after January 2007 and reestimating the model. Because the underlying trend is not dynamically modeled, the forecast is an extrapolation of the logistic curve that has been fitted with the available data. As a result the credible bands do not expand in the forecast period–forecasted posterior densities do not enter the calculation for subsequent forecasted posterior densities.

4.4.3 Model fit

A common and general measure for evaluating model fit in the context of Bayesian modeling is the Deviance Information Criterion (DIC). It is a generalization of the Akaika and Bayesian information criteria (AIC and BIC) for hierarchical models, and is calculated as: $DIC = \overline{D} + p_D$, i.e. the mean deviance plus the effective number of parameters p_D . Deviance is defined as $-2 \times log(p(y|\theta))$, where $p(y|\theta)$) is the likelihood of the data given the model parameters. The effective number of parameters is estimated as the variance of the deviance, and penalizes for model complexity (Spiegelhalter et al., 2002). The DIC can be interpreted as the ability of the model to predict new cases, with lower values indicating higher accuracy (Gelman and Hill, 2007, 525), and it also appropriate for evaluating the comparative performance of non-nested models, as is the case here. The DIC values for the log-normal exponential average and logistic models respectively are 1730 and 1950, indicating that the log-normal exponential average model has a much better model fit.

However, the graphs of the forecasts for January 2007 on in Figures 4.7 and 4.8 suggest that out-of sample, the forecasts created by the logistic model are much more accurate. The estimated levels of violence in the logistic model are lower than observed

¹⁰The three JAGS calls through R take 66, 68, and 32 seconds respectively using R 2.13 and JAGS 3.1 on a 2.7 GHz Intel Core i5 running Mac OS X 10.7 in 2012.

violence up to mid-2007, and overestimate violence subsequently. However, the model does correctly predict a reduction in violence after 2007, unlike the LN-EWMA forecast. With a better estimate of the underlying trend in violence, this model seems better suited for long-range forecasting than a moving average driven model.

Since the deviance is calculated on the basis of the likelihood of a fitted model, it is inappropriate for evaluating forecast accuracy. Although there does not seem to be a generally accepted measure of forecast accuracy for count data, a relatively simple measure gives inaccuracy as: $I = \hat{y}/y - 1$, i.e. forecasted over observed counts minus 1 (Parthasarathi and Levinson, 2010). Positive and negative values indicate the factor by which the forecasts over- or under-predict respectively, while a zero indicates perfect accuracy. The range of possible values is bound at the lower end by -1, and has no positive boundary. Figure 4.9 shows the forecast inaccuracy of the two models, as well as, at the bottom, the forecast inaccuracy of an AR(1) linear model. The latter is included to serve as a baseline model, and is easily estimated in R using the arima function. The graphs share common axes, and the logistic model has significantly less forecast inaccuracy than either the LN-EWMA or a simple AR(1) linear model. In fact, an AR(1) linear model does slightly better than the LN-EWMA.

In order to be able to compare forecast accuracy between the models, I calculate the mean absolute forecast inaccuracy for both models. This measure does not reflect the associated forecast uncertainty (Herron, 1999), but since the implementation of the logistic model here does not adequately reflect forecast uncertainty, a summary measure based on the mean forecast seems reasonable. The logistic model has a mean inaccuracy of 0.9, compared to 6.9 and 4.2 for LN-EWMA and an AR(1) model, indicating that while it still overpredicts violence, the logistic model forecasts are much more accurate than the alternatives.

4.4.4 Peak violence and the troop surge

One of the parameters of the fitted logistic model estimates the timing of peak violence. Looking first at the raw data in Figure 4.1, violence in Baghdad started to generally decrease from December 2006 or July 2006 on, depending on whether you want to consider the string of car bombings in Baghdad in July 2006 as an anomaly or not. The most violent month for all of Iraq, other than during the initial invasion, occurred in July 2006 on, with generally decreasing levels of violence after than. The estimated time of peak violence from the logistic model is December 2006, with a narrow posterior density (t = 46.6 with a 95% interval from 46.1 to 47.1).

The surge was announced in January 2007 and consisted of a redeployment of additional troops to Baghdad and its vicinity, changes in leadership, and changes in counterinsurgency tactics (Petraeus, 2007). It culminated in a series of offensives in June 2007, after which attacks recorded by the U.S. military in Baghdad started to decrease, as outlined in a report by Gen. Petraeus to Congress in the aftermath of the surge (Petraeus, 2007). Notably, a large proportion of these attacks consist of bombs (IEDs) and indirect fire, and therefore were likely directed at U.S. and Iraqi security forces. Judging from the IBC data, which has been corrected to include the civilian deaths in the U.S. military SIGACT data leaked by Wikileaks, the peak level of civilian deaths in Iraq and in Baghdad occurred in December 2006, before or just as the surge was announced. During the surge, i.e. in the first half of 2007, civilian deaths did increase in the rest of Iraq however, which would be expected if the surge and focus on Baghdad reduced security forces, both U.S. and Iraq, elsewhere.

An alternative argument is that violence in Baghdad decreased due to ethnic cleansing and eventual homogenization of neighborhoods. Agnew et al. (2008) examine nighttime light data gathered by satellite for Iraq's major cities, and find that it decreased from late 2006 until 2007, whereas it continued to increase in Iraq's other major cities. Furthermore, they find that the biggest disruption in electricity was in ethnically mixed neighborhoods of Baghdad, and interpret this as support for the argument that ethnic segregation was responsible for the reduction of violence.

Without data on Iraqi and U.S. troop levels in other parts of Iraq during this period, or data on refugee flows, it seems difficult to distinguish whether the troop surge or ethnic homogenization was responsible for the decrease in civilian deaths after December 2006. The timing of the decrease suggest ethnic homogenization, while the rise in civilian deaths in other parts of the time seems to be more consistent with a troop surge argument.

