A NUANCED STUDY OF POLITICAL CONFLICT USING THE GLOBAL DATASETS OF EVENTS LOCATION AND TONE (GDELT) DATASET

A Dissertation in
Political Science

by

James E. Yonamine

© 2013 James E. Yonamine

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013
The dissertation of James E. Yonamine was reviewed and approved* by the following:

Bumba Mukherjee  
Associate Professor of Political Science  
Dissertation Advisor and Co-Chair of Committee

Philip A. Schrodt  
Professor of Political Science  
Co-Chair of Committee

James Honaker  
Lecturer in Political Science

Le Bao  
Assistant Professor of Statistics

Lee Ann Banaszak  
Professor of Political Science  
Director of Graduate Studies of the Department of Political Science

*Signatures are on file in the Graduate School
ABSTRACT

As you read this sentence, know that a riot is occurring somewhere in the world. Elsewhere, or perhaps in the same location, government forces are constructing a barricade and a politician is being abducted. These types of politically motivated conflictual events are always occurring. In most places, accounts of these events quickly make their way online in the form of electronic news stories. In this dissertation, I utilize a new datasets called the Global Dataset of Events, Location, and Tone (GDELT) which contains almost 200 million politically relevant events that have been extracted from freely available online news articles. With this data, I analyze the effects of political violence on the Tel Aviv Stock exchange, measure how much civil war effects interstate war, and build an empirical model that generates accurate predictions of future conflict in Afghanistan.
TABLE OF CONTENTS

List of Tables v
List of Figures vi
Acknowledgements vii
Chapter 1: Motivation
   Introduction 1
   The Evolution of Political Conflict Data 2
   The Substance 19
   References 30
Chapter 2: Using Political Event Data to Analyze Variance in Tel Aviv 100 Index Returns 42-77
   Introduction 42
   The Literature 44
   Brief Explanation of Competing Hypotheses 47
   Data and Research Design 49
   The GARCH Models 55
   Results 57
   Robustness Checks Focusing on Insurance Equities 59
   Conclusion 61
   References 64
Chapter 3: Effects of Domestic Conflict on Interstate Conflict: An Event Data Analysis of Monthly Levels of Onset and Intensity 78-114
   Introduction 78
   Building Hypotheses from the Literature 80
   Research Design 83
   Empirical Tests 89
   Conclusion 98
   References 100
Chapter 4: Predicting Future Levels of Violence in Afghanistan Districts 115-146
   Introduction 115
   Literature Review 117
   Research Design 120
   Forecasting Approach 121
   Results 127
   Future Directions 129
   Conclusion 133
   References 135
Chapter 5: Concluding Remarks 147-153
   Conclusion 147
   Empirical Findings and future avenues of research 148
   The Future of event data and political violence research 152
   Final Thoughts 153
   References 156
Appendix: A Guide to Aggregation Choices when Working with Event Data 157-176
Chapter 1: Motivation
Table 1: Example of a COW interstate war observation 39
Table 2: Example of a MID observation 39
Table 3: Example of a UCDP observation 39
Table 4: Example of an ACLED observation 39
Table 5: Example of a COW interstate war observation 40
Table 6: Example of a UCDP-GED observation 40
Table 7: Example of a KEDS Event 40
Table 8: Example of a GDELT Event Data Event 40

Chapter 2: Using Political Event Data to Analyze Variance in Tel Aviv 100 Index Returns
Table 1: Correlation Matrix of Counts 67
Table 2: Correlation Matrix of material conflict MAs 67
Table 3: GARCH models of daily TA100 with DJIA control 68
Table 4: GARCH models of daily TA100 without DJIA control 69
Table 5: GARCH models of daily MGDL with DJIA control 70
Table 6: GARCH models of daily CLIS with DJIA control 71

Chapter 3: Effects of Domestic Conflict on Interstate Conflict: An Event Data Analysis of Monthly Levels of Onset and Intensity
Table 1: The Effects of Lagged Domestic Conflict Onset on Interstate Conflict Onset with 1-month Lag 103
Table 2: The Effects of Lagged Domestic Conflict Onset on Interstate Conflict Onset with 2-month Lag 104
Table 3: The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestics Conflicts on Interstate Conflict Onset 105
Table 4: The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestics Conflicts on Interstate Conflict Onset 106
Table 5: The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestics Conflicts on Whether an Ongoing Interstate Conflict Becomes more Intense 107
Table 6: The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestics Conflicts on Whether an Ongoing Interstate Conflict Becomes more Intense 108
Table 7: The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestics Conflicts on Changes in Interstate Conflict Intensity 109
Table 8: The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestics Conflicts on Changes in Interstate Conflict Intensity 109
Table 9: How much more likely is an interstate conflict onset in month $t$ relative to neither state experiencing an onset of domestic conflict in month $t - 1$ 110
Table 10: How much more likely is an interstate conflict onset in month $t$ relative to neither state experiencing an onset of domestic conflict in month $t - 2$ 110
Table 11: How many times more likely is an interstate conflict onset in month $t$ relative to neither state experiencing a more intense ongoing domestic conflict in month $t - 1$ 110
Table 12: How many times more likely is an interstate conflict onset in month $t$ relative to neither state experiencing a more intense ongoing domestic conflict in month $t - 2$ 110

Chapter 4: Predicting Future Levels of Violence in Afghanistan Districts
Table 1: Assessing Accuracy at the District Level 141
Table 2: Assessing Accuracy at the Province Level 142
Table 3: Assessing Accuracy at the Country Level 143
LIST OF FIGURES

Chapter 1: Motivation
   Figure 1: The Number of GDELT-derived Violent Events in Homs and Aleppo from January 2012 through June 2012 40
   Figure 2: The Number of Ground-truthed Violent Events in Homs and Aleppo from January 2012 through June 2012 41

Chapter 2: Using Political Event Data to Analyze Variance in Tel Aviv 100 Index Returns
   Figure 1: Total number of material conflict events, daily from 1992 to 2012
   Figure 2: The number of material conflict events with the time trend removed, daily from 1992 to 2012 73
   Figure 3: Raw, first-differenced, and logged first-differenced of TA100 prices 74
   Figure 4: Comparison of closing TA100, CLIS, and MGDL prices 75
   Figure 5: Raw, first-differenced, and logged first-differenced MGDL prices 76
   Figure 6: Raw, first-differenced, and logged first-differenced CLIS prices 77

Chapter 3: Effects of Domestic Conflict on Interstate Conflict: An Event Data Analysis of Monthly Levels of Onset and Intensity
   Figure 1: The Effects of Domestic Conflict Onset in Month t-1 on the Likelihood of Interstate Conflict Onset in Month t 111
   Figure 2: The Effects of Domestic Conflict Onset in Month t-2 on the Likelihood of Interstate Conflict Onset in Month t 112
   Figure 3: The Effects of an Increasingly Severe Domestic Conflict in Month t-1 on the Likelihood of Interstate Conflict Onset in Month t 113
   Figure 4: The Effects of an Increasingly Severe Domestic Conflict in Month t-2 on the Likelihood of Interstate Conflict Onset in Month t 114

Chapter 4: Predicting Future Levels of Violence in Afghanistan Districts
   Figure 1: The Number of Material Conflict events per Afghani District from 2001 to 2012 144
   Figure 2: One-month Forecast of the # of Material Conflict Events in Bughran District using ‘arfima’ package, with mean, 90%, and 95% confidence intervals. 145
   Figure 3: Average Farm-Gate Prices for Dry Opium in Afghanistan, September 2004-March 2012 146
ACKNOWLEDGEMENTS

Throughout my academic career, I have received tremendous support by a host of individuals. First and foremost, I would like to thank my advisor Phil Schrodt. When deciding where to attend graduate school, Phil was the primary factor that caused me to choose Penn State and throughout my time here, I cannot overstate how valuable Phil has been in educating, mentoring, and motivating me. Additionally, James Honacker has provided invaluable methodological training and support. Likewise, Bumba Mukherjee has consistently supported my academic development.

And of course, I must thank Robert Packer, who initially sparked my interest in the empirical study of conflict in 2004 and has been a steady mentor and friend since then.

Lastly, a number of other people deserve mention, including Chris Zorn, Burt Monroe, Doug Lemke, Glenn Palmer, Scott Bennett, Le Bao, Daniel Kifer, Kevin Jones, and Burt Johnson.
For Mom, Dad, Grandma, and Grandpa.
CHAPTER 1. MOTIVATION

1. INTRODUCTION

For better of worse, human history has been defined by the violent competition for political power. Given its ubiquity and importance, scholars have been interested in understanding and even predicting political violence for centuries. If we conservatively assume Thucydides to be the first rigorous student of political conflict, then scholars have been at this for over 2,500 years. For the better part of those years, scientific progress was relatively non-existent, as we could prove little more in the 1960s about the causes and effects of political violence than we could in 400 B.C.. But why? Especially when most other disciplines made such tremendous progress in the same time frame: in 400 B.C., leading physical “scientists” thought that the world was flat and that the sun was a chariot of fire, but by 1959 they had landed a man on the moon. For one, unlike physical particles, humans are complex creatures that exhibit free will, meaning that rules and laws governing their behavior are less concrete and more difficult to empirically demonstrate.

Additionally, and perhaps more importantly, unlike other disciplines, the amount and quality of data on political violence was basically stagnant for a millennium. Historically, experimental data had been scarce, and whereas it has been feasible for scholars of the biological or physical science to conduct experiments, this has not been possible for scholars interested in dynamics of large-scale political violence. Moreover, nuanced observational data had been similarly difficult to collect on a large scale since this would have required unfeasibly large numbers of highly educated and disciplined manpower to witness, record, store, and then distribute data about political events, as well as the technology to do so. Fast-forward to today, and experiments to learn about political violence are still not feasible (which is probably a good thing), but our ability to collect, store, and analyze observational data has grown exponentially. It is this rapid and relatively recent explosion in both the volume and quality of data on political conflict that makes this dissertation possible.

The overarching goal of this dissertation is to illustrate how we are able to perform increasingly nuanced empirical analyses of political conflict as our data simultaneously becomes increasingly voluminous and fine-grained. Released in 2012, the Global Database of Event Location and Tone (GDELT) dataset is the largest and arguably the most technologically advanced publicly available,
political-conflict dataset. Equally importantly, this dissertation is the first ever rigorous, empirical analysis of the GDELT dataset. This introductory chapter proceeds in two main sections. In Section 2, I provide a history of political conflict data collection efforts, culminating with a discussion of GDELT. In Section 3, I frame the three distinct substantive chapters of this dissertation within the relevant literature on political conflict and provide brief overviews of findings.

2. THE EVOLUTION OF POLITICAL CONFLICT DATA

Large-scale, systematic efforts at representing political violence as data did not occur until the 20th century.\(^1\) In the last 50 years, researchers have created dozens, if not hundreds of datasets about political violence. As with any historical account, there are a number of ways to structure this discussion of the history of political violence data. I choose to structure my discussion in terms of three conceptual “schools” of development. Two of these “schools” actually center around specific universities, while the third was decentralized across institutions and organizations. As we will see, the most commonly used political violence datasets fit cleanly into one of these three categories, with little overlap.\(^2\)

First, is the University of Michigan school, which produced the Correlates of War (COW) and Militarized Interstate Dispute (MID) data sets; second, is what I deem the Scandinavian school, comprised of the Peace Research Institute of Oslo (PRIO) and Uppsala University. The Scandinavian school produced the Uppsala Conflict Data Program (UCDP), UCDP/PRIO, UCDP Georeferenced Event Dataset (UCDP-GED), and the Armed Conflict Location Event Dataset (ACLED) datasets, as well as motivating additional, ACLED-like offshoots. Third, is what I call the Machine-coded school since unlike the Michigan school and the Scandinavian school, this lineage of data collection efforts have been decentralized. The Machine-coded school is responsible for the Kansas Event Dataset (KEDS), Virtual Research Associates (VRA), the 10 Million International Dyadic Events dataset, the Integrated Conflict and Early Warning System (ICEWS), and GDELT.

\(^1\)In order to qualify as “political violence”, the goal of the violent act, which may either be carried out, attempted, or simply threatened, must be to intimidate or coerce a government or civilian population in furtherance of social objective or access to positions of power. This definition is borrowed from http://legal-dictionary.thefreedictionary.com/Political+violence.

\(^2\)I do not include in my discussion datasets that focus on a specific subset of political conflict, such as terrorism or ethnic conflict.
To facilitate discussion of the history of data collection efforts of these three schools, I use some additional terminology. First, Schrodt (2012) provide a useful scheme for differentiating different types of data based on the degree of aggregation: ³

(1) Episodic
   - Episodic data are those coding characteristics of an extended set of events, such as a war or a crisis: the COW project [discussed in detail below] is the archetype.

(2) Composite
   - Composite events are those which occur in a relatively short period of time and limited geographical space—for example, a specific skirmish that occurs during a war—and multiple characteristics of the incident are coded.

(3) Atomic
   - Atomic events are basic units of political interaction—date, source, target, event—found in classic event data sets as World Event Interaction Survey (WEIS) and Conflict and Peace Data Bank (COPDAB), and in contemporary coding schemes such as Integrated Data for Event Analysis (IDEA) and Conflict and Mediation Event Observations (CAMEO) [which are discussed in detail below].

Second, from the earliest attempts nearly 100 years ago to the most cutting edge programs today in 2013, scholars have followed three main steps in order to build the types of datasets above:

(1) Build coding rules and ontology
(2) Obtain sources
(3) Generate data by applying coding rules and ontology to sources

As I discuss the history of political violence data collection across the three schools, I provide as detailed information as possible about this three-step data generating process.

2.1. University of Michigan School. In 1963, David Singer launched the COW project at the University of Michigan, which produced the first conflict dataset widely used in empirical analysis (see Singer (1972)). Although COW is often credited with producing the first dataset on political conflict, it was actually preceded by at least four earlier efforts, including Woods and Baltzly (1915), Sorokin (1937), Wright (1942), and Richardson (1960). According to Geller (2004) and Ward et al.

³I borrow identical language from Schrodt (2012), with one minor difference. Where I write “a specific skirmish that occurs during a war”, Schrodt (2012) writes “a terrorist attack”. I believe this change provides additional conceptual clarity.
(2012), these previous projects, and especially Richardson (1960), helped to motivate and shape the COW project. The primary breakthrough of the COW project was the rigor of its coding rules and ontology (i.e. Step 2) for defining exactly what must occur for a series of events to be considered a “war”, which is still employed (with a few changes) today. For an interstate war, a minimum of 1,000 battle fatalities between two officially recognized armies within a 12-month period, with at least 100 fatalities occurring on both sides. Thus, COW provided episodic data, meaning that it thought about political conflict in terms a broader “war”, rather than the specific battles comprising the war or the individual events making up the battles.

The release of Singer (1972) and the accompanying dataset, which provided detailed dyadic-level data on all inter-state wars from 1816-1946, ushered in a new era in the study of conflict, enabling scholars to use empirical models to test hypotheses about the causes and consequence of wars. Subsequent COW projects expanded beyond inter-state wars and provided data on intra-state(Small and Singer (1980)), non-state (Sarkees and Wayman (2010)), and extra-state (Sarkees and Wayman (2010)) wars. In the 40 years since the COW’s first published dataset, it has been the most heavily used source of political conflict data.

The process of building the COW datasets is similar today to how it was in 1963, and likely to how it was while Woods and Baltzly (1915) were working almost 100 years ago. First, COW researchers initially constructed an ontology of conflict, specifying rigorous requirements that must be met in order to political violence to qualify as various types (inter-, intra-, non-, and extra-state) of wars. Next, researchers (i.e. low paid or volunteer undergraduate/graduate students) combine Step 2 and Step 3 of the conflict-data generating process, pouring over various sources, such as newspapers, books, microfilms, and (more recently) online documents in order to determine whether historical events qualify as a type of war. These researchers then discuss amongst themselves, and eventually reach agreement about how to code the wars.4 Below, I provide an example of how the COW database presents an interstate war:5

Table 1 reflects that a war according to COW definition is a dyadic event occurring between two states. Each row reflects one of the two states involved in the war. In the entry above, the “2”

4The resulting data are available at www.correlatesofwar.org.
5In the actual dataset, the day, month, and year are provided in separate columns.
entry in the *WhereFought* reflects that the bulk of the fighting occurred in the second state listed, which in this instance is France.

Despite the popularity of COW data among empirical studies of political conflict, the coarse, episodic nature of the data constrained the types of questions that researchers could ask. Three aspects of COW were particularly limiting. First, the requirement of 1,000 battle-related fatalities excluded a large number of smaller-scale, yet important conflicts. Second, since the unit of analysis of the COW datasets is a war, COW does not provide specific, sub-state level information regarding where the war or its component battles occurred. Third, since COW uses the “war” as its unit of analysis, it is impossible to study the more intricate dynamics of violence that occur during the fighting.\(^6\)

Cognizant of the limitations of using COW to study conflictual activities short of war (as defined by COW), Gochman and Maoz (1984) introduced the MID dataset (also housed at the University of Michigan), which provides data regarding three types of dyadic level (i.e. occurring between two states in the international system) conflictual events: the threat of force, the display of force, and the use of force.\(^7\) The initial MID dataset included 886 MIDs that occurred from 1816 to 1975, while the most recent published iteration (see Ghosn, Palmer and Bremer (2004)) extends coverage through 2001, and the forthcoming MID 4.0 will cover 2002-2010. Additionally, whereas the original MID dataset contained three levels of dispute, MID 3.0 contains five categories:

1. No militarized action
2. Threat to use force
3. Display of force
4. Use of force
5. War [following COW guidelines]

Each MID contains information regarding the states involved, the type of event, and the specific day of the onset and termination of the dispute. The types of events coded as MIDs vary greatly. For example, the model length of all MIDs is one day, though some MIDs last longer than 10 years. The figure below provides an example of a MID observation: \(^8\)

\[\text{[INSERT TABLE 2 HERE]}\]

---

\(^6\)See Moore (2005) for a more rigorous critique of the COW datasets.

\(^7\)See Gochman and Maoz (1984) page 588 for greater detail about these events.

\(^8\)This example is from the participant level dataset. A dispute level dataset provides data in a different format, though the information contained is identical. See www.correlatesofwar.org for the data.
In Table 2, DispNum is the dispute number, StateAbb reflects the abbreviate name of the states, and HostLev reflects the highest level of hostility reached, on the scale of one to five as listed above.

Overall, there are two primary contributions of the MID dataset, neither of which were possible using only the COW datasets. First, it allowed researchers to systematically analyze conflictual inter-state behavior short of war. Second, it enabled the study of conflict escalation, from an initial threat, to the display of force, to the use of force, all the way to large-scale war. Although MID has enabled many interesting research agendas, the data still contain some potential shortcomings. For example, each MID is assigned an initiator and a target. However, it is often difficult to discern which state in an ongoing dyadic dispute made the first “threat”, and ultimately subjective judgment is often required. Additionally, a dispute between states can last many years, but the MID dataset only provides information on the initiation and termination of the dispute. Thus, MID, like COW, prevents scholars from analyzing conflict dynamics that occur during a dispute. Finally, as discussed in the following paragraph, MIDs are tedious to code, which prevents analysis of political conflict in real- or near real-time.

The data-collection process of the early MID efforts closely resembled that of COW, with all three steps in the process being performed without meaningful computational assistance. In the nearly 30 years since the initial release of MIDs, the coding rules and ontology have remained mostly consistent. Additionally, Step 3 (i.e. generating data by applying coding rules and ontology to sources) has likewise unchanged. However, the process by which MID researchers access relevant information (i.e. Step 2), has changed greatly. Originally, MID researchers would manually search newspapers, microfilms, etc. without technological support. In the mid 1990s, MID began to utilize assisted search engines like LexisNexis. MID 4.0 introduced automated text classification techniques, first introduced by Schrodt, Palmer and Hatipoglu (2008) and later implemented D’Orazio et al. (2012), to isolate electronic news stories likely to contain relevant information form those containing unimportant information. According to D’Orazio et al. (2012), this has saved considerable time and enhanced coding accuracy. However, even with these increases in efficiency, the process of updating the MID database is still tedious, meaning that new version are released sporadically rather than in real time.

2.2. The Scandinavian school. Like MID, the main goal of the Scandinavia school has been to move beyond the coarse structure of COW in order to generate increasingly nuanced political conflict data, though the focus has been primarily on domestic conflict. Wallensteen and Axell
(1993) introduced Scandinavia school’s first dataset, called the Uppsala Conflict Data Program (UCDP) dataset. The UCDP created five different categories of conflict based on the number of battle fatalities, and like COW, UCDP required that fighting occur between the state and an armed rebel group:

1. **armed conflicts**, >25 total casualties
2. **minor armed conflicts**, >25 but <1,000 total casualties
3. **intermediate conflicts**, >1,000 total casualties but <1,000 in a year
4. **wars**, >1,000 casualties in a year
5. **major armed conflicts**, all intermediate conflicts and wars

The most notable contribution of this initial UCDP dataset was to provide data on armed conflicts and minor armed conflicts, which were not included in COW datasets. The lower threshold of 25 battle deaths for armed conflicts measure meant that a single, small-scale battle between the military and a rebel group occurring in a specific location over the course of a single day would gain inclusion into the UCDP dataset. Thus, although UCDP provided information about “episodic” events like “major war”, it also contained more fine-grained, “composite” events that start and finish on the same day. Wallensteen and Axell (1993) initially provided data on armed conflicts (as well as larger scale “intermediate” and “major” armed conflicts) occurring globally from 1989 to 1992, and subsequent versions were released annually in the *Journal of Peace Research*. Table 3 provides an example of a UCDP observation. The “type” columns takes on an integer one through four, corresponding to the four types of conflict listed above. Also, note that the “location” is at the country level:

[INSERT TABLE 3 HERE]

In 2002, Gleditsch et al. (2002) back-coded the UCDP to 1946, resulting in a complete dataset of low, intermediate, and major conflicts from 1946-2001. Although Gleditsch et al. (2002) improved the UCDP dataset by broadened the temporal coverage, the resulting UCDP/PRIO continued to follow UCDP coding procedure, which meant that data continued to be presented at the binary, state-year measure. Moreover, the UCDP/PRIO dataset continued to require the conflicts occur between a government and rebel actor, meaning that conflict occurring between two rebels groups would not gain inclusion into the dataset.

---

9This dataset is referenced under multiple names (Prio-Uppsala, Uppsala/UCDP/PRIO, etc.), but I will refer to it exclusively as UCDP/PRIO.
In order to provide increasingly comprehensive and detailed information about political conflict beyond what was available in the UCDP/PRIO dataset, Raleigh et al. (2010) released the ACLED dataset with the goal of enabling more micro-level analyses of political conflict. Unlike UCDP/PRIO, which provides data in “episodic” and “composite” form, ACLED provides data in “atomic” form, and focuses exclusively on eight types of violence:\(^\text{10}\)

1. Battle - No Change of Location
2. Battle - Rebels Control Location
3. Battle - Government Regains Control
4. Battle - Headquarters of Base Established
5. Non-violent Conflict Event
6. Rioting/Protesting
7. Violence Against Civilians
8. Non-Violent Transfer of Location Control

All of the event types above, with the exception number seven, must occur between actors deemed rebels and actors affiliated with the official government, while number seven can clearly involve civilians. This meant that if rebels attacked and killed 25 civilians in a politically motivated act, ACLED would code this event, even though the COW intrastate conflict dataset and UCDP/PRIO would ignore it because a government actor was not involved.

Additionally, and perhaps more importantly, ACLED was the first dataset to provide specific sub-state location information regarding where the conflict events occurred. Each event in the ACLED dataset, be it “composite” or “atomic”, The original ACLED dataset provided limited temporal coverage for 50 countries. As of February 2013, it provides 75,000 “atomic” events for 60 countries, with key aspects of an observation from ACLED presented below:\(^\text{11}\) Table 4 provides an example observation from the ACLED dataset.

[INSERT TABLE 4 HERE]

Only a year after the release of ACLED, Melander and Sundberg (2011) released a new geo-coded dataset contains both “composite” and “atomic” conflict events called UCDP-GED (see Sundberg and Lindgren (2011) for coding rules). Like ACLED, the UCDP-GED codes events according to

\(^{10}\)See Raleigh et al. (2010) page 656 for a detailed list of the eight types of events coded in the ACLED dataset.

\(^{11}\)ACLED also provide the number of casualties as well as other information. See www.acleddata.com for more information about ACLED data.
the same four “types” discussed above, and provides data in a highly similar format to ACLED, as illustrated in Table 5:12

[INSERT TABLE 5 HERE]

Currently, the UCDP-GED dataset contains approximately 24,000 events for all African countries from 1998 to 2010. One major difference is that UCDP-GED contains “composite” events that occur across multiple days, while ACLED only focuses on specific, “atomic” events that occur on a single day. Despite the similarities, Eck (2012) has found that considerable differences exist between ACLED and UCD-GED codings of the same countries during the same time periods.

To generate their data, the UCDP/PRIO, UCDP-GED, and ACLED project utilize a similar approach to MID, with researching searching various electronic sources of data in order to obtain news stories likely to contain relevant information about political conflict. According to Chojnacki et al. (2012), the UCDP-GED dataset collects data exclusively from five news wires: BBC Monitoring, Reuters News, Agence France Presse, Dow Jones International News, and Xinhua. Additionally, Chojnacki et al. (2012) reports that ACLED varies its sources based on the country in focus. After obtaining sources, the researchers apply the coding rules to the sources, using their best subjective judgement to code new events. Although human-coding has a number of strengths, the primary benefit for ACLED and UCDP-GED has been human’s ability to connect events with places. According to Raleigh et al. (2010), this allowed ACLED to be the only dataset at the time of its release to provide geo-coded political conflict data. But, like all human-coded datasets, the process of generating data for these datasets is slow, which limits the spatial and temporal domain of coverage, the volume of total events, and the abilities to produce data in real or near real time.

2.2.1. More focused extensions. More recently in 2012, scholars released four additional human-coded “atomic” and “composite” event data datasets on political conflict. Although these were not created at one of the institutions comprising the Scandinavian school, they fit most cleanly into this section since they use similar data-generating approaches as ACLED and UCDP-GED, but do so for more focused spatial coverage. First, Urdal and Hoelscher (2012) generated a dataset of conflict events, focusing exclusively on 55 major cities in Asia and Africa. The dataset contains approximately 4,000 events geo-coded to the city level, and is generated based on human coding of Keesings Record of World Events (KRWE). Second, Daly (2012) introduced a dataset that codes for 4,003

12The UCDP-GED also provides information on number of fatalities, the source of the the article used to code the event, and a number of other attributes. See www.ucdp.uu.se/ged/data.php for the data.

2.3. The Machine-coded school. Unlike the University of Michigan and the Scandinavian schools, the Machine-coded school has not been centered at specific university/universities, but rather decentralized throughout a number of institutions and organizations. Additionally, whereas the previously discussed datasets have all focused specifically on aspects of political conflict, the Machine-coded datasets account for a wider range of politically relevant events, including cooperative acts like sending aid or releasing prisoners, while also providing rigorous coverage of political conflict events. Indeed, the majority of studies utilizing machine-coded event data focus primarily on the conflictual events.

Although machine-coded event data has many similarities to ACLED, SCAD, and EDACS, machine-coded event data has evolved on a considerably different tract largely independent of the human-coded datasets. Whereas the most recent human-coded efforts like SCAD and EDACS trace their roots to ACLED and UCDP-GED, which in turn used COW as a foundation, the most recent machine-coded efforts like ICEWS and GDELT stem back to WEIS and COPDAB. Thus, despite similarities, the differences are sufficiently strong to warrant discussion of machine-coded event data collection efforts separately from human-coded. Additionally, since this dissertation utilizes machine-coded event data, I give the history of the evolution of machine-coded event data more attention that given to the University of Michigan or Scandinavian school.

In my discussion of the machine-coded school, I first discuss McClelland (1976)’s WEIS and Azar (1980)’s COPDAB. Although these projects were human-coded, they provided a foundation on which later machine-coded efforts were built. Second, I discuss the rise of modern, machine-coded event data datasets that followed the computer / internet revolution of the late 1980s and
early 1990s, like the Kansas Event Dataset (KEDS) and the Integrated Conflict Early Warning System (ICEWS).

2.3.1. WEIS and COPDAB. In the late 1970s and 1980s, McClelland (1976)’s WEIS and Azar (1980)’s COPDAB began an alternative type of “atomic” event data (henceforth referred to as “event data”) as a more nuanced alternative to the coarse “episodic” COW style data. Whereas the “episodic” COW project generally utilized a war as its unit of analysis, the WEIS and COPDAB projects were interested in capturing the specific, daily-level events (such as attacks, protests, demonstrations, meetings, etc.) that taken together, may form a battle or a way but occur in real-time as independent events. Thus, much like how COW focused heavily on a clearly defined set of rules to define both a domestic and inter-state war, WEIS and COPDAB similarly required robust ontologies and coding rules that would allow them to extract specific “atomic” events in a [subject-verb-object] format from news sources in a consistent and replicable way.

Here, the formal definition of an event provided in Gerner et al. (1994) is useful: “An event is an interaction which can be described in a natural language sentence which has as its subject and direct or indirect object an element of a set of actors, and as the verb an element of a set of actions, all of which are transitive verbs, and which can be associated with a specific point in time. Gerner et al. (1994) further specifies that, “In event coding, the subject of the sentence is the source of the event, the verb determines the event code, and the object of the verb is the target.

Coding ontologies or schemes are the rules by which the source, the object, and the verb presented in natural language in articles are converted into categorical actor and event codes suitable for empirical aggregation and analysis. For WEIS, COPDAB, and all subsequent event data efforts, development of the ontologies was critical, since these form the rules by which the source, the object, and the verb presented in natural language in text are converted into categorical actor and event codes suitable for empirical aggregation and analysis. Since event data projects capture a broad range of events, from meetings to bombings to the provision of aid, they require far more detailed ontologies than the University of Michigan or Scandanavian school datasets.

Generally, two ontologies are used in machine-coded event data datasets, one for actors that informs that source and the target, and one for verbs that informs the actions. Other ontologies have been developed to code other characteristics of the sentence for example, COPDAB and the Protocol for the Assessment of Nonviolent Direct Action (PANDA) coded for political issues and CAMEO and IDEA have ontologies for general agents.

\[^{13}\text{In this instance, WEIS and COPDAB are both the names of the event data projects as well as the name of the ontologies.}\]

\[^{14}\text{Other ontologies have been developed to code other characteristics of the sentence for example, COPDAB and the Protocol for the Assessment of Nonviolent Direct Action (PANDA) coded for political issues and CAMEO and IDEA have ontologies for general agents.}\]
COPDAB were the first event data ontologies. Reflecting the status of international relations at the time, both followed the realist tradition in assuming that states operated as unitary actors. This meant that all events between individuals were treated as occurring between the states of each individuals respective citizenship. For example, if a group of Pakistani rebels attacks Indian civilians across this border, both WEIS and COPDAB treat this as an attack of Pakistan against India. Thus, officially recognized states in the international system were the only actors in the WEIS and COPDAB ontologies.

Consequently, both the WEIS and COPDAB event coding ontologies were also structured to capture important inter-state interactions. The WEIS event ontology is based on 22 distinct cue, or parent categories of actions (such as Consult, Reward, Warn, etc.), which take on 2-digit codes, and 63 sub categories. The sub-categories indicate additional information beyond the parent category. For example, Threaten is one of WEIS’s cue categories and its 2-digit code is 17. However, when more information is presented in the article regarding the type of These systems did code a small number of militarized non-state actors such as the Irish Republican Army (IRA) and the Palestinian Liberation Organization (PLO), as well as the United Nations, but the overall focus was on nation-states. COPDAB utilizes a similar verb typology that focuses on capturing interstate events but instead of WEISs 22 cue categories, COPDAB uses 16 and places these on a conflict-cooperation continuum to facilitate empirical analyses.

In comparison, recall that COW codes for five types of events (interstate war, intrastate war, non-state war, and extra-state war), MID codes for five types of events (no militarized action, threat to use force, display of force, use of force, and war), UCDP/PRIO codes 4 types of events (armed conflicts, minor armed conflicts, intermediate conflicts, and wars), and ACLED codes for nine (four types of battles, non-violent conflict events, rioting/protesting, violence against civilians, and non-violent transfers of location control).

While WEIS and COPDAB were the most commonly used ontologies from the first phase of event data analysis, quite a few additional systems were developed that never gained a foothold. For example, the Behavioral Correlates of War (BCOW) data set coded historical as well as contemporary crises and had more than 100 distinct event codes, including Assume foreign kingship; the Comparative Research on the Events of Nations (CREON) data set [(Harmann et al. (1973))](#)
was customized for coding foreign policy behaviors, and the Sherman Facts (SHERFACTS) (Sherman and Neack (1993)) and Computer-Aided System for Analysis of Local Conflicts (CASCON) (Bloomfield and Moulton (1989)) data sets coded crisis behavior using a crisis-phase framework.

Although COW and the WEIS/COPDAB projects utilized considerably different coding rules and ontologies, *Step 1* and *Step 3* of the conflict-data-generating process were quite similar and done almost entirely by human hand. Generally, graduate students would scour books, newspapers, and microfilms in order to procure as many sources as possible that may yield relevant information. Although coders were instructed about what types of articles to gather, they relied on their subjective judgment to determine whether an article was relevant and warranted inclusion into the archive of articles from which events were derived. *Step 3* was also quite straightforward and similar. Researchers applied the coding rules and ontologies to the corpus of sources. For WEIS and COPDAB, researchers would manually recorded dozens of relevant events of interest a day, which were then transferred to punch cards and eventually to magnetic tape.

### 2.3.2. The computer revolution and rise of machine-coded event data.

Unfortunately, the actual WEIS and COPDAB datasets were never transferred to electronic records. Thus, the resulting datasets are inaccessible, and have not been used in the empirical study of conflict in decades. However, the WEIS and COPDAB are nonetheless highly important, since they served as a launching pad for future, machine-coded projects. Throughout the 1980s, all three steps of the data-building process were done entirely by humans, with virtually no computational assistance. Then, in the late 1980s and early 1990s, major technological innovations made possible to rise of machine-coded data. To most efficiently discuss the evolution of machine-coded event data, I divide my discussion into the three component parts of the conflict-data generating process: ontology development, obtaining news stories, and processing.

Ontology and coding rules. Although WEIS and COPDAB deserve much credit for spearheading the entry of event data into the mainstream of political science, a number of shortcomings became apparent over time. Gerner et al. (2002) report that the state-centric focus of WEIS and COPDAB made them ill-suited to account for sub-state level events between domestic actors. Additionally, Gerner et. al. [2002] explain that both WEIS and COPDABs verb typologies contained too few event categories: For instance, WEIS has only a single cue category of Military engagement that must encompass everything from a shot fired at a border patrol to the strategic bombing of
cities...COPDAB contains just 16 event categories, spanning a single conflict-cooperation continuum that many researchers consider inappropriate.

Reacting to these shortcomings, a group of researchers led by Doug Bond constructed the first version of the PANDA dataset in 1988. The leading motivation behind PANDA was to more thoroughly account for domestic events, especially non-violent “direct action found in protests and demonstrations but overlooked by the WEIS and COPDAB schemes. Ten years later in 1998, Bond et. al. built upon PANDA to create the more comprehensive Integrated Data for Event Analysis (IDEA), adding incorporating codes from Taylor and Jodice (1983)’s World Handbook of Social and Political Indicators, WEIS, and MID. Furthermore, IDEA created additional event codes for economic events, biomedical phenomena such as epidemic disease, and various additional jurisprudence and electoral events [see Bond et al. (2003) for a further discussion of PANDA and IDEA]. In building the 10 Million International Dyadic Events dataset, King and Lowe (2004) also utilize the IDEA action ontology.

In 2002, Gerner et al. (2002) released the Conflict and Event Mediation Event Observation (CAMEO) coding framework. Like PANDA and IDEA, CAMEO was designed to capture sub-state events and capture nuanced attributes of the actors. However, there are two differences between CAMEO and IDEA. First, while IDEAs extensions preserved backwards compatibility with multiple earlier systems, CAMEO started only with the WEIS system (plus some of the IDEA extensions) and combined WEIS categories such as WARN/THREATEN and PROMISE/REWARD that were difficult to disambiguate in machine coding. Second, CAMEOs actor codes utilize a hierarchical structure of one or more three-character codes, which reflect the country or nation of origin and as much supplementary information as the article provides regarding region, ethnic/religious group, and domestic role (military, government, etc.). Recently in 2010, the ICEWS projectusing a variety of sources such as the national government lists of the CIA World Factbook [www.cia.gov/library/publications/the-world-factbook/] and lists of IGOs, NGOs, multinational corporations, and militarized groups, built on CAMEOs actor dictionary, eventually collecting over 40,000 names of important political figures in any countries in the world who had a position of prominence anytime from 1990 to 2011. This was a considerable improvement over previous

\[15\text{The PANDA dataset is no longer publicly available. See Bond et al. (2003) for a discussion of the PANDA ontology from its creators.}\]
projects like the 10 Million International Dyadic Events dataset, which only included 450 sub-state actors.\textsuperscript{16}

Obtaining stories. Due largely to technological limitations of the era (i.e. the lack of electronic articles and computational power), the WEIS and COPDAB projects relied on human analysts to physically collect newspaper clippings, press reports, and summary accounts from Western news sources to obtain news stories. Although coders were instructed about what types of articles to gather, they relied on their subjective judgment to determine whether an article was relevant and warranted inclusion into the archive of articles from which events were derived.

This manual approach began to be replaced with automated coding with the first iteration of the Kansas Event Data Set (KEDS) project in the late 1980s (see Schrodt (1990) and Schrodt (1994)). By this time, two major computing developments had occurred. First, the rise of the internet and the advent of large-scale data aggregators such as Lexis-Nexis allowed news reports to be obtained in machine-readable form. Second, computational power and natural language processing methods had advanced to the point where processing of large quantities of information was possible using personal computers. In its earliest version, the KEDS project automatically downloaded and archived Reuters leads from the NEXIS (precursor to Lexis-Nexis) service into an electronic database, then coded these using a custom computer program. Following the success of KEDS, other event data programs, such as the PANDA project adopted an automated data collection process.