4.5 Conclusion

The goal of this paper was to present statistical models that fit well in sample, to facilitate hypothesis testing, and that predict well out of sample. The two models in practice trade off usefulness for in-sample hypothesis testing and the ability to provide longrange forecasts. The issue at the core of this trade off is in how the underlying conflict dynamics are modeled. The log-normal exponential moving average (LN-EWMA) model makes relatively flexible assumptions about the underlying conflict trend and hence fits given data well. It is less useful for forecasting beyond the immediate future for the same reason, while a log-normal logistic (LN-logistic) model that assumes conflict will follow a logistic density over time fits the data less well but provides more accurate long-range forecasts.

The logistic model is based on a logistic density with elapsed time as a parameter. A better way to do this would be to create a dynamic growth model in which time is eliminated as a parameter and instead past observed or latent violence forms the basis for a set of state equations. In population ecology, a fairly simple set of differential equations can be used to model population growth over time in what is commonly called the "logistic equation".¹¹ The resulting function is in fact similar to the cumulative logistic density, the derivative of which, i.e. the logistic probability density function, is used as the basis of the current model. A rewritten version of the log-normal logistic model with

¹¹The logistic equation can be written as $dN/dT = rN \times (1 - N/K)$, where *N* is the current population, *r* is the growth rate, and *K* is the carrying capacity (Hoppensteadt, 2006).

a dynamic trend model would allow forecast uncertainty to enter long-term forecasts, unlike in the current model, and it would also make it easier to make changes to the way trend is modeled. Unfortunately such a version is also more difficult to implement in JAGS.

Furthermore, looking again at the series of monthly civilian deaths in Figure 4.1, it looks like violence increased at a slower rate up to a peak in late 2006 than it decreased subsequently. A model that fits two or more partial logistic curves with different growth rates, also known as "loglet analysis" is a practical option (Meyer, Yung and Ausubel, 1999). This seems theoretically somewhat agnostic however and it would make forecasting more challenging. A more attractive alternative could be to create an explicit model of the growth rate itself. One could for example argue that changes in Iraqi state capacity over time or growing counterinsurgency experience in the U.S. and Iraqi militaries should decrease this over time. In fact, a logistic density with a shape parameter that decreases as a function of time, when plotted, shows a shape similar to the plotted civilian deaths, with a sharper decrease after the peak.¹²

Since neither of the two models presented here meets both desirable criteria, i.e. insample fit and long-range forecasting accuracy, another extension might be based on a combination of the two models presented here, with a moving average component but also an explicitly assumption about the trend. This could be as simple as pooling the posterior predictive densities from two separately estimated models to create one averaged estimate or using bayesian model averaging to combine them in a weighted average (Montgomery and Nyhan, 2010).

Despite these shortcoming, the two models presented here each should allow for the evaluation of a policy's likely short-term impact on violence as well for long-range forecasting, respectively. Although the covariates included in the model here reflect exogenous factors, i.e. elections and religious dates, they can accommodate variables related to policy decisions, e.g. US troops levels, development spending, Iraqi Army and Police levels, etc. Disaggregated conflict data are increasingly becoming available, and I would argue that in most situations we can reasonably expect violence during the conflict to follow a general curve like shape with a peak period of conflict and longer periods of escalation and deescalation. Instead of considering this particular non-monotonic trend a nuisance to be treated with fixed effects in a panel regression, why not explicitly model it instead?

This question, how to dynamically model violence in civil conflicts, leads to a number of interesting avenues for future research. How do government and insurgent numbers evolve over time and what is their relationship to popular support and violence? Predator-prey and other models from population ecology (Hoppensteadt, 2006) use sets of differential equations to model the levels of competing species, and thus may provide interesting approaches to modeling the interaction between government, civilians, and insurgents in a civil conflict. With information or estimates of the actual size of gov-

¹²Rewriting the logistic density as $sech^2(t)$, a plot similar to $v(t) = sech^2(\frac{t}{2+0.1 \times t})$ will produce such a shape (link to plot).

ernment or insurgent forces and observed violence, and their growth rates, it might be possible to estimate parameters that govern the interaction between these factors, like government effectiveness.



Figure 4.1: Civilian deaths in Iraq, 2003 to 2012. The black line shows total deaths, red are deaths in Baghad, and blue are deaths in the rest of Iraq. Source: Iraq Body Count.



Figure 4.2: Histograms of monthly, Iraq-wide civilians deaths and log-transformed monthly counts.

Province	Min. Dead		Max. Dead	
Anbar	6,350	6	7,041	6
Arbil	212	<1	242	<1
Babil	2,789	3	3,080	3
Baghdad	55,028	55	60,214	55
Basra	3,667	4	4,123	4
Dakuk	2	<1	2	<1
Dhi Quar	1,175	1	1,199	1
Diyala	9,075	9	9,744	9
Karbala	1,883	2	2,064	2
Kirkuk	2,885	3	3,048	3
Maysan	262	<1	284	<1
Muthanna	195	<1	202	<1
Najaf	1,048	1	1,584	1
Nineveh	8,806	9	9,263	8
Qadisiyah	582	<1	607	<1
Salah ad Din	5,285	5	5728	5
Sulaymaniyah	61	<1	66	<1
Wasit	1,534	2	1616	1
Total	100,839		110,107	

Table 4.1: Civilian deaths from March 2003 to February 2012 by province



Figure 4.3: Time series of civilian deaths, 2003 to 2012, by province. The plots do not share a common scale, allowing the dynamics of civilians deaths to be compared across provinces.



Figure 4.4: Time series of civilian deaths, 2003 to 2012, by province. The plots share a common scale, and show that relative levels of violence across Iraq were minor compared to the scale of deaths in Baghdad.



Figure 4.5: Autocorrelation functions for Iraq-wide deaths and the first difference in logged deaths.