By 2000, virtually all large-scale event data projects in political science relied on automated collection of news stories. In addition to the data collection efforts becoming almost exclusively electronic and automated, the scope of media coverage also increased. However, until recently, academic projects like KEDS and the 10 Million International Dyadic Events dataset with global coverage relied on a small number of sources, including Reuters (or Reuters Business Briefing) and Agence France Presse (AFP) for news content. Only with the creation of the Defense Advanced Research Projects Agency (DARPA)-funded Integrated Conflict Early Warning System (ICEWS; see O’Brien (2010)) project in 2009, which draws articles form 29 international and regional news sources, did an event dataset with global coverage attempt to utilize a more comprehensive list of global news outlets. The key difference between the ICEWS event data coding efforts and those of earlier NSF-funded efforts was the scale. As O’Brien (2010) notes:

\textsuperscript{16}This was established by aggregating the raw dataset and counting the number of unique secondary actor codes.
...the ICEWS performers used input data from a variety of sources. Notably, they collected 6.5 million news stories about countries in the Pacific Command (PACOM) AOR [area of responsibility] for the period 1998-2006. This resulted in a dataset about two orders of magnitude greater than any other with which we are aware. These stories comprise 253 million lines of text and came from over 75 international sources (AP, UPI, and BBC Monitor) as well as regional sources (India Today, Jakarta Post, Pakistan Newswire, and Saigon Times).

As the name suggests, the most important innovation of the machine-coded school of conflict data collection was the introduction of computer software that could replace humans and fully automate the Step 3 of the conflict-data generating process. As discussed, in the early stages of event coding, the lack of readily available electronic news stories and sufficient computing power to support machine coded efforts meant that human coding was the only viable coding option. Although human coding was initially the only available way to code events, it has three main shortcomings: it is slow, expensive, and subjective. The average human coder can code around six to ten stories an hour on a sustained basis, and very few people can reliably code more than a few hours a day because the process is so mind-numbingly boring. At that rate, it takes a team of 10 coders at least three person-years to code 80,000 news stories. Paying coders $10 an hour would cost 100,000, and the costs to training, re-training, cross-checking and management would at least double that investment. Additionally, due to the inherently subjective nature of human analytical processes, interoperability between analysts rarely exceeded 70% and often falls in the 30%-40% range (see Mikhaylov, Laver and Benoit (2012) and King and Lowe (2004)) particularly when coding is done across institutions and over long periods of time.

By the late 1980s, computational power had advance to the point that it was possible to run automated coding software from personal computers. The KEDS project was the first attempt within academia to use a computer to parse through electronic text and code relevant events into an event data database, relying on dictionary-driven sparse parsing based on the WEIS typology. The sparse parsing relies primarily on simple pattern matching on the text of an article to find specific words (i.e. “Israel, “attack, “bomb) or sets of words (“United Nations Secretary General; “promised to provide aid, “promised to seek revenge) that match entries in dictionaries corresponding to the actor and event ontologies. In addition, the system knows some basic rules of English grammar: for example a phase of the form Representatives of the US and France will meet with Israeli negotiators
involves two events “US meets Israel and “France meets Israel and the passive voice construction “A U.S. convoy was attacked by Iraqi insurgents reverses the usual subject-verb-object ordering of English sentences so that this corresponds to “Iraq insurgents-attack-USA. Consider the following hypothetical sentence:

- March 12, 1998 Israeli troops launched offensive attacks against Palestinian insurgents on Monday, in the first of what is expected to be a new wave of counter-terrorism efforts.

Using the CAMEO verb typology and actor dictionaries, as well as rules that automatically concatenate the proper nouns “Israeli and “Palestinian with the generic agents “troops and “insurgents, the TABARI-derived output for the example is presented in Table 6:

| INSERT TABLE 6 HERE |

By the late 1990s, machine-coding had become increasingly popular, and almost all time and costs were upfront in the dictionary and software development phase. Because these were open source, they were easily adopted and upgraded. In 2000, the KEDS projects launched the Textual Analysis By Augmented Replacement Instructions (TABARI) software, which became the dominant machine coding system in event data. Subsequent automated-coding software, like the proprietary VRA Reader used to build the King and Lowe (2004)’s 10 Million International Dyadic Events datasets, was modeled off of TABARI. Automated event coding has proven to be fast, accurate and replicable, inexpensive, and easily updatable. As of November 2011, TABARI was able to code 26 million stories for the ICEWS project in 6 minutes using a small parallel processing system. Since computers are able to rigidly apply coding rules, results are perfectly replicable. Moreover, because TABARI is open source, it is free to install and is easily manipulated to include customized dictionaries or coding rules, which made possible the creation of the GDELT dataset.

2.4. GDELT. In 2012, Kalev Leetaru released a new, cutting edge event dataset called the Global Database of Events, Location, and Tone (GDELT). This new dataset not only combined the strengths of King and Lowe (2004) dataset (i.e. global coverage) and the ICEWS dataset (detailed sub-state actor dictionaries, but it also uses an advanced natural language processing (NLP) program to provide latitude and longitude coordinates for each event, which not only combines
the strengths of the 10 Dyadic Events (i.e. global coverage) with ICEWS (i.e. robust sub-state actor coverage), but also provides latitude and longitude coordinates for the events. Thus, a typical GDELT event data observation provides latitude and longitude coordinates, as illustrated in Table 7:

[INSERT TABLE 7 HERE]

Part of the GDELT coding scheme is still proprietary, but here is what we know. GDELT uses TABARI and the CAMEO ontology to machine-code the entire content of electronic news stories. Additionally, GDELT obtains news stories from four sources: LexisNexis, Agence France Presse, Reuters, Associated Press, and Xinhua. Two key aspects of GDELT remain proprietary. First, it is unclear what GDELT is using for actor dictionaries, though we can be highly confident that they are of similar richness as the dictionaries built for the ICEWS project. Second, the process by which GDELT assign specific latitude and longitude coordinates to each event is still unclear.

The final output is 200+ million events from 1979 to February 2013 (and at the time of writing this, it is being updated daily). Each observation contains up to 70 columns of additional information regarding the actors and location of the event.

This dataset allows me two major advancements: first, it combines the strengths of the 10 Million International Dyadic Events dataset and the ICEWS project, by providing global event coverage with nuanced sub-state actor coverage. Second, it is the first machine-coded dataset to provide location information for events. Prior to GDELT, ACLED and UCDP-GED were the only large geo-coded political conflict dataset, with ACLED (the larger of the two) providing 75,000 events recorded across 60 countries. Since this dissertation contains the first rigorous analyses of the GDELT data, few tests have been performed to assess the external validity of the data.

However, two anecdotes suggest a high degree of external validity. First, I used the GDELT data to calculate a time-series reflecting the number of violent events that occurred per week in Aleppo and Homs during 2011 and 2012. I did this by selecting all material conflict events that occurred within the latitude and longitude coordinates surrounding Aleppo and Homs, and then calculating the sum of these event by week. Next, I plotted these time-series and visually cross referenced these values with a ground-truthed database from a Syrian NGO. The GDELT derived time series from both Aleppo and Homs appeared nearly identical to that of the ground-truthed dataset.

[INSERT FIGURE 1 HERE]
Second, similar maps and figures reflecting violence in Afghanistan built with GDELT were presented to U.S. government officials, and they were sufficiently similar to maps built with classified ground-truthed datasets as to warrant accusations that the GDELT-derived maps were plagiarized version of the classified ground-truth maps.

3. The Substance

At its core, this dissertation is a study of political violence. While all three chapters provide an empirical analysis of political violence, they address three different literatures. In Chapter 2, I focus on the costs of conflict, in Chapter 3, I analyze the causes of conflict, and in Chapter 4, I attempt predict conflict.

3.1. Chapter 2. There is a general assumption within the empirical study of conflict literature that war is costly. The opening sentence of one of the most heavily cited articles on war written in the last 20 years states, Fearon (1995), states: “The central puzzle about war, and also the main reason we study it, is that wars are costly but nonetheless recur.” Why is it important to understand the costs of conflict? The game theoretic literature provides rigorous theoretical analysis of factors impacting various aspects (i.e. onset, duration, termination) of war. Critical to many of these models is a parameter reflecting the cost of fighting, which affects whether or not actors decide to engage in conflict.

For example, Fearon (1995) applies the bargaining model of war framework to demonstrate that all else being equal, two states are more likely to reach a negotiated settlement short of war as their expected costs of fighting increase. Gartzke (1999) builds on Fearon (1995), further stressing the importance of actors’ expectations about cost in deciding whether to fight or negotiate. Similarly, Mesquita and Siverson (1995) theoretically argue and empirically demonstrate that leaders are more likely to avoid war when the expected costs of fighting increase. All else being equal, conflict becomes more appealing as the expected costs decrease. Conversely, Wagner (2000) shows that the side facing higher expected costs will be more likely to initiate conflict, and additionally argues that state’s increase their odds of reaching a favorable negotiated settlement when they are able to make their opponent believe that his costs of continued fighting will increase. Overall, it is difficult to find any game-theoretic model of war with equilibrium outcomes not affected by costs – either observed or expected – of fighting. Thus, the game-theoretic literature has strongly
demonstrated the extent to which costs – both expected and observed – affect all aspects bargaining both preceding the onset of a conflict and also during fighting.

How are the costs of conflict measured? The measurement of some costs, such as government expenditures or damage to infrastructure that results from fighting is straightforward. Most governments maintain detailed accounts of expenditures related to a conflict, and it is fairly easy to appraise how much a bridge or factory would cost to rebuild after being destroyed. However, other costs of fighting are impossible to measure directly and require empirical estimation. For example, much empirical literature interested in analyzing the costs of war have attempted to empirically tests its effects on trade and GDP. For example, Collier (1999) and Kang and Meernik (2005) find that civil wars, on average, have strong negative effects on GDP. Focusing exclusively on civil war in Sri Lanka, Grobar and Gnanaselvam (1993) likewise finds civil war significantly decreased GDP performance.

Furthermore, traditional wisdom would suggest that interstate war should lead to decrease in trade, and domestic conflict should lead to decreased in GDP. Any number of anecdotal examples could easily support this belief; the United State and Japan had been vibrant trading partners prior to the outbreak of WWII, and bilateral trade ceased after 1939. Russett and Oneal (2001) and Hegre, Oneal and Russett (2010) find support for this example, contending that conflict does in fact lead to lower levels of trade. Conversely, Barbieri and Levy (1999) find that on average, conflict does not impact trade.

Although the majority of scholars interested in analyzing the costs of war have focused on its effects on GDP and trade, a smaller yet important literature analyzes the effects of political conflict on other important economic areas, such as commodity prices, government bonds, currencies, and equities. For example, Frey and Kucher (2000), Frey and Kucher (2001), Frey and Waldenstrom (2004), analyze the effects of WWII on U.S. government bond yields, while Ferguson (2008) performs a similar analysis but focuses on the effects of WWI. Eldor and Melnick (2004), Fratianni and Kang (2006), and Bolbol (1999) study the effects of war on currency exchange rates, Eldor and Melnick (2004) do not find that foreign exchange rates on the Israeli foreign exchange market tend to respond to terrorist attacks, whereas Bolbol (1999) demonstrates that the civil war in Lebanon that lasted until 1990 led to devaluation of the Lebanese pound. Rigobon and Sack (2003), Zussman, Zussman and Orregard (2008), Schneider and Troeger (2006) all analyze the effect of terrorist attacks on equity markets, focusing primarily on the effects of conflict in the Middle East on equities traded
on the Tel Aviv Stock Exchange (TASE), London Stock Exchange, and New York Stock Exchange (NYSE).

In Chapter 2, I contribute to the broader literature interested in assessing the costs of conflict by providing an empirical analysis measuring the effects of political violence on financial markets. Drawing on the bargaining model of war framework, understanding whether conflict has a meaningful effect on financial markets is potentially important. Governments of countries that house publicly traded equity markets, such as the NYSE in the United States, the TASE in Israel, or the LSE in the United Kingdom, have interests in maintaining strong market performance. Not only do governments in these countries generate tax revenues from corporate profits, meaning that government revenues increase as equity prices rise, but members of government also receive considerable financial contributions from members of the finance sector. Additionally, since financial markets are so interconnected, poor performance of equity markets can increase the costs of government borrowing and adversely affect currency rates. Thus, all else being equal, governments would prefer to avoid the costs incurred with poor market performance.

As previously discussed, a number of studies have analyzed the effects of political conflict on equity markets. Among these, every single study finds that political conflict (primarily operationalized “terrorist” attacks) has a statistically significant effect on either equity market returns or variance in returns. It is easy to find anecdotal support for these findings, since equity markets have tended to respond negatively to major conflict events (think 9/11 of the London bombings on July 7, 2005). However, it is important to note that these studies tend to analyze the effects of conflict on financial markets located in countries that generally exist devoid of political conflict, like the United States and the United Kingdom. Thus, a logical question emerges: is the consistency of findings suggestions that equity markets do meaningful respond to political conflict reflective of a true relationship that holds across time and space, or is it more a function of biased case selection? To address this, I provide a rigorous empirical analysis of the effects of variation in the level of conflict directly at Israel on variance of returns of the Tel Aviv 100, which is an index comprised of the largest 100 companies traded on the TASE.

Israel, unlike other countries that house highly liquid, publicly traded equity markets, regularly experiences high levels of political conflict. Thus, it may be the case the the regularity with which violence occurs in Israel has caused investors to price equities under the assumption that the future business climate in Israel will experience violence. If this is occurring, then equity prices should not
meaningfully vary when violent events occur, because these were largely expected. Following the logic of the game theoretic models discussed above, whether Israeli equities meaningfully respond to violence directed towards Israel should have considerable effects on conflict dynamics. For example, if we extend Fearon (1995)’s bargaining model of war, Israel should be less likely to initiate a conflict with Palestinians if they believe that retaliatory attacks from the Palestinians will negatively affect financial markets, since this would impart additional costs on the Israeli government. On the other hand, if the Israeli government is confident the equities on the TA100 are more or less immune from political conflict, then that is simply one less obstacle to a more hawkish position.

The central question that I address in Chapter 3 is: does variance in returns of the TA 100 index meaningfully respond to variation in levels of conflictual events targeted as Israel? To empirically test this, I utilize the GDELT dataset to construct a daily level measure reflecting the number of conflict events initiated by actors out of Israeli, Palestinian-occupied, or Lebanese territory against Israeli actors. Then, I follow common protocol within econometrics literature and utilize a series of generalized autoregressive conditional heteroskedasticity (GARCH) models to test the extent to which variance in the TA100 meaningfully responds to variation in daily conflict. Additionally, I replicate this procedure to analyze the effects of conflict on variance in equity returns of the two largest insurance companies traded on the TA100, Migdal Insurance and Financial Holdings Ltd. (MGDL) and Clal Insurance Enterprises Holdings Ltd. (CLIS). I find that on average, the TA100 does not meaningfully respond to violent attacks. However, political conflict achieve weakly significance effects on variance of MGDL and consistent and highly significant effect on CLIS returns.

3.2. Chapter 3. Few, if any topics in international relations have received more attention than the causes of interstate conflict. In fact, a desire to better understand interstate conflict is what motivated the earliest political conflict datasets, from Woods and Baltzly (1915) to to COW. Over the past 50 year scholars have found empirical support suggesting dozens, if not hundreds of different variables meaningfully affect various aspects of interstate conflict. For example, at the dyadic level, the following factors are a sample of some that have been empirically demonstrated affect the likelihood of interstate conflict. In one most consistent finding in all of the empirical study of war, dozens of scholars, including Doyle (1986), Maoz and Abdolali (1989), Maoz and Russett (1992), have found empirical support for the “democratic peace” theory, which argues that wars do not occur between democracies. Perhaps even more robust than the effects of joint democracy
is the effect of distance. As the distance between states increases, the likelihood of interstate war
decrease, and sharing a border dramatically increases the chances of conflict (see Wesley (1962),
Vasquez (1993), Lenke and Reed (2001) and Starr and Thomas (2005)). The effects of dyadic trade
levels has also received considerable attention, with many scholars studies like Russett and Oneal
(2001), Bennett and Stam (2000), Gartzke, Li and Boehmer (2000), and Oneal (1996) finding that
the likelihood of interstate conflict is lower between states that trade with each other, though others
like Barbieri and Schneider (1999) find the opposite effect. More subtle factors, such as whether
two states shared ethnic groups ( see Davis (1997) and Woodwell (2004)), have also been studied
in detail.

Additionally, scholars have analyzed domestic level conditions that a
fect the likelihood of in-
terstate conflict. For example, Mansfield and Snyder (1995), Mansfield and Snyder (2002), and
Mansfield and Snyder (2009) focus on domestic democratic transitions, finding that states are more
likely to engage in interstate conflict during and after a democratizing movement. Chiozza and
Goemans (2004), Wolford (2007), Gelpi and Grieco (2001) and Bak and Palmer (2010) focus on
aspects of states’ leaders, finding that leadership tenure can increase the likelihood of conflict.
Furthermore, additional scholars have tested aspects of the diversionary diversionary theories of
war, which is a general theory that leaders will often seek out international crises or conflict in
order to focus domestic audiences attentions away from domestic issues. For example, Morgan and
Anderson (1999),Baker and Oneal (2001), and Kisangani and Pickering (2007) focus on approval
ratings, Russett (1990), Fordham (1998b), Fordham (1998a), and DeRouen (2000) test the e-
effects of economic conditions and Hess and Orphanides (1995), Smith (1996), and Tir (2010) analyze the
effects of domestic elections on interstate conflict.

Given the massive number of studies attempting to isolate conditions that a
fect interstate con-
lict, it is somewhat surprising that such little empirical research has analyzed the effects of domestic
conflict on interstate conflict. This is even more surprising given the large number of historical ex-
amples of domestic conflict influencing interstate conflict. For example, spillover effects from the
civi war in Rwanda in 1994 led to broader interstate conflict amongst states in the Great Lakes
region. More recently, the ongoing civil war in Syria has led to interstate conflict, as Israel has
begun launching intermittent missile attacks against Syrian government forces.
Despite the historical precedence of domestic conflicts influencing interstate conflicts, relatively little empirical work has attempted to analyze the effect that domestic conflicts can have on interstate conflict. Moreover, the studies that do attempt to empirically test linkages between domestic and interstate conflict tend to focus on either a specific sub-set of all possible types of domestic conflict and interstate conflict. For example, Trumbore (2003) focuses exclusively on the effect of domestic ethnic conflict on MID initiation, Davies (2002) analyzes the effects of riots and protests on MID initiation, and (Elbadawi and Sambanis (2002), Gleditsch (2007), Regan (2000) all analyze the extent to which a domestic conflict makes a state more likely to be the target of an intervention. Although these studies provide a theoretical and empirical foundation, they all exclude certain types of domestic conflict or interstate conflict from their analyses. For example, Trumbore (2003) and Davies (2002) focus exclusively on the initiation of interstate conflict, while Elbadawi and Sambanis (2002), Gleditsch (2007), Regan (2000) only measure when a state becomes the target of interstate conflict.

Gleditsch, Salehyan and Schultz (2008) largely overcomes the problem of only focusing on limited types of domestic and interstate conflict by analyzing the effects of domestic conflicts according to the comprehensive, UCDP/PRIO coding rules, on involvement in a MID. Gleditsch, Salehyan and Schultz (2008) and all of the studies referenced in the previous paragraph have two things in common. First, they all find similar results to the aforementioned studies – regardless of operationalizations of conflict, a country experiencing a domestic conflict is more likely to be involved in an interstate conflict. Second, they all use binary, annual measures of domestic and interstate conflict. The first similarity is positive, since consistent empirical findings of factors influencing interstate conflict are rare in the literature. The second similarity leads to considerable shortcomings in empirically testing for relationship between domestic and interstate conflict.

For example, consider tests of the effect of onsets of domestic conflict on the likelihood of an onset of interstate conflict. In this context, binary data is not problematic, since the concept of an “onset” is conducive to a yes/no framework. However, the “annual” aspect of data is more difficult for two reasons. First, it is impossible to tell whether a domestic conflict onset actually precedes an interstate conflict if both onsets occur within the same calendar year. Thus, if scholars do not lag the variable reflecting domestic conflict, they run the risk of conflating interstate conflicts that lead to domestic conflicts (think the civil conflict that followed the U.S. Invasions of Iraq and Afghanistan) with domestic conflict that lead to interstate conflict (think the U.S. attacks on
Ghadaffi forces in response to civil conflict in Libya). As a result, lagging the variable reflecting domestic conflict onset is the preferred method, since this ensures that the domestic conflict onset temporal precedes the interstate conflict onset. However, this approach effectively drop instances of interstate conflict onsets that quickly follow onsets of domestic conflict in the same calendar year, as often occurs.

Although the “binary” nature of the data employed in the extant literature does allow for crude tests of onset, it completely eliminates the possibility of tests of conflict intensity, either in the domestic or interstate conflict. Thus, scholars relying on the binary, annual level conflict data are simply unable to ask interesting questions like, does an onset of interstate conflict become more likely as ongoing domestic conflicts becoming more severe? Or, does an onset of domestic conflict affect the intensity of an ongoing interstate conflict?

The central goal of Chapter 3 is to move beyond the binary, annual measures in order to provide more nuanced analyses of the effects of domestic conflict on interstate conflict. To accomplish this, I obviously need more nuanced data, which I am able to build using the GDELT data. With GDELT dataset, I derive monthly, continuous measures reflecting the number of domestic conflict events each month for over 150 countries from 1979 to 2004 and the number of inter-state conflict events per month for all non-directed dyads for the same time period, which results in over 4 million observations.

This allows me to make two major advances on the current literature. First, I move beyond the annual level to provide monthly level analyses of the effects of domestic conflict onset on the likelihood of domestic conflict conflict onset. Consistent with the existing literature, I find that onsets of domestic conflict in month $t$ in one or both states comprising a dyad tend to increase the likelihood of interstate conflict onset within that dyad in month $t+1$ and month $t+2$. Second, for the first time, I am able to provides tests accounting for domestic and interstate conflict intensity. I find that as ongoing domestic conflict becomes more intense in one or both states of a dyad, the likelihood of an interstate conflict onset increase. Moreover, when two states are engaged in an ongoing interstate conflict, that conflict intensity tends to lessen if both states also experience an onset of domestic conflict.

3.3. Chapter 4. In the previous two sections of this introductory chapter, I have cited over 60 studies performing empirical analyses of political conflict, and these are merely a small sample of the thousands of articles that have been published in peer-reviewed journals, presented at academic
conferences, or submitted as a PhD dissertation to collect dust in a back corner of a university library. An obvious, yet rarely asked question, is: what is the ultimate goal of all of these studies that comprise the subfield of political science that focuses on the quantitative study of conflict? Two sentences from Karl Deutsch’s introduction to Wright (1942)’s *Study of War* provide a commonly cited answer: “war, to be abolished, must be understood. To be understood, it must be studied.”

This statement contains two important points. First, we study war so that we can understand it. Second, we wish to understand war so that we can decrease its future occurrence. These seem like legitimate goals with which few scholars of war would likely disagree.

But, how do we actually know if our empirical models are enhancing our understanding of war, and how can we use these empirical models to actually prevent the occurrence of war, or at least provide some insight about future war dynamics as to lessen human suffering? Drawing on past scientists/philosophers like Sir Francis Bacon (Bacon (1602)), Sir David Hume (Hume (1748)), Sir Karl Popper (Popper (1934)) and current political scientists like Michael Ward (Ward, Greenhill and Bakke (2010)), Phil Schrodt (Schrodt (2010)), Gary King (King and Zeng (2001), I argue that the answer to both of these questions rests on models that focus on prediction, rather than explanation. Prediction is a vital tool for both of the two “big picture” goals of the empirical study of conflict. First, to the extent that an empirical model has actually enhanced our understanding of war, that model will be able to better predict war. If it cannot, then it is likely that statistically significant relationships it found either only hold for a limited spatial or temporal range and no longer apply to the current world, or, the relationships were simply noisy anomalies in the data. Either way, it is unlikely that a model unable to enhance predictive accuracy of war actually increases our understanding of war, and this model will certainly be unable to contribute to the future prevention of war.

Methodological approaches to forecasting conflict can be generally divided into two camps: game-theoretic and data-driven. Game-theoretic approaches generally focus on predicting a single outcome, such as will two sides reach a negotiated settlement during a given series of peace talks (see Bueno de Mesquita and Lalman (1992), Bueno de Mesquita (2002), and Bueno de Mesquita (2009)). To build these style models, researchers must first determine the relevant actors cable of influencing the outcome, second estimate each of each actor’s preferences regarding potential outcomes, and lastly mathematically solve for the equilibrium solutions given each actor’s preferences.

\[18\] I draw this example from Ward et al. (2012).
Bueno de Mesquita has used this approach successfully in a number of contexts (see Bueno de Mesquita (2002) and Bueno de Mesquita (2009)). However, it is slow and must be recalculated on a case-by-case basis. For example, Bueno de Mesquita often conducts rigorous interviews with relevant actors, and when this is not possible, spends considerable time reading about actors’ past behaviors. Thus, game-theoretic approaches are appropriate in some circumstances, such as trying to determine whether Iran will obtain a nuclear weapon, but less so in others, such as building real-time forecasts of local-levels of violence in multiple countries.

The data-driven method is a far more common technique to building predictions of political conflict. This approach involves three general steps, which are often immensely complicated in actual practice: first, collect data on an outcome of interest (say, a binary measure of war onset) and ideally additional covariates that may effect the outcome of interest; second, train an empirical model on a subset of the data in order to hopefully identify empirical patterns; third, use patterns found on the training set to build predictions on a hold-out portion of the data in order to determine predictive accuracy. To complicate matters further, an additional division exists among studies using data to attempt to predict political conflict, largely defined as “structural” and “dynamic” models.

Structural models, like those employed in Gurr and Harff (1996), King and Zeng (2001), Fearon and Laitin (2003), and Goldstone et al. (2010), rely on coarse, annual level data. This means that structural models are only capable of building forecasts of conflict at the state- or dyad year. Again, this is useful in some contexts, such as theory-testing for academics or defense budget allocation for major-power governments, but not capable of providing sub-annual or sub-state forecasts.

The bulk of data-driven approaches utilize dynamic models, focusing on sub-annual level variation in conflict. The vast majority of dynamic forecasting models utilize fine-grained, machine-coded event data, such as Schrodt (1999), Pevehouse and Goldstein (1999), and Shellman (2004). These studies have demonstrated are capable of providing accurate, sub-annual level forecasts of political conflict. However, the lack of geo-location information means that these studies have been unable to generate predictions desegregated to sub-state geographic units. A smaller number of dynamic forecasting models, like Weidmann and Ward (2010), utilize human-coded data that does provide geo-location information, and as a result, are able to build sub-annual level forecasts of levels of violence at the sub-state administrative unit. However, since human-coded datasets provided limited spatial coverage, it is not possible to extend these models to all countries.
The central goal of Chapter 4 is to build accurate forecasts of future levels of political violence at a sub-state and sub-annual level of temporal nuance, and do so in a way that could be applied to any future conflict occurring in any country in the world in real-time. Since this requires making many predictions in a short period of time with limited (or no) information about the preferences of the actors involved, a game-theoretic approach is not feasible. Thus, I implement a data-driven model. Historically, a tradeoff existed between using machine-coded datasets, which could provide global coverage but no geo-coded information, or human-coded datasets, which contained geo-coding but were difficult (or impossible) to maintain in real-time for a large number of countries. The use of the GDELT dataset allows me to overcome both of these shortcomings.

Despite the considerable benefits resulting from the scope and detail of GDELT’s 200+ million observations, this also creates technical difficulties. One major current challenge is aggregating GDELT data to sub-state geo-spatial units. Currently, this requires the combination of a number of computational scripts to first pull relevant GDELT observations, and then aggregate these using shape file and Geographic Information software (GIS). Given the time-intensity of this process, I build predictions for a single country to serve as a proof-of-concept for an eventual model with global coverage. I choose to focus on Afghanistan, since it experienced high levels of regionally dispersed violence, and existing studies have demonstrated strong local-level predictive accuracy using human-coded data (see Mangion-Zammit et al. (2012)).

Using GDELT and GIS, I calculate the number of conflict events that occur at the district, province, and country level for each month from 2001 April 2012. I focus primarily on building forecast at the district-month level (number of districts = 317), since the district is Afghanistan’s smaller administrative unit. However, I likewise build forecasts at the province-month (number of province = 32)) and country-month level. This allows me to speak to the effects of geo-spatial aggregation on predictive accuracy. Empirically, I build predictions using an autoregressive fractionally integrated moving average (ARFIMA) model. To assess predictive accuracy, I set aside the final 48 months from April 2009 to April March 2012 as out-of-sample test months, and iteratively build unique forecasts for each geo-spatial unit (district, province, country) for each of these months using a one-month-in-advance framework. This results in 317, 32, and 1 prediction at the district-, province-, and country-month unit of a analysis, respectively, for all 48 out-of-sample months. For each month, I calculate whether the ARFIMA model’s prediction generate lower mean absolute error (MAE) across the spatial units relative to a naive model that simply predicts that
the number of conflict events in month $t = \text{month } t - 1$. The ARFIMA model outperforms this naive model in 47 out of 48 months at the district-month level, 42 out of 48 at the province-month level, and 40 out of 48 at the country-month level. Additionally, I experiment with feature building, alternative forecasting algorithms, and the inclusion of exogenous drug price variables, but none of these approaches improve predictive accuracy achieved by the univariate ARFIMA model.


4. Appendix

Table 1. Example of a COW interstate war observation

<table>
<thead>
<tr>
<th>WarNum</th>
<th>StateName</th>
<th>Start</th>
<th>End</th>
<th>WhereFought</th>
<th>BatDeaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spain</td>
<td>4/7/1823</td>
<td>11/13/1823</td>
<td>2</td>
<td>600</td>
</tr>
<tr>
<td>1</td>
<td>France</td>
<td>4/7/1823</td>
<td>11/13/1823</td>
<td>2</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 2. Example of a MID observation

<table>
<thead>
<tr>
<th>DispNum</th>
<th>StateAbb</th>
<th>Start</th>
<th>End</th>
<th>HostLev</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>YUG</td>
<td>5/2/1913</td>
<td>10/25/1913</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>AUH</td>
<td>5/2/1913</td>
<td>10/25/1913</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3. Example of a UCDP observation

<table>
<thead>
<tr>
<th>Location</th>
<th>SideA</th>
<th>SideB</th>
<th>1</th>
<th>Start</th>
<th>End</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>United Kingdom</td>
<td>Albania</td>
<td>10/22/1946</td>
<td>12/31/1946</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Example of an ACLED observation

<table>
<thead>
<tr>
<th>Date</th>
<th>Actor 1</th>
<th>Actor 2</th>
<th>Event Types</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
</table>
Table 5. Example of a UCDP-GED observation

<table>
<thead>
<tr>
<th>SideA</th>
<th>SideB</th>
<th>Type</th>
<th>Start</th>
<th>End</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
</table>

Table 6. Example of a KEDS Event

<table>
<thead>
<tr>
<th>Date</th>
<th>Source</th>
<th>Target</th>
<th>CAMEO code</th>
<th>CAMEO event</th>
</tr>
</thead>
<tbody>
<tr>
<td>19980312</td>
<td>ISRMIL</td>
<td>PALINS</td>
<td>190</td>
<td>(Use conventional military force)</td>
</tr>
</tbody>
</table>

Table 7. Example of a GDELT Event Data Event

<table>
<thead>
<tr>
<th>Date</th>
<th>Source</th>
<th>Target</th>
<th>Action</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>20040507</td>
<td>THAMIL</td>
<td>THAREB</td>
<td>20</td>
<td>13.7500</td>
<td>100.4833</td>
</tr>
</tbody>
</table>

Figure 1. The Number of GDELT-derived Violent Events in Homs and Aleppo from January 2012 through June 2012
Figure 2. The Number of Ground-truthed Violent Events in Homs and Aleppo from January 2012 through June 2012
CHAPTER 2. USING POLITICAL EVENT DATA TO ANALYZE VARIANCE IN TEL AVIV 100 INDEX RETURNS

INTRODUCTION

Scholars and practitioners alike have long been interested in better understanding the effects of political events on financial markets. Largely after 9/11, researchers began to give considerable attention to how markets respond to political violence. Among the dozens of empirical studies that have emerged analyzing the effects of various forms of political violence (though primarily focusing on terrorism) on financial markets (including commodities, bonds, currencies, and equities), not a single study has failed to reject a null hypothesis that markets do not significantly respond to variations in violence over time. These findings tend to match much of the anecdotal evidence. For example, on the first day of trading following 9/11, the Dow Jones Industrial Average (DJIA) fell 7%.

But are these results a function of research design and case selection? Should we expect these results to hold across time and space? At least one major theory – the efficient markets hypothesis — would suggest that we should only expect to see consistent, meaningful reactions in financial markets to political violence events when these events reveal new information. To put it more directly, traders should only respond to a political violence event if they believe that the event reveals new information about the future profitability of an asset. In some cases, like 9/11, this was clearly the case as it was first large-scale, foreign attack on the contiguous United States in nearly 200 years. But in other countries like Israel, where political violence is common, it is possible that political violence actually reveals no new information since investors are already operating in a climate where political violence is the norm.

Consider what occurred in early November, 2012 in Israel, when Hamas militants operating out of the Gaza Strip escalated rocket attacks against targets in southern Israel. In response, Israel launched Operation Pillar of Defense on November 14 with 20 air strikes, killing high-profile Hamas leader Ahmed Jabari. From November 14 to the November 21 cease fire, Hamas launched over 1,000 small rockets into Israeli territory, and Israel carried out approximately 1,500 targeted...
air strikes. All together, November 14 to November 21 was the most conflictual week in the Arab-Israeli conflict in nearly a decade. Meanwhile, Israeli financial markets seemed unaffected. The Tel Aviv 100 (TA100) – an index of the 100 largest equities traded on the Tel Aviv Stock Exchange – opened on November 14 at $1061.79 and closed on November 21 at $1064.45, reflecting a trivial 0.25% gain for the seven day period. Though only an anecdote, this suggests the possibility that in Israel, financial markets may be fairly immune to variation in levels of political violence.

Thus, while the extant literature has provided consistent findings that financial markets in stable countries respond significantly to political violence, less is known about how financial markets in conflictual countries respond to violence. Therefore, in this paper, I provide rigorous empirical tests to measure the extent to which variation in the level of violent attacks against Israel affects variance in returns of the TA100.

Before proceeding, I address two questions underlying the importance of this paper: why focus on Israel and what is so important about financial markets? First, Israel is an ideal case study because it possesses an uncommon combination of a large, highly liquid financial market and high levels of political violence.

Second, understanding the effects of political violence on financial markets is important for both practical and theoretical reasons. For example, from a practical investment perspective, if you are a mutual fund manager wishing to reduce your emerging market fund’s volatility, should you avoid markets with political violence, like Israel, India, or Nigeria, based on the expectation that should future violence occur, market volatility will increase? Additionally, for traders, do short term opportunities for profit exist amidst political violence? If we know markets tend to not respond to political violence, but experience an abnormal negative shock following a particularly large attack, is this a strong buying opportunity? Additionally, the effect of violence on financial markets has tremendous influences on conflict dynamics. Imagine a rebel group with political demands that engages in missile attacks, bombings, riots, etc.. In Scenario 1, these actions have considerable negative effects on financial markets. Traders panic, volatility spikes, and prices plummet. In Scenario 2, traders basically ignore these rebel actions because they have been used off and on for the last 50 years and all political risk has already been priced into financial assets. Rebel bargaining power is likely considerably higher in Scenario 1, especially if the target state is a democracy and accountable to the domestic investors and corporations who are incurring the losses. In Scenario
2, the government would be under far less pressure to act since it is “business as usual” in the markets even amidst the increases in violence.

To address the central question of this paper — to what extent does variation in the level of violent attacks against Israel affects variance in returns of the TA100 — I utilize the GDELT event data dataset. From this, I build daily level features that reflect the number and severity of violent events committed against Israel by relevant local actors. Next, I employ generalized autoregressive conditional heteroskedasticity (GARCH) models with a number of financial control variable to see whether the political violence events have a significant effect on variation in TA100 returns. Contrary to the existing literature, I find that the TA100 does not meaningful respond to political violence committed against Israel. As an additional robustness check, I additionally test the extent to which political violence events affect variance of daily returns of the two largest insurance companies in the TA100, with the rationale being that if any companies are likely to respond to variation in violence, it is likely to be insurance companies, who’s profits fluctuate based on material damages. Using the same GARCH approach, I find that MGDL does not respond but CLIS does.

1. The Literature

A number of recent studies across a variety of disciplines have attempted to analyze the effects of various forms of political violence on different types of financial markets. Since “political violence” is a vague concept, scholars operationalize it in different ways. In general, the extant literature focuses on three different types of political violence:

- **terrorist attacks**
  - Though definitions vary between studies and datasets, these tend to be unannounced acts of violence committed against civilians.

- **episodic event history**
  - A series of actions that, taken together, make up a broader event, such as the start of a campaign or war.

- **atomic event data**
  - The specific actions of a conflict treated as unique events.
First, the majority of existing literature relating political violence to financial markets focuses on effects of terrorism (Eldor and Melnick (2004), Chen and Siems (2004), Arin, Ciferri and Spagnolo (2008), Kollias et al. (2011), Chesney, Reshetar and Karaman (2010), Johnson and Nedelecscu (2005)). Due to both the highly subjective natures of the term “terrorism” as well as the multidisciplinary backgrounds of the scholars writing on this subject (including finance, economics, and political science), no established benchmarks exist for how to operationalize “terrorism”. As a result, each of the above cited studies use terrorism datasets consisting of different events, many of which provide limited (or no) justification for why certain acts of terror warranted inclusion while others failed to make it into the dataset. For example, Chen and Siems (2004) select 14 major terrorist/military events that occurred between the sinking of the Lusitania in 1915 and the attacks against the World Trade towers in 2001; Kollias et al. (2011) analyze the effects of 15 and 21 incidents of terror in the United Kingdom and Greece, respectively; Arin, Ciferri and Spagnolo (2008) uses the MIPT terrorism data to analyze variance in equity market returns in Indonesia, Israel, Spain, Thailand, Turkey, and the U.K.; and Chesney, Reshetar and Karaman (2010) analyze the effects of 77 major terrorist attacks occurring around the world. In each of these cases, the scholars chose to focus on a small subsection of all possible acts of terror that occurred within their spatial and temporal domains of interest. More importantly, these subsamples of “terrorist” attacks are not selected at random, but rather, selected by intensity, with only the largest scale events gaining inclusion. As a result, the universal findings that terrorism events negatively affect financial markets do not speak to the effects of terrorism in general, but rather large scale and generally unexpected acts of terror.

With this in mind, the results become less insightful since they are both readily obvious and not generalizable to a country like Israel, which is often the target of hundreds of small-scale missile attacks a year. Among the studies analyzing the effects of terrorism on financial markets, Eldor and Melnick (2004) provides the most rigorous research design by analyze the effects of 639 “terrorist” attacks against Israel from 1990 to 2003 on the Tel Aviv 100 index and Shekel-Dollar exchange rates. However, like the previous studies, Eldor and Melnick (2004) does not provide an explanation of the coding rules used to determine a “terrorist” attack. Additionally, although 639 seems like a sufficiently large number to eliminate the biasing effects of selecting only the most severe attacks, consider that from December 2011 to November 2012 alone, over 600 rockets were launched into
southern Israel from Gaza. Thus, it is likely that 639 attacks in a 13 year span represents only a subsection of total attacks that occurred during that time period, potentially biasing results.

With regard to the central question of this chapter—does variation in levels of political violence have a meaningful effect on variance in TA100 returns?—“terrorist” attacks only represent a small portion of events that comprise the broader concept of political violence. Other forms of violence—such as riots, protests, or traditional military exchanges—may also be important but tend to be excluded by studies focusing exclusively on “terrorism”. To this end, a number of studies utilize a more inclusive “event history” operationalization of political violence. For example, Frey and Kucher (2000) and Frey and Waldenstrom (2004) focus on the effects of major events during WWII on European bond rates; Rigobon and Sack (2003) analyzes the effects of various events preceding the U.S. invasion of Iraq in 2003 including key addresses by George W. Bush and meetings and activities of U.N. weapons inspectors; and Zussman, Zussman and Orregard (2008) looks at the effects of a broad range of events in Israel on TASE prices, including meetings, cease fires, and key military outcomes.

Although these “event history” studies account for more inclusive forms of political violence than just terrorist attacks, they are still highly subjective and inherently ad hoc. The subjectivity arises from informally choosing select “conflict” events. For example, Frey and Kucher (2000) provide no justification for only including a handful of key events from WWII and omitting hundreds of others. Additionally, Zussman, Zussman and Orregard (2008) clearly highlights the presence of severe “hindsight bias” among these “event history” event studies. They suggest that a cut in interest rates in August, 1998 increased TASE equity returns, but they do not mention that interest rates were cut seven other times in 1998 alone. Unless the interest rate cut in August was a unique case (it was not), it is likely that the true relationship between the August 1998 interest rate cut and the increase in TASE returns was corollary, not causal. Given the ad hoc nature of the “event history” approaches, it is impossible to differentiate correlation from causality.

Schneider and Troeger (2006) overcomes the shortcomings of both the “terrorism” and the “event history” studies by operationalizing the concept of political violence with the use of event data. In their study, Schneider and Troeger (2006) build a measure of political conflict using data from King and Lowe (2003)’s 10 Million International Dyadic Events datasets, which incorporates events ranging from cooperative meetings and negotiations to conflictual threats, bombings, and artillery attacks. Further, this approach overcomes the shortcomings of the “terrorism” and “event history”
approaches in a number of ways. First, it does not require making judgment calls about whether an attack qualifies as a “terrorist” act – a bombing is simply a bombing, regardless of whether it was intended to instill fear and targeted civilians. Second, unlike the event history approach, event data records atomic, rather than aggregate conflict events. This approach incorporates only the information available in real time, thereby avoiding hindsight bias by retrospectively clustering individual events into an aggregate event. For example, event data does not provide events like “Germany invaded France”. Instead, it would report that German troops crossed the border, German planes bombed French cities, and German and French troops exchanged gunfire. This also avoids hindsight bias – after the fact it may be easy to attribute financial market movements to the aggregate “German Invasion” but in real time events occur atomically, not cumulatively.

Though Schneider and Troeger (2006) make a convincing case for the use of event data and find interesting results, this chapter builds on their work in two important ways. First, whereas Schneider and Troeger (2006) are interested in the effects of levels of violence in Israel on global financial markets (the DJIA and FTSE), I focus on the effects on the domestic TASE. I argue that this is important for a number of reasons. For example, although Israeli policymakers work closely with politicians in the United States (who may alter their positions towards Israel based on the effects of conflict in Israel on U.S. financial markets), it is likely that Israeli policy makers take into much greater consideration the effects of conflict on their own domestic markets. Imagine if the Hamas rocket attacks from December 2011 through November 2012 had dramatic and detrimental effects on the TASE, either in the form of lower returns or greater market volatility.\(^1\) Additionally, it is interesting from a theoretical position because Israel represents one of the only cases of a well-functioning, highly liquid financial market in a highly conflictual region. Given its uniqueness, findings in the extant literature that consistently find that financial markets do respond to political conflict may not hold for Israel.

Second, I utilize a newer, more comprehensive event data set and utilize the event data to build more logical measures of political violence. In Section 3, I outline the use of event data in Schneider and Troeger (2006) highlighting shortcomings, introduce the GDELT dataset, and discuss how my aggregation techniques overcome these shortcomings. Before proceeding to Section 3, I first briefly outline theoretical arguments underlying competing hypotheses arguing that the TA100 should and should not respond to violence.

\(^1\)All else being equal, many fund managers prefer less volatility, meaning that enhanced volatility in the TA 100 would likely discourage investment.
2. Brief explanation of competing hypotheses

Much of the extant literature analyzing the relationship between political violence and financial markets tend to suggest the similar, intuitive explanation that markets respond negatively to political violence events since these disrupt economies both physically and mentally. Indeed, implicit in some of the most canonical models of political conflict is that all else being equal, violence is costly to a state (see Fearon (1995)). Physically, violence can damage the means of production and the infrastructure needed to transport goods. Mentally, violent events both cause fear among investors and buyers of goods (especially acts of terrorism) and generate feelings of ill will between leaders thereby disrupting trade and other forms of mutually beneficial commerce (see Anderton and Carter (2001) and Anderton and Carter (2003) for empirical evidence suggesting war disrupts interstate trade). Any number of case studies – 9/11 world trade attacks, Germany’s invasion of France in 1940, July 7th bombings in London — support the straightforward argument that financial markets tend to respond significantly and negatively to political violence. This all suggests that the same relationship should hold true in Israel, which leads to the sole hypothesis of this paper:

**Hypothesis 1**: Variation in the level of violence towards Israel will have a statistically significant impact on variance in returns of the TA100.

However, theoretical arguments suggesting that political violence has a meaningful effect on financial markets tend to conceptualize violence and consequently, peace, in black and white terms with peace being the natural state of the world, intermittently disrupted by violent events. We see this play out in the majority of the previously discussed “terrorism” and “event history” studies, which tend to subjectively analyze the effects of large scale (economically disruptive) and unexpected (unpredictable) instances of political violence. But what about in areas where political violence is common and tends to be low-scale, as is the case in many countries that would serve as interesting case studies for the effects of violence on financial markets (India, Nigeria, Indonesia, etc.)? In these cases, the relationship may not be as straightforward.

According to financial theory, the price of an equity reflects investors’ perceptions of the underlying company’s ability to profit in the future. As Robock (1971) argues, the extent to which a future event is likely to affect the price of a financial asset is directly related to extent to which the that event is predictable, because according to the “efficient market hypothesis”, if the future event is predictable, then the risk of that event occurring at time $t + n$ will be priced into the asset at
time $t$ (see Fama (1970)). Cosset and Doutriaux de la Rianderie (1985) extends this even further to assert that, “events that are either expected or easy to anticipate do not constitute political risk.” This line of argument fully supports much of the previous findings in the extant “terrorism” and “event history” studies, which find that political violence significantly affects financial markets because these studies tend to focus on rare, unanticipated events.

Unlike large-scale terrorist attacks against western industrialized democracies, violence directed towards Israel is frequent and tends to be low scale. Thus, the efficient market hypothesis suggests that equity prices already assume that domestic corporations will exist in a climate of violence in the future. This means that during escalations in violence directed towards Israel (as occurred in November 2012), we should not expect to see markets meaningfully respond.2 Li and Sacko (2002) provide empirical support for this line of argument in the context of interstate trade. They find that levels of trade between two states tend to decline after a military dispute between countries that do not usually engage in military disputes, but is unaffected when a dispute occurs between states that consistently experience conflict. These empirical findings along with the theoretical arguments of the efficient markets suggest the null – that variation in levels of violence against Israel should not significantly affect variance in TA100 returns. Thus, both theoretical and empirical support exists for both Hypothesis 1 as well as a null finding.

**Null:** Because traders efficiently incorporate future risk in today’s prices, variation in levels of violence directed against Israel should not have a significant effect on variation in TA100 returns.

3. **Data and Research Design**

3.1. **The Event Data.** Recognizing that collecting and utilizing event data is difficult, the critiques I level against the data and data treatment of Schneider and Troeger (2006) are more the result of the improvements in the field across the last 6 years than mistakes made in 2006. Nevertheless, I am able to improve on operationalizing political violence with event data in a number of ways.

First, in terms of event data quality, Schneider and Troeger (2006) use the King and Lowe (2003) 10 Million International Dyadic Events dataset, which relies exclusively on Reuters and Agence France Press (AFP) newswires. Like most automated coding systems, King and Lowe

---

2This assumes that the level of violence is relatively in line with past levels. If a “black swan” event should occur – such as an ICBM exploding in Tel Aviv –, then we should expect to see a meaningful market response because this level of attack would have been unexpected and therefore not priced into assets.
(2003)’s approach does not measure the scope or intensity associated with each event (i.e. it is
unable to extract information about whether a bombing kills 1 person of kills 1000). Because of
the reliance on just the title and lead sentence of two newswires, events tend to only generate one
(or a few) article a day, regardless of their importance/scope. Thus, during periods of intensified
conflict between Israeli’s and Palestinians, it is likely that the event data does not fully capture the
variation in actual levels of conflict at the daily level (e.g. the temporal unit of analysis employed
in Schneider and Troeger (2006)) , since the event data generating process used to build to 10
Million International Dyadic Events dataset is inherently smoothing. Additionally, the 10 Million
Dyad Dataset was designed to capture inter-state interactions. As a result, it only recognized 450
sub-state domestic actors meaning that it is likely to miss a large number of conflictual events that
target specific individuals or sub-state groups.

Second, regarding treatment of the event data, Schneider and Troeger (2006) utilize questionable
aggregation approaches. As discussed in the Appendix to this dissertation, studies utilizing event
data must, at minimum, clearly specify their aggregation choices regarding actors, actions, and
temporal range. Schneider and Troeger (2006) do clearly explain their action (counts of conflictual
and cooperation events) and temporal (daily) choices. However, they make no mention of the
actors between whom actions must occur in order for the event to make it into the dataset. This
both renders replication impossible and inhibits our ability to interpret the empirical findings.

Third, Schneider and Troeger (2006) build two measures that reflect political conflict: the count
of cooperation events and the count of conflictual events. While this is not problematic, the
way that they use the two counts in the time-series models is. First, these two count variables are
highly correlated. In the Israel-Palestine conflict, cooperation events (including meetings and verbal
dialogue) tend to follow shortly after almost all attacks (see Table 2). Simultaneously estimating two
highly corollary variables in the same model can severely inhibit our ability to conduct inferences
on coefficients and standard errors. Additionally, and perhaps more importantly, is that Schneider
and Troeger (2006) only consider the events that occur at time $t$ and time $t - 1$. However, we
know that political events do not occur in a vacuum. Rather, they occur within a broader political
climate. Thus, it is not enough to simply account for the number of events occurring at time $t$
and $t - 1$ because investors are likely to interpret these events differently based on recent events.
Imagine a missile exploding in Tel Aviv that kills three civilians. We should expect that traders
would react differently to this missile attack if it were to occur during a peaceful period than if
it occurred amidst ongoing fighting and daily missile attacks. Thus, I argue that it is important to construct event data derived measures that reflect both the ongoing political climate as well as more immediate, short-term shocks. Existing studies that have analyzed financial markets with daily level event data have done a poor job at constructing appropriate measures.

With these shortcomings in mind, I utilize the GDELT dataset and detail my aggregation choices and variable construction techniques. Following the advice of the Appendix, I describe the actor, action, and temporal choices I make during the aggregation process, and then describe further data manipulation to build meaningful measures of political violence.

3.1.1. Actors: GDELT contains events occurring between tens of thousands of actors, most of whom have no influence on Israeli equity markets (unless you believe that a rebel group flapping its wings in Mozambique cause a market crash in Israel, which I do not). I focus on events initiated by actors whose primary affiliation is with Israel, Palestine, or Lebanon against an actor whose primary affiliation is with Israel. Like all event data studies, the choice of actors is a subjective call, and all else being equal, parsimony is preferred. Arguments could be made to include actors from the United States, Saudi Arabia, Jordan, Egypt, Syria, etc.. However, from 1992 to the 2012 (i.e. the years of active trading in the TA100 index), the most pressing security concerns for Israel have arisen from agents operating from the Palestinian occupied territories (Gaza and the West bank), Lebanon, and from within Israel itself. I feel confident that focusing strictly on events initiated by actors identified as Lebanese, Palestinian, and Israeli towards Israeli actors captures the large percentage of the political events that affect Israeli financial markets.

3.1.2. Actions: GDELT uses the 20-cue category CAMEO coding system, which uses 1-4 digit numerical codes to reflect a broad spectrum of relevant political events. As the Appendix illustrates, scholars utilizing CAMEO-coded data tend to aggregate the raw codes into a more meaningful format. Although scaling CAMEO codes is the most popular action aggregation technique within the event data literature, this leads to major shortcomings (see the “sum” and “mean” problems in the Appendix. Instead of scaling, I implement a highly simplistic measure of the level of political violence by simply counting the number of events that qualify as “material conflict”.

3.1.3. Temporal: GDELT provides the specific day on which every event occurs. Since I use daily level financial data as my dependent variable, I keep the GDELT data in its daily form as opposed to aggregating into more coarse weekly or monthly totals. Daily level temporal aggregation is
employed in a number of relevant studies, including Schneider and Troeger (2006), Leblang and Mukherjee (2005), Freeman, Hayes and Stix (2000), and Hammoudeh, Yuan and McAleer (2009).

3.1.4. Measuring scope from duplicates. Automated event data extraction has advanced considerably in recent years, but still struggles to extract measures of scope. One way to measure the severity of an event is to count the number of times various news outlets report the same event. This assumes that large-scale events – for example, a bombing that kills 100 civilians – will receive more media attention than a smaller-scale event, such as a similar bombing that only kills one civilian. Since Schneider and Troeger (2006) use the 10 Million Dyad Dataset which only codes Reuters and AFP stories, even large scale events are likely to receive one or a few stories a day. Thus, they are unable to differentiate between a bombing that kills and a bombing that kills 100. Since the GDELT dataset codes hundreds of local, regional, and international news stories, the same large-scale event can be reported dozens or hundreds of times. In the absence of software able to extract measures of severity from the content of the articles, the number of times that a specific event is reported on a given day is the best approximation of the scope of the event. It is common practice among event data studies to eliminate duplicate entries. However, in this study, I keep all duplicates. This allows me to overcome the shortcoming in Schneider and Troeger (2006) by utilizing event data to better capture variation in the intensity of violent events. This information is especially important when analyzing consistently conflictual daily level data, which often contains minimal day-to-day variation when using the reduced form of the data.

In total, my final event dataset consists of 473,197 events across the 5,148 days on which a recorded event occurred from December 31 1991 to August 31 2012 (coded events do not occur on every day).

3.1.5. Additional manipulation of daily level count data. The process by which an observation gains inclusion into an event data dataset generally involves two broad components: 1) the event occurs in the real world; 2) the event is reported in open-source, readily accessible electronic news stories. The dramatic increased in the volume of online journalism in past decades has meant that even if the first component were to hold steady, the number of those events being coded into automated event datasets would steadily increase. Thus, a month in the GDELT dataset in 2011 with 100 material conflict events likely experienced less severe conflict in reality than a month in the GDELT
dataset with 100 material conflict events in 1992. The graph below illustrated the steady rise in the total number of events reported to have occurred between ISR-PAL-LEB from 1992 to 2012.

[INSERT FIGURE 1 HERE]

We know with virtual certainty that the large spikes in 2006, 2008, and 2010 did not actually experience 50 times more conflictual events that the most conflictual day from 1992 to 1995, despite the fact that 50 times more accounts of violence are in the GDELT dataset. Thus, left untreated, our data lacks external validity and would almost certainly lead to biased estimated when modeled in a time-series framework. In order to adjust the data to control for changes in the second component (i.e. the amount of online reporting) of the data generating process, I eliminate the increasing time trend through the following process.

First, I regress the daily event counts by time. Second, I use the stored regression coefficient and constant term to generate predicted values for the 1991 to 2012 time period. Third, I divide the counts by the fitted values. Minor adjustments needed to be to the fitted values during the first years of the time series made (setting the minimum divisor to equal the fitted value from early 1994), since they were in some instances negative, \( \leq 1 \), sufficiently small that 1992-1993 adjusted values became disproportionately large and variant to the rest of the time series. Additionally, I experimented with calculating two best-fit lines, one for 1992-2001, and one for 2002-2012 to reflect a major shift in Reuters policies around 2001. However, this had minimal effect on the final, de-trended values, but it did generate a strange shock at the 2001-2002 break. Thus, I use all data from 1992-2012 to generate the best fit line. The graph below illustrates the total sum of events after de-trending the data by dividing by fitted value.

[INSERT FIGURE 2 HERE]

As the graph indicates, considerable variation still exists in the series, but the increasing trend has been removed.

3.1.6. **Converting counts to meaningful information:** As previously discussed, political events do not occur in a vacuum. Rather, they occur within a broader political climate. Thus, it is important to construct event data derived measures that reflect both the ongoing political climate as well as more immediate, daily level shocks. Existing studies that have analyzed financial markets with daily level event data have done a poor job at constructing appropriate measures. For example, Schneider and Troeger (2006) simply account for the number of events occurring at time \( t \) and
time \( t - 1 \), which ignores longer term trends in the level of conflict. Consequently, this approach ignores events occurring more than one day in the past. This lacks considerable external validity, as we know with complete certainty that investors account for a longer temporal lag than one unit. There are two general techniques that are able to account for events beyond a one unit lag. First, it is possible to simply add more temporal components to the time-series model. Though feasible, this complicates both model estimation and interpretation. Second, we can construct new variables that incorporate longer term trends as well as short term shocks. This approach allows for a more parsimonious model (since one variable is able to reflect \( N \) number of lags, rather than having to include all \( N \) lags in the model) and is easier to substantively interpret. As such, I choose this latter option. Since the event data literature does not provide any additional established techniques for manipulating data to capture trends, I construct my own measures, as listed and defined below:

- **one_week_MA** = \( \frac{1}{7} \sum_{i=0}^{6} \text{violence}_{t-i} \)
  - The unweighted average number of violent events that occurred during the prior 7 days
- **two_week_MA** = \( \frac{1}{14} \sum_{i=0}^{13} \text{violence}_{t-i} \)
  - The unweighted average number of violent events that occurred during the prior 14 days
- **four_week_MA** = \( \frac{1}{28} \sum_{i=0}^{27} \text{violence}_{t-i} \)
  - The unweighted average number of violent events that occurred during the prior 28 days
- **\( \Delta \) one_week** = \( \text{violence}_i - \text{one_week}_{MA} \)
  - The change in the number of events occurring today from the unweighted average number of violent events that occurred during the prior 7 days
- **\( \Delta \) two_week** = \( \text{violence}_i - \text{two_week}_{MA} \)
  - The change in the number of events occurring today from the unweighted average number of violent events that occurred during the prior 14 days
- **\( \Delta \) four_week** = \( \text{violence}_i - \text{four_week}_{MA} \)
  - The change in the number of events occurring today from the unweighted average number of violent events that occurred during the prior 28 days

3.2. **The Dependent variable.** The TA100 is an index of the 100 largest companies traded on the Tel Aviv Stock Exchange, with a mean trading volume in 2012 of over 200 million shares per
day. The TA100 index was first introduced on December 31, 1991 at $100, and closed on November 17 at $1,045.13 (see Plot 1). The TA100 is an ideal for the purposes of this study since it reflects companies from an area with considerable variation in the level of conflict, high media attention, and sufficient liquidity to respond to short term shocks. In order to convert the raw TA100 time series into a more appropriate format for time-series analysis, I follow common protocol and calculate the first difference of logged returns. This is commonly employed when modeling financial time series because, as is the case with the TA100, simply taking the first differences generates a series with steadily increasing variances, as apparent in Plot 2. After taking the first difference of logged returns, the variance appears consistent throughout the series (see Plot 3), and a a Dickey-fuller test allows us to reject the null of a unit root. Additionally, I run a Philips-Perron test, which further rejects the null that a unit root exists with the first difference of the logged returns and indicates that this series is non-integrated.

[INSERT FIGURE 3 HERE]

4. THE GARCH MODELS

I choose to model model variance in TA100 returns using a GARCH model for two main reasons. First, it is an empirically justified approach. Like many other high frequency financial time-series, the TA100 index contains high degrees of volatility that tend to cluster rather than follow a random distribution. Notice in plot three that small returns tend to follow small returns (in the absolute value) and large returns tend to follow large returns to a greater extent than if returns were generated randomly. In his seminal study, Engle (1982) provides the first statistical approach able to account for what he deems autoregressive conditional heteroscedasticity (ARCH). To empirically test for what visually appears to be the presence of an ARCH process in the TA100 data, I run a Lagrange multiplier test, which allows me to reject the null hypothesis that no ARCH process exists with nearly 100% confidence. Instead of using an ARCH model, which tends to require a high order of autoregressive error terms, I utilize Bollerslev (1986)’s Generalized ARCH, or GARCH model.

Second, the majority of recent studies attempting to analyze the effects of exogenous variables (not only political violence but also elections, domestic policies, etc) on daily level financial market data employ a GARCH model or one of its variants. This includes Schneider and Troeger (2006) as well as Leblang and Mukherjee (2005), Freeman, Hayes and Stix (2000), Hammoudeh, Yuan and McAleer (2009), Bernhard and Leblang (2002), Arin, Ciferri and Spagnolo (2008), Dhankar
and Chakraborty (2007), and Mun (2008). Additionally, Alberg, Shalit and Yosef (2008) focus exclusively on univariate analyses of the TA100 and find that the GARCH model (and its variants) best fit the time series.\(^3\)

4.1. **Specifying the model.** A GARCH\((p,q)\) model consists of a conditional mean and conditional variance equation, both of which allow for the inclusion of exogenous variables. For consistency, I follow adopt the notation from Leblang and Mukherjee (2005). The conditional mean is:

\[
\Delta (\ln(TA100_t)) = \lambda + \psi Z_t + \epsilon_t
\]

where \(\Delta (\ln(TA100_t)) = \ln(TA100_t) - \ln(TA100_{t-1})\), \(\lambda\) is a constant that is approximately 0 due to the first differencing, \(Z_t\) is a vector of exogenous variables, \(\psi\) is a vector of estimated coefficients, and \(\epsilon_t\) is the error term distributed \((0, \sigma^2_t)\).

The conditional variance is:

\[
\sigma^2_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon^2_{t-i} + \sum_{i=1}^{p} \beta_i \sigma^2_{t-i} + \delta_i I_{i,t}
\]

where \(\omega\) is a constant, \(\epsilon_{t-i}\) is the lagged error, \(\sigma_{t-i}\) is the lagged variance, \(I_{i,t}\) is a matrix of exogenous variables, and \(\alpha_i, \beta_i,\) and \(\delta_i\) are estimated parameters.

4.1.1. **Control variables.** Decades of literature on financial markets has uncovered hundreds of factors that influence financial markets. Although the causal direction is often difficult to uncover, we know that statistically significant relationships exist between equity markets and commodity prices (especially oil), inflation rates, trade, other global markets, domestic regime types, etc. However, within the relevant literature interested in the relationship between political events and financial markets, little established precedent exists in terms of appropriate control variables. For example, most studies focusing on the effects of terrorism on equity markets do not control for any other financial variables, which likely leads to considerably underspecified models thereby systematically overestimating the effects of the terrorist events. In two of the most methodologically sophisticated relevant studies, Leblang and Mukherjee (2005) control for trading volume, inflation, and interest

\(^3\)It is highly likely that a variant of the GARCH, such as the T-GARCH or M-GARCH, may be a better fit for the data. However, comparing all possible GARCH models far exceeds the scope of this paper. For an application of APARCH, and EGARCH to the TA100 index, see Alberg, Shalit and Yosef (2008).
rates when testing for the effects of parties on variation in DJIA returns, and Schneider and Troeger (2006) control for major global financial markets when analyzing the effects of political violence on U.S. and British equity markets. Drawing on these studies, I choose two sets of control variables:

Control 1:

- DJIA – The first difference of the logged Dow Jones Industrial average, which reflects the performance of the global economy
- Crude oil – The first difference of logged daily Brent crude oil prices, which reflects global commodity markets
- Inflation – The first difference of logged inflation rates, reported monthly by the Israel Central Bureau of Statistics. Monthly values are extended to the daily level by assuming constant rates for every day in a given month.

Because Schneider and Troeger (2006) find that variation in the intensity of violence in Israel significantly affects variation in the DJIA, I implement a second set of control variables – Control 2—that excludes the DJIA. If variation in levels of violence in Israel does, in fact, affect both variation in the DJIA and TA100, then the GARCH model may be unable to uncover the relationship between variation in violence and the TA100 due to endogeneity issues with the DJIA. If the political violence variables fail to achieve statistical significant with Control 2 variables, we can be highly confident that no true effect exists.

5. Results

In Models 1-7 in Table III, I use the Control 1 set of controls in both the mean and variance equation of a GARCH (1,1) model. Additionally, in Model 1-6, I test for the effects of moving averages of the level of violence across different temporal ranges (1 week in Model 1 and Model 2, two weeks in Model 3 and Model 4, and four weeks in Model 5 and Model 6). None of the six event-data derived measures of political violence achieve statistical significance in the variance equation, which suggests that traders are not, on average, highly responsive to variations in the level of violence directed at Israel. Additionally, in Model 7, I omit all measures of violence. The AIC and BIC scores, which provide alternative measures of model fit that penalize for extra parameters, suggest that Model 7—which contains no measures of political violence—best fits the data. These findings suggest that we are unable to reject the null hypothesis that the TA100

---

4Due to the fat tails of the distribution of returns, I assumes a Student’s t, rather than a normal distribution.
does not meaningfully respond to variation in the level of intensity of violence against Israel. This finding also suggests that equity prices in Israel may already price in future political violence, and to the extent that the level of violence is not dramatically more intense than expected, markets will not take notice. Instead, it appears that the variance in TA100 returns is driven primarily by previous error and variance rates of the previous trading day, as well returns in the DJIA.\footnote{I also re-run Models 1, Model 2, Model3, Model 4, Model 5, and Model 6 using the one-unit lagged values of the political violence variables. The political violence variables continue to fail to achieve p-values of \( \leq 1 \).}

Second, because Schneider and Troeger (2006) find that violence in Israel affects variance in the DJIA, I rerun all analyses in Table III after omitting the DJIA as a control variable in both the mean and variance equation of a GARCH (1,1), again using Student’s t distribution of errors. Model 1 and Model 2 of Table IV suggest that change from one week MA does have a meaningful effect on variation in the TA100 at a modest \( p = 0.063 \) value. However, the AIC and BIC scores in Model 7, which omits all political violence variables, are considerably lower, suggesting that the low level of statistical significance in Model 1 and Model 2 may not actually improve model fit. Further, as Model 3, Model 4, Model 5, and Model 6 in Table IV show, the four additional measures of political violence do not approach statistical significance. This suggests that even in a potentially underspecified model, variation in levels of political violence do not meaningfully impact variance of TA100 prices. Taken together, findings in Table III and Table IV suggest that we are unable to reject the null hypothesis with any meaningful degree of confidence. Consequently, this indicates that the causal argument outlined in Hypothesis 1 is not present.

Realizing that it is dangerous to provide ad-hoc explanations of statistical findings, I cautiously argue that these findings suggest that traders already price in future violence so that when it occurs, they tend to not react. This finding has a number of important practical implications. First, and most importantly to conflict dynamics in Israel, it suggests that it is highly difficult for opponents of Israel to adversely affect Israeli equity markets through violence. This likely affects bargaining dynamics, since all else being equal, the costs that Israel incurs from being the target of violence is lower than more stable countries like the United States or England, where political violence is not already priced into assets. Second, it suggests that findings in the extant literature about the effects of political violence on various financial markets may not be generalizable to all countries. Alternatively, financial markets in countries with long histories of political violence, like Israel, may be more resilient to future violence than in more stable countries. Third, from a purely capitalist perspective, this finding suggests that any abrupt changes in Israeli equity markets
following escalations of attacks may reflect a buying opportunity, since the effects of those attacks (in terms of decreased corporate profitability) are likely already priced into the equity. Thus, abrupt change following violence may indicate exploitable mis-pricings.

6. **Robustness checks focusing on insurance equities**

This chapter, like many other studies interested in the effects of political violence on equity markets, has thus far focused on an index – the TA100 – as the dependent variable. Although the preceding sections have found almost no empirical support suggesting that variance in TA100 returns is driven by changes in levels of violence directed against Israel, it does not mean that violence has no effect on Israeli financial markets. Rather, it simply suggests that the average effect of variation in levels of violence on the 100 largest companies tends to not be statistically significant. However, it is feasible that certain sectors or companies are more vulnerable to political violence than others. For example, consider that airline and insurance stocks suffered the largest average losses on the DJIA on the first day of trading after both the September 11 attacks as well as the London bombings on Jul 7, 2005.\(^6\)

In this section, I repeat the research design and empirical testing outlined in Section 3 and Section 4, except instead of focusing on the TA100 index, I analyze the effects of variation in levels of violence against Israel on variance in returns of the two largest insurance companies publicly traded on the TASE: Migdal Insurance and Financial Holdings ltd. (MGDL) and Clal Insurance Enterprise Holdings (CLIS). MGDL is the largest insurance company traded on the TASE with a market cap of over 6 billion USD, with 69.3% of shares are owned by the Eliahu Insurance company and the remaining 30.7% held by public investors. CLIS is the second largest insurance company, though considerably smaller than MGDL with a market cap of 3.3 billion USD. Unlike MGDL, CLIS’ ownership is more diversified, with the largest shareholder owning only 10.7% of outstanding shares. Daily closing prices were obtained using a Bloomberg Terminal, and were available from June 1997 to January 2010 for MGDL and January 1995 to January 2010 for CLIS.

As illustrated in the graph, CLIS and MGDL appear to follow a similar general pattern as the TA100 of a general increasing time trend with a sharp decline during the 2008 global recession, with the correlations of .83 and .72 for CLIS:TA100 and MGDL:TA100, respectively. Additionally,\(^6\)

correlation between CLIS and MGDL daily closing prices is high, at .89, which is common among stocks operating within the same sector.

Following the approach in Section 3.2, prior to estimating GARCH models, I first calculate the first difference of the logged daily closing prices for both MGDL and CLIS. Figure 5 and Figure 6 illustrate the raw closing prices, the first differences of closing prices, and then the first difference of the logged prices. As the tables illustrates, the first difference of the logged prices for both equities appear to exhibit desirable stationary processes.

[INSERT FIGURE 5 HERE]

[INSERT FIGURE 6 HERE]

In Table 5, I replicate Table 3, except use the logged first difference of MGDL instead of the TA100. In Model 1, Model 2, and Model 6, none of the variables reflecting the number of conflictual events targeted towards Israel achieve statistical significance, which is consistent with the findings in Table 3 and Table 4. However, in Model 3, Model 4, and Model 5, the variables reflecting the change in the number of conflictual events targeted towards Israel today relative to the moving average of events across the past two and four weeks does achieve moderate levels of statistical significance. The positive coefficients of the statistically significant variables indicates that increase in the level of conflict relative to the moving averages tends to increase variance in MGDL returns. This means that we are unable to reject the null hypothesis – that investors do not respond to variation in the level of conflictual acts against Israel – with a high degree of confidence.

Table 6 replicates Table 5, but analyzes the first difference of logged returns for CLIS instead of MGDL. Across Model 1 through Model 6, the variables reflecting the change in conflictual events relative to the moving averages are consistently significant, with four of the five conflict variables generating p-values < .01. The robustness of these findings across the three different lengths of moving averages (i.e. one, two, and four weeks) used to calculate $\Delta_{\text{one\_week}}$, $\Delta_{\text{two\_week}}$, and $\Delta_{\text{four\_week}}$ allow us to confidently reject the null of Hypothesis 1. As in Table 5, the coefficients on the statistically significant variables are all positive, indicating that increases in conflict levels tend to increase variance of equity returns.

Although these results clearly suggest that investors in CLIS stock do meaningfully responding to variation in the level of attacks against Israel when making trading decisions, the likelihood, AIC, and BIC scores indicate that accounting for political conflict may not actually increase model fit. In Model 2, Model 4, and Model 6, $\Delta_{\text{one\_week}}$, $\Delta_{\text{two\_week}}$, and $\Delta_{\text{four\_week}}$ all generate highly
significant p-values, but the AIC and BIC scores are still considerably higher than for Model 7, which omits all political conflict variables. This indicates that according to AIC and BIC measures, Model 7 actually provides a better fit of the data than Model 1 through Model 7. This tempers our interpretation of the statistical significance of the $\Delta_\text{one\_week}$, $\Delta_\text{two\_week}$, and $\Delta_\text{four\_week}$ variables.

Although any number of factors could explain why CLIS stock appears to respond more strongly to variation in conflict levels than MGDL, one potential explanatory factor may be differences in the distribution of ownership of MGDL and CLIS stock. Theoretically, Aggarwal and Rao (2005) find that as the percentage of stock owned by institutional investors increases, the variance in returns of the stock tends to decrease. This is because institutional investors tend to have longer-term horizons than individual investors and pay less attention to day-to-day events. The 69.3% of MGDL owned by Assicurazioni Generali S.p.A. are not actively traded, meaning that variation in MGDL is driven exclusively by trading among the remaining 30% of shares. Thus, if all active traders of CLIS and MGDL responded similarly to conflictual events targeting Israel, the effect on CLIS shares would be greater.

7. Conclusion

On November 13, 2012, the Haaretz Daily Newspaper (the most prominent Israeli newspaper printed in English) reported that the “Tel Aviv Stock exchange ended lower again on Monday amid violence in Gaza. TA-100 dropped .6% to 1,061.38”, attributing the decline on mortar shells that fell on Israeli settlements in the Golan Heights.7 Only six days later on November 19, the TA-100 rose 1.1%, prompting the same newspaper to report, “TASE ignores war, cheers low CPI.” In that article, Haaretz reports:

“Israel was bombarded by a barrage of rockets Sunday as Operation Pillar of Defense entered its fifth day and Prime Minister Benjamin Netanyahu told ministers at the weekly cabinet meeting that Israel was prepared to significantly expand its operation in the Gaza Strip. Nevertheless, Hadar Oshart, head of the equities trading desk at Deutsche Bank in Israel, said foreigners had not been deterred by the fighting. ‘The reaction of foreigners has, all told, been restrained. The sense is that the impact of

---

7http://www.haaretz.com/business/border-tensions-worry-tase.premium-1.477322
the operation for now is limited and doesn’t demand any reassessment by overseas investors of their investments or positions.”\footnote{http://www.haaretz.com/business/market-report-tase-ignores-war-cheers-on-low-cpi-1.478936}

These two headlines underscore the central question of this paper: does the TA-100 meaningfully respond to variation in the level of violence target at Israel?

As of November 2012, the TA-100 had an approximate market cap of $114.5 billion USD. If we are to believe the November 13 Haaretz article, then missile attacks against Israel cost TA-100 investors approximately $87 million USD. Consider the implications of this purported causal argument – that the TA-100 meaningfully responds to violence against Israel – actually being supported by the data. Politically, this would drastically increase the costs of fighting against Hamas and other opponents of Israeli, likely encouraging more extreme measures by Israeli politicians to prevent future attacks. Economically, this would likely discourage investment in Israel given the vulnerability of equities to relatively common missile attacks.

This paper is the first attempt to provide a rigorous and objective analysis of the extent to which the TA-100 responds to violence against Israel. To achieve this, I utilize the GDELT event data data set, which allows me to calculate daily counts reflecting the intensity of attacks committed against Israel. Using this data, I follow common practice and perform a series of multivariate \textit{GARCH} models to test for the extent to which variance in TA-100 returns is explained by variation in the level of intensity of violent events committed against Israel. I find that on average, variance in returns of TA-100 taken as a whole are not significantly driven by levels of violence committed against Israel.

Additionally, I perform similar analyses on the returns of the two largest insurance companies on the TA100, Migdal Insurance and Financial Holdings Ltd. (MGDL) and Clal Insurance Enterprises Holdings Ltd. (CLIS). Variables reflecting changes in the level of conflictual attacks against Israel achieve inconsistent, moderate statistical significance explaining variance in MGDL return, and strong and consistent significance when modeling CLIS returns. This strongly suggests that while the TA100 may not meaningfully respond to variation in attacks against Israel, specific companies that comprise the TA100 do.

The findings in this paper warrant two caveats as well as three logical extensions. In terms of caveats, when working with finely grained political event data and financial data, a number of aggregation choices must be made. Although I draw on both theory and the extant literature when...
making aggregation choices in this chapter, it is feasible the different aggregations may have led to different empirical findings. For example, working with weekly level averages (as opposed to daily level data), may alter findings. Additionally, it is possible that the timing of violent attacks against Israel matters. If this is true, then both the November 13 and November 19 Haaretz articles may actually be correct: investors may have actually responded to the initial attacks on November 13, but by November 19, investors may have already priced in the elevated level of violence and therefore ignored the ongoing attacks. Taken together, the empirical empirical findings along with the two caveats suggest that it i that some equities respond to some operationalizations of violence some of the time.

The caveats above suggest at least three useful extensions. First, analyses may be re-run using different temporal aggregations. Although more coarse aggregations (i.e. weekly) may yield meaningful results, I believe that more interesting findings would result from finer-grained temporal analyses. As event data collection becomes increasingly fine-grained, it may be feasible in the near future to obtain data at the hourly or minute level. This data, leverage with second-to-second financial tick data, could allow researchers to test for more immediate effects of political conflict on equity markets. Second, in Section 6, I demonstrate that the same political attacks have different effects on returns of two different specific equities. Repeating similar empirical tests on the other 98 companies that comprise the TA100 could provide more comprehensive insight into the types of companies that then to be affected by political conflict. Third, further analyses of the effects of political violence may vary cross sectionally (i.e. between companies), but as the second caveat suggests, they also may vary over time or by the type of conflictual act. Researchers may be test for changing effects over time testing for the effects of initial conflictual events separately from subsequent violence that follows. Additionally, it would be feasible to test whether certain types of conflictual events have different substantive effects by disaggregating “material conflict” events into sub-categories, such as attacks targeting specific political leaders or those more indiscriminately targeted as unaffiliated civilians.

Hopefully, the research design and results in this chapter can serve as a foundation for future research to test for more nuanced relationships between political violence and equity markets returns.