Figure 4.6: Parameter estimates for LN-EWMA and LN-logistic models. The parameter d[1] is for national constitutional or parliamentary elections, and d[2] is for calendar months that include part of the Arab month of Ramadan.



Figure 4.7: LN-EWMA in-sample fit (top) and forecast (bottom). While the model matches observed data well, its forecasts are essentially a continuation of the last observed value, depending on the estimated value of ω and accounting for a slight upward drift due to the non-symmetrical log-normal outcome equation.



Figure 4.8: LN-logistic model in-sample fit (top) and forecast (bottom). The in-sample fit is decent for observed data but worse than the LN-EWMAs. However, the forecast from January 2007 on is over the long-run more useful. Although it does not match realized outcomes closely, it is more accurate than LN-EWMA forecasts and does accurately predict a downward trend in violence.



CHAPTER 5

CONCLUSION

This dissertation presented three essays predicting interstate war deaths, the spatial distribution of civilian deaths in the Bosnia War, and forecasting civilian deaths in the Iraq War. Aside from predictive rather than causal modeling, a common theme is an emphasis on theoretically-driven model development and structure. Quantitative analyses in international relations and conflict research using regression models like the various binary response and count models readily available in statistical software use theory to inform what sets of exogenous covariates (independent variables) are included, while the choice of regression model itself is oftentimes driven by the nature of the dependent variable. The models in all three essays here reflect "theory" more in the structure of the statistical model itself, as part of the data generating process, rather than in the set of exogenous covariates for any given model (King, 1998, 13).

A significant issue not addressed is measurement error in the data on political violence. Any single data source on violence is likely a convenience sample, i.e. a nonrandom sample in which there are systematic differences in the rate with which cases are included. For example, media-based sources are more likely to include events in easily accessible and prominent urban areas than remote rural areas. Davenport and Ball (2002) examine three different sources regarding violence during the Guatemalan civil war: newspapers, human rights documents, and eyewitness interviews. They find that the three sources differ substantially not only in the magnitude of reported violence, but also in the patterns of reported violence. As a result, statistical analysis using any single source can provide substantially different results and false inferences (Ball, Kobrak and Spirer, 1999).

The data on political violence used in these three essays is almost certainly a convenience sample that inaccurately measures actual violence across the involved conflicts. Data for battle deaths in interstate wars comes from historical sources and the data on civilian deaths during the Bosnian War is based on post-war documentation. The Iraq Body Count data for civilian deaths in Iraq is based on media reports, although it does included additional cases of civilian deaths recorded by the U.S. military. Thus a significant caveat to this research is that the results are likely influenced by systematic inaccuracies in the data used.

Multiple systems estimation (MSE) is a statistical technique to correct for measure-

ment error (Ball et al., 2002, 2003). It uses the information on how cases in 3 or more data sources overlap to estimate the number of missing cases, thus providing a correction for measurement error. MSE data for Guatemala shows that media reports significantly misrepresent the actual magnitude and pattern in violence, and that a significant proportion of violence targeting Mayans in rural areas was underreported (Ball, Kobrak and Spirer, 1999). In Peru, MSE data shows that, contrary to prior estimates, the Shining Path was responsible for the majority of deaths, and that deaths among rural and indigenous populations were also undercounted (Ball et al., 2003). In general, it seems that MSE can significantly improve accuracy provided that at least 3 data sources covering a conflict are available.

Aside from necessarily improving the quality of the data, there are two approaches that I would like to pursue in the future. The first is to integrate the ability to predict across space and across time into one framework, i.e. a forecasting model appropriate for panel data of violence. The goal would be to create a set of statistical models that can provide forecasts of future levels of violence with some spatial resolution, akin to weather prediction for conflicts. The second is to create more complex models of the interactions within a conflict, between actors like the state and insurgents, that play a significant part in generating the violence we observe. The last essay references systems of differential equations from population ecology that depict complex systems, and this seems like a fruitful area for modeling the interactions between government, insurgents, and civilians in a conflict.

APPENDIX A

REPLICATION CODE FOR BOSNIA CHAPTER

A.1 JAGS code: Constant-only negative binomial model

The JAGS code below shows an implementation of a basic negative binomial regression model with only a constant term. The number of civilian deaths by municipality, killed[i], is modeled with a negative binomial distribution with mean lambda[i] and dispersion parameter r. Because the negative binomial density is implemented in JAGS using probability and scale parameters, an additional section of code transforms the mean into the equivalent probability p. A constant, b[1], is introduced in exponentiated form via a log link function.

```
model {
    # Model
    for(i in 1:N) {
        killed[i] ~ dnegbin(p[i],r)
        p[i] <- r/(r+lambda[i])
        lambda[i] <- theta[i]
        log(theta[i]) <- b[1]
        }
    # Prior
    b[1] ~ dnorm(0,5)
    r ~ dgamma(1,1)
    }
</pre>
```

A.2 JAGS code: Constant-only spatial count model

The JAGS code below shows the negative-binomial spatial model with a population offset and constant only. Building on the negative-binomial model above, this version has a spatial lag s.killed and population offset pop91 that directly enter the mean function lambda. Equivalently, logged versions of them could be included as additive terms in theta, but for aesthetic reasons are coded as here. Other coefficients, in this case a constant, b[1], are introduced in exponentiated form via theta.

```
model {
    # Model
    for(i in 1:N) {
        killed[i] ~ dnegbin(p[i],r)
        p[i] <- r/(r+lambda[i])
        lambda[i] <- theta[i] * s.killed[i]^rhos * pop91[i]
        log(theta[i]) <- b[1]
        }
    # Prior
    b[1] ~ dnorm(0,5)
    rhos ~ dnorm(0,1)
    r ~ dgamma(1,1)
    }
</pre>
```

APPENDIX B

REPLICATION CODE FOR IRAQ CHAPTER

Below is the R code for coding IBC incident provinces, as well as the JAGS code for the two models presented in the paper.