8. Appendix

Table 1. Correlation Matrix of Counts

<table>
<thead>
<tr>
<th></th>
<th>mater_conf</th>
<th>mater_coop</th>
<th>verb_conf</th>
<th>verb_coop</th>
</tr>
</thead>
<tbody>
<tr>
<td>mater_conf</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mater_coop</td>
<td>.66</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>verb_conf</td>
<td>.78</td>
<td>.68</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>verb_coop</td>
<td>.73</td>
<td>.81</td>
<td>.82</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 2. Correlation Matrix of material conflict MAs

<table>
<thead>
<tr>
<th></th>
<th>1_week_MA</th>
<th>2_week_MA</th>
<th>4_week_MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_week_MA</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2_week_MA</td>
<td>.86</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>4_week_MA</td>
<td>.71</td>
<td>.86</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 3. GARCH models of daily TA100 with DJIA control

<table>
<thead>
<tr>
<th>Mean Equation</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>-.020</td>
<td>-.019</td>
<td>-.020</td>
<td>-.019</td>
<td>-.020</td>
<td>-.020</td>
<td>-.020</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.013)</td>
<td>(.014)</td>
<td>(.013)</td>
<td>(.014)</td>
<td>(.013)</td>
<td>(.014)</td>
</tr>
<tr>
<td>inflation</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
</tr>
<tr>
<td>crude</td>
<td>.021**</td>
<td>.022**</td>
<td>.021**</td>
<td>.021**</td>
<td>.021**</td>
<td>.021**</td>
<td>.022**</td>
</tr>
<tr>
<td></td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.010)</td>
</tr>
<tr>
<td>DJIA</td>
<td>.182***</td>
<td>.183***</td>
<td>.182***</td>
<td>.184***</td>
<td>.182***</td>
<td>.183***</td>
<td>.182***</td>
</tr>
<tr>
<td></td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.019)</td>
</tr>
<tr>
<td>Variance Equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>.245***</td>
<td>.247***</td>
<td>.244***</td>
<td>.247***</td>
<td>.244***</td>
<td>.246***</td>
<td>.244**</td>
</tr>
<tr>
<td></td>
<td>(.036)</td>
<td>(.036)</td>
<td>(.036)</td>
<td>(.037)</td>
<td>(.037)</td>
<td>(.037)</td>
<td>(.036)</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>.714***</td>
<td>.717***</td>
<td>.714***</td>
<td>.716***</td>
<td>.714***</td>
<td>.716***</td>
<td>.716***</td>
</tr>
<tr>
<td></td>
<td>(.032)</td>
<td>(.032)</td>
<td>(.032)</td>
<td>(.032)</td>
<td>(.032)</td>
<td>(.033)</td>
<td>(.035)</td>
</tr>
<tr>
<td>inflation</td>
<td>.626</td>
<td>.381</td>
<td>.636</td>
<td>.459</td>
<td>.641</td>
<td>.556</td>
<td>.702</td>
</tr>
<tr>
<td></td>
<td>(25.58 )</td>
<td>(46.97 )</td>
<td>(24.81 )</td>
<td>(45.70 )</td>
<td>(24.42 )</td>
<td>(45.75 )</td>
<td>(40.92 )</td>
</tr>
<tr>
<td>crude</td>
<td>-.204</td>
<td>4.95</td>
<td>-.745</td>
<td>4.15</td>
<td>-1.28</td>
<td>3.57</td>
<td>.534</td>
</tr>
<tr>
<td></td>
<td>(15.01 )</td>
<td>(15.57 )</td>
<td>(15.09 )</td>
<td>(15.34 )</td>
<td>(14.99 )</td>
<td>(14.99 )</td>
<td>(13.74 )</td>
</tr>
<tr>
<td>DJIA</td>
<td>-139.99***</td>
<td>-136.17***</td>
<td>-140.17***</td>
<td>-133.30***</td>
<td>-140.10***</td>
<td>-132.185***</td>
<td>-127.56***</td>
</tr>
<tr>
<td></td>
<td>(42.28 )</td>
<td>(42.91 )</td>
<td>(41.88 )</td>
<td>(41.51 )</td>
<td>(41.81 )</td>
<td>(41.11 )</td>
<td>(38.80 )</td>
</tr>
<tr>
<td>dt_1_mater_conf</td>
<td>-5.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_one_mater_conf</td>
<td>-5.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.281)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_two_mater_conf</td>
<td></td>
<td>-5.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_two_mater_conf</td>
<td></td>
<td>-5.25</td>
<td>-1.177</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_four_mater_conf</td>
<td></td>
<td></td>
<td></td>
<td>-5.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_four_mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-8.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.800)</td>
<td></td>
<td></td>
<td></td>
<td>(0.975)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.06 )</td>
<td>(1.509 )</td>
<td>(1.05 )</td>
<td>(1.58 )</td>
<td>(1.05 )</td>
<td>(1.44 )</td>
<td>(1.37 )</td>
</tr>
<tr>
<td>N</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>8802</td>
<td>8799</td>
<td>8801</td>
<td>8799</td>
<td>8801</td>
<td>8798</td>
<td>8798</td>
</tr>
<tr>
<td>AIC</td>
<td>7.83</td>
<td>5.84</td>
<td>7.83</td>
<td>5.84</td>
<td>7.83</td>
<td>5.84</td>
<td>3.84</td>
</tr>
<tr>
<td>BIC</td>
<td>85.58</td>
<td>77.60</td>
<td>85.56</td>
<td>77.60</td>
<td>85.56</td>
<td>77.60</td>
<td>69.62</td>
</tr>
</tbody>
</table>

***, **, *: 1%, 5%, and 10% level. Coefficients with standard errors in (). Distribution = Student’s t
Table 4. GARCH models of daily TA100 without DJIA control

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Equation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>-.027**</td>
<td>-.027**</td>
<td>-.027*</td>
<td>-.027*</td>
<td>-.027*</td>
<td>-.027*</td>
<td>-.027*</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.013)</td>
<td>(.014)</td>
<td>(.014)</td>
<td>(.014)</td>
<td>(.014)</td>
<td>(.014)</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
<td>(.006)</td>
</tr>
<tr>
<td>crude</td>
<td>.033***</td>
<td>.033***</td>
<td>.034***</td>
<td>.034***</td>
<td>.034***</td>
<td>.034***</td>
<td>.034***</td>
</tr>
<tr>
<td></td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
</tr>
<tr>
<td><strong>Variance Equation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>.264***</td>
<td>.265***</td>
<td>.267***</td>
<td>.266***</td>
<td>.267***</td>
<td>.267***</td>
<td>.267***</td>
</tr>
<tr>
<td></td>
<td>(.038)</td>
<td>(.038)</td>
<td>(.039)</td>
<td>(.039)</td>
<td>(.039)</td>
<td>(.039)</td>
<td>(.039)</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>.682***</td>
<td>.677***</td>
<td>.665***</td>
<td>.669***</td>
<td>.669***</td>
<td>.669***</td>
<td>.669***</td>
</tr>
<tr>
<td></td>
<td>(.043)</td>
<td>(.046)</td>
<td>(.053)</td>
<td>(.051)</td>
<td>(.052)</td>
<td>(.053)</td>
<td>(.056)</td>
</tr>
<tr>
<td>inflation</td>
<td>1.045</td>
<td>.984</td>
<td>.897</td>
<td>.975</td>
<td>.958</td>
<td>.904</td>
<td>.925</td>
</tr>
<tr>
<td></td>
<td>(15.207)</td>
<td>(13.76)</td>
<td>(11.03)</td>
<td>(11.19)</td>
<td>(11.48)</td>
<td>(10.38)</td>
<td>(10.26)</td>
</tr>
<tr>
<td>dt_1_mater_conf</td>
<td>-.570</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.307)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_one_mater_conf</td>
<td>1.659*</td>
<td>1.458*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.936)</td>
<td>(.784)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_two_mater_conf</td>
<td>.635</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.792)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_two_mater_conf</td>
<td>.947</td>
<td>.990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.786)</td>
<td>(.818)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_four_mater_conf</td>
<td>.524</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_four_mater_conf</td>
<td>.197</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.819)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(1.967)</td>
<td>(1.941)</td>
<td>(1.931)</td>
<td>(1.890)</td>
<td>(1.878)</td>
<td>(1.938)</td>
</tr>
<tr>
<td>N</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
<td>2923</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>8748</td>
<td>8748</td>
<td>8748</td>
<td>8747</td>
<td>8747</td>
<td>8747</td>
<td>8746</td>
</tr>
<tr>
<td>AIC</td>
<td>7.84</td>
<td>5.85</td>
<td>7.84</td>
<td>5.85</td>
<td>7.84</td>
<td>5.85</td>
<td>3.85</td>
</tr>
<tr>
<td>BIC</td>
<td>85.59</td>
<td>77.61</td>
<td>85.59</td>
<td>77.61</td>
<td>85.59</td>
<td>77.61</td>
<td>69.63</td>
</tr>
<tr>
<td>df</td>
<td>6.52</td>
<td>6.51</td>
<td>6.48</td>
<td>6.48</td>
<td>6.48</td>
<td>6.49</td>
<td>6.45</td>
</tr>
</tbody>
</table>

***, **, *: 1%, 5%, and 10% level. Coefficients with standard errors in (). Distribution = Student’s t distribution.
Table 5. GARCH models of daily MGDL with DJIA control

<table>
<thead>
<tr>
<th>Mean Equation</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>.021</td>
<td>.021</td>
<td>.021</td>
<td>.021</td>
<td>.021</td>
<td>.021</td>
<td>.021</td>
</tr>
<tr>
<td></td>
<td>(.022)</td>
<td>(.022)</td>
<td>(.022)</td>
<td>(.022)</td>
<td>(.022)</td>
<td>(.022)</td>
<td>(.022)</td>
</tr>
<tr>
<td>inflation</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.011)</td>
<td>(.012)</td>
<td>(.012)</td>
<td>(.012)</td>
<td>(.011)</td>
<td>(.011)</td>
</tr>
<tr>
<td>crude</td>
<td>.035*</td>
<td>.035*</td>
<td>.037*</td>
<td>.036*</td>
<td>.037*</td>
<td>.035*</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.021)</td>
</tr>
<tr>
<td>DJIA</td>
<td>.358***</td>
<td>.357***</td>
<td>.361***</td>
<td>.362***</td>
<td>.364***</td>
<td>.368***</td>
<td>.357***</td>
</tr>
<tr>
<td></td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.042)</td>
</tr>
<tr>
<td>Variance Equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>.051***</td>
<td>.054***</td>
<td>.047***</td>
<td>.052***</td>
<td>.050***</td>
<td>.054***</td>
<td>.055***</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.012)</td>
<td>(.011)</td>
<td>(.012)</td>
<td>(.012)</td>
<td>(.012)</td>
<td>(.012)</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>.920***</td>
<td>.915***</td>
<td>.932***</td>
<td>.921***</td>
<td>.907***</td>
<td>.913***</td>
<td>.912***</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.019)</td>
<td>(.015)</td>
<td>(.017)</td>
<td>(.016)</td>
<td>(.019)</td>
<td>(.020)</td>
</tr>
<tr>
<td>inflation</td>
<td>2.966</td>
<td>2.56</td>
<td>4.611</td>
<td>3.509*</td>
<td>3.350</td>
<td>2.496</td>
<td>2.595</td>
</tr>
<tr>
<td>crude</td>
<td>20.008***</td>
<td>19.612*</td>
<td>17.608***</td>
<td>-3.509*</td>
<td>3.350</td>
<td>2.496</td>
<td>2.595</td>
</tr>
<tr>
<td>DJIA</td>
<td>39.529***</td>
<td>37.301***</td>
<td>60.238***</td>
<td>54.745***</td>
<td>56.636***</td>
<td>37.820***</td>
<td>33.772***</td>
</tr>
<tr>
<td>dt_1_mater_conf</td>
<td>1.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.136)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_one_mater_conf</td>
<td>-1.972</td>
<td>1.274*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.026)</td>
<td>(.733)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_two_mater_conf</td>
<td>2.038**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.960)</td>
<td>(.733)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt_c_two_mater_conf</td>
<td>-1.567</td>
<td>1.532**</td>
<td>1.651</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.549)</td>
<td>(.741)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.392)</td>
<td>(.401)</td>
<td>(.418)</td>
<td>(.432)</td>
<td>(.382)</td>
<td>(.395)</td>
<td>(.398)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>4553</td>
<td>4553</td>
<td>4553</td>
<td>4554</td>
<td>4554</td>
<td>4554</td>
<td>4554</td>
</tr>
<tr>
<td>AIC</td>
<td>9.15</td>
<td>7.15</td>
<td>9.15</td>
<td>7.15</td>
<td>9.15</td>
<td>7.15</td>
<td>5.15</td>
</tr>
<tr>
<td>BIC</td>
<td>81.75</td>
<td>74.16</td>
<td>81.75</td>
<td>74.16</td>
<td>81.75</td>
<td>74.16</td>
<td>66.58</td>
</tr>
</tbody>
</table>

***, **, *: 1%, 5%, and 10% level. Coefficients with standard errors in (). Distribution = Student’s t.
### Table 6. GARCH models of daily CLIS with DJIA control

<table>
<thead>
<tr>
<th>Mean Equation</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>.074***</td>
<td>.074***</td>
<td>.075***</td>
<td>.075***</td>
<td>.074***</td>
<td>.075***</td>
<td>.073***</td>
</tr>
<tr>
<td>inflation</td>
<td>.027**</td>
<td>.027**</td>
<td>.026**</td>
<td>.026**</td>
<td>.026**</td>
<td>.026**</td>
<td>.027**</td>
</tr>
<tr>
<td>crude</td>
<td>.032</td>
<td>.033</td>
<td>.033</td>
<td>.033</td>
<td>.033</td>
<td>.033</td>
<td>.031</td>
</tr>
<tr>
<td>DJIA</td>
<td>.302***</td>
<td>.302***</td>
<td>.301***</td>
<td>.302***</td>
<td>.302***</td>
<td>.302***</td>
<td>.304***</td>
</tr>
<tr>
<td>Variance Equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>.054***</td>
<td>.055***</td>
<td>.053***</td>
<td>.055***</td>
<td>.053***</td>
<td>.055***</td>
<td>.056***</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>.917***</td>
<td>.916***</td>
<td>.917***</td>
<td>.914***</td>
<td>.916***</td>
<td>.914***</td>
<td>.918***</td>
</tr>
<tr>
<td>inflation</td>
<td>-.631</td>
<td>-.645</td>
<td>-.476</td>
<td>-.476</td>
<td>-.476</td>
<td>-.476</td>
<td>-.256</td>
</tr>
<tr>
<td>DJIA</td>
<td>.302***</td>
<td>49.913***</td>
<td>51.988***</td>
<td>50.237***</td>
<td>51.465***</td>
<td>49.831***</td>
<td>45.351***</td>
</tr>
<tr>
<td>dt, l mater_conf</td>
<td>-.079</td>
<td>(1.083)</td>
<td>1.682**</td>
<td>1.643***</td>
<td>(7.91)</td>
<td>(.533)</td>
<td></td>
</tr>
<tr>
<td>dt, c, one mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt, two mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt, c, two mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt, four mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dt, c, four mater_conf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2337</td>
<td>2337</td>
<td>2337</td>
<td>2337</td>
<td>2337</td>
<td>2337</td>
<td>2337</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>5291</td>
<td>5291</td>
<td>5293</td>
<td>5292</td>
<td>5293</td>
<td>5292</td>
<td>5288</td>
</tr>
<tr>
<td>BIC</td>
<td>81.45</td>
<td>73.86</td>
<td>81.45</td>
<td>73.86</td>
<td>81.45</td>
<td>73.86</td>
<td>66.28</td>
</tr>
<tr>
<td>df</td>
<td>5.219</td>
<td>5.220</td>
<td>5.251</td>
<td>5.249</td>
<td>5.250</td>
<td>5.250</td>
<td>5.193</td>
</tr>
</tbody>
</table>

***, **, *: 1%, 5%, and 10% level. Coefficients with standard errors in (). Distribution = Student’s t.
Figure 1. Total number of material conflict events, daily from 1992 to 2012.
Figure 2. The number of material conflict events with the time trend removed, daily from 1992 to 2012
Figure 3. Raw, first-differenced, and logged first-differenced of TA100 prices
Figure 4. Comparison of closing TA100, CLIS, and MGDL prices
Figure 5. Raw, first-differenced, and logged first-differenced MGDL prices
Figure 6. Raw, first-differenced, and logged first-differenced CLIS prices
CHAPTER 3. EFFECTS OF DOMESTIC CONFLICT ON INTERSTATE CONFLICT: AN EVENT DATA ANALYSIS OF MONTHLY LEVEL ONSET AND INTENSITY

1. INTRODUCTION

That domestic conflicts can affect interstate conflict is clear. Consider a few examples. In the first months of 2001, fighting between Burmese government troops and domestic rebels intensified, with much of the violence occurring near the border between Burma and Thailand. During this same period, interstate conflict between Burmese and Thai military forces reached their highest levels in decades, as troops from both sides clashed over control of strategic locations near the border and engaged in shelling and small arms fire resulting in scores of civilian deaths. In this case, both the presence and intensity of the domestic conflict in Burma that spread across the border into Thailand led directly to the interstate conflict events that transpired between official military personnel of each state. More recent events in Libya provide another example. After more than a month of unrest in the region, the first substantial demonstrations in Libya occurred on February 15, 2011, leading to approximately 15 deaths by February 17, 2011. Within a week, anti-Gaddafi rebels had mobilized and the country was engaged in severe revolutionary fighting. Seeking to take advantage of a window of opportunity that the domestic conflict provided to attempt to help extricate Gaddafi from power, United States and British forces began an international campaign against Gaddafi by firing over 100 Tomahawk cruise missiles against Libya’s key air defense installations on March 19, 2011. Again, the existence of the domestic conflict influenced an interstate conflict, albeit through different mechanisms than through the previous example.

The central goal of this chapter is to provide a thorough and nuanced analysis of the effects of both the onset and intensity of domestic conflict on interstate conflict—an area of research that is surprisingly underdeveloped in the extant literature. One potential reason for the lack of relevant empirical analyses may be due to the coarseness of existing data on both domestic and interstate

conflicts. In terms of domestic conflict, UCDP/PRIO (see Gleditsch et al. (2002)) and Correlates of War (COW) datasets (see Sarkees and Wayman (2010)) predominate in the literature. However, scholars primarily use these datasets to provide a binary measure of whether or not conflict/war occurred in a given state-year. According to the UCDP/PRIO dataset, Burma experienced domestic conflict every year from 1997 to 2011, but according to COW, domestic conflict did not reach sufficiently high thresholds to become a “war”, so every year in that period receives a “0” on the dichotomous scale. Thus, it is impossible to test for the effects of variation in the intensity of domestic conflict in Burma on the onset or intensity of interstate conflict, even though real world examples suggest such relationships might exist. Furthermore, studies analyzing interstate conflict also tend to rely on dichotomous, annual level measures such as militarized interstate dispute (MIDs see Ghosn, Palmer and Bremer (2004)) or COW measures. The dichotomized and annual-level nature of these measures inhibits the ability of researchers to test for more subtle variations in levels of intensity during and between years, yet this is what we expect to see in Libya, as NATO forces varied the intensity of bombings based on the success of the rebels.

These shortcomings in existing data all suggest that temporally nuanced measures of both domestic and interstate conflict are needed in order to appropriately test for the range of potential effects that domestic conflict may have on interstate conflict. This chapter will addresses this problem by generating monthly level measures reflecting the number of conflictual events at both the state-month (for domestic events) and the dyad-month (for interstate events) level based on the GDELT dataset. Using this data, I perform numerous empirical tests for the effects of domestic conflict onset and intensity on the onset and intensity of interstate conflict across a range of operationalizations of “onset” and “intensity” at the monthly level. To the best of my knowledge, this chapter provides the first empirical test for the effects of domestic conflict intensity on both the likelihood of interstate conflict onset and the intensity of ongoing conflicts.

This chapter proceeds in four sections: first, I provide a brief review of relevant literature and from that literature develop my testable hypotheses; second, I explain my use of event data; third, I outline my variable operationalization and research design; fourth, I provide empirical models and results. Lastly, I conclude with a discussion of future extensions.

4These datasets also provide estimates of the total number of battle fatalities, but those figures reflect the duration of the conflict and cannot be disaggregated to smaller temporal units. As a consequence, in many cases, it is impossible to determine even annual level variation in conflict intensity.
2. Building Hypotheses from the Literature

Although the examples of Burma and Libya are recent, similar cases are pervasive throughout history. For example, over a century ago in 1911, the Russian Bolsheviks were entrenched in civil war against reactionists and lacked sufficient resources to defend Russia’s external border. Aware of this weakness, Japan attacked northern Siberia with 70,000 troops in an attempt to acquire Russian territory (see Humphrey (1995)). Despite both the historical occurrence of domestic conflicts affecting interstate conflicts as well as studies calling for more comprehensive analyses of potential relationships (see Sambanis (2002), and Chiozza, Gleditsch and Goemans (2006)), this topic has received relatively little attention in the literature.\(^5\) Since theorizing and testing for a full range of potential relationships between the onset and intensity of domestic conflict on the onset and intensity of interstate conflict exceeds the scope of this chapter, I build four hypotheses derived from the related conflict literature.

The most relevant extant empirical studies tend to focus only on the effects of a domestic conflict onset on the likelihood of an interstate conflict onset. For example, Davies (2002) finds that certain contentious domestic events such as protests or riots may increase the likelihood of initiating a MID; Walt (1996) argues through an opportunism framework that states undergoing domestic conflict make more attractive targets for interstate attacks; Trumbore (2003) finds that domestic ethno-political rebellion may increase the likelihood of MID initiation; a number of scholars have illustrated that domestic conflicts increase the likelihood of third-party interventions, which may or may not be welcome by the host government (Elbadawi and Sambanis (2002), Gleditsch (2007), Regan (2000)); and interstate conflict that can result from foreign support of rebels (Schultz (2010)).

Gleditsch, Salehyan and Schultz (2008) provide a more comprehensive argument that onsets of domestic conflict increase the likelihood of interstate conflict by outlining the five main mechanisms through which this occurs:\(^6\)

- **Opportunism:** Civil wars and insurgencies expose and exacerbate weaknesses in a state’s military capabilities and divert resources away from defenses against foreign enemies, thereby increasing the expected utility of attacking a state with domestic conflict.

\(^5\)This is even more surprising given the large number – from which I select a small number of of “diversionary war” studies analyzing the effects domestic economic (inflation (Mitchell and Prins (2004)), inflation and unemployment (Fordham (2002)), GDP (Bennett and Nordstrom (2000)) and political (regime type, leader approval ratings (Ostrom and Job (1986)), election cycles (Smith (1996))) conditions as well as the presence of a number of studies addressing ways in which interstate conflicts can affect domestic conflicts (Thyne (2006), Akcinaroğlu and Raziszewski (2005)).

\(^6\)Gleditsch, Salehyan and Schultz (2008) treat opportunism and diversion as one conceptually unique category, I divide them because I believe that they are conceptually unique concepts.
Diversion: Faced with domestic conflict, a leader may intentionally seek out interstate conflicts in order to divert attention away from domestic issues and generate a rally around the flag effect.

Intervention: States can intervene either on the side of the government or the side of the rebels during a domestic conflict.

Externalization: During domestic conflicts, rebels and government forces may cross interstate borders in order to find safe havens or more favorable territory from which to launch attacks.

Spillover effects: Domestic conflicts often lead to enhanced troop movements near borders, cross-border refugee flows, and regional economic disruptions that can all increase the likelihood of interstate conflict.

Overall, the relevant extant literature—including all five of Gleditsch, Salehyan and Schultz (2008)’s mechanisms—is in agreement that the occurrence of domestic conflicts increases the likelihood of an interstate dispute onset and they also provide real-world examples and empirical testing to support this proposition. These theoretical arguments and empirical findings lead to my first hypothesis:

**Hypothesis 1:** The likelihood of interstate conflict onset should increase after an onset of domestic conflict in one or both states comprising a dyad.\(^7\)

Although Gleditsch, Salehyan and Schultz (2008) and others provide a clear testable hypothesis for the effects of an onset of domestic conflict on the likelihood of an onset of interstate conflict, their theory and research design do not directly address how fluctuations in the intensity of an ongoing domestic conflict might affect either the likelihood of an onset or variation in intensity of an interstate conflict. Indeed, in many conflict prone countries, such as the case of Burma or the Democratic Republic of the Congo (DRC), levels of domestic conflict are very rarely zero. Despite this, considerable variation tends to exist in the intensity of ongoing domestic conflicts. By focusing exclusively on the effect of domestic conflict onsets, as Gleditsch, Salehyan and Schultz (2008) and others, countries like Burma are necessarily omitted from empirical testing since there is always an ongoing conflict. Therefore, in countries like Burma and the DRC, changes in intensity, rather

---

\(^7\)In all four hypotheses, I focus on the effects of key independent variables occurring in month \(t-1\) and month \(t-2\) on the dependent variable in month \(t\). This ensures that the independent variables is temporally preceding the dependent variable, which dismisses the possibility of the two events – in the case of Hypothesis 1, a domestic conflict onset and an interstate conflict onset – occurring in the same month but in reverse order (i.e. the interstate conflict preceding the domestic conflict onset)
than the occurrence or the onset of domestic conflicts that should matter most. Unfortunately, the literature is sparse with respect to the effects of variation in domestic conflict intensity on interstate conflict. However, following the logic behind Hypothesis 1, we could expect that as the intensity of a domestic conflict increase, the degree of opportunism, diversion, intervention, externalization, and spillover should also increase. This leads to my second testable hypothesis:

**Hypothesis 2:** The likelihood of interstate conflict onset in a dyad should increase after the intensity of ongoing domestic conflict in one or both of the states comprising the dyad increases.

This line of reasoning could also be applicable to changes in the predicted levels of ongoing interstate conflicts, as increases in intensity in domestic conflict could lead to increases in the expected intensity of interstate conflict. Despite a dearth of relevant empirical literature, numerous case studies support the argument that the intensity of a domestic conflict has a positive effect on levels of externalization, spillover, and opportunism. For example, more brutal conflicts tend to have higher levels of externalization and spillover, as violence in Darfur, the DRC, and Rwanda illustrate. An estimated 1.8 million refugees have fled Darfur amidst violence that has led to 300,000 deaths; in the DRC, civil violence since 1996 resulted in an estimated 5.4 million deaths and 3.4 million refugees; and in Rwanda, an estimated 2 million Hutus fled after approximately 800,000 deaths. In many cases—illustrated by Hutu militiamen fleeing to Zaire and other neighboring states—rebels are among the refugees, which has tended to increase interstate conflict. Clearly, these conflicts have broad negative local and regional economic consequences that nearby states would like avoid. Thus, it is logical to assume that externalization and spillover become more pronounced as the severity of domestic conflict increases. Following this line of argument, it is reasonable to suggest that as the intensity of domestic conflict increases, so does the likelihood of an interstate conflict.

**Hypothesis 3:** The intensity of an ongoing interstate conflict should increase after the intensity of ongoing domestic conflicts in one or both of the states comprising the dyad increases.

These first three hypotheses all suggest a positive relationship between domestic conflict and interstate conflict, both in terms of onset and intensity. Despite this, a considerably different strain of logic developed in the bargaining model of war literature suggests alternative relationships. Consider a baseline bargaining model of war in which the probability or winning a potential interstate conflict is a function of both states’ capabilities (Fearon (1995)). If two states in a dyad (State A and State B) both have capabilities totaling 100 units each, the probability that either wins.

---

8The preponderant measure of capabilities is the COW CINC score.
the conflict is approximately 50/50 under the assumption that both states are able to commit all capabilities to a potential interstate conflict. However, as Gleditsch, Salehyan and Schultz (2008) articulates, “Civil wars and insurgencies(...)divert resources away from defense against foreign enemies.” This implies that if State A is engaged in a domestic conflict, it is forced to funnel a portion of the 100 units of total capabilities towards the domestic conflict meaning that its real capabilities to be used in the event of interstate conflict is 100. Consequently, the probability that a domestically conflicted State A wins an interstate conflict with State B is lower than it would be if it were not also fighting in a domestic conflict.

Additionally, this rationale is extendable to cases in which the level of ongoing domestic conflict in both (or either) State A and State B varies after a conflict is already initiated. A number of cases exist to provide a clear illustration of two states simultaneous engaged in domestic and interstate conflict–such as Angola and the DRC, Burma and Thailand, and India and Pakistan. For example, consider a state with 100 units of capabilities is engaged in both an interstate and a domestic conflict. As the percentage of resources that it chooses to dedicate towards one of the ongoing conflicts increases, the amount of resources available to fight the other conflict decreases, meaning that the probability of winning also decreases. Focusing on enduring rivalries–conceptually akin to continuing conflict–Bennett and Nordstrom (2000) posit that by reducing or ending interstate conflict with a rival, a state becomes able to “free up important resources that may be reallocated to the domestic economy.” It is likely that these resources could also be used to reinforce efforts to put down domestic conflict. For example, in 1905, increasing levels of domestic unrest played into Russia’s calculi during negotiations to end the Russo-Japanese war:

“They (Russian negotiators’) country was in the first throes of a slow revolution that they knew to be unstoppable. At best, the revolution could be postponed if they could negotiate a foreign peace [with Japan] that would enable the Tsar’s ministers to deal undistractedly with the war developing in the streets and basements back home.”

Furthermore, in the post WWII era, the likelihood of losing territory, rents, and control over government is higher in modern domestic conflicts than interstate conflicts due largely to evolving international norms, meaning that it is logical to assume that a state would place greater emphasis on winning the domestic conflict. Thus, if domestic intensity increases in both states and causes

9Excerpt taken from Theodore Roosevelt biographer Edmund Morris’ account of negotiations between Russian and Japanese diplomats, overseen by then President Roosevelt.
10See Zacher (2001) for a discussion of how the territorial integrity norm that has increased the rarity of transfers of territory through interstate war; Collier (1999) for an illustration of the negative effects of civil war on domestic
them to funnel resources away from the interstate conflicts, it follows that the intensity of the interstate conflict should decrease. These arguments lead to my final hypothesis:

**Hypothesis 4**: The intensity of an ongoing interstate conflict should decrease after the intensity of domestic conflict increases in one or both states of a dyad that are engaged in domestic conflict.

In the following section, I outline how I outline my research design, focusing primarily on how I use the GDELT event dataset to comprise measures reflecting onset and changes of intensity of both domestic and interstate conflicts.

3. **Research Design**

As previously mentioned, existing literature interested in the effects of domestic conflict on interstate conflicts has been restricted by its use of annual level measures of conflict. In order to allow for sufficiently nuanced analyses to test my four hypotheses, I utilize the dyad-month unit of analysis for all empirical testing. Since the existing conflict datasets (UCDP, COW, MIDs, etc.) are aggregated to the yearly level, I am required to build my own measures reflecting onsets and changes in intensity of both domestic and interstate conflict. To do so, I utilize the GDELT event data dataset. With this data, I construct domestic and interstate conflict variables for over 4 million dyad-month observations for all countries from 1979 to 2012.

3.1. **Constructing Independent Variables.** In order to test my four hypotheses, I need independent variables that reflect both onsets of domestic conflicts as well as variation in the intensity of ongoing domestic conflicts. To build these measures, I first use the GDELT data to calculate how many “domestic material conflict” events occur in each country per month. In order to qualify as “domestic material conflict”, the event must be between two actors whose primary affiliation is with the same country. Additionally, one actor’s secondary affiliation must be with the government, either as a member of government, a member of the armed forces, or a member of the policy. The other actor’s secondary affiliation must be as a rebel, separatist, or insurgent.

3.1.1. **Measures of Domestic Conflict Onset.** At the most basic level, the concept of an “onset” assumes that an event that was not previously occurring suddenly begins. In theory, this should mean that an operationalization of domestic conflict onset should require a period devoid of domestic conflict during which an onset may occur. In reality, however, states with a history of economic sectors and GDP; and Le Billon (2001) for a discussion of domestic rebels’ ability to siphon resource rents from governments.
domestic conflict are rarely entirely devoid of conflictual events. As a result, current intra-state conflict datasets such as COW and UCDP use a cutpoint (1,000 fatalities for COW, 25 fatalities for UCDP) and assume any state with fewer than the cutpoint number of deaths is at peace and a state with more than the cutpoint number of fatalities in a given time period is at conflict. To operationalize civil conflict onset with event data, I follow this cutpoint approach. However, unlike COW and UCDP, which both use a single cutpoint, I test across three cutpoints since there is no single theoretically justified cutpoint, regardless of COW or UCDP procedure. This allows me to conduct robustness checks, which helps ensure that statistically significant findings are not merely a function of a certain cutpoint, but rather are consistent across various cutpoints. In total, I build six binary measures of civil conflict onset - three to reflect domestic conflict onset in only one of the states per dyad-month, and three to reflect onset in both of the states per dyad-month.

- **one_domestic_20** – A “1” if only one of the states in each dyad-month experiences > 20 “domestic material conflict” events in month $t$ and both states experienced fewer than 20 “domestic material conflict” events between the government and rebel groups in month $t - 1$, and “0” otherwise.
- **one_domestic_40** – identical to **one_domestic_20**, except the cutpoint is set at 40 material conflict events.
- **one_domestic_60** – identical to **one_domestic_20**, except the cutpoint is set at 60 material conflict events.
- **both_domestic_20** – A “1” if both of the states in each dyad-month experiences > 20 “domestic material conflict” events in month $t$ and both states experienced < 20 “domestic material conflict” events in month $t - 1$, and “0” otherwise.
- **both_domestic_40** – identical to **both_domestic_20**, except the cutpoint is set at 40 material conflict events.
- **both_domestic_60** – identical to **both_domestic_20**, except the cutpoint is set at 60 material conflict events.

3.1.2. Measures of Domestic Conflict Intensity. Empirically testing for the effects of variation in the severity of domestic conflict on the level of interstate conflict at the monthly level is a difficult task that the extant literature has yet to address. Consequently, I am unable to draw upon existing methodological approaches to operationalize variation in the intensity of domestic conflict intensity. Given the dearth of precedent, I build the following six binary measures reflecting changes
in intensity of ongoing domestic conflicts in an attempt to construct as straightforward measures as possible. As in section 3.1.1, I build variables across the three different cutpoints, which allows me to test for the robustness of findings.

- **one_worse_20** – A “1” If the state in the dyad experienced >20 “domestic material conflict” events in month $t-1$ and more “domestic material conflict” events in month $t$ than in month $t-1$. “0” otherwise.

- **both_worse_20** – A “1” if both states in the dyad experienced >20 “domestic material conflict” events in month $t-1$ and more “domestic material conflict” events in month $t$ than in month $t-1$, or, one state in month $t-1$ experienced >20 “domestic material conflict” events in month $t-1$ and more “domestic material conflict” events in month $t$ than month $t-1$, and the other state experienced $\leq 20$ “domestic material conflict” events in month $t-1$ but $> 20$ “domestic material conflict” in month $t$. “0” otherwise.

- **one_worse_40** – Identical to `emphone_worse_20`, except the cutpoint is set at 40 material conflict events.

- **both_worse_40** – Identical to `emphboth_worse_20`, except the cutpoint is set at 40 material conflict events.

- **one_worse_60** – Identical to `emphone_worse_20`, except the cutpoint is set at 60 material conflict events.

- **both_worse_60** – Identical to `emphboth_worse_20`, except the cutpoint is set at 60 material conflict events.

### 3.2. Constructing Dependent Variables.

To build measures reflecting onsets and variation in intensity of interstate conflicts, I follow a similar process to that used to build domestic conflict measures. First, I construct a measure called “interstate material conflict” for each dyad-month, which reflects the number of material conflict events occurring each month between two actor’s whose primary affiliation is with different states comprising each dyad, and whose secondary continuous affiliation is with the government, military, or police. For example, consider the Thailand-Vietnam dyad. For an event to qualify as an “interstate material conflict” event, one actor’s primary affiliation is required to be Thailand and the other required to be Vietnam, and both of their secondary affiliations must be either government, military, or police. I require these restrictive secondary commands to maintain consistency with COW definitions of interstate conflict (i.e. that it occurs between official state forces). Whereas COW uses a single cutpoint (1,000 battle
fatalities) to qualify as an interstate conflict, I again employ three different cutpoints since there is neither theoretical nor empirical justification to chose a single cutpoint. Using the “interstate material conflict” and the 20, 40, 60 event cutpoints, I build three binary variables reflecting an onset of interstate conflict:

- \textit{interstate}_{20} – A “1” if greater than 20 “interstate material conflict” events occur in month \(t\) and fewer than 20 “interstate_material_conflict” events occured in month \(t - 1\), and “0” otherwise.
- \textit{interstate}_{40} – identical to \textit{interstate}_{20}, except the cutpoint is set at 40 “interstate material conflict” events
- \textit{interstate}_{60} – identical to \textit{interstate}_{20}, except the cutpoint is set at 60 “interstate material conflict” events

To measure the change in intensity of interstate conflict, I build three binary and three continuous variables reflecting the change in intensity of ongoing interstate conflicts. Again, I calculate measures across the three different cutpoints –20, 40, and 60 events – in order to facilitate robust checks in Section 4.3.

- \textit{interstate}_{worse}_{20} – A “1” if the dyad experienced greater than 20 interstate material conflict events in month \(t - 1\) and more interstate material conflict events in month \(t\) than in month \(t - 1\). A “0” otherwise.
- \textit{interstate}_{worse}_{40} – A “1” if the dyad experienced greater than 40 interstate material conflict events in month \(t - 1\) and more interstate material conflict events in month \(t\) than in month \(t - 1\). A “0” otherwise.
- \textit{interstate}_{worse}_{60} – A “1” if the dyad experienced greater than 60 interstate material conflict events in month \(t - 1\) and more interstate material conflict events in month \(t\) than in month \(t - 1\). A “0” otherwise.
- \textit{interstate}_{change}_{20} – The (number of interstate material conflict events in month \(t\)) - (the number of interstate material conflict events in month \(t - 1\)), calculated when month \(t - 1\) experienced greater than 20 interstate material conflict events.
- \textit{interstate}_{change}_{40} – The (number of interstate material conflict events in month \(t\)) - (the number of interstate material conflict events in month \(t - 1\)), calculated when month \(t - 1\) experienced greater than 40 interstate material conflict events.
- **interstate_change_60** – The (number of interstate material conflict events in month $t$) - (the number of interstate material conflict events in month $t - 1$), calculated when month $t - 1$ experienced greater than 60 interstate material conflict events.