B.1 Coding incident location

The function below uses the location string from the IBC data and parses it for matches in a dictionary of city province pairs. The function takes the location string, splits it into separate words, and deletes idiosyncratic characters and phrases using the function DelChar to pass a candidate word on. Starting with the last word in the location phrase, GetProvince looks for a match in a dictionary of city and province pairs, and assigns a return value as appropriate.

The function codes approximately 94% of IBC incidents. The remainder are either because a genuine place name is not in the dictionary, e.g. small villages, or because of spelling variations for a place in the dictionary. A function cycling through some of these common misspellings would improve coding accuracy.

```
# Parse location for matched in the city.prov dictionary.
GetProvince <- function(location, city.prov) {
    # Function for deleting "al-" and other special characters
    DelChar <- function(x) {
        word <- sub("al-", "", x, ignore.case = TRUE)
        word <- gsub("'[s]", "", word)
        word <- gsub("[?,']", word)
        return(word)
    }
    # Function for looking up in city-province list
    ProvLook <- function(x) {
        province <- NULL
        repeat {
            if (x %in% city.prov[, 2]) {
                province <- x</pre>
```

```
break
    } else if (x %in% city.prov[, 1]) {
      province <- city.prov[(x==city.prov[, 1]), 2]</pre>
      break
    } else {
      province <- "no match"
      break
    }
  }
  return(province)
}
# Split up into separate words and get number of words
words <- str_trim(location)</pre>
words <- unlist(strsplit(words, " "))</pre>
words <- unlist(strsplit(words, "/"))</pre>
n <- length(words)</pre>
# Loop over words to find match
province <- ""
while (province == "") {
  # Exception for zero length location
  if (location == "") {
    province <- "no location"
    break
  }
  # Loop to get candidate string
  for (i in n:1) {
    # Get a candidate value for province
    p.cand <- DelChar(words[i])</pre>
    # Look up candidate value
    province <- ProvLook(p.cand)</pre>
    if (province != "no match") { break }
  }
  # Note if no matches
  if (province == "no match") {
    province <- paste("Z:", location)</pre>
  }
}
return(province)
```

}

B.2 JAGS code: LN-EWMA

Log-normal exponentially weighted moving average model, similar to the Poisson exponentially weighted moving average (PEWMA) model (Brandt et al., 2000).

```
model {
  # Initial state
  a[1] <- 2000
  b[1] <- 4
  xd[1] <- d[1] * election[1] + d[2] * ramadan[1]
  mu.start ~ dgamma(a[1], b[1])
  mu[1] <- mu.start * exp(xd[1])</pre>
  # State transition model
  for (t in 2:N) {
    mu[t] <- mustar[t-1] * exp(xd[t])</pre>
    mustar[t-1] ~ dgamma(a[t-1], b[t-1])
    a[t] \leq deaths[t] + w * a[t-1]
    b[t] \le exp(xd[t]) + w * b[t-1]
    xd[t] <- d[1] * election[t] + d[2] * ramadan[t]
  }
  # Observation model
  for (t in 1:N) {
    deaths[t] ~ dlnorm(lmu[t], tau)
    yhat[t] ~ dlnorm(lmu[t], tau)
    lmu[t] <- log(mu[t])</pre>
  }
  # Parameter priors
  tau <- 1/pow(sigma, 2)</pre>
  sigma ~ dunif(0.01, 0.2)
  w ~ dbeta(0.5, 0.5)
  d[1] ~ dnorm(0, 100)
  d[2] ~ dnorm(0, 100)
}
```

B.3 JAGS code: LN-logistic

Log-normal logistic model. This model fits a latent mean of violence, mu, that consists of exogenous variables, xd, and a latent logistic trend with parameters C as a constant, K as scaling parameter, peak at time p and shape r. The logistic trend is a function of time t, and past observed outcomes do not enter the latent trend estimation except through the observation model. The latter describes observed outcomes (deaths) as a stochastic log-normal function of the log latent mean lmu with variance tau.

```
model {
  for (t in 1:N) {
    # Observation model
    deaths[t] ~ dlnorm(lmu[t], tau)
    yhat[t] ~ dlnorm(lmu[t], tau)
    lmu[t] <- log(mu[t])</pre>
    # State model
    mu[t] <- trend[t] * exp(xd[t])</pre>
    trend[t] <- C + K * dlogis(t, p, r)</pre>
    xd[t] <- d[1] * election[t] + d[2] * ramadan[t]</pre>
  }
  # Priors
  tau <- 1/pow(sigma, 2)</pre>
  sigma ~ dunif(0.01, 0.2)
  K ~ dnorm(80000, 30000)
  C ~ dnorm(300, 100)
  p ~ dnorm(50, 10)
  r ~ dgamma(2, 10)
  d[1] ~ dnorm(0, 100)
  d[2] ~ dnorm(0, 100)
}
```

REFERENCES

- Agnew, J., T. Gillespie, J. Gonzalez and B. Min. 2008. "Baghdad nights: evaluating the US military 'Surge' using nighttime light signatures." California Center for Population Research online working paper series.
- Ball, Patrick, Jana Asher, David Sulmont and Daniel Manrique. 2003. How many Peruvians have died? An estimate of the total number of victimes killed or disappeared in the armed internal conflict between 1980 and 2000. Technical report American Association for the Advancement of Science.
- Ball, Patrick, Paul Kobrak and Herbert F. Spirer. 1999. State violence in Guatemala, 1960-1996: a quantitative reflection. Technical report American Association for the Advancement of Science.
- Ball, Patrick, Wendy Betts, Fritz Scheuren, Jana Dudukovich and Jana Asher. 2002. Killings and refugee flow in Kosovo, March – June 1999: a report to the International Criminal Tribunal for the Former Yugoslavia. Technical report American Association for the Advancement of Science.
- Bamgbose, J.A. 2009. "Falsification of population census data in a heterogeneous Nigerian state: the fourth republic example." *African Journal of Political Science and International Relations* 3(8):311–319.
- Bartlett, A.A. 2000. "An analysis of US and world oil production patterns using Hubbertstyle curves." *Mathematical Geology* 32(1):1–17.
- Beath, A., F. Christia and R. Enikolopov. 2011. "Winning hearts and minds through development aid: evidence from a field experiment in Afghanistan." Center for Economic and Financial Research (CEFIR) Working Paper.
- Belasco, A. 2009. Troop levels in the Afghan and Iraq Wars, FY 2001-FY 2012: Cost and other potential issues. Technical report Congressional Research Service.
- Bennett, D. Scott and A.C. Stam. 2006. "Predicting the length of the 2003 US–Iraq War." *Foreign Policy Analysis* 2(2):101–116.
- Bennett, D. Scott and Allan C. Stam III. 1996. "The duration of interstate wars, 1816-1985." *American Political Science Review* 90(2):239–257.

- Berman, E., Jacob N. Shapiro and Jacob H. Felter. 2011. "Can hearts and minds be bought? the economics of counterinsurgency in iraq." *Journal of Political Economy* 119(4):766.
- Berman, E., M. Callen, Jacob H. Felter and Jacob N. Shapiro. 2011. "Do working men rebel? insurgency and unemployment in Afghanistan, Iraq, and the Philippines." *Journal of Conflict Resolution* 55(4):496–528.
- Blimes, R.J. 2006. "The indirect effect of ethnic heterogeneity on the likelihood of civil war onset." *Journal of Conflict Resolution* 50(4):536–547.
- Box, G.E.P., G.M. Jenkins and G.C. Reinsel. 1994. *Time series analysis: forecasting & control.* 3rd ed. Upper Saddle River, NJ: Prentice Hall.
- Brandt, Patrick T. and John R. Freeman. 2006. "Advances in Bayesian time series modeling and the study of politics: theory testing, forecasting, and policy analysis." *Political Analysis* 14(1):1–36.
- Brandt, Patrick T., John R. Freeman and Phil A. Schrodt. 2011. "Real time, time series forecasting of inter-and intra-state political conflict." *Conflict Management and Peace Science* 28(1):41–64.
- Brandt, Patrick T. and John T. Williams. 2001. "A linear Poisson autoregressive model: the Poisson AR (p) model." *Political Analysis* 9(2):164–184.
- Brandt, Patrick T., John T. Williams, Benjamin Fordham and Brian Pollins. 2000. "Dynamic modeling for persistent event-count time series." *American Journal of Political Science* 44(4):823–843.
- Bringa, T.R. 2010. "Nationality categories, national identification and identity formation in 'multinational' Bosnia." *Anthropology of East Europe Review* 11(1):80–89.
- Bueno de Mesquita, Bruce and Randolph M. Siverson. 1995. "War and the survival of political leaders: a comparative study of regime types and political accountability." *American Political Science Review* pp. 841–855.
- Buhaug, Halvard and Jan Ketil Rød. 2006. "Local determinants of African civil wars, 1970-2001." *Political Geography* 25(3):315–335.
- Buhaug, Halvard and Päivi Lujala. 2005. "Accounting for scale: measuring geography in quantitative studies of civil war." *Political Geography* 24(4):399–418.
- Buhaug, Halvard and Scott Gates. 2002. "The geography of civil war." *Journal of Peace Research* 39(4):417–433.
- Cameron, A. Colin and Frank A.G. Windmeijer. 1996. "R-squared measures for count data regression models with applications to health-care utilization." *Journal of Business & Economic Statistics* 14(2):209–220.
- Cameron, A. Colin and P.K. Trivedi. 1998. *Regression analysis of count data*. Cambridge University Press Press.
- Carson, J.L., J.A. Jenkins, D.W. Rohde and Mark A. Souva. 2001. "The impact of national tides and district-level effects on electoral outcomes: the US Congressional elections of 1862-63." *American Journal of Political Science* 45(4):887–898.
- Cederman, Lars-Erik. 2003. "Modeling the size of wars: from billiard balls to sandpiles." *American Political Science Review* 97(1):135–150.
- Cioffi-Revilla, C. 1991. "On the likely magnitude, extent, and duration of an Iraq-UN war." *Journal of Conflict Resolution* 35(3):387.
- Clauset, A., C.R. Shalizi and M.E.J. Newman. 2007. "Power-law distributions in empirical data." *Arxiv preprint arXiv:0706.1062*.
- Coghlan, B., R.J. Brennan, P. Ngoy, D. Dofara, B. Otto, M. Clements and T. Stewart. 2006. "Mortality in the Democratic Republic of Congo: a nationwide survey." *The Lancet* 367(9504):44–51.
- Collier, P. and A. Hoeffler. 2004. "Greed and grievance in civil war." *Oxford Economic Papers* 50(4).
- Condra, Luke N. and Jacob N. Shapiro. 2010. "Who takes the blame? The strategic effects of collateral damage." *American Journal of Political Science*.
- Crost, Benjamin, Jacob H. Felter and Patrick Johnston. 2012. "Government transfers and civil conflict experimental evidence from the Philippines." Manuscript. **URL:** *http://patrickjohnston.info/materials/4Ps.pdf*
- Crost, Benjamin and Patrick Johnston. 2010. Aid under fire: development projects and civil conflict. Technical report Belfer Center Discussion Paper.
- Davenport, C. and P. Ball. 2002. "Views to a Kill." *Journal of Conflict Resolution* 46(3):427–450.
- DeMaris, A. 2004. *Regression with social data: modeling continuous and limited response variables.* Vol. 417 Wiley-Blackwell.
- DiCiccio, Thomas J. and Bradley Efron. 1996. "Bootstrap confidence intervals." *Statistical Science* 11(3):189–212.
- Duan, N. 1983. "Smearing estimate: a nonparametric retransformation method." *Journal of the American Statistical Association* 78(383):605–610.
- Dupuy, R.E. and T.N. Dupuy. 1986. *The encyclopedia of military history: from 3500 BC to the present*. 2nd ed. New York: HarperCollins.