3.3. **Control Variables.** In addition to the event data-derived measures of domestic conflict, I follow Russet and Oneal (2001) and employ their eight baseline variables used to explain MID involvement (Table A5.1, p. 316), which are all aggregated at the yearly level:

- Non-Contiguity: 0 reflecting a shared land border or fewer than 150 miles of water separating the nearest borders, and 1 indicating non-contiguous borders (Stinnett et al. (2002)).
- Power Ratio: The ratio of CINC scores between the two states in the dyad, with the lower score serving as the numerator (Singer, Bremer and Stuckey (1972)).
- Minor Powers: A binary measure taking on 1 if neither of the two states in the dyad are considered major powers. In my dataset, all dyads are comprised of minor powers with the exception of those containing either China or Japan.
- Log Distance: COW data reflecting distance between capitals in miles, logged (Stinnett et al. (2002)).
- Democ$_L$: Polity IV data, which reflects the autocracy-democracy score of the lesser democratic state in the dyad on the 21-point, -10 (fully autocratic) to +10 (fully democratic) scale (Marshall and Jaggers (2009)).
- Depend$_L$: Bilateral trade data, calculated to reflect the percentage of both states in the dyads total trade comprised by the dyadic trade. The dyad receives the lower of the two state scores (Barbieri, Keshk and Pollins (2009)).
- IGO: A count of that reflects the number of shared dyadic IGO membership (Pevehouse and Nordstrom (2004)).
- Alliance_membership: An ordinal measure of 0, 1, or 2, reflecting the highest degree of dyadic alliance (Gibler and Sarkees (2004)).

To construct my complete dataset, I first use **EUGene** (see Bennett and Stam (2000)) to build a dyad-year time-series cross section dataset with the eight annual level controls from Russet and Oneal (2001) from 1979 to 2004 for all possible dyads. Next, since my analysis focuses on monthly level variation, I must convert these yearly scores to the monthly level. To do so, I assume that the value of the control variables in each month is the same as their yearly total. For example, if

---

1. The control variables are only collected through 2004.
trade the $Depend_L$ score for the China-Japan dyad in 2004 is .42, I set the $Depend_L$ score for all 12 months in the Japan-China dyad at .42. Lastly, I merge the event data derived domestic/interstate conflict onset/intensity variables. The final dataset is a time-series cross section dataset at the dyad-month level with over 4 million observations. With this data, I use logistic and ordinary least squares (OLS) regression to test all four hypotheses.

I perform all empirical tests twice, once lagging the domestic conflict variables by one month, and then a second time lagging these variables two months. This is useful for two reasons. First, using lags of my domestic conflict variables helps ensure that the domestic conflict onsets actually precede the interstate conflicts. If I were to use unlagged measures, a domestic conflict onset could occur after an interstate conflict onset but within the same calendar month, meaning that it would have been impossible for onset of domestic conflict to have “caused” the onset of interstate conflict. Second, the four hypotheses suggest that a relationship between domestic conflict and interstate conflict may exist, but do not suggest how quickly the relationships unfold. Thus, testing for a one- and two-month lags allow the empirical models to capture effects that occur quickly as well as ones that take longer (i.e up to two months) to unfold.

4. Empirical Tests

4.1. Tests of the effects of domestic conflict onset on the likelihood of interstate conflict onset.

My first empirical test addresses Hypothesis 1, analyzing whether an onset of domestic conflict affects the likelihood of an onset of interstate conflict.

[INSERT TABLE 1 HERE]

In Table 1, I test for the effects of onsets of domestic conflict in month $t - 1$ on the likelihood of an onset of interstate conflict in month $t$. Since I am interested in modeling onsets, both states in the dyad must have experienced an absence of domestic conflict in the previous month. Additionally, the dyad must not have experienced an interstate conflict in the previous month. I run three separate logistic regressions, which all account for the same set of Russet and Oneal (2001) control variables, but utilize difference cutpoints – 20, 40, and 60 conflictual events– needed to qualify as a domestic and interstate conflict. To further explain the use of cutpoints, each model in Table 1 corresponds to a single logistic regression. In Model 1, the dependent variable is $interstate_{20}$ and the key independent variables reflecting domestic conflict onsets are $one_{domestic}_{20}$ and
*both\_domestic\_20*, meaning that Model 1 assumes that both a state with fewer than 20 “domestic material conflict” events and a dyad with fewer than 20 “interstate material conflict” events are at peace, but once the 20 material conflict cutpoint is passed, the state and dyad are treated as being at conflict.\textsuperscript{12}

In all three models, an onset of domestic conflict in one or both of the two states comprising each dyad in month $t - 1$ increases the likelihood of an interstate conflict onset in month $t$ at statistically significant levels. Additionally, the relative size of coefficients in Table 1 suggest that dyads in which *both* states experiencing an onset of domestic conflict in month $t - 1$ have a higher likelihood of an interstate conflict onset in month $t$ than dyads in which only *one* state experiences a domestic conflict onset.

To facilitate interpretation of the empirical findings in Table 1, I calculate the marginal effects of each of the key domestic conflict variables. To do so, I build three “average” dyads according to the eight control variables (by taking the mean of the continuous measures and the mode of the binary measures) – one with no onsets of domestic conflict in month $t - 1$, one with one onset of domestic conflict in $t-1$, and one with two onsets of domestic conflict in month $t - 1$. Then, I use the mean and standard errors of coefficient estimates from the logistic regressions to calculate the mean predicted probabilities with 95% confidence intervals reflecting the likelihood of an onset of interstate conflict onset. I repeat this across the three cutpoints – 20, 40, and 60 – to serve as a robustness check.\textsuperscript{13}

[INSERT FIGURE 1 HERE]

Figure 1 is divided into three columns, cutpoint\_20, cutpoint\_40, and cutpoint\_60, which correspond to Model 1, Model 2, and Model 3 of Table 1, respectively. Along the x-axis in each column are labels “neither”, “one”, and “both”, which reflect whether neither, one, or both of the states comprising each dyad experienced an onset of domestic conflict in month $t - 1$. The y-axis reflects the predicted probability of an onset of interstate conflict occurring in month $t$. The box plots reflect the mean and 95% confidence intervals of the estimates. Though difficult to visually recognize due to the small predicted probabilities, the 95% confidence interval do not overlap from the “neither” (i.e. neither of the two states in the dyad experience an onset of domestic conflict in

\textsuperscript{12}I perform additional robustness check would be mixing cutpoints, meaning use *interstate\_20* as a dependent variable but use *one\_domestic\_40* and *both\_domestic\_40* or *one\_domestic\_60* and *both\_domestic\_60* as the key independent variables. Results are consistent.

\textsuperscript{13}I also calculate marginal effects for a hypothetical “dangerous dyad” by assuming either *Minor\_power*=0 or *Non\_contiguity*=0. Findings are consistent with those reported in Figure 1.
month $t-1$) to the “one” (i.e. one of the two states in the dyad experiences an onset of domestic conflict in month $t-1$) across any of the three cutpoints. This reflects the statistically significant coefficients in Table 1, which strongly suggest that all else being equal, a dyad with one domestic conflict onset in month $t-1$ is more likely to experience an onset of interstate conflict in month $t$. Additionally, Figure 1 clearly illustrates that dyads in which both states experience an onset of domestic conflict are considerable more likely to experience an onset of interstate conflict than dyads in which either one or neither states experienced an onset of domestic conflict. Note that the large confidence intervals around the “both” box plots is due to the relatively small number of observations, as presented in Table 1.

In Table 9, I report the odds ratio, which reflects how much more likely a dyad is to experience an onset of interstate conflict in month $t$ following a month in which either “one” or “both” of the states experiences an onset of domestic conflict in month $t-1$ relative to a the likelihood of an interstate conflict onset in month $t$ when neither of the two states experienced an onset of domestic conflict in month $t-1$. As Table 9 indicates, a dyad month following one domestic conflict onset in month $t-1$ is between 2.38 and 3.85 times more likely to experience an onset of interstate conflict in month $t$ across the three cutpoints, relative to a dyad-month with no domestic conflict onsets in month $t-1$.

Taken together, both the consistency of findings across cutpoints as well as the high degree of statistical significance in Table 1 provide strong support for Hypothesis 1. We can say with a high degree of confidence that domestic conflict onsets in one or (especially) both states in a dyad dramatically increases the likelihood that that dyad will experience an onset of interstate conflict in the following month.

In Table 2, I repeat this process but lag the domestic conflict variables two months, as opposed to the one-month lag used in Table 1, Figure 1, and Table 9. The results using two-month lags are largely consistent with the findings when using a one-month lag.

As Table 2 illustrates, an onset of domestic conflict in one or both states comprising each dyad occurs in month $t-2$, the likelihood of an interstate conflict onset in month $t$ increases.\(^{14}\) This

\(^{14}\)In Model 3 of Table 2, a standard logistic regression omits the $both\_domestic\_60$. To overcome this problem, I instead implement a Firth Logit, as recommended by Zorn (2005).
relationship holds across all three cut points and consistently achieves >95% confidence. Although it seems logical to test for the effects of domestic conflict onset in month $t - 1$ and month $t - 2$ simultaneously in the same model, this is not possible, since by definition, if a conflict onset occurs in month $t - 2$, month $t - 1$ is dropped from the empirical model because, an onset can not occur after an onset occurred in the previous month.

\[\text{INSERT FIGURE 2 HERE}\]

Figure 2 provides the marginal effects of onsets of domestic conflict in one or both states in each dyad in month $t - 2$ (as opposed to month $t - 1$ in Figure 1) on the likelihood of interstate conflict onset in month $t$, calculated with the same approach used to build Figure 1, as detailed above.\(^{15}\)

As in Figure 1, dyads that experience an onset of domestic conflict in one or both states are more likely to experience an onset of interstate conflict. Additionally, the likelihood of an interstate conflict onset is greater when both states experience an onset of domestic conflict in month $t - 2$ than when only one state experiences an onset. Furthermore, like in Figure 1, the large confidence intervals across all three columns when both states experience an onset of domestic conflict is the result of only a small number of dyads, as reported in Table 2.

\[\text{INSERT TABLE 10 HERE}\]

Table 10 is identical to Table 9, except it reflects the marginal effects of a two-month, rather than one-month lag. As Table 10 indicates, dyads in which one state experiences an onset of domestic conflict in month $t - 2$ is between 2.4 and 3.9 times more likely to experience an onset of interstate conflict in month $t$. Furthermore, the likelihood of interstate conflict becomes between 4.4 and 26.8 times more likely when both states experience an onset of domestic conflict in month $t - 2$.

4.2. Tests of the effects of domestic conflict intensity on the likelihood of interstate conflict onset.

In Table 3, I test for the effects of whether increasing intensity of ongoing domestic conflicts in month $t - 1$ increases the likelihood of an onset of an interstate conflict in month $t$. Since I am interested in the effects of domestic conflict intensity on interstate conflict onset, at least one of the two states in the dyad must have experienced a domestic conflict in month $t - 1$ and the dyad must not have experienced an interstate conflict in month $t - 1$ gain inclusion into the regression.

\(^{15}\)Note that in Figure 2, I cap the upper bound on the “both” boxplot in the cutpoint at column 0.005 in order to facilitate interpretation. The actual upper bound is 0.0022.
As in Table 3, I utilize logistic regression while accounting for the eight Russet and Oneal (2001) controls and test across the three different cutpoint values.

[INSERT TABLE 3 HERE]

In both Model 1 and Model 2 of Table 3, the key variables reflecting increasing intensity in domestic conflicts – one_worse_20 and both_worse_20 in Model 1 and one_worse_40 and both_worse_40 in Model 2 – achieve statistical significance at 95% confidence, suggesting that the likelihood of an interstate conflict onset increases as the level of domestic conflict intensity in one or both states comprising the dyad increases. However, in Model 3, only one_worse_60 achieves relatively weak statistical significance, with the estimated effect of both_worse_60 being statistically indistinguishable from zero.

[INSERT FIGURE 2 HERE]

Figure 2 presents the estimated marginal effects of increases in intensity of domestic conflicts across the three cutpoints on the likelihood of an interstate conflict. The box plots – calculated following the same procedure used in Figure 1 – visually represent the empirical findings in Table 2. As the “cutpoint_20” and “cutpoint_40” columns indicate, the likelihood of an interstate conflict onset increases slightly as the intensity of domestic conflict increase in one of the states, and dramatically increases in months following an increase in domestic conflict intensity in both states. Based on these estimated marginal effects, when the cutpoint = 20, the likelihood that a dyad experiences an onset of interstate conflict in month \( t \) is approximately 50% greater when one state experiences an increase in intensity of a domestic conflict in month \( t - 1 \), and 70% greater when both states experience increasing intensity of domestic conflict in month \( t - 1 \). In the third column, labeled “cutpoint_60”, the mean predicted probability of interstate conflict onset when “one” and “both” states experienced an increase in intensity in ongoing domestic conflict is within the 95% confidence interval of the predicted probability of an onset of interstate conflict when neither of the two states comprising the dyad experienced increasing intensity in a domestic conflict, which indicates that increasing intensity in domestic conflicts does not have statistically significant impact on the likelihood of interstate conflict.

[INSERT TABLE 11 HERE]

Table 11 follows the same procedure used to build Table 9, this time reporting how much more likely an interstate onset becomes between a dyad in month \( t \) as one or both of the states comprising
the dyad experience a worsening domestic conflict in month $t - 1$. Column 1 and Column 2 indicate that across the first two cutpoints, a dyad becomes between 1.28 and 1.92 times more likely to experience an interstate conflict onset in month $t$ as one or both of the states comprising the dyad experience a more severe domestic conflict in month $t - 1$. As reflective of the findings in Table 3 and Figure 2, the marginal effects at cutpoint.60 do not achieve statistical significance at a meaningful level, which reduces our overall confidence in the strong findings across cutpoint.20 and cutpoint.40.

In Table 4, I test for whether changes in domestic conflict intensity during month $t - 2$ affect the likelihood of an interstate conflict during month $t$.

[INSERT TABLE 4 HERE]

The findings are similar to those in Table 3, but the statistical significance is more consistent across cut points and the marginal effects are even stronger. Using 1-month lag in Table 3, one_worse.60 is weakly significant and one_worse.60 fails to achieve a p-value of .1. However, in Table 4, both one_worse.60 one_worse.60 generate strong statistical significance at the .01 level.

[INSERT FIGURE 3 HERE]

Figure 3 visualizes the results in Table 4, demonstrating that across all cut points, an onset of interstate conflict becomes considerably more likely in month $t$ when one or both states comprising the dyad experience an increasingly intense domestic conflict in month $t - 1$. Unlike in Table 3 and Figure 2, these results hold when the cutpoint=60.

[INSERT TABLE 12 HERE]

Table 12 presents how much more likely an interstate conflict onset during month $t$ becomes as one or both of the states in each dyad experience an increasingly intense domestic conflict during month $t - 2$. Interestingly, all of the marginal effects are stronger than they are in Table 10, which reflects the marginal effects at a one-month lag. For example, at cutpoint=40, an interstate conflict onset in month $t$ becomes 128% more likely as both states experience increasing intense domestic conflicts in month $t - 1$, but 220% more likely when both states experiencing increasingly intense domestic conflicts in month $t - 2$. These findings, interpreted jointly with the results in Table 10, are interesting for a number of reasons. First, it suggests that it takes interstate dynamics varying amounts of time to respond to domestic conditions. For example, Table 10 demonstrates that states tend to meaningfully respond to events occurring in the previous month, but Table 11 shows
that interstate dynamics respond even stronger to events occurring two months ago. Though it is outside the focus of this chapter to rigorously analyze why this is the case, it seems feasible that domestic institutional processes may slow response times to domestic crises, meaning that states’ foreign policies often take longer than one-month to respond to important events. Second, the tests of the one-month lag highlight the importance of using different cutpoints as a robustness check. Since the empirical findings are not consistent across all three cutpoints (like they were in Model 1 and Figure 1), we must be less confident about the strength of findings. With that in mind, Table 2 and Figure 2 still provide general support for Hypothesis 2 – that the likelihood of an onset of interstate conflict increase in months following increases in intensity in domestic conflicts.

4.3. Tests of the effects of domestic conflict intensity on interstate conflict intensity. Thus far, I have analyzed the effects of domestic conflict onset on the likelihood of interstate conflict onset and the effects of domestic conflict intensity on the likelihood of an interstate conflict onset. Finally, in this section, I address whether changes in domestic conflict intensity affect the intensity of ongoing interstate conflicts. In Table 3, I test for the effects of whether one or both states experienced more intense domestic conflicts in month \(t - 1\) on whether the ongoing interstate conflict becomes more intense in month \(t\) than in month \(t - 1\). Given this focus, I only model dyad-months in which at least one of the two states experienced greater than the cutpoint number of material conflict events in month \(t - 1\) and a >cutpoint number of interstate material conflict events in month \(t - 1\). Of the over 4 million dyad-month observations in my dataset, only 3,791, 1,074, and 458 dyad-months meet this requirement across the three cutpoints, respectively. As in Table 1 through Table 4, I run a series of logistic regressions while accounting for eight Russet and Oneal (2001) controls.

\[\text{INSERT TABLE 5 HERE}\]

As Table 5 reflects, none of six measures reflecting whether one or both countries in each dyad experienced more intense domestic conflict achieved statistical significance. In Table 6, I repeat the analysis using a two-month lag of the domestic conflict variables.

\[\text{INSERT TABLE 6 HERE}\]

Results are similar to Table 5, with the single exception that one_worse_60 achieves moderate statistical significance. However, it seems unlikely that this relationship is robust based on the lack of consistency across Model 1 and Model 2 in Table 6. Rather, it is more likely that the
statistical significance of one\_worse\_60 in Model 3 is simply fitting “noise” in the dataset. Overall, a joint interpretation of the results in Table 5 and Table 6 jointly provide suggest that the causal mechanisms purported in Hypothesis 3 and Hypothesis 4 may not be present in the data, though I provide further testing of these hypotheses in Table 7 and Table 8 below.

[INSERT TABLE 7 HERE]

In Table 7, I provide an additional test of the effects of changes in the intensity of ongoing domestic conflicts on changes in the intensity of ongoing interstate conflicts, but this time I utilize the three continuous measures that reflect changing intensity of interstate conflict (\textit{interstate\_change\_20}, \textit{interstate\_change\_40}, and \textit{interstate\_change\_60}) as dependent variables. Since my dependent variables are now continuous, I utilize a basic OLS regression. In Model 1 and Model 2 of Table 4, dyads in which both states experienced worsening domestic conflict in month $t-1$ tend to engage in fewer interstate material conflict events in month $t$. Interpreting the coefficients, Model 1 suggests that as both states experience more intense domestic conflicts in month $t-1$, these two states tend to engage in approximately 4.6 fewer “interstate material conflict” events in month $t$. When the cutpoint shifts to 40 in Model 2, we should expect to see over 7 fewer “interstate material conflict” events in month $t$ if both states experienced more intense domestic conflict in month $t-1$. However, the failure of both\_worse\_60 to achieve statistical significance in Model 3 tempers our confidence in the the significant and negative relationship between two countries experiences more intense domestic conflict and the dyad experience less intense interstate conflict.

[INSERT TABLE 8 HERE]

As in Section 4.1 and Section 4.2, I rerun the empirical models in Table 7 using a two-month lag of the domestic conflict variables, with results presented in Table 8. Although results are not perfectly consistent across Model 1, Model 2, and Model 3 in Table 8, a joint interpretation of the three models suggests that increases in intensity of domestic conflicts in month $t-2$ tends to decrease the intensity of ongoing interstate conflicts in month $t$. Additionally, unlike in Section 4.3, the lack of consistency in Table 7 and Table 8 makes it difficult to compare the substantive effects of the one- and two-month lag. One potentially interesting comparison is that in Table 7, none of the three one\_worse variables achieve statistical significance, but two of the both\_worse variables do. In Table 8, this is reversed, with only one of the three both\_worse variables achieving a p-value<1 but all three of the one\_worse variables generating p-values<0.05. While this may suggest that interstate conflict dynamics may respond more rapidly to changes in domestic conflict intensity in both states.
than to changes in domestic conflict intensity occurring in one state, the lack of robust empirical support across cut points prevents me from asserting this relationship with meaningful confidence.

Overall, across the 12 empirical tests run in Table 5, Table 6, Table 7, and Table 8, I find no support for Hypothesis 3 – increases in intensity of domestic conflict should increase the intensity of ongoing interstate conflicts. Conversely, Table 7 and Table 8 provide some support for Hypothesis 4–increase in intensity of domestic conflicts should lead to decreases in intensity of ongoing interstate conflicts.

5. Conclusion

In the real world, we know that the levels of domestic and interstate conflict can fluctuate rapidly. One month, Rwanda is at relative domestic peace, the next three months, it experiences a horrific genocide, and the next month it returns to relative peace. Similarly in the interstate conflict context, one month India and Pakistan are at peace, the next month there is an escalation in violence, and the following month they return to peace. Also based on real-world observations, we have a strong expectation that often times, domestic conflicts tend to affect interstate relations. Given these observations, a number of interesting questions emerge regarding potential relationships between domestic conflict and interstate conflict. Do onsets of domestic conflict in a state increase the likelihood that it will engage in interstate conflict with a neighbor? If two states are engaged in interstate conflict are also fighting domestic conflicts, do increases in intensity of the domestic conflicts tend to lead to increases in intensity in the interstate conflict as well? Despite the massive number of quantitative studies of both domestic and interstate conflict, studies testing for relationships between domestic and interstate conflict are scarce, primarily due to a lack of appropriate data. For example, Gleditsch, Salehyan and Schultz (2008) is the most comprehensive study to date analyzing the relationship between domestic and interstate conflict, but their use of state-level binary measures of both domestic and interstate conflict (as is ubiquitous throughout the related literature) inhibit their ability to both construct measures of conflict intensity and test for sub-annual level variation that tends to be ubiquitous in all conflicts.

The key advancement of this study is my use of the GDELT event data to build measures reflecting levels of domestic and interstate conflict at the monthly level for all countries and dyads in the world. This allows me to test four hypotheses that the extant literature has theorized to be
true but never before been able to test empirically. Based on various logistic and OLS regressions, I find strong support for Hypothesis 1, moderate support for Hypothesis 2, no support for Hypothesis 3, and moderate support for Hypothesis 4. Additionally, increases in domestic conflict intensity in both states in month \( t - 1 \) tending to lead to less intense interstate conflicts in month \( t \).\(^{16}\)

Overall, I believe that this study has two major takeaways as well as two clear paths for future research. First, we now have sufficiently nuanced data to move beyond coarse, yearly level binary measures of conflict. With event data, we can build monthly (or even weekly or daily) level measures that are able to capture the sub-annual variation in the level of domestic and interstate conflicts. This should allow researchers to test for a host of theoretically expected relationships that have heretofore been difficult or impossible to empirically test given a lack of data. Second, the empirical evidence seems to suggest that if you are interested in analyzing interstate conflict onsets, you should account for whether one or both states in the dyad recently experienced an onset of interstate conflict or increasingly intense conflicts if one or both states have ongoing domestic conflicts. It is possible that by accounting for various measures of domestic conflicts, inferences drawn on other variables may change.

In terms of future research, this chapter has provided some preliminary answers to hypotheses framed in a basic “do these relationships exist?” format. For some proposed relationships, including the one suggested in Hypothesis 1, my empirical results provide strongly suggests that the answer is “yes”. The next logical test is to test for hypotheses that propose why these relationships exist. Again, the major obstacle to asking these more difficult why questions has been a lack of data. However, as I attempt to highlight throughout this chapter, the 200 million (and counting) events in the GDELT dataset make it possible to ask increasingly nuanced questions, and I am confident that moving forward, scholars will be able to use GDELT to isolate specific causal mechanisms – such as intervention or diversion – that may be responsible for the strong effect that domestic conflicts seems to have on interstate conflicts.

Additionally, this chapter provides an initial framework for analyzing whether the effects of domestic conflict events on interstate conflict dynamics change as the amount of time since the occurrence of the domestic events increases. For example, I find some support suggesting that domestic conflict onsets in month \( t \) tend to increase the likelihood of an interstate conflict onset more in month \( t + 2 \) than in month \( t + 1 \). Future studies could focus more heavily on this and

\(^{16}\)strong>moderate>weak>no.
similar findings and potentially isolate factors, such as institutional design or regime type, that may affect the time that amount of time that elapses before an interstate conflict reflects changes occurring at domestic levels.
References


### Table 1. The Effects of Lagged Domestic Conflict Onset on Interstate Conflict Onset with 1-month Lag

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>-1.11***</td>
<td>-1.10***</td>
<td>-.98***</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.03***</td>
<td>.31***</td>
<td>.04***</td>
</tr>
<tr>
<td>Alliance</td>
<td>.35***</td>
<td>.36***</td>
<td>.51***</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-1.28***</td>
<td>-1.55***</td>
<td>-2.31***</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>.29**</td>
<td>.55***</td>
<td>1.10***</td>
</tr>
<tr>
<td>Depend_L</td>
<td>.02</td>
<td>-.23</td>
<td>-1.24**</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.08***</td>
<td>-.12***</td>
<td>-.17***</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>4.55***</td>
<td>4.33***</td>
<td>3.51***</td>
</tr>
<tr>
<td>l.one_domestic_20</td>
<td>.99***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.both_domestic_20</td>
<td>2.36***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.one_domestic_40</td>
<td></td>
<td>1.49***</td>
<td></td>
</tr>
<tr>
<td>l.both_domestic_40</td>
<td></td>
<td>4.35***</td>
<td></td>
</tr>
<tr>
<td>l.one_domestic_60</td>
<td></td>
<td></td>
<td>1.72***</td>
</tr>
<tr>
<td>l.both_domestic_60</td>
<td></td>
<td></td>
<td>4.71***</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.25***</td>
<td>-5.06***</td>
<td>-6.23***</td>
</tr>
<tr>
<td>N</td>
<td>3,989,306</td>
<td>4,313,227</td>
<td>4,427,683</td>
</tr>
<tr>
<td># of Interstate onsets</td>
<td>559</td>
<td>279</td>
<td>176</td>
</tr>
<tr>
<td># of one_domestic onsets</td>
<td>147,798</td>
<td>65,794</td>
<td>36,473</td>
</tr>
<tr>
<td># of both_domestic onsets</td>
<td>2,065</td>
<td>387</td>
<td>130</td>
</tr>
</tbody>
</table>

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
Table 2. The Effects of Lagged Domestic Conflict Onset on Interstate Conflict Onset with 2-month Lag

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>-1.14***</td>
<td>-1.10***</td>
<td>-1.02***</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.03***</td>
<td>.03***</td>
<td>.04***</td>
</tr>
<tr>
<td>Alliance</td>
<td>.31***</td>
<td>.377***</td>
<td>.49***</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-1.30***</td>
<td>-1.60***</td>
<td>-2.42***</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>.45***</td>
<td>.58**</td>
<td>1.14***</td>
</tr>
<tr>
<td>Depend_L</td>
<td>-1.33</td>
<td>-.19</td>
<td>-.88*</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.08***</td>
<td>-.11***</td>
<td>-.17***</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>4.96***</td>
<td>4.35***</td>
<td>3.89***</td>
</tr>
<tr>
<td>l2.one_domestic_20</td>
<td>.84***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.both_domestic_20</td>
<td>1.50***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.one_domestic_40</td>
<td></td>
<td>1.32***</td>
<td></td>
</tr>
<tr>
<td>l2.both_domestic_40</td>
<td></td>
<td>2.16***</td>
<td></td>
</tr>
<tr>
<td>l2.one_domestic_60</td>
<td></td>
<td></td>
<td>2.01***</td>
</tr>
<tr>
<td>l2.both_domestic_60</td>
<td></td>
<td></td>
<td>3.21**</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.22***</td>
<td>-4.99***</td>
<td>-6.10***</td>
</tr>
<tr>
<td>N</td>
<td>3,976,179</td>
<td>4,297,435</td>
<td>4,410,831</td>
</tr>
<tr>
<td># of Interstate onsets</td>
<td>531</td>
<td>274</td>
<td>173</td>
</tr>
<tr>
<td># of one_domestic onsets</td>
<td>147,377</td>
<td>65,471</td>
<td>36,126</td>
</tr>
<tr>
<td># of both_domestic onsets</td>
<td>2,072</td>
<td>393</td>
<td>130</td>
</tr>
</tbody>
</table>

Coefficients with p-values reflected by: ***(.01), **(.05), *.10)
Table 3: The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestic Conflicts on Interstate Conflict Onset

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>-.12***</td>
<td>-.19***</td>
<td>-.25***</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.03***</td>
<td>.03***</td>
<td>.04***</td>
</tr>
<tr>
<td>Alliance</td>
<td>.05*</td>
<td>.15**</td>
<td>.31**</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-.87***</td>
<td>-.78***</td>
<td>-.97***</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>.45**</td>
<td>.47***</td>
<td>.60*</td>
</tr>
<tr>
<td>Depend_L</td>
<td>.46**</td>
<td>.82**</td>
<td>.06</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.04***</td>
<td>-.06***</td>
<td>-.07***</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>.01</td>
<td>.67</td>
<td>1.15</td>
</tr>
<tr>
<td>l.one_worse_20</td>
<td>.30***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.both_worse_20</td>
<td>.76***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.one_worse_40</td>
<td></td>
<td>.35***</td>
<td></td>
</tr>
<tr>
<td>l.both_worse_40</td>
<td></td>
<td>1.06***</td>
<td></td>
</tr>
<tr>
<td>l.one_worse_60</td>
<td></td>
<td></td>
<td>.30*</td>
</tr>
<tr>
<td>l.both_worse_60</td>
<td></td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.57***</td>
<td>-3.88***</td>
<td>-4.79***</td>
</tr>
</tbody>
</table>

N 580,094 257,577 143,008
# of Interstate onsets 1458 389 155
# of one_worse 187,679 88,530 49,320
# of both_worses 8,438 1,909 620

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
Table 4. The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestic Conflicts on Interstate Conflict Onset

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>-1.6***</td>
<td>-.14***</td>
<td>-.16</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.03***</td>
<td>.05***</td>
<td>.07***</td>
</tr>
<tr>
<td>Alliance</td>
<td>.11***</td>
<td>.33***</td>
<td>.93***</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-1.25***</td>
<td>-1.45***</td>
<td>-1.53***</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>.38***</td>
<td>.49*</td>
<td>.26</td>
</tr>
<tr>
<td>Depend_L</td>
<td>.42**</td>
<td>.33</td>
<td>-.48</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.05***</td>
<td>-.08***</td>
<td>-.10***</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>.00</td>
<td>-.13</td>
<td>-.36</td>
</tr>
<tr>
<td>l2.one_worse_20</td>
<td>.64***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.both_worse_20</td>
<td>.83***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.one_worse_40</td>
<td></td>
<td>.84***</td>
<td></td>
</tr>
<tr>
<td>l2.both_worse_40</td>
<td></td>
<td>.79***</td>
<td>.55***</td>
</tr>
<tr>
<td>l2.one_worse_60</td>
<td></td>
<td></td>
<td>.55***</td>
</tr>
<tr>
<td>l2.both_worse_60</td>
<td></td>
<td></td>
<td>1.433***</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.19***</td>
<td>-5.50***</td>
<td>-8.62***</td>
</tr>
<tr>
<td>N</td>
<td>589,448</td>
<td>271,226</td>
<td>158,598</td>
</tr>
<tr>
<td># of Interstate onsets</td>
<td>1,466</td>
<td>390</td>
<td>162</td>
</tr>
<tr>
<td># of one Worse</td>
<td>186,350</td>
<td>87,821</td>
<td>48,965</td>
</tr>
<tr>
<td># of both Worses</td>
<td>8,471</td>
<td>1,919</td>
<td>628</td>
</tr>
</tbody>
</table>

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
Table 5. The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestics Conflicts on Whether an Ongoing Interstate Conflict Becomes more Intense

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>.10*</td>
<td>.04</td>
<td>-.02</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.00</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>Alliance</td>
<td>.26***</td>
<td>.35***</td>
<td>.22</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>.01</td>
<td>-.24</td>
<td>-.55</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>-.18</td>
<td>-.84**</td>
<td>-.67</td>
</tr>
<tr>
<td>Depend_L</td>
<td>-.24</td>
<td>-.89**</td>
<td>-.34</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.01</td>
<td>-.02</td>
<td>-.02</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>-.82**</td>
<td>-.43</td>
<td>-.37</td>
</tr>
<tr>
<td>l.one_worse_20</td>
<td>-.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.both_worth_20</td>
<td>-.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.one_worse_40</td>
<td></td>
<td>-.15</td>
<td></td>
</tr>
<tr>
<td>l.both_worse_40</td>
<td></td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>l.one_worse_60</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>l.both_worse_60</td>
<td></td>
<td></td>
<td>.36</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.09***</td>
<td>-2.34***</td>
<td>-2.87***</td>
</tr>
</tbody>
</table>

N 3791 1,074 458

# of more intense interstate conflicts 957 264 114
# of one_worse 1,712 488 209
# of both_worse 600 161 67

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
### Table 6. The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestic Conflicts on Whether an Ongoing Interstate Conflict Becomes more Intense

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (interstate_20)</th>
<th>Model 2 (interstate_40)</th>
<th>Model 3 (interstate_60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>.11*</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.00</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>Alliance</td>
<td>.23***</td>
<td>.45***</td>
<td>.36</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>.07</td>
<td>-.24</td>
<td>-.42</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>-.20</td>
<td>-.62*</td>
<td>.72</td>
</tr>
<tr>
<td>Depend_L</td>
<td>-.05</td>
<td>-.88</td>
<td>-.61</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.01</td>
<td>-.01</td>
<td>-.01</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>-.96**</td>
<td>-.54</td>
<td>-.31</td>
</tr>
<tr>
<td>l2.one_worse_20</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.both_worth_20</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.one_worse_40</td>
<td></td>
<td>-.14</td>
<td></td>
</tr>
<tr>
<td>l2.both_worse_40</td>
<td></td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>l2.one_worse_60</td>
<td></td>
<td></td>
<td>-.52**</td>
</tr>
<tr>
<td>l2.both_worse_60</td>
<td></td>
<td></td>
<td>.35</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.92***</td>
<td>-2.54***</td>
<td>-0.91***</td>
</tr>
</tbody>
</table>

| N                         | 3,529                    | 964                     | 395                     |
| # of more intense interstate conflicts | 935 | 247 | 109 |
| # of one_worse            | 1,577                    | 451                     | 186                     |
| # of both_worses          | 515                      | 138                     | 54                      |

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
**Table 7. The Effects of Changes in Intensity of 1-month Lagged Ongoing Domestics Conflicts on Changes in Interstate Conflict Intensity**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>.42</td>
<td>1.11</td>
<td>.28</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.11**</td>
<td>.34**</td>
<td>.23</td>
</tr>
<tr>
<td>Alliance</td>
<td>1.55***</td>
<td>2.67</td>
<td>2.64</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-1.08</td>
<td>-.44</td>
<td>-4.9</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>-3.50</td>
<td>-19.99***</td>
<td>-28.54**</td>
</tr>
<tr>
<td>Depend_L</td>
<td>-2.84</td>
<td>-12.51</td>
<td>-18.97</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.07</td>
<td>-.31</td>
<td>-.35</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>-3.11</td>
<td>7.70</td>
<td>-11.83</td>
</tr>
<tr>
<td>l.one_worse_20</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.both_worse_20</td>
<td>-4.63***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.one_worse_40</td>
<td>-1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.both_worse_40</td>
<td>-7.18*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.one_worse_60</td>
<td></td>
<td></td>
<td>2.61</td>
</tr>
<tr>
<td>l.both_worse_60</td>
<td>-16.94***</td>
<td>-33.38***</td>
<td>-27.00***</td>
</tr>
<tr>
<td>Constant</td>
<td>-16.94***</td>
<td>-33.38***</td>
<td>-27.00***</td>
</tr>
</tbody>
</table>

| N          | 3791 | 1,074 | 458 |
| # of one_worse | 1,712 | 488 | 209 |
| # of both_worses | 600 | 161 | 67 |

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)

**Table 8. The Effects of Changes in Intensity of 2-month Lagged Ongoing Domestics Conflicts on Changes in Interstate Conflict Intensity**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_distance</td>
<td>.31</td>
<td>.40</td>
<td>.93</td>
</tr>
<tr>
<td>IGO_count</td>
<td>.06</td>
<td>.19</td>
<td>.20</td>
</tr>
<tr>
<td>Alliance</td>
<td>1.19**</td>
<td>1.23</td>
<td>.44</td>
</tr>
<tr>
<td>Minor Powers</td>
<td>-.60</td>
<td>-1.61</td>
<td>-2.17</td>
</tr>
<tr>
<td>Power Ratio</td>
<td>-2.39</td>
<td>-13.78**</td>
<td>-29.40**</td>
</tr>
<tr>
<td>Depend_L</td>
<td>-1.71</td>
<td>-15.39</td>
<td>-24.83</td>
</tr>
<tr>
<td>Democ_L</td>
<td>-.06</td>
<td>-.10</td>
<td>-.20</td>
</tr>
<tr>
<td>Non-Contiguity</td>
<td>-4.24</td>
<td>-9.39</td>
<td>-16.25</td>
</tr>
<tr>
<td>l2.one_worse_20</td>
<td>-2.27**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.both_worse_20</td>
<td>-3.05***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.one_worse_40</td>
<td>-8.04***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.both_worse_40</td>
<td>-2.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l2.one_worse_60</td>
<td></td>
<td></td>
<td>-14.02**</td>
</tr>
<tr>
<td>l2.both_worse_60</td>
<td></td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>Constant</td>
<td>-10.91***</td>
<td>-11.83***</td>
<td>-6.30***</td>
</tr>
</tbody>
</table>

| N          | 3,529 | 964 | 395 |
| # of one_worse | 1,577 | 451 | 186 |
| # of both_worses | 515 | 138 | 54 |

Coefficients with p-values reflected by: ***(.01), **(.05), *(.10)
Table 9. **How much more likely is an interstate conflict onset in month \( t \) relative to neither state experiencing an onset of domestic conflict in month \( t - 1 \)**

<table>
<thead>
<tr>
<th>cutpoint 20</th>
<th>cutpoint 40</th>
<th>cutpoint 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>2.38x*</td>
<td>3.78x*</td>
</tr>
<tr>
<td>both</td>
<td>2.81x*</td>
<td>4.56x*</td>
</tr>
</tbody>
</table>
* indicates statistically significant at 95% confidence

Table 10. **How much more likely is an interstate conflict onset in month \( t \) relative to neither state experiencing an onset of domestic conflict in month \( t - 2 \)**

<table>
<thead>
<tr>
<th>cutpoint 20</th>
<th>cutpoint 40</th>
<th>cutpoint 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>2.29x*</td>
<td>3.67x*</td>
</tr>
<tr>
<td>both</td>
<td>4.41x*</td>
<td>8.68x*</td>
</tr>
</tbody>
</table>
* indicates statistically significant at 95% confidence

Table 11. **How many times more likely is an interstate conflict onset in month \( t \) relative to neither state experiencing a more intense ongoing domestic conflict in month \( t - 1 \)**

<table>
<thead>
<tr>
<th>cutpoint 20</th>
<th>cutpoint 40</th>
<th>cutpoint 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>1.33x*</td>
<td>1.39x*</td>
</tr>
<tr>
<td>both</td>
<td>1.92x*</td>
<td>1.28x*</td>
</tr>
</tbody>
</table>
* indicates statistically significant at 95% confidence

Table 12. **How many times more likely is an interstate conflict onset in month \( t \) relative to neither state experiencing a more intense ongoing domestic conflict in month \( t - 2 \)**

<table>
<thead>
<tr>
<th>cutpoint 20</th>
<th>cutpoint 40</th>
<th>cutpoint 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>1.90x*</td>
<td>2.32x*</td>
</tr>
<tr>
<td>both</td>
<td>2.28x*</td>
<td>2.20x*</td>
</tr>
</tbody>
</table>
* indicates statistically significant at 95% confidence
Whether 'neither', 'one', or both states experienced a domestic conflict onset in month t-1

Figure 1. The Effects of Domestic Conflict Onset in Month t-1 on the Likelihood of Interstate Conflict Onset in Month t
Whether 'neither', 'one', or both states experienced a domestic conflict onset in month t-2 (with "both" for cutpoint_60 capped)

Figure 2. The Effects of Domestic Conflict Onset in Month t-2 on the Likelihood of Interstate Conflict Onset in Month t
Whether 'neither', 'one', or both states experienced a more severe domestic conflict in month t-1 than month t-2

Figure 3. The Effects of an Increasingly Severe Domestic Conflict in Month t-1 on the Likelihood of Interstate Conflict Onset in Month t
Whether 'neither', 'one', or both states experienced a more severe domestic conflict in month t-2 than month t-3.

Figure 4. The Effects of an Increasingly Severe Domestic Conflict in Month t-2 on the Likelihood of Interstate Conflict Onset in Month t.
CHAPTER 3. PREDICTING FUTURE LEVELS OF VIOLENCE IN AFGHANISTAN DISTRICTS

1. Introduction

For centuries, key pillars of the philosophy of science like Francis Bacon and David Hume, have stressed that scientific progress occurs through the development of consistently accurate, replicable, and falsifiable predictive models. Building on these argument, numerous scholars of political conflict, including Choucri (1974), Singer and Wallace (1979), Beck, King and Zeng (2000), Bueno de Mesquita (2002), and Ward, Greenhill and Bakke (2010), have similarly stressed the importance of predictive models for two main reasons. First, as Beck, King and Zeng (2000), Weidmann and Ward (2010), and others convincingly argue, predictions are vital for the development of theories about the causes of violence, since the most rigorous way to test whether an empirical model is actually reflecting a real-world data generating process, or simply fitting “noise”, is to measure its forecast accuracy.\(^1\) Second, accurate conflict forecasts can be tremendously useful in the real world – they can help peacekeepers allocate scarce resources, inform Non-governmental Organizations (NGOs) on potential hot-spots to avoid, and even provide speculative investment opportunities. Although the majority of empirical studies of conflict continue to focus on “explanation” – primarily in the form of interpreting coefficients and standard errors established through in-sample testing – a smaller though considerable number papers and projects exist with the explicit goal of building dynamic forecasts of future levels of violence. Likewise, the goal of this chapter is to build a forecasting model, though not for theory-building or hypothesis-testing, but rather to create a proof of concept tool for real-time, policy relevant decision making.

Extant empirical forecasting studies focusing on domestic conflict range tremendously in terms of data, methods, and scope. The most coarse studies build forecasts at that state-year level using primarily structural variables like GDP per capita, ethnic diversity, and infant mortality (see Gurr and Harff (1996), King and Zeng (2001), Fearon and Laitin (2003), and Goldstone et al. (2010)), which are useful in some contexts but unable to build predictions beyond the state-year unit of analysis. The majority of studies attempting to build empirical forecasts of violence use

\(^1\)I use “prediction” and “forecast” interchangeably throughout this chapter.
more fine grained, event data coded at the daily and sometimes local level, as these data allow scholars to capture more dynamic patterns of violence and ultimately build more detailed forecasts than those using state-year, structural data. Historically, scholars building empirical forecasting models of violence have used either machine-coded (like KEDS, ICEWS, 10 Million International Dyadic Events Dataset) or human-coded event data datasets (like ACLED, KOSVED, SCAD, etc.) built from open-source text, with the majority of scholars utilizing the machine-coded option. Recently, however, WikiLeaks has provided an alternative data set of conflict events that previously required security clearance from the United States Government to access, but have subsequently been illegally obtained and distributed to the public. The logical question, then, is which of these sources of data is more appropriate for this study? Given the goal of this chapter, an ideal dataset would contain the following five key attributes:

1. **Broad spatial coverage**: Global coverage is preferable to one with country or region-specific coverage as it would enable a forecasting model to be built for any global location.

2. **Density**: Predictive algorithms tend to perform better with more data, meaning that many fine-grained events is preferable to fewer larger-scale events.

3. **Geo-coding**: Sub-state, geo-spatial predictions require sub-state, geo-coded events.

4. **Accuracy**: The data should accurately reflect the events as they occur in reality in order to build relevant predictions.

5. **Future availability in real-time**: If the data are not accessible in the future in real or near real-time, then it becomes highly difficult to build actionable predictions.

As discussed in Chapter 1 in greater detail, the GDELT dataset provides greater spatial coverage, event density, and prospects for future availability in real-time than either the human-coded datasets or the WikiLeaks datasets. Accuracy is likely greater for the WikiLeaks dataset since it is based on first-hand accounts, and ongoing debate exists regarding the accuracy of human-coded and machine-coded datasets suggesting that neither may have a clear advantage (see King and Lowe (2004), O’Brien (2010), O’Loughlin et al. (2010), Schrodt (2012), Chojnacki et al. (2012), and Eck (2012) for discussions of the accuracy of machine-coded and human-coded event data datasets). However, because the ultimate goal is building policy relevant predictions in real (or near real) time, the fifth attribute is a necessary condition that neither the human-coded nor WikiLeaks dataset meet. Thus, GDELT is the most appropriate dataset.
This is the first study to ever use open-source, machine-coded event data to build forecasts of political violence at a sub-state level of geospatial aggregation. Since the process of aggregating conflict events into sub-state units based on latitude and longitude is currently time and computationally intensive, doing so on a global scale exceeds the scope of a dissertation chapter. Thus, I focus on forecasting conflict in sub-state geospatial units in a single country: Afghanistan. I choose Afghanistan for two reasons. First, there is dense political violence across a long time-frame (2001-2012) with considerable variation at local levels. Second, Mangion-Zammit et al. (2012) have demonstrated the ability to build forecasts with the WikiLeaks data, meaning that to the extent it is possible at all to build temporally and geo-spatially nuanced forecasts of political violence using open source, machine-coded event data, it should be feasible in Afghanistan.

Although I focus primarily on building predictions one-month in advance at the district-month unit of analysis (Afghanistan’s smallest administrative unit, N=317), I also build forecasts at the province-month (N=32) and the country-month (N=1) level, which provides a rudimentary test of the effects of geo-spatial aggregation on forecast accuracy. Empirically, I use an autoregressive fractionally integrated moving average (ARFIMA) model, which builds forecasts of levels of material conflict one-month-in-advance that consistently outperforms a naive model assuming that the level of violent in a location during a month will be the same as it was in the same location in the previous month. The ARFIMA model performance decrease relative to the naive model at each additional level of geo-spatial aggregation, suggesting further justification for the use of fine-grained geo-spatial analyses. Additionally, I implement two logical extensions to the univariate ARFIMA model, first by building and modeling additional features, and second by incorporating exogenous drug price data to ARFIMA model, though neither enhance predictive accuracy. The remainder of this chapter provides a review of relevant literature, details my research design and ARFIMA forecasting model, discusses two logical extensions, and lastly concludes.

2. Literature Review

To facilitate this review of relevant literature, I organize studies that forecast domestic political violence into the three general types of data that they use: machine-coded, human-coded, and WikiLeaks.

2.1. Machine-coded data. Although a large number of studies utilize machine-coded event data (see Appendix), a much smaller subset of these studies build forecasts: Schrodt and Gerner (1997)
use discriminant analysis to predict conflict phases in the Levant, Schrodt (1999) uses HMMs to forecast conflict in southern Lebanon, Pevehouse and Goldstein (1999) use time-series to predict events in the Serbia-Kosovo conflict, Schrodt and Gerner (2000) forecast unique clusters of conflict in the Levant from 1979 to 1997, Schrodt (2000) uses HMM’s to forecast conflict dynamics in the Levant form 1979 to 1997, Bond et al. (2004) forecast conflict in Indonesia, Shellman (2004b) forecasts conflicts between government and dissident actors in Chile and Venezuela, Brandt and Freeman (2005) use Bayesian time-series to forecast dynamics between the United States, Israel, and Palestine, Schrodt (2006) forecasts conflict in the Balkans using HMMs, Shearer (2006) uses HMMs to forecast conflict between Israel and Palestine, Bagozzi (2011) uses zero-inflated count models and D’Orazio, Yonamine and Schrodt (2011) use sequence analysis to forecast domestic conflict in 29 Asian countries, and Brandt, Freeman and Schrodt (2011) employ Markov Switching Bayesian Vector Autoregression (MS-BVAR) for forecast domestic and inter-state conflict in the Levant in 2010. Although these and other scholars demonstrate the ability to generate accurate forecasts of when and between whom conflict will occur in the future using open-source, machine-coded event data, they have been unable to predict where this conflict will occur at a sub-state level since none of the relevant machine-coded event data datasets provided geo-location information prior to GDELT.

2.2. Human-coded data. A number of geo-located, human-coded event data datasets exist that could allow researchers to build forecasts of violence at specific sub-state geographic units. For example, the Armed Conflict Location and Event Dataset (ACLED), which provides over 75,000 geo-coded violent events with (both atomic and composite) for approximately 60 countries, including all of Africa, and other, conflict-prone countries throughout the world (see Raleigh et al. (2010)), Daly (2012) provides a dataset with 7,729 geo-coded acts of violence in Colombia from 1964-1984, Schneider, Bussman and Ruhe (2012) presents the Konstanz One-Sided Violence Event Dataset (KOSVED) with 21,458 attacks against civilians in Bosnia, Urdal and Hoelscher (2012) introduces a dataset of 4,003 events occurring in 55 major cities in Asia and sub-Saharan Africa from 1960 to 2008, and Salehyan et al. (2012) introduce the Social Conflict in Africa Database (SCAD), which contains 7,200 events of political unrest occurring in 47 African countries from 1990-2010.

Despite the geospatial nuance of these datasets, it is somewhat surprising that only Weidmann and Ward (2010) uses one of the aforementioned datasets (ACLED) in order to build predictions, whereas dozens of other articles dimly focus on explanation. Weidmann and Ward (2010) use
ACLED’s Bosnia dataset in order to build a model that predicts a binary measure of whether a given municipality-month in Bosnia. In total, 4,796 municipality months exists (109 municipalities form March 1992 to October 1995), of which 301 experienced an ACLED conflict event and are treated as a “1”. They build a model based on exogenous variables (population, ethnic diversity, borders, and mountains) as well as various endogenous lags of the dependent variable, and utilize a Markov Chain Mote Carlo (MCMC) technique to estimate a logistic regression which is then used to calculate predictions in a rigorous out-of-sample framework, which I discuss in greater detail in Section 4.2.

Despite making major theoretical and empirical contributions to the study of political violence, the fact that the only study to build out-of-sample forecasts using human-coded event data (e.g. Weidmann and Ward (2010)) did so for a conflict that ended five years prior to the release of the study underscores the slow, tedious nature of building human-coded datasets that makes them extremely difficult to update sufficiently close to real time as to build policy-relevant forecast actually for the future.

2.3. WikiLeaks data. On July 25, 2010, WikiLeaks publicly released the majority of classified documents comprising both the Afghan War Diary (containing 91,731 documents) and the Iraq War Log (containing 391,832), which contain classified documents that provide a highly detailed account of events occurring in Afghanistan and Iraq from January 2004 through December 2009. Additionally, in 2010, the United States government declassified subsections of the Afghan War Diary and the Iraq war log, called Significant Acts (SIGACT). Although both the WikiLeaks and SIGACT datasets have become difficult to obtain, a number of academic studies have been published that empirically model these data for both Iraq and Afghanistan. Like studies discussed in Section 2.2, the majority of studies using the WikiLeaks and SIGACT data focus on explanation, rather than prediction.

For example, Berman et al. (2011) analyze the effects of sub-state level unemployment data for 297 district-quarters (3 quarters for 99 districts) for Iraq and 2,160 district-months (6 months for between 363 and 365 districts) for Afghanistan on levels of violence using the SIGACT data; Weidmann and Salehyan (2011) use the SIGACT data to analyze the effects of the U.S. surge in Iraq on levels of violence in 85 neighborhoods in Baghdad; O’Loughlin et al. (2010) use hotspot and cluster analysis to compare the Afghan War Diaries data to ACLED’s Afghanistan data; Linke, Witmer and O’Loughlin (2012) model violence dynamics between the U.S-led coalition forces and
insurgent by analyzing 301,374 violent events aggregated at the three-day, 30-by-30 second grid-cell level, and although the authors do assess their model’s predictive accuracy, this is done only using in-sample findings as opposed to a proper in-sample/out-of-sample break, meaning that the model is not actually building predictions. Among studies drawing on the WikiLeaks or SIGACT datasets, Mangion-Zammit et al. (2012) is the only to actually build out-of-sample forecasts. To do so, Mangion-Zammit et al. (2012) first use the WikiLeaks data to calculate the number of violent events at the province-month level in Afghanistan from 2004 to 2009, which serves as the in-sample training set. Second, they construct and train a point-process model on the 2004-2009 training data. Third, they build future predictions at the province-year level for 2010, based purely on information from 2004-2009. Since WikiLeaks only provides data through 2009, Mangion-Zammit et al. (2012) evaluate their model’s predictive accuracy based on data provided by the Afghan NGO Safety Office (ANSO), and find that 62.5% of actual levels of violence fall within 95% confidence intervals of predicted levels.

Although these studies apply innovative methods to address interesting questions, they highlight two major shortcomings to working with WikiLeaks-style of data. First, even when it can be acquired, it does not provide real or near-real time updates. As a result, Mangion-Zammit et al. (2012) needed to use a different data source to obtain data from 2010 since WikiLeaks only covered 2004-2009. Second, all of the studies discussed in Section 2.3 focus on either Iraq or Afghanistan since WikiLeaks only provided dense data for those countries, which clearly means that WikiLeaks data is unsuitable to build predictions for any other states in the world.

The research design I outline in the following sections using the GDELT dataset not only overcome the shortcomings WikiLeaks-style data, but also those of the extant literature relying on human-coded and pre-GDELT machine-coded datasets. In the following section, I outline how I use GDELT to build state- and sub-state levels of political conflict in Afghanistan and discuss my forecasting approach.

3. Research Design

3.1. Constructing material conflict counts. As previously mentioned, Afghanistan is spatially divided into 32 provinces and 317 sub-provincial-level districts. Using the GDELT data in conjunction with GIS software, I calculate the number of material conflict events that occur from February 1, 2001 through April 30, 2012 between all actors in each month at three (country, province, and
district) geo-spatial levels of analysis. To accomplish this, I first select all material conflict events for which either the source or target actor’s primary affiliation (i.e. the first three characters of their actor identification) was with Afghanistan. I use a version of the GDELT data that has duplicate entries eliminated, as my goal in this chapter is to forecast actual the occurrence of events, rather than the perception or intensity of events. This step generates 139,915 material conflict events, each of which contains a specific latitude and longitude coordinate reflecting where the event occurred. Next, using shape files and GIS software, I calculate the the number of events that occur within each district and province in each month. I choose to use the month as my level of temporal aggregation because this provides sufficient variation throughout the time-series while reducing the level of noise that is present at daily or weekly levels. Largely for those reasons, the monthly level aggregation is the most commonly used in the relevant literature, employed by Goldstein (1991), Schrodt (1997), Schrodt and Gerner (1997), Schrodt and Gerner (2000), Schrodt and Gerner (2001), Shellman (2004a), Shellman (2004b), Gleditsch and Beardsley (2004), Schrodt (2007), Brandt, Colaresi and Freeman (2008), Weidmann and Ward (2010), Ward, Greenhill and Bakke (2010), Shellman, Hatfield and Mills (2010), Brandt, Freeman and Schrodt (2011), D’Orazio, Yonamine and Schrodt (2011), and Mangion-Zammit et al. (2012). District- and province-months with no material conflict events are assigned a “0”. This results in 43,746 district months, 4,352 province months, and 136 country months.\[^2\]

\[\text{INSERT FIGURE 1 HERE}\]

Figure 1 provides a visual overview of the data, illustrating changes in the number of material conflict events from 2001 to 2012 that occur in each district-year.

4. Forecasting Approach

In this section, I outline my forecasting approaches using the univariate data comprised solely of the counts of material conflict events. To facilitate discussion, I detail my forecasting approach as applied to the district-month level-of-analysis, though the approach is identical at the province-month and country-month levels-of-analysis as well. Since the structure of the data is time-series cross sectional at highly nuanced unit of analysis – i.e Afghani districts – I am unable to find

\[^2\] This was done with substantial assistant form John Bieler as well as Josh Steven, who completed all geo-spatial aggregation using GIS.
appropriate exogenous variables to help predict future levels of material conflict. As such, the district-month dataset contains 317 univariate time-series of the count of material conflict events at the district-month level, and I reflect the number of material conflict events occurring in a single district month with the notation $District_{it}$.

Since accurate forecasts are so useful across academia, government, and private sectors, there are many different empirical approaches to building forecasts. No one-size-fits all model exists, and it is impossible to know ahead of time which algorithm will generate the greatest degree of predictive accuracy. Due primarily to the large number of observations and amount of information (i.e. location, actors, date, etc.) contained in most event data datasets, including machine-coded, human-coded, and WikiLeaks data, researchers have applied a large number of different forecasting models.

D’Orazio, Yonamine and Schrodt (2011) report that models forecasting domestic conflict largely fall into three general categories: time series (Shellman (2004a), Shellman (2007), Harff and Gurr (2001)), vector auto regression (VAR) (Pevehouse and Goldstein (1999), Goldstein (1992), Freeman (1989), Brandt, Freeman and Schrodt (2011)), and HMMs (Schrodt (1999), Bond et al. (2004), Shearer (2006), Schrodt (2000), and Schrodt (2006), Petroff, Bond and Bond (2012)). Additionally, other studies using event data have employed additional methods, such as linear models (Weidmann and Ward (2010), Fearon and Laitin (2003), Gurr and Harff (1996)), clustering algorithms (Schrodt and Gerner (2000), and point-process modeling (Mangion-Zammit et al. (2012)). Even after choosing a base algorithm, a number of choices must still be made regarding tuning parameters. For example. In addition, a number of techniques, like bagging and boosting can be applied to most of these algorithms (see Schrodt, Yonamine and Bagozzi (2012) for a discussion of these techniques in the context of political violence forecasting). As if that did not provide enough choices, a number of approaches combine multiple algorithms into model averaging methods, such as bayesian model averaging (BMA) (Montgomery, Hollenbach and Ward (2012)).

Despite the nearly infinite number of plausible forecasting approaches, the structure of my data is highly constraining for two main reasons. First, it is a univariate time series, meaning that it does not contain exogenous covariates. Most of the methods above specifically designed for datasets with many covariates and are less relevant for my data. Second, my data is temporal. This

---

3Exogenous variables on employment and drug prices exist for select districts for select months, but neither variable is available with sufficient coverage to include in an empirical forecasting model at the district-month level. I discuss this further in Section 6.2
restricts how I am able to divide my training and test set, since that training set must exclusively
contain observations that preceded the test set. This greatly inhibits re-sampling techniques like
boosting as a way of enhancing predictive accuracy. In the following section, I outline a forecasting
model that achieves highly accurate predictions using a univariate time-series, discuss my out-of-
sample forecasting framework, and detail how I build a benchmark to assist with evaluating forecast
accuracy.

4.1. The ARFIMA model. To build forecasts with the univariate time-series, I implement an
Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which models all uni-
variate time-series (317 at the district-level, 32 at the province level, and 1 at the country-level)
independently of each other. Though this is the first time an ARFIMA model has been used to fore-
cast political conflict, a number of studies have demonstrated its ability to generate more accurate
and consistent forecasts than other time-series models across various substantive fields. For exam-
ple, Siew, Chin and Wee (2008) demonstrate that an ARFIMA model consistently outperforms a
traditional ARIMA model in forecasting air pollution rates, Chu (2009) generates more accurate
forecasts of tourism levels in Asia with an ARFIMA model than with seasonal ARIMA (SARIMA)
models, Barkoulas and Baum (2006) illustrates how ARFIMA models outperform other autoregres-
sive models in forecasting U.S. monetary indices, and Bhardwaj and Swanson (2006) show that the
ARFIMA model outperforms both ARIMA models and GARCH models in forecasting returns in
the S&P500.

To introduce the ARFIMA model, first consider an ARIMA (p,d,q) model for a univariate time
series \( X(t, x_{t-2}, x_{t-3}, ..., x_{t-n}) \) with \( d=0 \), which we can write as:

\[
\begin{align*}
    x_t &= \omega + \epsilon + \sum_{i=1}^{p} \beta_i x_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i} \\
\end{align*}
\]

(1)

where \( \omega \) is a constant, \( x_{t-i} \) is the lagged dependent variable, \( \epsilon_{t-i} \) is the lagged error, \( \epsilon_t \) is the
current error, and \( \beta_i \) and \( \alpha_i \) are estimated parameters. When a time-series is non-stationary, first-
differencing or “integrating” the series can help achieve stationarity. This generates a new time
series, \( \Delta x_t \), calculated by the following formula:

\[
\Delta x_t = x_t - x_{t-1}
\]

(2)
Thus, we can convert the ARIMA(p,d,q) model with $d=0$ to an ARIMA(p,d,q) model with $d=1$ by replacing the $x$ characters with $\Delta x$, as done in the following formula:

$$\Delta x_t = \omega + \epsilon + \sum_{i=1}^{p} \beta_i \Delta x_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_t + \epsilon_{t-i}$$

Although the ARIMA(p,d,q) model is among the most commonly used time-series models and has been used successfully to forecast with event data (see Shellman (2007)), it is rigid in that $d$ must be an integer. The key innovation of the ARFIMA model is that it allows for $d$ to take on any real number, which need not be an integer (hence the name “fractionally integrated”). Thus, when $d = 0$, the ARFIMA model becomes an ARMA model, and when $d$=any positive integer, the ARFIMA is a simply an ARIMA model. Mathematically, Granger and Joyeux (1980) demonstrates that by allowing $d<1$, the ARFIMA model is able to efficiently account for a long memory process, which occurs when the time-series tends to revert to a historical mean. Parke (1999) provides a thorough explanation of the fractional integration process, and demonstrates how a key innovation of the ARFIMA model is that it allows the effects of past errors on current observations to vary, whereas AR, MA, and ARMA models force this past errors to have uniform effects across the duration of the time-series. Importantly, the ARFIMA model is capable of accounting for the long memory process even without increasing the number of p and q lags.

To implement a flexible ARFIMA(p,d,q) model, I utilize the ‘arfima’ package in r, which automatically establishes values for the p, d, and q parameters of a univariate time series by determining the estimates for these parameters that maximize the likelihood function. This means that the researcher does not need to pre-specify the number of autoregressive components, moving average components, or degree of fractional integration. I treat each cross-section as a unique time-series, meaning that I train and build forecasts with the ARFIMA model one district and one province at a time through a looping function. The ‘forecast’ function in the ‘arfima’ package allows the user to build a prediction N units into the future and provides a mean prediction along with 95% confidence intervals. To establish predictions, I use the mean of the one-month-ahead prediction rounded to the nearest integer. Figure 2 demonstrates the use of the ‘arfima’ package to build a prediction of the number of material conflict events in Bughran province in April, 2009 using data.

\[\text{Many districts have long periods of consecutive months with "0" material conflict events, which causes the 'arfima' package to crash. To allow the 'arfima' package to properly converge, I generate a random number from a uniform distribution from 0 to .1 for each district-month, and add that value to the count of material conflict events.}\]
from February 2001 through March 2009. The prediction in Figure 2 provide the mean (the circle) as well as 90 and 95% confidence intervals, indicated by the light and darker vertical shading.

4.2. **Out-of-sample framework.** In order to calculate out-of-sample performance accuracy of the ARFIMA model, I utilize the same approach implemented by Weidmann and Ward (2010), which I implement on my data according to the steps outlined below, using the district-level model as an example:

- Train the model on an initial in-sample set containing all data from February 2001 until April 2008.
- Predict (and store) the number of material conflict events for May 2008 (i.e. a one-month-ahead out-of-sample forecast).
- Incorporate May 2008 into the in-sample set.
- Retrain the model on this new in-sample set, which now includes all data from February 2001 to May 2008.
- Predict (and store) the number of material conflict events for June 2008.
- Repeat until a final prediction is made for April 2012 (i.e. the last month in the data set), using a model trained on February 2001 through March 2012.

This results in 48 out-of-sample, one-month-ahead forecasts for each of the 317 municipalities. At the province-month level, this approach yields 48 out-of-sample, one-month-in-advance forecasts for each of the 32 provinces, and at the country-month level, this results in 48 one-month-in-advance forecasts for Afghanistan as a whole.

4.3. **Establishing a benchmark.** Since this is the first paper to build nuanced predictions of political conflict in Afghanistan at the monthly level, no existing appropriate benchmark of predictive accuracy exists. Without an appropriate benchmark, it is difficult to assert whether an alternative predictive model is performing well. The literature provides two plausible approaches to assessing how well a predictive model is performing in the absence of other models attempting to predict the same outcome. First, Gurr and Lichbach (1986) provides a strong theoretical argument called “the conflict persistence model”, which suggests that in the absence of an existing benchmark, it is logical to build a naive model that assumes conflict in the future will be the same in a given location as it is today. Second, Mangion-Zammit et al. (2012) reports the percentage of times that
the true number of violent events fall within the 95% and 99% confidence intervals of predicted levels of violence. I choose to follow Gurr and Lichbach (1986)’s approach, and construct a naive model that predicts the number of material conflict events in $\text{District}_{it} = \text{District}_{i,t-1}$, for three reasons.

First, Mangion-Zammit et al. (2012)’s approach tells actually tells us little about a model’s predictive accuracy because it does not penalize for large confidence intervals. Imagine that the true number of violence events occurring in $\text{District}_{it}$ is 75. Now, consider two models. Model 1 generates a prediction for the number of violent events in $\text{District}_{it}$ with 95% confidence intervals at 12 and 162, while Model 2’s prediction for $\text{District}_{it}$has 95% confidence intervals at 68 and 74. Mangion-Zammit et al. (2012)’s approach would report that Model 1 is accurate and Model 2 is inaccurate, when in reality, it is difficult to imagine a scenario in which we would prefer Model 1’s prediction to that of Model 2. Second, and directly related to the first point, is that Gurr and Lichbach (1986) approach generates a specific point prediction as a benchmark, which creates greater flexibility in assessing model performance. For example, Gurr and Lichbach (1986)’s approach allows me to calculate Mean Absolute Error (as detailed below), which is impossible using Mangion-Zammit et al. (2012)’s approach. Lastly, in many forecasting contexts (especially predicting civil conflict at the state-year level), the Gurr and Lichbach (1986) approach achieves almost perfect accuracy – countries at peace tend to stay at peace and countries at conflict tend to stay at conflict. This naive approach often works so well that it occasionally outperforms far more sophisticated forecasting models.

For example, Montgomery, Hollenbach and Ward (2012) introduce Bayesian Model Averaging (BMA) approach, and demonstrate how they are able to leverage the predictions of three separate models in order to build accurate forecasts that outperform all of the three component models. Montgomery, Hollenbach and Ward (2012) report that their BMA technique outperforms all of the three component models, accurately predicting 13 of 35 conflict onsets (“1’s”) and all 313 of the 313 non-onsets (“0’s” ) in their dataset. While these may appear strong at first, Gurr and Lichbach (1986)’s naive benchmark approach accurately predicts 33 of the 35 conflict onsets and 310 of the 313 non-onsets, which is a dramatic improvement over the not only the BMA, but also the three component predictive models. Based on this, I assume that any model that consistently outperforms the naive t=t-1 assumption to be accurate.
4.4. **Calculating accuracy.** For each of the 48 months that iteratively serve as the out-of-sample test, I calculate the error rates for the naive model (naive_error) and the ARFIMA model (arfima_error rate), which reflect the MAE across the N cross-sections (N=317 for the district-month model, N=32 for the province-month model, and N=1 for the country-month model) according to the Formula (4) and Formula (5).

\begin{align*}
\text{naive_error}_m &= \frac{1}{N} \sum_{i=1}^{N} |\text{naive_prediction}_{i,m} - \text{true_count}_{i,m}| \\
\text{arfima_error}_m &= \frac{1}{N} \sum_{i=1}^{N} |\text{naive_prediction}_{i,m} - \text{true_count}_{i,m}|
\end{align*}

These formulas result in a naive_error and arfima_error rate for the district-level, province-level, and country-level models for each of the 48 months that serve as the test-month allowing me to determine the extent to which the ARFIMA model outperforms the naive model across the three levels of geo-spatial aggregation (district, province, and country) in the following section.

5. **Results**

Table 1 provides the arfima_error rate, naive_error rate, and a TRUE/FALSE label indicating whether the ARFIMA forecasts are more accurate on average across all 317 districts for the given month.

[INSERT TABLE 1 HERE]

As Table 1 indicates, the ARFIMA model outperforms the naive model in 47 out of 48 of the out-of-sample months. Additionally, the ARFIMA model reduces the sum of the 48 monthly MAE’s by over 16%. Taken together, these are highly impressive finding, especially when considering that naive models (that assume \(t=t-1\)) of conflict tend to perform well in forecasting.\(^5\)

\(^5\)A potential critique of these results is that I do not perform any rigorous external validity check, meaning that I may simply be predicting the event-data generating process, rather than actual levels of violence. I believe that this is not overly problematic for two main reasons. First, many other forecasting studies likewise rely exclusively on event data and do not perform rigorous external validity checks, which has set a precedent that this is generally accepted practice. Second, the anecdotal story discussed in Chapter 1 serves as an informal external validity check that suggests the GDELT data is accurate.
Table 2 provides the arfima.error rate, naive.error rate, and a TRUE/FALSE label calculated from province-level geo-spatial aggregations, meaning that each of the 48 arfima.error and naive.error rates reflect their respective means across the 32 provinces. At the province-month level, the ARFIMA does not perform as well as at the district-month level, but it still outperforms the naive model in 40 of the 48, or approximately 83% months that serve as the test month. Furthermore, the ARFIMA model reduces the sum of the 48 month MAE by approximately 13%. Even though the ARFIMA performs slightly worse at the province-level than the district-level, it still achieves a respectable level of enhanced accuracy relative to the naive benchmark.

Table 3 replicates Table 1 and Table 2, except it reflects the arfima.error rate, naive.error rate, and the TRUE/FALSE label based on a single country-level forecast per month. Table 3 illustrates that at the country-month level, the ARFIMA still outperforms the naive model, but does so at a lower margin than at the district-month or province-month level. Of the 48 months that test sample, the ARFIMA model outperforms the naive model 30 times, or 62.5%. Additionally, the ARFIMA model generates a lower sum of MAE’s, but only by approximately 1%, which suggests that the increase in predictive accuracy of the ARFIMA model at the country-month level may be largely meaningless.

Across the district-, province-, and country-month forecasts, the key aspect of the ARFIMA model is that it tends to build forecasts that are between the naive model forecast and a longer term moving average. Exactly how much the ARFIMA model shifts forecasts away from the naive forecasts and towards the longer term moving average varies based from by month and by cross-section, but in effect, the ARFIMA acts like a smoothing function. Figure 2 visually demonstrates this. The last observed number of material conflict events is approximately 280 in month 99, meaning that the naive model would predict 280 events for the month 100. However, we can see that the average number of material conflict events in the previous months is less than 280, so the mean ARFIMA forecast (represented by the black dot) is less than 280. To the extent that the ARFIMA model outperforms the naive model, it suggests that levels of future violence tend to exhibit mean reverting characteristics.
6. Future directions

Although the ARFIMA model outlined above largely accomplishes the goal of this paper, in this section I provide preliminary analysis of two logical extensions for the finding in the previous section: first, building features from the univariate time-series to allow for other types of predictive algorithms; second, incorporating exogenous information, such as drug prices.

6.1. Building features and implementing a stacking method. A common approach when building forecasting models is to manipulate existing data in order to build additional features, or covariates, which may uncover meaningful patterns in the data that are hidden in other variables. In many contexts across disciplines, building additional features leads to enhanced predictive accuracy. Note that building features can also decrease predictive accuracy because the additional dimensionality increases the likelihood of over fitting a model. To overcome this, I employ the same out-of-sample predictive framework as previously outline in Section 4.2.

Just like there there is no definitive way to pick the best forecasting algorithm, there are no rules for constructing features. As such, I build 11 new features below, all from the univariate time series, in an attempt to enhance predictive accurate beyond the univariate ARFIMA model outlined in the previous section.

- 2\_month\_MA = (\text{count}_t + \text{count}_{t-1})/2
- 3\_month\_MA = (\text{count}_t + \text{count}_{t-1} + \text{count}_{t-2})/3
- 4\_month\_MA = (\text{count}_t + \text{count}_{t-1} + \text{count}_{t-2} + \text{count}_{t-3})/4
- 5\_month\_MA = (\text{count}_t + \text{count}_{t-1} + \text{count}_{t-2} + \text{count}_{t-3} + \text{count}_{t-4})/5
- 6\_month\_MA = (\text{count}_t + \text{count}_{t-1} + \text{count}_{t-2} + \text{count}_{t-3} + \text{count}_{t-4} + \text{count}_{t-5})/6
- \Delta2\_month\_MA = \text{count}_t - 2\_month\_MA
- \Delta3\_month\_MA = \text{count}_t - 3\_month\_MA
- \Delta4\_month\_MA = \text{count}_t - 4\_month\_MA
- \Delta5\_month\_MA = \text{count}_t - 5\_month\_MA
- \Delta6\_month\_MA = \text{count}_t - 6\_month\_MA
- monthly\_sum = the sum of all material conflict events occurring across all spatial units each month

With these additional covariates, I build a number of additional predictive models following the general approach in Section 4.2. Using the ‘glm’ package in R, I build predictions using linear models
comprised of various combinations of the 11 additional covariates above (all lagged one-unit) as well as a one-unit lag of the dependent variable, trying both “gaussian” and “poisson” distributions. I am unable to find a linear combinations of the covariates above (including the lagged dependent variable) capable of outperforming the naive benchmark at the district-month level in more than 35 out of the 48 district-months that serve as the out-of-sample set. Motivated by the enhanced predictive accuracy of the approach in Montgomery, Hollenbach and Ward (2012), I also implement a stacking approach. To build a stacking prediction, I build use two component models, Model_1 and Model_2, which are specified below and estimated using the ‘glm’ package in R with a gaussian distribution.

\[ (6) \]
\[ \text{Model}_1 \]
\[ \text{District}_{it} = \beta_0 + \beta_1 2\text{month} \_ \text{MA}_{i(t-1)} + \beta_2 3\text{month} \_ \text{MA}_{i(t-1)} + \beta_3 4\text{month} \_ \text{MA}_{i(t-1)} + \beta_4 5\text{month} \_ \text{MA}_{2i(t-1)} + \beta_5 6\text{month} \_ \text{MA}_{i(t-1)} + \beta_7 \text{monthly} \_ \text{sum}_{i(t-1)} + \beta_8 \text{District}_{i(t-1)} \]

\[ (7) \]
\[ \text{Model}_2 \]
\[ \text{District}_{it} = \beta_0 + \beta_1 \Delta 2\text{month} \_ \text{MA}_{i(t-1)} + \beta_2 \Delta 3\text{month} \_ \text{MA}_{i(t-1)} + \beta_3 \Delta 4\text{month} \_ \text{MA}_{i(t-1)} + \beta_4 \Delta 5\text{month} \_ \text{MA}_{2i(t-1)} + \beta_5 \Delta 6\text{month} \_ \text{MA}_{i(t-1)} + \beta_7 \text{District}_{i(t-1)} \]

Using these two models, I build an ensemble forecasting model according to the six steps below:

1. Estimate two models on the same in-sample set as in Section 4.2, which contains all data from February 2001 until April 2008, and generate predictions for these in-sample months and store coefficient estimates
2. Train the Ensemble model using the ‘glm’ function in R on the in-sample predictions from Model_1 and Model_2 according to the formula below, and store coefficient estimates:

---

6I follow the stacking approach suggested by Hastie, Tibshirani and Friedman (2009) on pages 289-290.
7Although the dependent variable is a count, predictions made with the ‘glm’ package using the gaussian distribution consistently outperforms those build with the ‘poisson’ distribution.
8This is conceptually similar to Montgomery2012, but install of updated posteriors, I simply weight each component model based on OLS.
(8) \[ \text{Ensemble} = \hat{\text{District}}_{it} = \beta_0 + \beta_1 \text{Model}_1_{it} + \beta_2 \text{Model}_2_{it} \]

(3) Build predictions for May 2008 (i.e. one-month ahead out-of-sample forecast) for \text{Model}_1\ and \text{Model}_2 by matrix multiplying the coefficient estimates from Step 1 and the covariates for May 2008, which have been lagged one-month to simulate an actual prediction.

(4) Calculate and store an \text{Ensemble} prediction by matrix multiplying the predicted values for \text{Model}_1\ and \text{Model}_2\ by their coefficient estimates from the \text{Ensemble} model trained on the in-sample set in Step 3.

(5) Incorporate May 2008 into the in-sample set.

(6) Repeat Step 1 through Step 4.

(7) Repeat Step 1 through Step 6 until a final prediction is made for April 2012 (i.e. the last month in the data set), using a model trained on February 2001 through March 2012.