- Fearon, James D. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88(3):577–592.
- Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3):379–379.
- Fearon, James D. and David D. Laitin. 2003. "Ethnicity, insurgency, and civil war." American Political Science Review 97(1):75–90.
- Felter, Jacob H. 2005. "Taking guns to a knife fight: counterinsurgency in the Philippines." Unpublished Doctoral Dissertation, Stanford University.
- Filson, Darren and Suzanne Werner. 2002. "A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war." *American Journal of Political Science* 46(4):819–837.
- Filson, Darren and Suzanne Werner. 2004. "Bargaining and fighting: the impact of regime type on war onset, duration, and outcomes." *American Journal of Political Science* 48(2):296–313.
- Galula, D. 1964. *Counterinsurgency warfare: theory and practice.* Santa Barbara, CA: Praeger Security International.
- Galula, D. 2006. Pacification in Algeria, 1956-1958. Vol. 478 RAND Corporation.
- Gartner, S.S. 2008. "The multiple effects of casualties on public support for war: an experimental approach." *American Political Science Review* 102(01):95–106.
- Gartzke, E. 1999. "War is in the error term." International Organization 53(3):567–587.
- Gelman, A. and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Ghosn, F. and A. Khoury. 2011. "Lebanon after the civil war: peace or the illusion of peace?" *The Middle East Journal* 65(3):381–397.
- Gill, J. 2002. *Bayesian methods: A social and behavioral sciences approach*. Boca Raton: CRC Press.
- Glenny, M. 2001. *The Balkans: nationalism, war, and the Great Powers, 1804-1999.* London: Penguin Books.
- Goldstein, Joshua S. and Jon C. Pevehouse. 1997. "Reciprocity, bullying, and international cooperation: time-series analysis of the Bosnia conflict." *American Political Science Review* pp. 515–529.

Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, Jay Ulfelder and Mark Woodward. 2010. "A global model for forecasting political instability." *American Journal of Political Science* 54(1):190–208.

Government Accountability Office. 2007. "Securing, stabilizing, and rebuilding Iraq: Iraqi government has not met most legislative, security, and economic benchmarks." Report to Congressional Committees. URL: http://www.gao.gov/new.items/d071195.pdf

Greene, HW. 2008. Econometric Analysis. 6th ed. Upper Saddle River, NJ: Prentice Hall.

- Gupta, S., M. Shuaib, S. Becker, M.M. Rahman and D.H. Peters. 2011. "Multiple indicator cluster survey 2003 in Afghanistan: outdated sampling frame and the effect of sampling weights on estimates of maternal and child health coverage." *Journal of health, population, and nutrition* 29(4):388.
- Gurr, Ted Robert and Barbara Harff. 1994. *Ethnic conflict in world politics*. Boulder, CO: Westview Press.
- Harvey, A.C. and C. Fernandes. 1989. "Time series models for count or qualitative observations." *Journal of Business & Economic Statistics* 7(4):407–417.
- Hegre, Håvard, G. Østby and Clionadh Raleigh. 2009. "Poverty and civil war events." *Journal of Conflict Resolution* 53(4):598–623.
- Hensel, P.R., Sara McLaughlin Mitchell, Thomas E. Sowers and Clayton L. Thyne. 2008."Bones of contention: comparing territorial, maritime, and river issues." *Journal of Conflict Resolution* 52(1):117–143.
- Herron, Michael C. 1999. "Postestimation uncertainty in limited dependent variable models." *Political Analysis* 8(1):83–98.
- Hicks, M.H.R., H. Dardagan, G.G. Serdán, P.M. Bagnall, J.A. Sloboda and M. Spagat. 2011.
 "Violent deaths of Iraqi civilians, 2003–2008: analysis by perpetrator, weapon, time, and location." *PLoS Medicine* 8(2):e1000415.
- Holsti, K.J. 1991. *Peace and war: armed conflicts and international order, 1648-1989.* Vol. 14 New York: Cambridge University Press.

Hoppensteadt, F. 2006. "Predator-prey model." Scholarpedia 1(10):1563.

- Hubbert, M.K. 1956. Nuclear energy and the fossil fuel. In *Drilling and production practice*. American Petroleum Institute. URL: http://energycrisis.biz/hubbert/1956/1956.pdf
- Hultman, Lisa. 2012. "COIN and civilian collaterals: patterns of violence in Afghanistan, 2004–2009." *Small Wars & Insurgencies* 23(2):245–263.

- International Security Assistance Force. 2009. "Tactical directive." Memorandum. URL: http://www.nato.int/isaf/docu/official_texts/Tactical_Directive_090706.pdf
- International Security Assistance Force. 2010. "COMISAF's COIN guidance." Memorandum.
 - **URL:** *http://www.isaf.nato.int/images/stories/File/COMISAF/COMISAF_COIN_Guid-ance_01Aug10_.doc*
- Judah, T. 2000. *The Serbs: history, myth, and the destruction of Yugoslavia*. New Haven, CT: Yale Nota Bene.
- Kalyvas, Stathis N. 2006. *The logic of violence in civil war*. New York: Cambridge University Press.
- King, Gary. 1998. Unifying political methodology: the likelihood theory of statistical inference. Ann Arbor, MI: University of Michigan Press.
- Lacina, Bethany. 2006. "Explaining the severity of civil wars." *Journal of Conflict Resolution* 50(2):276.
- Lacina, Bethany and Nils Petter Gleditsch. 2005. "Monitoring trends in global combat: a new dataset of battle deaths." *European Journal of Population/Revue Européenne de Démographie* 21(2):145–166.
- Lambert, D.M., J.P. Brown and R.J.G.M. Florax. 2010. "A two-step estimator for a spatial lag model of counts: theory, small sample performance and an application." *Regional Science and Urban Economics* 40(4):241–252.
- Leventoğlu, B. and Branislav L. Slantchev. 2007. "The armed peace: a punctuated equilibrium theory of war." *American Journal of Political Science* 51(4):755–771.
- Levy, Jack S. and T. Clifton Morgan. 1984. "The frequency and seriousness of war: an inverse relationship?" *Journal of Conflict Resolution* 28(4):731–749.
- Lewis-Beck, M.S. 2005. "Election forecasting: principles and practice." *The British Journal of Politics & International Relations* 7(2):145–164.
- Lyall, Jason and Isaiah Wilson. 2009. "Rage against the machines: explaining outcomes in counterinsurgency wars." *International Organization* 63(1):67–106.
- Makul, Anes and Heather McRobie. 2011. "Yugoslavs in the twenty-first century: 'erased' people." Online.