This \text{Ensemble} model outperforms the naive benchmark in 33 out of 48 months. Although this is not a terrible result, it does not approach the accuracy of the more straightforward, univariate ARFIMA model discussed in the previous section. However, given the large number of predictive algorithms and the infinite number of features that can be built from a univariate time-series, scholars in the future may be able to build on my ensemble approach and build a model that eventually outperforms the predictive accuracy of my straightforward ARFIMA model.\textsuperscript{9}

6.2. **Incorporating drug prices.** In addition to building features from the univariate time-series as performed in the previous section, another way of potentially improving forecast accuracy is to incorporate exogenous variables. Although a large number of studies have found empirical relationships between many exogenous variables and political conflict, most operate at a state-year level of analysis. Finding relevant exogenous variables at sub-annual and sub-state levels is far more difficult. Even studies that do utilize fine-grained exogenous variables, like Weidmann and Ward (2010) and Berman et al. (2011) face considerable limitations.

For example, Weidmann and Ward (2010) analyze future violence at the municipality-month unit of analysis as a function of past violence as well as a set of exogenous variables comprised of population, ethnic diversity, terrain, and whether the municipality is on an international border.

\textsuperscript{9}I tried additional algorithms, including a number of random forest variations as well as additional combinations of component models within various ensembles, and none enhanced predictive accuracy beyond the ARFIMA model.
However, these exogenous variables vary cross-sectional (i.e. between municipalities) but not temporally (i.e. from month-to-month for the same municipality), which reduces the extent to which they can improve predictive accuracy. Additionally, Berman et al. (2011) collect unemployment statistics at the province-month level for Afghanistan, Iraq, and the Philippines that do vary at a province-month unit of analysis, but the difficulty in collecting such data limit their temporal domain to just six months in the case of Afghanistan, which also inhibits their effectiveness at enhancing predictive models. Therefore, an ideal set of exogenous variables would vary at a fine grained unit of analysis and span a long temporal range, but these are difficult to collect, especially for conflict-prone countries like Afghanistan.

For Afghanistan, one potential source of an exogenous variables come from the Afghanistan Opium Survey 2012, which is published by the United Nations Office on Drugs and Crime (UNODC). This document provides considerable information at the district-level regarding opium and cannabis prices as well a dataset containing average opium prices at the country-month unit of analysis from September 2004 through March 2012, as illustrated below in Figure 3. Unfortunately, similarly complete time-series data are not publicly provided at the province- or district-month level.

Given the number of empirical studies that either theoretically suggest or empirically demonstrate relationships between drug prices and conflict (see Palmer (1994), Buhaug and Gates (2002), Ross (2003), Ross (2004), and Collier, Hoeffler and Soderbom (2004)) it seems reasonable that the addition of opium prices as an exogenous variable may enhance predictive accuracy at the country-month unit of analysis. To test this, I repeat the six steps outlined in Section 4.2 in order to compare the predictive accuracy of the naive model with the original univariate ARFIMA model outlined in Section 4 and Section 6.2 as well as the ARFIMA model that includes the exogenous opium data, which I call the ARFIMA\_opium model. Since the opium price data spans a smaller temporal range than my GDELT-derived data on political violence, I set September 2004 through March 2010 as the initial in-sample training set, and use April 2010 through March 2012 as the out-of-sample test months. As Table 3 indicates, the ARFIMA model outperforms the Naive model in 18 of the 24 months that serve as the out-of-sample test months. Interestingly, the ARFIMA\_opium

---

model only outperforms the Naive model in 17 out of 24 months. Although this suggests that the inclusion of the drug price data may not actually enhance predictive accuracy, it does not rule out the possibility that more nuanced data on drug prices at the province- or district-level of analysis could lead to more accurate predictions.

7. Conclusion

This chapter is the first to build temporally and geo-spatially nuanced forecasts of future levels of violence relying exclusively on open-source, machine coded event data. The release of the GDELT dataset made this chapter possible. Before GDELT, the leading open-source, machine-coded datasets did not provide location information, and the hand-coded datasets that did provide location information were too sparse for rigorous empirical forecasting. The Afghan War Diary that was released as part of WikiLeaks provided a notable exception, but this data is not only of questionable legality but also unlikely to be replicable for future conflicts, meaning that forecasting models built from WikiLeaks data may lack real-world applicability moving forward.\footnote{Standard questioning when applying to positions require top-secret clearance is whether you have accessed and used Wikileaks data.}

Using nothing but GDELT data, I build an ARFIMA model capable of providing forecasts at the district month level that nearly always outperform a naive model that simply assumes that the level of conflict tomorrow will be the same as it is today. My empirical findings suggest three major takeaways: First, it appears that it is feasible to build accurate and nuanced predictions at a sub-state level using only open source, machine-coded event data. Second, the level of forecast accuracy decreased as the degree of geo-spatial aggregation increases: forecasts at the district-month (N=317), province-month (N=32) and country-month (N=1) level outperform their naive benchmarks in 47 out of 48, 40 out of 48, and 30 out of 48 month, respectively. It appears that patterns in violence that are discernible at fine-grained levels of geo-spatial aggregation (i.e. the district-level in Afghanistan) become increasingly noisy at higher levels of geo-spatial aggregation. This strongly suggests that researchers attempting to build empirical forecasts of violence should use as finely grained geo-spatial aggregations as possible. Third, the fact that the ARFIMA model tends to outperform the naive model suggests that patterns of violence tend to be mean reverting. This means that when we see a major spike in violence during a specific period of time in a specific
sub-state location, we should expect violence in the following time period to be more subdued. Conversely, when we see a sudden drop in the level of violence, we should expect a rebound-effect.

Moving forward, a number of logical extensions to this chapter exist. First, researchers could use the GDELT data to further explore whether the mean-reversion properties present in the levels of violence in Afghanistan hold across other countries. Mean-reversion properties, as first identified by Galton (1886) in his seminal analysis of human heights, are common and influential across other substantive fields like biology and economics. Determining whether local levels of violence in other states also tend to be mean-reverting could be a major theoretical advancement to the study of conflict dynamics.

Second, Section 6.1 provides a basic framework for building additional features from the univariate time series and using these features to construct alternative forecasting algorithms to the ARFIMA model. Although my attempts at enhancing predictive accuracy through this approach were unsuccessful, other scholars find greater success by building additional features and experimenting with other predictive algorithms. Similarly, the inclusion of additional exogenous variables, such as drug prices at finer grained spatial coverage than the country-level data modeling in Section 6.2, terrain, or measures of reflecting potential geo-spatial correlation (i.e. a count of the number of conflictual events occurring in neighboring districts or provinces) may also be helpful.

Third, since GDELT provides event data for all countries in the world (as opposed to WikiLeaks, which only provides detailed data for Afghanistan) researcher could apply a similar forecasting model to that outlined in this chapter to build geo-spatially and temporally nuanced forecasts of future levels of violence any number of countries with ongoing domestic conflicts, like India or the Democratic Republic of the Congo.

Lastly, since the GDELT data is updated daily, the forecasting approach outlined in this chapter could be implemented in near real-time. This could provide real-world guidance to a host of potential benefactors, ranging from military leaders hoping to more efficiently allocate resources, to Afghani businessmen trying to identify the safest routes to transport goods. Overall, I hope that this chapters serves as a foundation for further forecasting efforts at fine-grained temporal and geo-spatial scales.


Schrodt, Philip A. 2012. “Response to BBN evaluations of TABARI.”.

138


### Table 1. Assessing Accuracy at the District Level

<table>
<thead>
<tr>
<th>m</th>
<th>month</th>
<th>arfima_error</th>
<th>naive_error</th>
<th>arfima_error &lt; naive_error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>May 2008</td>
<td>2.73</td>
<td>3.59</td>
<td>TRUE</td>
</tr>
<tr>
<td>2</td>
<td>June 2008</td>
<td>2.75</td>
<td>3.21</td>
<td>TRUE</td>
</tr>
<tr>
<td>3</td>
<td>July 2008</td>
<td>3.76</td>
<td>4.41</td>
<td>TRUE</td>
</tr>
<tr>
<td>4</td>
<td>August 2008</td>
<td>2.40</td>
<td>3.66</td>
<td>TRUE</td>
</tr>
<tr>
<td>5</td>
<td>September 2008</td>
<td>2.58</td>
<td>3.13</td>
<td>TRUE</td>
</tr>
<tr>
<td>6</td>
<td>October 2008</td>
<td>3.51</td>
<td>3.60</td>
<td>TRUE</td>
</tr>
<tr>
<td>7</td>
<td>November 2008</td>
<td>2.24</td>
<td>2.87</td>
<td>TRUE</td>
</tr>
<tr>
<td>8</td>
<td>December 2008</td>
<td>1.71</td>
<td>2.58</td>
<td>TRUE</td>
</tr>
<tr>
<td>9</td>
<td>January 2009</td>
<td><strong>2.72</strong></td>
<td><strong>2.63</strong></td>
<td><strong>FALSE</strong></td>
</tr>
<tr>
<td>10</td>
<td>February 2009</td>
<td>2.11</td>
<td>2.53</td>
<td>TRUE</td>
</tr>
<tr>
<td>11</td>
<td>March 2009</td>
<td>2.53</td>
<td>3.11</td>
<td>TRUE</td>
</tr>
<tr>
<td>12</td>
<td>April 2009</td>
<td>2.43</td>
<td>2.85</td>
<td>TRUE</td>
</tr>
<tr>
<td>13</td>
<td>May 2009</td>
<td>3.58</td>
<td>4.02</td>
<td>TRUE</td>
</tr>
<tr>
<td>14</td>
<td>June 2009</td>
<td>2.85</td>
<td>4.02</td>
<td>TRUE</td>
</tr>
<tr>
<td>15</td>
<td>July 2009</td>
<td>3.56</td>
<td>4.07</td>
<td>TRUE</td>
</tr>
<tr>
<td>16</td>
<td>August 2009</td>
<td>5.50</td>
<td>6.25</td>
<td>TRUE</td>
</tr>
<tr>
<td>17</td>
<td>September 2009</td>
<td>4.15</td>
<td>5.00</td>
<td>TRUE</td>
</tr>
<tr>
<td>18</td>
<td>October 2009</td>
<td>4.18</td>
<td>4.79</td>
<td>TRUE</td>
</tr>
<tr>
<td>19</td>
<td>November 2009</td>
<td>3.38</td>
<td>4.09</td>
<td>TRUE</td>
</tr>
<tr>
<td>20</td>
<td>December 2009</td>
<td>3.07</td>
<td>3.09</td>
<td>TRUE</td>
</tr>
<tr>
<td>21</td>
<td>January 2010</td>
<td>2.50</td>
<td>3.62</td>
<td>TRUE</td>
</tr>
<tr>
<td>22</td>
<td>February 2010</td>
<td>3.82</td>
<td>4.11</td>
<td>TRUE</td>
</tr>
<tr>
<td>23</td>
<td>March 2010</td>
<td>3.37</td>
<td>4.15</td>
<td>TRUE</td>
</tr>
<tr>
<td>24</td>
<td>April 2010</td>
<td>1.66</td>
<td>1.70</td>
<td>TRUE</td>
</tr>
<tr>
<td>25</td>
<td>May 2010</td>
<td>1.99</td>
<td>2.12</td>
<td>TRUE</td>
</tr>
<tr>
<td>26</td>
<td>June 2010</td>
<td>2.02</td>
<td>2.25</td>
<td>TRUE</td>
</tr>
<tr>
<td>27</td>
<td>July 2010</td>
<td>1.77</td>
<td>2.09</td>
<td>TRUE</td>
</tr>
<tr>
<td>28</td>
<td>August 2010</td>
<td>3.63</td>
<td>4.06</td>
<td>TRUE</td>
</tr>
<tr>
<td>29</td>
<td>September 2010</td>
<td>3.28</td>
<td>3.28</td>
<td>TRUE</td>
</tr>
<tr>
<td>30</td>
<td>October 2010</td>
<td>2.05</td>
<td>2.57</td>
<td>TRUE</td>
</tr>
<tr>
<td>31</td>
<td>November 2010</td>
<td>1.77</td>
<td>2.22</td>
<td>TRUE</td>
</tr>
<tr>
<td>32</td>
<td>December 2010</td>
<td>2.00</td>
<td>2.31</td>
<td>TRUE</td>
</tr>
<tr>
<td>33</td>
<td>January 2011</td>
<td>2.02</td>
<td>2.44</td>
<td>TRUE</td>
</tr>
<tr>
<td>34</td>
<td>February 2011</td>
<td>1.91</td>
<td>2.34</td>
<td>TRUE</td>
</tr>
<tr>
<td>35</td>
<td>March 2011</td>
<td>1.89</td>
<td>2.18</td>
<td>TRUE</td>
</tr>
<tr>
<td>36</td>
<td>April 2011</td>
<td>3.47</td>
<td>4.03</td>
<td>TRUE</td>
</tr>
<tr>
<td>37</td>
<td>May 2011</td>
<td>2.91</td>
<td>3.64</td>
<td>TRUE</td>
</tr>
<tr>
<td>38</td>
<td>June 2011</td>
<td>2.32</td>
<td>3.07</td>
<td>TRUE</td>
</tr>
<tr>
<td>39</td>
<td>July 2011</td>
<td>3.06</td>
<td>3.60</td>
<td>TRUE</td>
</tr>
<tr>
<td>40</td>
<td>August 2011</td>
<td>2.64</td>
<td>3.30</td>
<td>TRUE</td>
</tr>
<tr>
<td>41</td>
<td>September 2011</td>
<td>3.02</td>
<td>3.47</td>
<td>TRUE</td>
</tr>
<tr>
<td>42</td>
<td>October 2011</td>
<td>1.97</td>
<td>2.79</td>
<td>TRUE</td>
</tr>
<tr>
<td>43</td>
<td>November 2011</td>
<td>2.32</td>
<td>2.68</td>
<td>TRUE</td>
</tr>
<tr>
<td>44</td>
<td>December 2011</td>
<td>1.81</td>
<td>2.10</td>
<td>TRUE</td>
</tr>
<tr>
<td>45</td>
<td>January 2012</td>
<td>2.21</td>
<td>2.42</td>
<td>TRUE</td>
</tr>
<tr>
<td>46</td>
<td>February 2012</td>
<td>2.09</td>
<td>2.41</td>
<td>TRUE</td>
</tr>
<tr>
<td>47</td>
<td>March 2012</td>
<td>2.81</td>
<td>3.08</td>
<td>TRUE</td>
</tr>
<tr>
<td>48</td>
<td>April 2012</td>
<td>2.98</td>
<td>3.58</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Total: May 2008 - Apr 2012 129.76 155.07 47 TRUE, 1 FALSE
Table 2. Assessing Accuracy at the Province Level

<table>
<thead>
<tr>
<th>m</th>
<th>month</th>
<th>arma_error</th>
<th>naive_error</th>
<th>arma_error &lt; naive_error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>May 2008</td>
<td>16.16</td>
<td>23.69</td>
<td>TRUE</td>
</tr>
<tr>
<td>2</td>
<td>June 2008</td>
<td>22.34</td>
<td>21.53</td>
<td>FALSE</td>
</tr>
<tr>
<td>3</td>
<td>July 2008</td>
<td>29.16</td>
<td>33.63</td>
<td>TRUE</td>
</tr>
<tr>
<td>4</td>
<td>August 2008</td>
<td>21.13</td>
<td>27.69</td>
<td>TRUE</td>
</tr>
<tr>
<td>5</td>
<td>September 2008</td>
<td>15.38</td>
<td>17.69</td>
<td>TRUE</td>
</tr>
<tr>
<td>6</td>
<td>October 2008</td>
<td>24.16</td>
<td>24.09</td>
<td>FALSE</td>
</tr>
<tr>
<td>7</td>
<td>November 2008</td>
<td>12.78</td>
<td>18.59</td>
<td>TRUE</td>
</tr>
<tr>
<td>8</td>
<td>December 2008</td>
<td>6.91</td>
<td>14.59</td>
<td>TRUE</td>
</tr>
<tr>
<td>9</td>
<td>January 2009</td>
<td>18.03</td>
<td>17.06</td>
<td>FALSE</td>
</tr>
<tr>
<td>10</td>
<td>February 2009</td>
<td>12.69</td>
<td>15.53</td>
<td>TRUE</td>
</tr>
<tr>
<td>11</td>
<td>March 2009</td>
<td>18.84</td>
<td>22.5</td>
<td>TRUE</td>
</tr>
<tr>
<td>12</td>
<td>April 2009</td>
<td>11.00</td>
<td>15.53</td>
<td>TRUE</td>
</tr>
<tr>
<td>13</td>
<td>May 2009</td>
<td>28.75</td>
<td>30.41</td>
<td>TRUE</td>
</tr>
<tr>
<td>14</td>
<td>June 2009</td>
<td>21.69</td>
<td>28.50</td>
<td>TRUE</td>
</tr>
<tr>
<td>15</td>
<td>July 2009</td>
<td>31.91</td>
<td>31.22</td>
<td>FALSE</td>
</tr>
<tr>
<td>16</td>
<td>August 2009</td>
<td>40.75</td>
<td>44.25</td>
<td>TRUE</td>
</tr>
<tr>
<td>17</td>
<td>September 2009</td>
<td>21.31</td>
<td>28.22</td>
<td>TRUE</td>
</tr>
<tr>
<td>18</td>
<td>October 2009</td>
<td>36.44</td>
<td>36.16</td>
<td>FALSE</td>
</tr>
<tr>
<td>19</td>
<td>November 2009</td>
<td>21.56</td>
<td>32.49</td>
<td>TRUE</td>
</tr>
<tr>
<td>20</td>
<td>December 2009</td>
<td>21.34</td>
<td>19.31</td>
<td>FALSE</td>
</tr>
<tr>
<td>21</td>
<td>January 2010</td>
<td>18.19</td>
<td>26.91</td>
<td>TRUE</td>
</tr>
<tr>
<td>22</td>
<td>February 2010</td>
<td>23.63</td>
<td>27.88</td>
<td>TRUE</td>
</tr>
<tr>
<td>23</td>
<td>March 2010</td>
<td>35.38</td>
<td>36.16</td>
<td>TRUE</td>
</tr>
<tr>
<td>24</td>
<td>April 2010</td>
<td>19.00</td>
<td>10.38</td>
<td>FALSE</td>
</tr>
<tr>
<td>25</td>
<td>May 2010</td>
<td>12.94</td>
<td>14.66</td>
<td>TRUE</td>
</tr>
<tr>
<td>26</td>
<td>June 2010</td>
<td>14.38</td>
<td>17.13</td>
<td>TRUE</td>
</tr>
<tr>
<td>27</td>
<td>July 2010</td>
<td>13.47</td>
<td>14.59</td>
<td>TRUE</td>
</tr>
<tr>
<td>28</td>
<td>August 2010</td>
<td>32.03</td>
<td>33.00</td>
<td>TRUE</td>
</tr>
<tr>
<td>29</td>
<td>September 2010</td>
<td>8.75</td>
<td>19.66</td>
<td>TRUE</td>
</tr>
<tr>
<td>30</td>
<td>October 2010</td>
<td>12.25</td>
<td>12.91</td>
<td>TRUE</td>
</tr>
<tr>
<td>31</td>
<td>November 2010</td>
<td>12.81</td>
<td>13.94</td>
<td>TRUE</td>
</tr>
<tr>
<td>32</td>
<td>December 2010</td>
<td>11.75</td>
<td>14.75</td>
<td>TRUE</td>
</tr>
<tr>
<td>33</td>
<td>January 2011</td>
<td>14.34</td>
<td>16.13</td>
<td>TRUE</td>
</tr>
<tr>
<td>34</td>
<td>February 2011</td>
<td>17.13</td>
<td>17.53</td>
<td>TRUE</td>
</tr>
<tr>
<td>35</td>
<td>March 2011</td>
<td>12.94</td>
<td>16.25</td>
<td>TRUE</td>
</tr>
<tr>
<td>36</td>
<td>April 2011</td>
<td>26.22</td>
<td>31.66</td>
<td>TRUE</td>
</tr>
<tr>
<td>37</td>
<td>May 2011</td>
<td>21.19</td>
<td>27.19</td>
<td>TRUE</td>
</tr>
<tr>
<td>38</td>
<td>June 2011</td>
<td>12.81</td>
<td>18.56</td>
<td>TRUE</td>
</tr>
<tr>
<td>39</td>
<td>July 2011</td>
<td>20.25</td>
<td>25.56</td>
<td>TRUE</td>
</tr>
<tr>
<td>40</td>
<td>August 2011</td>
<td>20.29</td>
<td>24.03</td>
<td>TRUE</td>
</tr>
<tr>
<td>41</td>
<td>September 2011</td>
<td>24.19</td>
<td>26.69</td>
<td>TRUE</td>
</tr>
<tr>
<td>42</td>
<td>October 2011</td>
<td>14.97</td>
<td>19.41</td>
<td>TRUE</td>
</tr>
<tr>
<td>43</td>
<td>November 2011</td>
<td>16.72</td>
<td>20.63</td>
<td>TRUE</td>
</tr>
<tr>
<td>44</td>
<td>December 2011</td>
<td>12.75</td>
<td>14.97</td>
<td>TRUE</td>
</tr>
<tr>
<td>45</td>
<td>January 2012</td>
<td>14.88</td>
<td>16.34</td>
<td>TRUE</td>
</tr>
<tr>
<td>46</td>
<td>February 2012</td>
<td>14.31</td>
<td>15.97</td>
<td>TRUE</td>
</tr>
<tr>
<td>47</td>
<td>March 2012</td>
<td>21.72</td>
<td>24.24</td>
<td>TRUE</td>
</tr>
<tr>
<td>48</td>
<td>April 2012</td>
<td>93.09</td>
<td>91.66</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

Total: May 2008 - Apr 2012  1,004.56  1,151.50  40 TRUE, 8 FALSE
<table>
<thead>
<tr>
<th>m</th>
<th>month</th>
<th>arfima_error</th>
<th>naive_error</th>
<th>arfima_error &lt; naive_error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>May 2008</td>
<td>52</td>
<td>94</td>
<td>TRUE</td>
</tr>
<tr>
<td>2</td>
<td>June 2008</td>
<td>462</td>
<td>393</td>
<td>FALSE</td>
</tr>
<tr>
<td>3</td>
<td>July 2008</td>
<td>474</td>
<td>410</td>
<td>FALSE</td>
</tr>
<tr>
<td>4</td>
<td>August 2008</td>
<td>358</td>
<td>426</td>
<td>TRUE</td>
</tr>
<tr>
<td>5</td>
<td>September 2008</td>
<td>96</td>
<td>238</td>
<td>TRUE</td>
</tr>
<tr>
<td>6</td>
<td>October 2008</td>
<td>383</td>
<td>277</td>
<td>FALSE</td>
</tr>
<tr>
<td>7</td>
<td>November 2008</td>
<td>5</td>
<td>81</td>
<td>TRUE</td>
</tr>
<tr>
<td>8</td>
<td>December 2008</td>
<td>135</td>
<td>231</td>
<td>TRUE</td>
</tr>
<tr>
<td>9</td>
<td>January 2009</td>
<td>293</td>
<td>204</td>
<td>FALSE</td>
</tr>
<tr>
<td>10</td>
<td>February 2009</td>
<td>39</td>
<td>25</td>
<td>FALSE</td>
</tr>
<tr>
<td>11</td>
<td>March 2009</td>
<td>156</td>
<td>78</td>
<td>FALSE</td>
</tr>
<tr>
<td>12</td>
<td>April 2009</td>
<td>51</td>
<td>23</td>
<td>FALSE</td>
</tr>
<tr>
<td>13</td>
<td>May 2009</td>
<td>709</td>
<td>629</td>
<td>FALSE</td>
</tr>
<tr>
<td>14</td>
<td>June 2009</td>
<td>393</td>
<td>444</td>
<td>TRUE</td>
</tr>
<tr>
<td>15</td>
<td>July 2009</td>
<td>455</td>
<td>309</td>
<td>FALSE</td>
</tr>
<tr>
<td>16</td>
<td>August 2009</td>
<td>845</td>
<td>754</td>
<td>FALSE</td>
</tr>
<tr>
<td>17</td>
<td>September 2009</td>
<td>135</td>
<td>227</td>
<td>TRUE</td>
</tr>
<tr>
<td>18</td>
<td>October 2009</td>
<td>220</td>
<td>311</td>
<td>TRUE</td>
</tr>
<tr>
<td>19</td>
<td>November 2009</td>
<td>235</td>
<td>417</td>
<td>TRUE</td>
</tr>
<tr>
<td>20</td>
<td>December 2009</td>
<td>484</td>
<td>336</td>
<td>FALSE</td>
</tr>
<tr>
<td>21</td>
<td>January 2010</td>
<td>633</td>
<td>693</td>
<td>TRUE</td>
</tr>
<tr>
<td>22</td>
<td>February 2010</td>
<td>71</td>
<td>20</td>
<td>TRUE</td>
</tr>
<tr>
<td>23</td>
<td>March 2010</td>
<td>1,035</td>
<td>1,087</td>
<td>TRUE</td>
</tr>
<tr>
<td>24</td>
<td>April 2010</td>
<td>213</td>
<td>244</td>
<td>TRUE</td>
</tr>
<tr>
<td>25</td>
<td>May 2010</td>
<td>244</td>
<td>259</td>
<td>TRUE</td>
</tr>
<tr>
<td>26</td>
<td>June 2010</td>
<td>137</td>
<td>146</td>
<td>TRUE</td>
</tr>
<tr>
<td>27</td>
<td>July 2010</td>
<td>139</td>
<td>143</td>
<td>TRUE</td>
</tr>
<tr>
<td>28</td>
<td>August 2010</td>
<td>1,037</td>
<td>1,028</td>
<td>FALSE</td>
</tr>
<tr>
<td>29</td>
<td>September 2010</td>
<td>515</td>
<td>519</td>
<td>TRUE</td>
</tr>
<tr>
<td>30</td>
<td>October 2010</td>
<td>85</td>
<td>85</td>
<td>TRUE</td>
</tr>
<tr>
<td>31</td>
<td>November 2010</td>
<td>170</td>
<td>200</td>
<td>TRUE</td>
</tr>
<tr>
<td>32</td>
<td>December 2010</td>
<td>185</td>
<td>174</td>
<td>FALSE</td>
</tr>
<tr>
<td>33</td>
<td>January 2011</td>
<td>160</td>
<td>204</td>
<td>TRUE</td>
</tr>
<tr>
<td>34</td>
<td>February 2011</td>
<td>220</td>
<td>223</td>
<td>TRUE</td>
</tr>
<tr>
<td>35</td>
<td>March 2011</td>
<td>253</td>
<td>270</td>
<td>TRUE</td>
</tr>
<tr>
<td>36</td>
<td>April 2011</td>
<td>539</td>
<td>507</td>
<td>FALSE</td>
</tr>
<tr>
<td>37</td>
<td>May 2011</td>
<td>144</td>
<td>66</td>
<td>FALSE</td>
</tr>
<tr>
<td>38</td>
<td>June 2011</td>
<td>135</td>
<td>230</td>
<td>TRUE</td>
</tr>
<tr>
<td>39</td>
<td>July 2011</td>
<td>386</td>
<td>298</td>
<td>FALSE</td>
</tr>
<tr>
<td>40</td>
<td>August 2011</td>
<td>126</td>
<td>33</td>
<td>FALSE</td>
</tr>
<tr>
<td>41</td>
<td>September 2011</td>
<td>120</td>
<td>202</td>
<td>TRUE</td>
</tr>
<tr>
<td>42</td>
<td>October 2011</td>
<td>314</td>
<td>371</td>
<td>TRUE</td>
</tr>
<tr>
<td>43</td>
<td>November 2011</td>
<td>214</td>
<td>266</td>
<td>TRUE</td>
</tr>
<tr>
<td>44</td>
<td>December 2011</td>
<td>55</td>
<td>57</td>
<td>TRUE</td>
</tr>
<tr>
<td>45</td>
<td>January 2012</td>
<td>5</td>
<td>23</td>
<td>TRUE</td>
</tr>
<tr>
<td>46</td>
<td>February 2012</td>
<td>112</td>
<td>145</td>
<td>TRUE</td>
</tr>
<tr>
<td>47</td>
<td>March 2012</td>
<td>453</td>
<td>475</td>
<td>TRUE</td>
</tr>
<tr>
<td>48</td>
<td>April 2012</td>
<td>2,759</td>
<td>2,737</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

Total: May 2008 - Apr 2012 16,439 16,612 30 TRUE, 18 FALSE
Figure 1. The Number of Material Conflict events per Afghani District from 2001 to 2012
Figure 2. One-month Forecast of the of Material Conflict Events in Bughran District using ‘arfima’ package, with mean, 90%, and 95% confidence intervals.
Figure 3. Average Farm-Gate Prices for Dry Opium in Afghanistan, September 2004-March 2012
CHAPTER 5. CONCLUDING REMARKS

1. Conclusion

Today is January 31, 2013. Five years ago, this dissertation would have been impossible. In fact, it would have been impossible one year ago as well, and not because of some theoretical advancement or statistical breakthrough, but rather due to a lack of requisite data. More specifically, the lack of an event data dataset sufficiently broad in spatial and temporal coverage but detailed in sub-state coverage and geo-spatial coding. In the introduction to this dissertation, I write that, “This dissertation is fundamentally about the study of political violence.” While this statement is true, the academic study of political violence has been fundamentally about data. Without data, we are unable to empirically falsify hypotheses, which serve as the cornerstone of scientific progress, and without continued innovation in tools used to generate data, we are unable to ask progressively nuanced questions and perform increasingly rigorous empirical testing.

If we think of the evolution of event data in waves, then the first wave occurred with the advent of the early hand coded efforts like WEIS and COPDAB in the 1970s. The second wave occurred in the 1990s when KEDS and other projects automated the event data generating process. The third wave happened in the 2000s with projects like the 10 Million International Dyadic Events and ICEWS dramatically expanded sub-state nuance and geo-spatial coverage. Finally, the introduction of GDELT in 2012 marked the fourth wave, which finally made this dissertation possible. The GDELT dataset made two major advancements to the leading event data datasets.

First, whereas the 10 Million International Dyadic Events contained global coverage, it does not have detailed information about sub-state actors. Conversely, the ICEWS dataset has tremendously nuanced information about sub-state actors, but geo-spatial coverage limited to 29 Asian countries and the temporal range is limited. GDELT combined these two strengths but providing ICEWS level sub-state actor nuance with 10 million dyadic event-level global coverage. Second, GDELT is the first open source, machine-coded event data dataset that provides information about the approximate location (in latitude and longitude coordinates) where each event occurs. I use the GDELT dataset to construct indicators of political violence in all three substantive chapters, meaning that the use of the GDELT data is one of the unifying threads of this dissertation.
The other consistent theme throughout the three substantive chapters is a focus on political violence in general, and the predictability of violence more specifically. The rationale behind my emphasis on prediction over explanation is straightforward and by no means novel. Drawing on the arguments of leading scholars past and present, I assert that models focusing on explanation of political violence (which dwarf the number of models attempting to predict) limit scholars ability to separate relationships that are only true of the past with those likely to maintain into the future. Without doubt, understanding the past for the sake of general knowledge is clearly a worthwhile task, but it is one that should be left to historians. As a political scientist, I am interested in identifying trends likely to hold in the future. This shaped the methodological structure of my substantive chapters to emphasize prediction over explanation.

Substantively, the questions that I address in this dissertation are fairly basic and focus on the costs, causes, and the predictability of political violence. Despite the seemingly straightforward nature of these questions, this dissertation makes two major contributions to the extant literature. First, it demonstrate how the GDELT dataset allows us to explain and predict political violence in ways not previously possible. Second, my empirical findings suggest that we should increasingly be moving to more and more disaggregated analyses of political violence.

1.1. **Empirical findings and future avenues of research.** Historically, empirical studies of political violence have tended to rely primarily on high-level aggregations, often at the state-year level. The logic has been that data at finer grained levels of nuance is noisy, but as you aggregate up to higher levels of analysis (like the state), interactions become dictated by rules and laws that can be identified through scientific methods. Through the words of his famous detective, Sir Arthur Conan Doyle (1890) succinctly summarizes this line of thinking:

> “Winwood Reade is good upon the subject,” said Holmes. ”He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.”

The findings in my three substantive chapters suggest an alternative to Winwood Reade’s argument – that explanatory and predictive power can be lost when combining nuanced level units
of analysis into an aggregate level. Below, I briefly summarize my findings and make the case for disaggregation.

In Chapter 2, I focus on the financial consequences of political violence. Whereas the majority of studies interested in measuring the costs of political violence focus on macro-level dependent variables like trade or GDP, I concentrate on the smaller scale effects of violent on equities markets. Building on the extant literature, I analyze the extent to which conflictual events targeting Israel effect variance in returns of the TASE. Although many studies have focused on the effects of political conflict on equity markets in politically stable countries (like the New York Stock Exchange or the London Stock Exchange), little literature has addressed the effects of political violence on equity markets operating in highly conflictual countries. Logically, arguments cold be made in either direction. On the one hand, equity markets might not meaningfully respond to variation in violence because violence is so common that it may already priced into the value of the asset or it might simply not affect corporate profitability. On the other hand, financial markets could be highly responsive to violence since traders may have developed distinct strategies for trading around violence.

I attempt to fill this gap in the existing literature by analyzing the extent to which conflictual events targeting Israel affects variance of returns of equities traded on the TASE. To do so, I use the GDELT dataset to construct a daily-level measure reflecting the intensity of violence committed against Israeli territories and actors. Using this conflict data and a GARCH model, I first find that contrary to results in much of the extant literature focusing on equity markets in stable countries like the United State and Britain, variation in the intensity of violence against Israel does not have a statistically meaningful effect on variance in returns in the TA-100 index (an index of the largest 100 companies traded on the Tel Aviv Stock Exchange). Additionally, I repeat the analyses but utilize returns of the two largest insurance companies on the TASE, ticker symbols MGDL and CLIS, instead of the aggregate TA-100 index. For these equities, I find relatively consistent support that political violence is significantly driving variance in returns. In the context of Israel, these findings suggest that the overall impact of political violence on equity returns may be insignificant, but certain companies or sectors – like insurance – do meaningfully respond. In terms of affecting conflict dynamics, these findings suggest that Israeli policy makers do not need to be overly worried about the effects of political conflict on financial markets, since the risk of future violence seems to already be accounted for in current equity prices. This is in line with the efficient market hypothesis.
This finding speaks directly to the role of aggregation. When I use an indicator that reflects the cumulative average returns of 100 equities, the effects of violence are insignificant, but at the individual equity level, violence does seem to drive variance in returns. That the effects of violence disproportionately effects different sectors of the economy is a finding largely matching our intuitive expectation, but one that is impossible to detect when analyzing cross-sector aggregated indices like the TA-100 index.

However, two caveats to my findings warrant mention. First, drawing on both theory and the extant literature, I analyze daily-level data and focus only on conflictual event. It is feasible that different research design choices could to different empirical findings. For example, working with weekly level averages (as opposed to daily level data) or the inclusion of threats of violence (as opposed to focusing exclusively on the actual violent events) could affect findings. Second, it is possible that the timing of violent attacks against Israel matters. For example, investors may respond to the first attacks of a new conflict, but ignore later fighting, washing out the empirical effects of the initial response. Taken together, these three caveats suggest that it is highly likely that some equities respond to some operationalizations of violence some of the time. Hopefully, the research design and results in this chapter can serve as a foundation for future research to test for more nuanced relationships between political violence and equity markets returns. It is important to note that both of these proposed topics of future research could be tested with the currently available data.

In Chapter 3, I analyze the relationship between domestic conflict and inter-state conflict. Despite the large number of case studies suggesting that domestic conflict has an effect on interstate conflict and considerable number of studies arguing for theoretical linkages between domestic and international politics, this topic has received relatively little empirical attention. Moreover, studies that do attempt to empirical measure how domestic conflict affect interstate conflicts tend to rely on crude, yearly level aggression which eliminate all potentially meaningful sub-annual level variation.

To overcome the shortcomings of relying on crude temporal aggregations and provide a more rigorous test of the relationships between domestic and inter-state conflicts, I use the GDELT dataset, which allows me to construct various measures reflecting the onset and intensity of both domestic and interstate conflicts with global coverage for the first time. I then convert the measures into dyad-month level analyses, which allow me to empirically test a number of relationships,
including whether onsets of domestic conflict affect the likelihood of an onset of interstate conflict, and whether variation in the intensity of ongoing domestic conflict affects onsets of variation in intensity of ongoing interstate conflicts.

I find strong empirical support that domestic conflict onsets in one or both countries comprising a dyad dramatically increases the likelihood of an onset of a dyadic conflict. This finding is considerably stronger and more consistent than much existing literature interested in the effects of domestic conditions on interstate conflict. Additionally, I find that the intensity of domestic conflicts increases, the likelihood of an interstate conflict further increase. Lastly, and perhaps most interesting, I find weak support that as levels of domestic conflict increases in two states in a dyad, the level of ongoing interstate conflict tends to reduce. The strength of my findings in Chapter 3 demonstrate the benefit of disaggregation by finding trends that would otherwise be indiscernible using the more commonly applied binary, annual level conflict data.

While I empirically demonstrated that these relationships between domestic and interstate conflict exist, and I believe that the next logical test is to test for hypotheses that propose why they exist. Again, the major obstacle to asking these more difficult why questions has been a lack of data. However, as I attempt to highlight throughout this Chapter, the 200 million (and counting) events in the GDELT dataset make it possible to ask increasingly nuanced questions, and I am confident that moving forward, scholars will be able to use GDELT to isolate specific causal mechanisms – such as intervention or diversion – that may be responsible for the strong effect that domestic conflicts seems to have on interstate conflicts.

In Chapter 4, I build the first temporally and geo-spatially nuanced forecasts of future levels of violence relying exclusively on open-source, machine coded event data. The release of the GDELT dataset made this chapter possible. Before GDELT, the leading open-source, machine-coded datasets did not provide location information, and the hand-coded datasets that did provide location information were too sparse for rigorous empirical forecasting. The Afghan War Diary that was released as part of WikiLeaks provided a notable exception, but this data is not only of questionable legality but also unlikely to be replicable for future conflicts, meaning that forecasting models built from WikiLeaks data may lack real-world applicability moving forward.

Using nothing but GDELT data, I build an ARFIMA model capable of providing forecasts at the district month level that nearly always outperform a naive model that simply assumes that the level of conflict tomorrow will be the same as it is today. My empirical findings suggests three
major takeaways: First, it appears that it is feasible to build accurate and nuanced predictions at a sub-state level using only open source, machine-coded event data. Second, the level of forecast accuracy decreased as the degree of geo-spatial aggregation increases: forecasts at the district-month (N=317), province-month (N=32) and country-month (N=1) level outperform their naive benchmarks in 47 out of 48, 40 out of 48, and 30 out of 48 month, respectively. It appears that patterns in violence that are discernible at fine-grained levels of geo-spatial aggregation (i.e. the district-level in Afghanistan) become increasing noisy a higher levels of geo-spatial aggregation. This strongly suggests that researchers attempting to build empirical forecasts of violence should use as finely grained geo-spatial aggregations as possible. Of all the empirical findings in the preceding chapters, this is the clearest example of the benefits of disaggregation. Third, the fact that the ARFIMA model tends to outperform the naive model suggests that patterns of violence tend to be mean reverting. This means that when we see a major spike in violence during a specific period of time in a specific sub-state location, we should expect violence in the following time period to be more subdued. Conversely, when we see a sudden drop in the level of violence, we should expect a rebound-effect.

Moving forward, a number of logical extensions to this chapter exist. First, researchers could use the GDELT data to further explore whether the mean-reversion properties present in the levels of violence in Afghanistan hold across other countries. Mean-reversion properties are common across other substantive fields like biology and economics. Moreover, the inertia of conflict (regions at conflict tend to stay at conflict, whereas regions at peace tend to stay at peace) suggests that mean-reversion may be common. Determining whether local levels of violence in other states also tend to be mean-reverting could be a major theoretical advancement to the study of conflict dynamics.

Second, I provide a basic framework for building additional features from the univariate time series and using these features to construct alternative forecasting algorithms to the ARFIMA model. Although my attempts at enhancing predictive accuracy through this approach were unsuccessful, other scholars may find greater success by building additional features and experimenting with other predictive algorithms. Similarly, although I found that the inclusion of drug prices at the country level did not enhance predictive accuracy, the inclusion of additional exogenous variables, including drug prices at finer grained spatial coverage like the district- or province-level, or measures of reflecting potential geo-spatial correlation (i.e. a count of the number of conflictual events occurring in neighboring districts or provinces) may also be helpful.
Third, since GDELT provides event data for all countries in the world (as opposed to WikiLeaks, which only provides detailed data for Afghanistan) researcher could apply a similar forecasting model to that outlined in this chapter to build geo-spatially and temporally nuanced forecasts of future levels of violence any number of countries with ongoing domestic conflicts, like India, the Democratic Republic of the Congo, or Iraq.

Lastly, since the GDELT data is updated daily, the forecasting approach outlined in this chapter could be implemented in near real-time. This could provide real-world guidance to a host of potential beneficiaries, ranging from military leaders hoping to more efficiently allocate resources, to Afghani businessmen trying to identify the safest routes to transport goods. Overall, I hope that this chapter serves as a foundation for further forecasting efforts at fine-grained temporal and geo-spatial scales.

1.2. The future of event data and political violence research. For decades, the empirical study of political violence has been constrained by data. State-year aggregations, coarse definitions of “war” or “disputes”, and limited country coverage are no longer limiting factors. GDELT is not the end of the story, it is simply a new dataset that signals the trend of the future, towards increasingly finely grained data. It seems likely that the growth in event data will come from two places.

First, natural language processing (NLP) technologies are becoming increasingly sophisticated, meaning that machine-coding software will continue to be able to not only extract richer information (i.e. more nuanced location, actors, and action information) from electronic texts but also do so more accurately. Second, the amount of information available online about politically relevant events continues to grow at an exponential pace. Moving forward, machine-coded events data projects will almost certainly look to expand beyond focusing exclusively on electronic news articles from established journalist sources to Twitter, Facebook, and the blogosphere (see King, Pan and Roberts (2013)). Additionally, the growth of visual media sharing sites like Youtube and Instagram may allow researchers to code political violence from mediums like video and photography. In the foreseeable future, coding visual media will likely continue to require a human eye. However, companies like Google and Facebook are investing tremendous resources in automated coding of video and photos. Regarding the study of political violence, these likely future advancements all mean that researcher will be able to conduct more nuanced analyses with more sophisticated
methods, which ideally lead to more robust findings and insights about the causes or effects of political violence.

Although GDELT is the most advanced machine-coded, political conflict event data dataset in existence today, it still provides fairly crude temporal and geo-spatial aggregations. For example, events do not occur on a day, they occur at a specific time, or over a specific span of time. By combining coding of traditional online media with Facebook and Twitter, automated coding techniques could likely parse out the specific time at the minute or second level that an event – such as a bombing – occurs (see Zeitzoff (2011) for an analysis of minute-level data). Moreover, events tend to not occur in an entire city, but rather in a specific neighborhood or in a specific building. Again, an automated coding platform capable of cross-referencing different new stories, blogs, or social media could likely provide far more detailed event location information than GDELT. Of course, increasingly geo-spatially nuanced data will bring new challenges, such as how to handle locations with the same name, but clever solutions will almost certainly exist.

Above, I outline a number of plausible extensions for each of the three substantive chapters in this dissertation, which are all feasible using the current GDELT dataset. Perhaps even more exciting are future extensions that will likely be feasible in the near future as machine-coded political event data continues to advance. In Chapter 2, I build daily-level measures of political violence against Israel. Once machine-coded event data is able to provide specific sub-day level time-stamps on events, an hourly or even minute-level measure of violence becomes feasible, and could likely yield interesting results about the immediate impact of violence on financial markets. In Chapter 3, I analyze how domestic conflicts affect inter-state conflicts, but am unable to account for the actual transfer of troops or supplies between domestic and inter-state battle fronts. In the future, automated coding may be able to actually track physical movements of resources, which could provide direct testing of domestic conflict dynamics affect inter-state conflicts. Lastly, in Chapter 4, I build forecasts of future violence at the district-month-level, which is highly geo-spatially nuanced for today’s standards. However, if the rate of progress in NLP continues, then machine-coding approaches will likely be capable of extracting the specific neighborhoods or buildings in which political violence occurs, though this may require cross-referencing a number of sources beyond just electronic new articles. With this data, far more geo-spatially nuanced forecasts become feasible.
1.3. **Final Thoughts.** In the 1970s, early hand coded event data datasets contained approximately 2,000 observations. In the early 1990s, breakthroughs in machine-coding produced datasets with between 200,000 and 500,000 observations. Further technological advances in the 2000s led to datasets with between 3,000,000 and 10,000,000 observations. Now, in 2013, the GDELT dataset exists, with over 200,000,000 observations and is growing daily. If we make the realistic assumption that the largest event data dataset in 1978 consisted of 2,000 hand-coded observations, then Moore’s law (that transistors on integrated circuits will double every two years) would predict an event data dataset would exist 34 years later in 2012 with 262,144,000 observation. The incredible accuracy of this prediction might surprise us had Moore’s law not proved so accurate in other domains. If the growth in event data continues to follow Moore’s law in the future, then we should expect to see event data datasets containing on the magnitude of 1 billion events in the next 4 to 6 years. In 10 years time, the current GDELT dataset will likely seem quaint, and just as we look back on empirical studies of conflict 30 years ago that ran OLS regressions on 14 observations with a combination of dismay and sympathy, it is almost certain that future scholars will look back on this dissertation with similar feelings. Please be sympathetic.
REFERENCES


APPENDIX: A GUIDE TO AGGREGATION CHOICES WHEN WORKING WITH EVENT DATA

1. Introduction

In its raw form, event data like GDELT is unsuitable for both theory building and for empirical models. Thus, researchers must convert raw event data to a more usable format. The fine grained structure of raw event data provides researchers with a tremendous amount of flexibility regarding aggregation techniques across the three primary dimensions of manipulation: actors, actions, and time. Additionally, geo-coded datasets like GDELT additionally allow researchers to aggregate geospatially, though this is not required in all case. Counting only the aggregation strategies used across actor, action, and temporal dimensions in the extant literature, over 500 potential theoretically justifiable combinations exist for every research design.\(^1\) Despite this, studies utilizing event data rarely provide rigorous discussion of aggregation techniques across all three dimensions.\(^2\) This is especially problematic, as Shellman (2004a); Alt, King and Signorino (2001); Freeman (1989); Thomas (2002) have all demonstrated that empirical findings in event data studies often change when analyzed at different levels of temporal aggregation. Given the large number of potential aggregation choices and their potential to effect empirical findings, it is critical studies using event data clearly articulate their aggregation choices.\(^3\) Below, I provide an overview of the existing aggregation techniques employed in extant studies using machine-coded event data, highlighting strengths, weakness, and best practices when applicable, focusing on event data presented in the GDELT format (which is used by the most common machine-coded datasets like KEDS, the 10 Million Dyadic dataset, and ICEWS), illustrated below.

Table 1. Example of a GDELT Event Data Event

<table>
<thead>
<tr>
<th>Date</th>
<th>Source</th>
<th>Target</th>
<th>Action</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>20040507</td>
<td>THAMIL</td>
<td>THAREB</td>
<td>20</td>
<td>13.7500</td>
<td>100.4833</td>
</tr>
</tbody>
</table>

\(^1\)This is a low-ball estimate and only assumes 10 types of actor aggregation when many multiples of that exist.

\(^2\)Some exceptions include Shellman and Stewart (2007); Moore (1998); Gleditsch and Beardsley (2004); D’Orazio, Yonamine and Schrodt (2011).

\(^3\)This articles covered in this paper are by no means the total population of event data studies. However, I did my best to include all published papers in the last 10 years that use machine-coded event data to forecast political outcomes.
Since GDELT is the first machine-coded event data dataset to provide geo-location information, I cite literature using human-coded datasets like ACLED and UCDP-GED when discussing geospatial aggregation techniques.

2. Actor Aggregation

Raw event data provide information regarding the actors involved with the action, generally in terms of a ‘source’ and a ‘target,’ although some actions are non-directional. Among the three major dimensions of raw event data aggregations (actors, actions, and dates), actor identification techniques have the highest degree of variation between different data sets and taxonomies. Some coding schemes, including the original COPDAB and WEIS, use a 3-character identifier for all unique actor groups while others utilize broad numerical codes. In this section, I focus on the CAMEO taxonomy (also used by the ICEWS project), which uses hierarchically constructed actor identification codes. Under this scheme, each actors receives a 3- to 12-character code. At a minimum, every actor receives the 3-character country code, but additional 3-character identification codes are appended in a hierarchical fashion based on the level of description provided in the text. Thus, actor codes in CAMEO range from the general [IND] (i.e. India) to the specific [INDGOV-OPPPTY] (i.e. Indian Government Opposition Party). Given this, researchers must determine the actors of interest between whom an event must occur in order for that event to be included in their study’s empirical models. Additionally, of the three areas of aggregation, scholars tend to provide the least information about their choices about actors. Though I am unaware of any study that replicates existing event data analyses using different actor aggregation techniques, it is likely that doing so could have dramatic effects of empirical findings.

At a minimum, it is critical that scholars focus on events involving at least one actor affiliated with a country of interest. Substantively, justification for this minimal level of actor aggregation is clear; a study focusing on Israeli-Palestinian conflicts would not want to include events between Aceh rebels and the Indonesian army, as these are not relevant to the conflict of interest. Although excluding Indonesian rebel activity is obvious, more difficult decisions exist for this example, such as whether or not to include events between members of the Lebanese and Syrian armies or between the governments of the United States and Iran. For example, studies interest in inter-state relations
should make explicit whether all actors form a certain country are included (i.e. rebels, civilians, police) or only specific, policy relevant position (i.e. legislators, president, party members).

In addition to limiting actors by country, scholars interested in intrastate dynamics should be explicit about the treatment of actor codes beyond the 3-character country identification. It is not sufficient for a study to merely state that it is focusing on “sub-state actors” or “relevant domestic actors”. Instead, it is crucial that scholars clearly state how they treat specific secondary, tertiary, or even quaternary 3-character codes. For example, a study interested in domestic rebel groups should clearly illustrate the specific actor identification codes that they treat assume to indicate a rebel. In some event data datasets, only one, catch-all code exists for all actors associated with a mobilized and armed opposition force, but for others, multiple 3-character codes exist.

Due to the lack of precedence of how to make and discuss actor aggregation choices, I provide a “best practice” discussion on actor aggregation from D’Orazio, Yonamine and Schrodt (2011) below.4

Every coded event in the ICEWS dataset contains two actors: a source and a target. Because ICEWS uses the CAMEO coding ontology, each actor is coded using a three-tiered scheme. The first tier is provided for all actors and reflects national identity (for example, [CHN] for an actor identified as Chinese), which we require to be identical for both the source and the target. This ensures that we only analyze events occurring domestically. Additionally, the ICEWS dataset often includes a second tier (or 3-letter code) of information includes many sub-national level descriptions and, where applicable, a third-tier sub-sub-national descriptions. We drop the third tier and select only a relevant selection of actors from the second tier. Specifically, we build three main “classes” of actors based on their second tier categories for the events that occur domestically:

- **Government**, which includes actors identified by ICEWS as:
  - [MIL] – Military
  - [POL] – Police
  - [BUR] – Bureaucrats
  - [POL] – Politicians

---

4Gleditsch and Beardsley (2004) and Moore (1998) also provide comprehensive discussions of actor aggregation choices.
• Rebels, which includes actors identified by ICEWS as:
  – [INS] – Insurgents
  – [SEP] – Separatists
  – [REB] – Rebels

• Other, which includes actors identified by ICEWS as:
  – [CIV] – Civilians unaffiliated with another group
  – [BUS] – Individuals identified as a business person
  – [EDU] – Students and teachers

The domestic events occurring between the government and rebel groups (whether labeled as insurgents, separatists, terrorists, or other categories) and between rebel groups and other non-governmental actors are the primary interactions comprising escalatory processes in intra-state conflict. Therefore, these are actors used. They comprise two classes of undirected actor dyads of interest: GOV-REB and REB-OTH. We omit all events not occurring between GOV-REB or GOV-OTHER actors with the same national identify.

Additionally, in Table 2, I provide the level of actor aggregation and spatial coverage for a number of select event data studies in order to facilitate literature reviews for future event data studies.

3. Action Aggregation

In order to indicate the specific actions that occur between actors, event data datasets utilize an action typology, which provides structured numerical codes that correspond to a series out politically relevant events (see column 4 in Table 1). Within the peer-review published event data studies, CAMEO, WEIS, and IDEA are the most commonly used action typologies. Since CAMEO and IDEA are built on the WEIS framework, these three typologies share similar characteristics. CAMEO, for example, utilizes 20 “cue categories”, or classes of events, which contain different sub- and sub-sub categories. The list below CAMEO’s hierarchical numerical code structure for COERCE, which is the 17th “cue category”. The two-digit, three-digit, and four-digit codes reflect the primary, secondary, and tertiary level of action code, respectively.

• 170: Coerce, not specified below
• 171: Seize or damage property, not specified below
  – 1711: Confiscate property
Table 2. Summary of level of actor aggregation in select studies

<table>
<thead>
<tr>
<th>Article</th>
<th>Level of Actor Aggregation</th>
<th>Country/Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fordham (2005)</td>
<td>State</td>
<td>US, North Korea, Vietnam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syria, Iraq, Egypt</td>
</tr>
<tr>
<td>Pevehouse (2004)</td>
<td>State</td>
<td>Politically relevant dyads</td>
</tr>
<tr>
<td>Schrodt and Gerner (1997)</td>
<td>State</td>
<td>Lebanon Conflict</td>
</tr>
<tr>
<td>Goldstein et al. (2001)</td>
<td>State</td>
<td>Middle Eastern states</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nicaragua</td>
</tr>
<tr>
<td>Schneider and Troeger (2006)</td>
<td>State</td>
<td>Israel-Palestine, Iraq war</td>
</tr>
<tr>
<td>Brandt and Freeman (2004)</td>
<td>State</td>
<td>Gulf war, Ex-Yugoslavia</td>
</tr>
<tr>
<td>Brandt, Colaresi and Freeman (2008)</td>
<td>State</td>
<td>Israel-Palestine</td>
</tr>
<tr>
<td>Shellman (2004a)</td>
<td>Sub-state</td>
<td>Indonesia, East Timor</td>
</tr>
<tr>
<td>Hümmerli, Gattiker and Weyermann (2006)</td>
<td>Sub-state</td>
<td>Cambodia</td>
</tr>
<tr>
<td>Goldstein (1997)</td>
<td>State and IO’s</td>
<td>Serbia, Bosnia, UN, NATO</td>
</tr>
<tr>
<td>Moore (1998)</td>
<td>Sub-state</td>
<td>Peru and Sri Lanka</td>
</tr>
<tr>
<td>Bond et al. (1997)</td>
<td>State and Sub-state</td>
<td>Poland, South Korea, China</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>Brandt, Freeman and Schrodt (2011)</td>
<td>State</td>
<td>Israel and Palestine</td>
</tr>
<tr>
<td>Shellman and Stewart (2007)</td>
<td>Sub-state</td>
<td>Haiti</td>
</tr>
<tr>
<td>Shellman (2007)</td>
<td>Sub-state</td>
<td>Afghanistan</td>
</tr>
<tr>
<td>Shellman (2004b)</td>
<td>Sub-state</td>
<td>Chile and Venezuela</td>
</tr>
<tr>
<td>Shearer (2006)</td>
<td>State</td>
<td>Israel and Palestine</td>
</tr>
<tr>
<td>Schrodt and Gerner (2000)</td>
<td>State</td>
<td>Levant states</td>
</tr>
<tr>
<td>Kovar et al. (2000)</td>
<td>State</td>
<td>US and Iraq and Levant states</td>
</tr>
<tr>
<td>Stoll and Subramanian (2006)</td>
<td>State</td>
<td>Levant states</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle East, Balkans</td>
</tr>
<tr>
<td>Schrodt, and Yilmaz (2003)</td>
<td>State</td>
<td>West Africa</td>
</tr>
<tr>
<td>Schrodt and Gerner (2001)</td>
<td>State</td>
<td>Middle East and Yugoslavia</td>
</tr>
<tr>
<td>D’Orazio, Yonamine and Schrodt (2011)</td>
<td>Sub-state</td>
<td>Asia</td>
</tr>
</tbody>
</table>

- 1712: Destroy property
- 172: Impose administrative sanctions, not specified below
  - 1721: Impose restrictions on political freedom
  - 1722: Ban political parties or politicians
  - 1723: Impose curfew
  - 1724: Impose state of emergency or martial law
- 173: Arrest, detain, or charge with legal action
• 174: Expel of deport individuals
• 175: Use tactics of violent repression

To the best of my knowledge, Kovar et al. (2000) is only one study has ever analyzed the complete, raw actions codes including the secondary and tertiary codes when provided. Aside from this, every study aggregates the action codes into some type of higher level variables. The vast majority of the extant event data literature converts the action codes for each event into coarse scale or count variables. Thus, this section is primarily dedicated to techniques used to build scale and count variables. However, it is important to note that a small number of studies using HMMs collapse secondary and tertiary action codes to the “cue category” level, then account for the number of each type of “cue category” event that occurs in a given temporal range (see Schrodt (1997), Bond et al. (2004), and Schrodt (2000)). Although this approach accounts for more information about the types of events that are occurring that any of the action aggregation techniques discussed below, it requires a methodological technique, like HMMs, particularly suited to handle a large number of sparsely populated independent variables.

3.1. Scale. Though rarely cited in modern event data publications, Azar and Sloan constructed the first event data scale (called the Azar-Sloan scale) in 1975 as part of the COPDAB project. This original scale was important because it placed events into a conflict-cooperation continuum and relied on a small sample of experts to determine degree of conflict/cooperation for each class of event — two ideas that are the foundation of the Goldstein Scale (discussed below), which dominates the current literature that uses scales. In brief, the Azar-Sloan scale consists of 15 cue categories, ranging from category 1, which reflects the most cooperational events to 15, which indicates the most conflictual. Azar and Sloan interviewed 18 international relations scholars to provided a weighted value for each cue category that reflects how may times more conflictual (for events 9-15) or cooperational (for events 1-7) each event is relative to category 8 events, deemed neutral. This resulted in categories 1-7 receiving values ranging from 6 to 92, with 92 indicating events that are 92 times more cooperational than the neutral category 8 events, and categories 9-15 receiving values ranging from 6 to 102.

In 1992, Goldstein used a similar operational approach (i.e. surveys of international relations professors) to build the conceptually similar Goldstein scale for the WEIS typology.\footnote{Prior scales for the WEIS typology proceeded the Goldstein scale but did not gain prominence in the literature. Goldstein (1992) provides a condensed summary of early efforts at building scales and counts from WEIS. Additionally, see Goldstein (1992) and Vincent (1983) for more thorough discussions of prior scaling approaches.} The Goldstein...
scale scores all WEIS actions on a -10 to +8.3 continuum, with -10 and +8.3 reflecting the most conflictual and cooperational events, respectively.\(^6\) In order to determine the specific conflict-cooperation score to assign to each WEIS action, Goldstein surveyed either international relations faculty at the University of Southern California. Thus, despite the ubiquity of the Goldstein scale in the extant event data literature, the approach used to build the scale are questionable. Not satisfied with the informal survey approach that Goldstein originally used to build the Goldstein scale, Bond et al. (2003) utilized a web-based survey in order to rank all terminal nodes within the IDEA action typology on a conflict-cooperation scale. The result was conceptually similar to the Goldstein scale, but action scores ranged from -13 (for the most conflictual) to +7 (for the most cooperational).\(^7\)

Despite the popularity of the Goldstein scale, scholars have recently built two additional scales based on coding typologies other than WEIS and WEIS compatible typologies like CAMEO and IDEA. First, the Intranational Political Interactions project is an alternative scaling approach that codes events into according to a unique typology of 10 classes of conflictual event and 10 classes of cooperational events. Each class of events takes on a score of +10 to +100 (in multiples of 10), which reflect the degree of cooperation conflict (see Shellman (2004b,a) for more information about the IPI scale). Second, the Violent Intranational Conflict Data Project (VICDP) collapses events into 15 categories, which escalate from most cooperational to most conflictual (see Moore and Lindström (1996) for additional information on the VICDP project). For example, category 1 is called “Agreement-Resolution” defined as the termination of the internal war, and category 15 is “Extensive war acts causing deaths, dislocation, and high strategic costs”, which requires no further explanation.

Scholars who choose to use a scale must still make another important decision regarding how to aggregate the actual scores.\(^8\) Common techniques are to calculate either the sum or the mean of the scaled scores, though both of these technique have potential shortcomings that I call the mean problem, sum problem, and the single scale problem.

- Mean problem:

\(^6\)Due to the similarities between WEIS and other taxonomies, the Goldstein scale is easily implemented on CAMEO and IDEA coded event data. See URL: http://eventdata.psu.edu/cameo.dir/CAMEO.scale.html for the Goldstein scale applied to CAMEO codes

\(^7\)See Hämmerli, Gattiker and Weyermann (2006) and Bond et al. (2003) for additional information.

\(^8\)Because the IPI scale only contains positive scores, the following discussion of sum and mean scales is not applicable to the IPI data.
− A month with three -10 events occurring every day would have the same mean score as a month with only 1 “-10 event per day. Since it is obvious that a month with 90 “-10 events” is more conflictual than a month with only 30 “-10” events, taking the mean score can lack external validity.

• **Sum problem:**
  − According to the Goldstein scale for WEIS data, “Issue order or command, insist, demand compliance” and “military attack; clash; assault” receives a -4.9 and -10 on the Goldstein scale, respectively. Consider two months, one with little dialogue but actual violence, and one with no violence but considerable negative dialogue. By summing the Goldstein values, the latter month could appear more conflictual than the first, even though it experience no actual conflict.

• **Single scale problem:**
  − Consider a day during which “noninjury desructive action” (a -8.3 on the Goldstein scale) and a “extend military assistance” (a +8.3 on the Goldstein scale) occur between the same actors. The sum and the mean of these two events equals 0, which is the same score that a day with no events receive. Theoretically, it is apparent that the nature of events occurring on a day comprised of purely neutral events and a day with a -8.3 event and a +8.3 event are fundamentally different than a day on which no relevant actions occurred. Additionally, many conflicts experience negotiation events (assigned small, positive scale values) during fighting (assigned large, negative scale values). However, these types of conflicts are impossible to differentiate from other conflicts with slightly lower levels of violence but no attempts at mediation or negotiation.

To overcome the single scale problem, some scholars divide actions into two separate classes, one for conflictual actions and another for cooperational. Next, they calculate the mean or sum score for conflictual and cooperational events separately. Although this is more logical than taking the sum or mean of all events together, it is still vulnerable to the sum problem and the mean problem defined above.

3.2. **Counts.** In an attempt to best overcome the three problems above, a number of studies utilize count data. “net cooperation” is the most straightforward count, which is simply the total number

9Unlike Goldstein and IDEA scale, the IPI action typology divides conflictual and cooperational events. Consequently, scholars choosing to scale or sum use IPI-coded event data must do so for conflictual and cooperational events separately.
of cooperational actions - number of conflictual actions according to the Goldstein scale. In order to attempt for most variation in the types of events that are occurring, a number of other scholars utilize Duvall and Thompson counts Duval and Thompson (1980), which place all events on the WEIS scale into the four conceptually unique categories below.\textsuperscript{10}

- **Verbal Cooperation**: The occurrence of dialogue-based meetings (e.g. negotiations, peace talks), statements that express a desire to cooperate or appeal for assistance (other than material aid) from other actors.
- **Material Cooperation**: Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc.
- **Verbal Conflict**: A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflict.
- **Material Conflict**: Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc.

Since I use counts of material conflict in all three substantive chapters of this dissertation, I list the CAMEO codes that comprise material conflict below:

- 150: Demonstrate military or police power, not specified below
- 151: Increase police alert status
- 152: Increase military alert status
- 153: Mobilize or increase police power
- 154: Mobilize or increase armed forces
- 160: Reduce relations, not specified below
- 161: Reduce or break diplomatic relations
- 162: Reduce or stop material aid, not specified below
  - 1621: Reduce or stop economic assistance
  - 1622: Reduce or stop military assistance
  - 1623: Reduce or stop humanitarian assistance
- 163: Impose embargo, boycott, or sanctions
- 164: Halt negotiations
- 165: Halt mediation
- 166: Expel or withdraw, not specified below

\textsuperscript{10}This approach is easily extendible to the CAMEO coding scheme since it is built on WEIS’ general framework.
– 1661: Expel or withdraw peacekeepers
– 1662: Expel or withdraw inspectors, observers
– 1663: Expel or withdraw aid agencies
• 170: Coerce, not specified below
• 171: Seize or damage property, not specified below
  – 1711: Confiscate property
  – 1712: Destroy property
• 172: Impose administrative sanctions, not specified below
  – 1721: Impose restrictions on political freedoms
  – 1722: Ban political parties or politicians
  – 1723: Impose curfew
  – 1724: Impose state of emergency or martial law
• 173: Arrest, detain, or charge with legal action
• 174: Expel or deport individuals
• 175: Use tactics of violent repression
• 180: Use unconventional violence, not specified below
• 181: Abduct, hijack, or take hostage
• 182: Physically assault, not specified below
  – 1821: Sexually assault
  – 1822: Torture
  – 1823: Kill by physical assault
• 183: Conduct suicide, car, or other non-military bombing, not specified below
  – 1831: Carry out suicide bombing
  – 1832: Carry out vehicular bombing
  – 1833: Carry out roadside bombing
  – 1834: Carry out location bombing
• 184: Use as human shield
• 185: Attempt to assassinate
• 186: Assassinate
• 190: Use conventional military force, not specified below
• 191: Impose blockade, restrict movement
• 192: Occupy territory
• 193: Fight with small arms and light weapons
• 194: Fight with artillery and tanks
• 195: Employ aerial weapons, not specified below
  – 1951: Employ precision-guided aerial munitions
  – 1952: Employ remotely piloted aerial munitions
• 196: Violate ceasefire
• 200: Use unconventional mass violence, not specified below
• 201: Engage in mass expulsion
• 202: Engage in mass killings
• 203: Engage in ethnic cleansing
• 204: Use weapons of mass destruction, not specified below
  – 2041: Use chemical, biological, or radiological weapons
  – 2042: Detonate nuclear weapons

Below, I provide a list that illustrates the type of action aggregation strategy utilized in a selection of prominent event data studies.

• **Scale**
  – Goldstein mean
    * Sky (2000); Fordham (2005); Brandt and Freeman (2004); Shellman (2004b); Shellman and Stewart (2007); Shellman, Hatfield and Mills (2010); Shellman (2007); Brandt, Colaresi and Freeman (2008)
  – Goldstein sum
    * Schrodt and Gerner (1997, 2000); Goldstein et al. (2001); Stoll and Subramanian (2006); Schrodt and Gerner (2001); Gleditsch and Beardsley (2004); Schneider and Troeger (2006); Shellman, Hatfield and Mills (2010)
  – IPI scale
    * Shellman (2004a,b)
  – IDEA scale
    * Hämmerli, Gattiker and Weyermann (2006), Bond et al. (2003)
  – VICDP scale
• Counts
  – Goldstein counts (positive and negative)
    * Pevehouse (2004); Shellman, Hatfield and Mills (2010)
  – Net cooperation
  – Duvall and Thompson counts
    * Brandt, Freeman and Schrodt (2011) (only material conflict events), Shearer (2006), Schrod,
      and Yilmaz (2003) (adds a fifth category called “Mediation and negotiation”), Schrod (2006); D’
      Orazio, Yonamine and Schrod (2011)

• Action codes
  – Kovar et al. (2000)
  – Schrod (2006)

• Other
  – Bond et al. (1997) – Conflict carrying capacity

4. Temporal Aggregation

After completing the first two steps of the aggregation process, researchers must determine how
to account for time. The vast majority of studies temporally aggregate event data into a man-
made length of time, commonly by day, week, month, quarter, or year. The choice of temporal
unit determines that length of time across which the action aggregation is performed. For example,
consider a scholar who is interested in measuring the Goldstein sum of events occurring between
the Indian military and Indian rebels. The actor and action aggregation steps (i.e. Section 1
and Section 2) would result in a dataset containing only events that occurred between the actors
of interest with an extra column indicating the Goldstein score for each action. To complete the
aggregation process, the researcher must determine a temporal unit of aggregation. If the researcher
chooses the weekly level, he/she would simply calculated the sum of the Goldstein scores for events
occurring in the same week.

As mentioned in the introduction to this study, a number of articles demonstrate that different
temporal aggregations can affect the empirical results (see Freeman (1989); Alt, King and Sig-
norino (2001); Dale (2002); Shellman (2004a)). These analyses suggest that it is important to both
theoretically justify the level of aggregation and, if possible, employ multiple levels as robustness checks. Below, I list the temporal aggregations employed in a number of prominent studies.\textsuperscript{11}

Although most scholars aggregate temporally, a small body of studies focus on the sequential order of events, irrespective of traditional units of time. Drawing on Marlin-Bennett, Rosenblatt and Wang (1991), Moore (1998) builds event sequence based on “moves” and “turns” as opposed to standard temporal units.\textsuperscript{12} Subsequence work, including Moore (2000); Shellman (2004\textsuperscript{a}, 2007) adopt a similar “move” and “turn” approach to model event data. Likewise, some studies using HMMs do not confine event patterns to traditional temporal units.\textsuperscript{13} For example, Schrodt (2000) focuses on identifying and predicting transitions between distinct phases of conflict irrespective of the amount of time that a state exists in a given phase.

Below, I illustrate the temporal aggregation choices employed within the event data literature.

- **Daily**
  - Goldstein (1991); Pevehouse and Goldstein (1999); Goldstein et al. (2001); Shellman (2004\textsuperscript{a},\textsuperscript{b}); Schrodt (2006); Schneider and Troeger (2006); Shearer (2006)
- **Weekly**
  - Goldstein (1991); Goldstein et al. (2001); Brandt and Freeman (2005); Shellman and Stewart (2007); Brandt and Freeman (2004); Shellman, Hatfield and Mills (2010); Goldstein (1997); D’Orazio, Yonamine and Schrodt (2011)
- **Bi-weekly**
  - Stoll and Subramanian (2006)
- **Monthly**
  - Goldstein (1991); Schrodt (1997); Sky (2000); Schrodt, and Yilmaz (2003); Schrodt (2007); Schrodt and Gerner (1997, 2000, 2001); Shellman (2004\textsuperscript{a,\textit{b}}); Ward, Greenhill and Bakke (2010); Gleditsch and Beardsley (2004); Shellman, Hatfield and Mills (2010); Brandt, Freeman and Schrodt (2011); D’Orazio, Yonamine and Schrodt (2011); Brandt, Colaresi and Freeman (2008)
- **Quarterly**
  - Jenkins and Bond (2001); Fordham (2005); Shellman (2004\textsuperscript{a})

\textsuperscript{11}Note that certain studies, like Shellman (2004\textsuperscript{a}) and D’Orazio, Yonamine and Schrodt (2011) follow best practice and perform analyses across different levels of temporal aggregation to serve as robustness.

\textsuperscript{12}For more information on the rationale behind analyzing events in terms of sequences, see Schrodt (2000).

\textsuperscript{13}This is not true of all event data studies using HMMs. For example, Shearer (2006) applies HMMs to temporally aggregated data to facilitate the substantive interpretation of his findings.
• Yearly
  – Pevehouse (2004); Bond et al. (1997)
• Non-temporally defined sequences
  – “Move” and “Turn” sequences
  – Markov transition sequences
    * Schrodt (2000)
• Unspecified
  – Hümerli, Gattiker and Weyermann (2006)

5. GEO-SPATIAL AGGREGATION

Although actor, action, and temporal aggregation choices are mandatory for all studies using event data, geo-spatial aggregation is not always required. For example, studies focusing solely on interactions between two actors over time might not be interested in where the interactions occurred, and therefore will not need geo-spatial aggregation. However, other studies, primarily though interested in sub-state dynamics, do require geo-spatial aggregation. Since GDELT is the first machine-coded event data dataset, I draw on examples from studies using human-coded datasets to discuss current geo-spatial aggregation practices, which can be divided into two main categories: administrative units and arbitrary units. GDELT, like geo-coded datasets including ADLED and UCDP/PRIO, provide specific latitude and longitude coordinates reflecting the specific point where each event occurs. Since there are a virtually infinite number of latitude and longitude coordinates in the world, scholars tend to aggregate up to a coarser level of geo-spatial aggregation to facilitate analysis.

5.1. Administrative units. Most countries in the world are divided into sub-state administrative units, such as municipalities, provinces, or districts. A common approach among studies performing sub-state analyses is to geo-spatially aggregate events according to these units. Below, I provide a list of a sample of these studies along with the country and sub-state administrative level to which they aggregate their data.

• Mangion-Zammit et al. (2012) – Province-level in Afghanistan
• Weidmann and Callen (2013) – District-level in Afghanistan
• Weidmann and Ward (2010) – Municipality-level in Bosnia
5.2. **Polygons.** A different approach is to ignore administrative units and construct sub-state, geo-spatial units centered around the specific location where an event occurs.

- Lujala, Rod and Thieme (2007) – 30-kilometer buffer Polygons surrounding exact event locations, bound by country borders for 48 African countries
- Buhaug, Gates and Lujala (2009) – Polygons surrounding exact event locations
- Weidmann (2009) – Polygons surrounding exact event locations
- Ostby, Nordas and Rod (2009) – Polygons surrounding exact event locations

6. **Conclusion**

Many other areas of political science have established clearly defined coding rules. For example, the Correlates of War provide clear definitions of how many specific actions (battle fatalities) must occur between specific actors (government forces and a domestic rebel group) within a set period of time (one calendar year) to qualify as an “intra-state war”. Despite the massive number of aggregation options and proven effect that even minor changes can have on empirical results, it is surprising that studies utilizing event data have yet to adopt a formal framework for selecting and documenting aggregation choices. This document provides a general overview of the aggregation techniques – across the actor, action, temporal, and geo-spatial dimensions used in the relevant literature – which help guide the aggregation choices I make in the substantive chapters of this dissertation.
References


Jay Yonamine
• 952-237-5881 • jay.yonamine@gmail.com • jayyonamine.com

Education

Pennsylvania State University: State College, PA
• PhD, Political Science (2013): Major Field: - International Relations; First Minor - Quantitative methods; Second Minor - Advanced computational methods.
• Master of Arts, Political Science (May 2011): Major Field: International Relations
• Relevant Coursework: Robust Models, Data Mining, Event Data Analysis, Geo-Visual Analytics, Maximum Likelihood Estimation, Game Theory 1, Game Theory 2, Multivariate Regression, Multivariate Regression, Machine Learning (sit-in)

Interuniversity Consortium for Political and Social Research (ICPSR) Ann Arbor, MI
• Coursework (Summer 2010): Introduction to Complex Systems, Time Series Analysis

Middlebury College: Middlebury, VT
• Bachelor of Arts (May 2007): Political Science, cum laude

Professional Experience

Allstate Insurance: Northbrook, IL
Predictive Analytics Associate (March 2013 – present)
• Implement Spark cluster computing system to increase speed of algorithms at scale both through EC2 and local servers
• Build frequency, severity, and pure premium loss models for auto coverage

Pennsylvania State University: State College, PA
Instructor: War in World Politics (Summer 2012)
• Taught advanced political science undergraduate course. Built syllabus and wrote all lectures.

National Geospatial Intelligence Agency (NGIA): State College, PA (remotely)
Statistical Consultant subcontracted through iSciences L.L.C. (June 2011-December 2011)
• Lead statistician responsible for manipulating and analyzing meteorological, macro level, and textual data
• Implemented machine learning and forecasting models in R and MATLAB to predict social events during/after hurricanes

Integrated Conflict Early Warning System (ICEWS) project: State College, PA
Research Assistant to Dr. Phil Schrodt (Fall 2009-Summer 2011)
• Applied innovative forecasting methodologies to machine-code Event Data to enhance predictive accuracy of civil wars
• Analyzed sparse parsing software coding accuracy and performed interoperability studies between software iterations

Georgetown University Global Surveillance of Emerging Threats: Washington, D.C.
Research Intern (February 2009-May 2009)
• Assisted with construction of a semi-automated early warning system for political conflict; focused on building an event typology, writing a coding manual, and training new employees how to properly code events based on the manual

Published Articles/Invited Lectures:
“Data-based Computational Approaches to Forecasting Political Violence” Handbook of Computational Approaches to Counterterrorism, Springer Press. With Philip A. Schrodt and Benjamin Bagozzi.
“Using Political Event Data to Analyze Variance in Tel Aviv 100 Index Returns” Forthcoming.

Event Data Workshop (December 2011): Istanbul, Turkey
• Designed and led 2-day workshop for 30 international academics and high ranking directors of the Turkish police

New Directions in Text as Data (July 2012): Cambridge, MA
• Delivered presentation on emerging technologies in machine-coding text to extract politically relevant events

Additional Skills and Experiences

Technology: Skilled in R, STATA; Introductory SAS, MATLAB, AWS, Spark
Coaching: Strength and conditioning coach of Penn State Boxing Team
Athletics: 4-year starting varsity golfer at Middlebury College, 2-0 in semi-professional mixed martial arts fights;