URL: *http://www.opendemocracy.net/heather-mcrobie-anes-makul/yugoslavs-in-twenty-first-century-'erased'-people*

Malcolm, N. 1996. Bosnia: A short history. New York: New York University Press.

- Marshall, Monty G. and K. Jaggers. 2002. "Polity IV Data Set." Available online. **URL:** *http://www.systemicpeace.org/polity/polity4.htm*
- Matis, J.H., T.R. Kiffe, T.I. Matis and D.E. Stevenson. 2005. "Nonlinear stochastic modeling of aphid population growth." *Mathematical biosciences* 198(2):148–168.
- Melander, Erik and Ralph Sundberg. 2011. "Climate change, environmental stress, and violent conflict: tests introducing the UCDP Georeferenced Event Dataset." Paper presented at the International Studies Association, March 16-19, Montreal, Canada.
- Meyer, P.S., J.W. Yung and J.H. Ausubel. 1999. "A primer on logistic growth and substitution: the mathematics of the loglet lab software." *Technological Forecasting and Social Change* 61(3):247–271.
- Montalvo, J.G. and M. Reynal-Querol. 2005. "Ethnic polarization, potential conflict, and civil wars." *The American Economic Review* 95(3):796–816.
- Montgomery, J.M. and B. Nyhan. 2010. "Bayesian model averaging: theoretical developments and practical applications." *Political Analysis* 18(2):245–270.
- Montgomery, J.M., F.M. Hollenbach and Michael D. Ward. 2012. "Improving predictions using ensemble bayesian model averaging." *Political Analysis*.
- Nation, R.C. 2003. War in the Balkans, 1991-2002. Strategic Studies Institute.
- O'Brien, Sean P. 2010. "Crisis early warning and decision support: contemporary approaches and thoughts on future research." *International Studies Review* 12(1):87–104.
- Parthasarathi, P. and D. Levinson. 2010. "Post-construction evaluation of traffic forecast accuracy." *Transport Policy* 17(6):428–443.
- Petraeus, David H. 2007. "Report to Congress on the situation in Iraq." House Committee on Foreign Affairs. URL: http://foreignaffairs.house.gov/110/pet091007.pdf
- Petrovic, R. 1992. "The national composition of Yugoslavia's population, 1991." *Yugoslav survey* 33(1):3.
- Pillar, P.R. 1983. *Negotiating peace: war termination as a bargaining process*. Princeton, NJ: Princeton University Press.
- Posen, Barry. 1993. "The security dilemma and ethnic conflict." Survival 35(1):27-47.
- Powell, Robert. 1996. "Stability and the distribution of power." World Politics 48:239–267.
- Powell, Robert. 2004. "Bargaining and learning while fighting." *American Journal of Political Science* 48(2):344–361.

- Powell, Robert. 2006. "War as a commitment problem." *International Organization* 60(1):169–203.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED: An armed conflict location and event dataset." *Journal of Peace Research* 47(5):651–660.
- Raleigh, Clionadh and H. Urdal. 2007. "Climate change, environmental degradation and armed conflict." *Political Geography* 26(6):674–694.
- Raleigh, Clionadh and Håvard Hegre. 2009. "Population size, concentration, and civil war. A geographically disaggregated analysis." *Political Geography* 28(4):224–238.
- Reed, William. 2003. "Information, power, and war." *American Political Science Review* 97(4):633–641.
- Reiter, Dan. 2003. "Exploring the bargaining model of war." *Perspective on Politics* 1(1):27–43.
- Said, S.E. and D.A. Dickey. 1984. "Testing for unit roots in autoregressive-moving average models of unknown order." *Biometrika* 71(3):599–607.
- Sarkees, M.R., F.W. Wayman and J. David Singer. 2003. "Inter-state, intra-state, and extrastate wars: a comprehensive look at their distribution over time, 1816–1997." *International Studies Quarterly* 47(1):49–70.
- Sarkees, M.R. and P. Schafer. 2000. "The Correlates of War data on war: an update to 1997." *Conflict Management and Peace Science* 18(1):123–144.
- Schneider, G., Nils Petter Gleditsch and S. Carey. 2011. "Forecasting in international relations: one quest, three approaches." *Conflict Management and Peace Science* 28(5):5– 14.
- Schrodt, Phil A. 2010*a*. "Seven deadly sins of contemporary quantitative political analysis." Paper presented at the 106th Annual American Political Science Association (APSA) Meeting and Exhibition, Washington, DC.
- Schrodt, Phil A. and D.J. Gerner. 1994. "Validity assessment of a machine-coded event data set for the Middle East, 1982-92." *American Journal of Political Science* pp. 825– 854.
- Schrodt, Philip. 2010b. "Automated production of high-volume, near-real-time political event data." Manuscript. URL: http://polmeth.wustl.edu/mediaDetail.php?docId=1253
- Schultz, K.A. 1999. "Do democratic institutions constrain or inform? Contrasting two institutional perspectives on democracy and war." *International Organization* 53(2):233–266.

- Schutte, Sebastian and Nils B. Weidmann. 2011. "Diffusion patterns of violence in civil wars." *Political Geography* 30(3):143–152.
- Selway, J.S. 2011. "The measurement of cross-cutting cleavages and other multidimensional cleavage structures." *Political Analysis* 19(1):48–65.
- Shapiro, Jacob N. and Nils B. Weidmann. 2011. "Talking about killing: cell phones, collective action, and insurgent violence in iraq." Manuscript (31 May 2011 version). URL: http://www.princeton.edu/jins/papers/SW_2011_Cell_Phones_Insurgency_31MAY11.pdf
- Shellman, S.M. 2008. "Coding disaggregated intrastate conflict: machine processing the behavior of substate actors over time and space." *Political Analysis* 16(4):464–477.
- Singer, David J. 1987. "Reconstructing the Correlates of War Dataset on material capabilities of states, 1816–1985." *International Interactions* 14(2):115–132.
- Singer, J. David, Stuart A. Bremer and John Stuckey. 1972. Capability distribution, uncertainty, and major power war, 1820-1965. In *Peace, war, and numbers*, ed. Bruce Russett. Beverly Hills: Sage pp. 19–48.
- Slantchev, Branislav L. 2003*a*. "The power to hurt: costly conflict with completely informed states." *American Political Science Review* 97(1):123–133.
- Slantchev, Branislav L. 2003b. "The principle of convergence in wartime negotiations." *American Political Science Review* 97(4):621–632.
- Slantchev, Branislav L. 2004. "How initiators end their wars: the duration of warfare and the terms of peace." *American Journal of Political Science* 48(4):813–829.
- Slantchev, Branislav L. 2010. "Feigning weakness." International Organization 64(3).
- Smith, A. 1998. "Fighting battles, winning wars." Journal of Conflict Resolution 42(3):301.
- Spiegelhalter, D.J., N.G. Best, B.P. Carlin and A. Van Der Linde. 2002. "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):583–639.
- Stam III, Allan C. 1999. *Win, lose, or draw: domestic politics and the crucible of war*. Ann Arbor, MI: University of Michigan Press.
- Stevanovic, R. and D. Breznik. 1991. "The population of Yugoslavia, 1948-1991." *Yugoslav* survey 32(3):3.
- Sundberg, Ralph, M. Lindgren and A. Padskocimaite. 2010. "UCDP Geo-referenced Event Dataset (GED) Codebook.".

Thiel, Joshua and Joyce Hogan. 2011. "The statistical irrelevance of American SIGACT data: Iraq surge analysis reveals reality." *Small Wars Journal*.

URL: *http://smallwarsjournal.com/jrnl/art/the-statistical-irrelevance-of-american-sigact-data-iraq-surge-analysis-reveals-reality*

- Thompson, Robert. 1966. *Defeating communist insurgency: the lessons of Malaya and Vietnam.* New York: F.A. Praeger.
- Toft, Monica Duffy and Yuri M. Zhukov. 2012. "Denial and punishment in the north Caucasus: evaluating the effectiveness of coercive counterinsurgency." *Journal of Peace Research*.
 URL: http://j.mp/u3xxNA
- Ulfelder, Jay. 2012. "Forecasting onsets of mass killing." Presented at the 2012 Northeast political methodology meeting (NEMP).
- U.S. Army. 2010. Intelligence Field Manual (FM 2-0). U.S. Government Printing Office.
- U.S. Army and Marine Corps. 2006. *Counterinsurgency Field Manual (FM 3-24)*. U.S. Government Printing Office.
- U.S. Marine Corps. 1940. Small Wars Manual. U.S. Government Printing Office.
- Wagner, R. Harrison. 2000. "Bargaining and war." *American Journal of Political Science* 44(3):469–484.
- Ward, Michael D., Brian D. Greenhill and Kristin M. Bakke. 2010. "The perils of policy by p-value: predicting civil conflicts." *Journal of Peace Research* 47(4):363–375.
- Ward, Michael D. and Kristian Skrede Gleditsch. 2008. *Spatial regression models*. Vol. 155 New York: SAGE Publications.
- Weidmann, Nils B. 2011. "Violence "from above" or "from below"? The role of ethnicity in Bosnia's civil war." *The Journal of Politics* 1(1):1–13.
- Weidmann, Nils B. and Michael D. Ward. 2010. "Predicting conflict in space and time." *Journal of Conflict Resolution* 54(6):883–901.
- Weinberger, Sharon. 2011. "Web of war." Nature 471.
- Wittman, Donald. 2009. "Bargaining in the shadow of war: when is a peaceful resolution most likely?" *American journal of political science* 53(3):588–602.
- Wolford, Scott., Dan Reiter and Clifford J. Carrubba. 2011. "Information, commitment, and war." *Journal of Conflict Resolution* 55(4):556–579.
- Wu, Shuo-sheng, Xiaomin Qiu and Le Wang. 2005. "Population estimation methods in GIS and remote sensing: a review." *GIScience & Remote Sensing* 42(1):80–96.

- Zeitzoff, Thomas. 2011. "Using social media to measure conflict dynamics." *Journal of Conflict Resolution* 55(6):938–969.
- Zhukov, Yuri M. 2012. "An epidemic model of violence and public support in civil war." *Conflict Management and Peace Science*.

BIOGRAPHICAL SKETCH

Andreas Beger was born in Germany to German and Yugoslavian parents and moved to the United States in 1999. He earned a Bachelor of Arts degree in International Studies from the University of North Florida in 2004. In 2007 he earned a Master of Science degree in Political Science from Florida State University and commissioned as an officer in the U.S. Army. His research interests include civil conflicts, counterinsurgency, and statistical forecasting.