

UCLA

UCLA Electronic Theses and Dissertations

Title

Essays on Networks, Dictatorships, and Political Violence

Permalink

<https://escholarship.org/uc/item/33c9t14j>

Author

Derpanopoulos, George

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Essays on Networks, Dictatorships, and Political Violence

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Political Science

by

George Derpanopoulos

2018

© Copyright by
George Derpanopoulos
2018

ABSTRACT OF THE DISSERTATION

Essays on Networks, Dictatorships, and Political Violence

by

George Derpanopoulos

Doctor of Philosophy in Political Science

University of California, Los Angeles, 2018

Professor Barbara Geddes, Chair

This dissertation contains three essays, each addressing a different question in political economy and comparative politics. The first essay speaks to the large literature arguing that dictatorships can achieve high levels of economic growth if dictators can commit to not expropriate elites. Extant research has focused on the role of formal institutions – legislatures and parties – in helping elites constrain dictators’ predation. I complement this literature by documenting the role of an informal institution, elite financial networks, in constraining the dictator. I argue that dense financial ties among elites diffuse private information on the state of the economy, hence facilitating elites’ monitoring – if the dictator reneges on his commitments to elites, informed elites are able to infer and punish his defection. Accordingly, I hypothesize that dictatorships with denser elite financial networks enjoy stronger property rights. To test my argument, I uncover networks of elites’ co-ownership of offshore companies – a strong type of financial tie – using a large, untapped leak of private financial information, the Panama Papers. A thorough regression analysis of almost all dictatorships in the period 2002 – 2013 supports my theory: a one-standard deviation increase in financial network density predicts a half- standard deviation decrease in expropriation risk.

The second essay asks: why have some countries counted hundreds of their citizens fleeing to fight in Syria, while other countries’ citizens have remained bystanders? There are three methodological challenges to answering this question. First, there may be two groups of countries: one at no risk of “supplying” foreign fighters and another supplying some pos-

itive amount. Second, there is no theory that specifies a functional form linking countries' features to foreign fighter supply. Third, existing models for predicting foreign fighter supply perform poorly out of sample or yield output that is not amenable to social-scientific interpretations. To solve these challenges, I augment a count regression model, the hurdle negative binomial, with two machine learning tools. Namely, I allow features to affect the response non-parametrically, by using kernel functions that represent expansions of the data. Furthermore, I add regularization terms that penalize complexity to mitigate overfitting. My approach combines the strengths of predictive and confirmatory models: it performs similarly to state-of-the-art machine learning algorithms in prediction while providing substantively interpretable output. Applying the model to data on 163 countries, I find that populous, developed countries, with a large Sunni population and proximity to Syria supply more fighters. These results lend themselves to viewing foreign fighter supply as largely driven by structural forces.

The third essay contributes to the literature on civil war, which has recently shifted its attention from state-rebel violence to rebel-rebel violence. I build on this work by applying tools from social network analysis to visualize, summarize, and model conflict among 22 rebel groups in Lebanon's Civil War, specifically in the period 1980 – 1991. Using a network graph and node-, dyad-, and network-level statistics, I find a conflict structure in line with historical accounts: a dense pattern of hostilities, high reciprocity in hostilities (i attacks $j \Leftrightarrow j$ attacks i), low transitivity in hostilities (the enemy of my enemy is my friend), infighting within religious sects, and the existence of 3 central groups. Furthermore, using regression models tailored to network data, I find that groups that command support from the ethno-religious sect they belong to, control valuable natural resources and territory, and use terrorist tactics are more likely to attack other rebels, while groups that are able to reach an agreement with the state are less likely to attack other rebels. Finally, using a clustering model, I detect 2 sub-conflicts: a narrow cluster that includes the infighting among Palestinian groups and their Sunni allies and a broader cluster that includes the hostilities between rival Shi'ite groups. My approach is relevant to policy-makers deciding which rebel groups to support, particularly in conflicts where opposition to the state is fragmented.

The dissertation of George Derpanopoulos is approved.

Mark Stephen Handcock

Ronald L. Rogowski

Daniel Simon Treisman

Barbara Geddes, Committee Chair

University of California, Los Angeles

2018

*To my parents, who made everything leading up to this degree possible
and to Marianna, who was always there for me*

TABLE OF CONTENTS

1 Elites, Financial Networks, and Commitment in Dictatorships:

Evidence from the Panama Papers	1
1.1 Introduction	1
1.2 Literature	4
1.3 Theory	6
1.4 Data: Elites & Financial Networks	11
1.5 Analysis	19
1.5.1 Approach	19
1.5.2 Baseline Results	25
1.5.3 Robustness Checks	30
1.6 Discussion	40
1.7 Conclusion	46

2 Predicting Foreign Fighter Flows to Syria Using Machine Learning:

An Introduction to Kernel Regularized Hurdle Negative Binomial	48
2.1 Introduction	48
2.2 Foreign Fighters in Syria	53
2.3 Predictors of Foreign Fighter Supply	55
2.4 Hurdles to Predicting Foreign Fighter Supply	58
2.5 The Model	63
2.6 Computation & Quantities of Interest	70
2.6.1 Optimization	70
2.6.2 Pointwise Marginal Effects	71

2.6.3	Estimating Uncertainty of Sample-Average Pointwise Marginal Effects	72
2.7	Performance Relative to Other Models	72
2.8	Results & Discussion	75
2.8.1	Data	75
2.8.2	Main Results	77
2.8.3	Effect Heterogeneity	81
2.9	Conclusion	84
2.10	Appendix 1: Results Table and Heterogeneous Effects	87
2.10.1	Full Results	87
2.10.2	Heterogeneous Effects	87
2.11	Appendix 2: Target Function Derivation	90
2.11.1	Appendix 2.1: Sample Log-Likelihood	90
2.11.2	Appendix 2.2: Using Mercer’s Theorem	92
3	Modeling Inter-Rebel Group Conflict with Network Analysis:	
	The Case of Lebanon’s Civil War	94
3.1	Introduction	94
3.2	Analysis	99
3.2.1	Data & Descriptive Statistics	99
3.2.2	Predicting Inter-Rebel Hostilities	104
3.2.3	Clustering	107
3.3	Discussion	109
3.4	Conclusion	113

LIST OF FIGURES

1.1	Two Examples of Elite Networks	10
1.2	Elite Financial Network in Chad 2015 (L) vs. Botswana 2015 (R)	16
1.3	Global Distribution of Elite Financial Network Density	22
1.4	Dependent and Independent Variable Means Across Time	23
1.5	Marginal Effect of Elite Network Density on Expropriation Risk by Regime-Type.	29
2.1	Global Distribution of Foreign Fighter Supply, Through 2014	54
2.2	Leave-one-out CV Predictions	75
2.3	Sample-Average Pointwise Marginal Effects	78
2.4	Distribution of Pointwise Marginal Effects	81
2.5	Effects Inside and Outside Europe	83
2.6	Effects of Government Regulation of Religion by Sunni Population Share	84
2.7	Interaction Effects	91
3.1	Inter-Rebel Hostilities in Lebanon’s Civil War, 1980 – 1991	101
3.2	Degree Histograms & Scatterplot	103
3.3	Latent Position Cluster Model	110

LIST OF TABLES

1.1	Expropriation Risk – Baseline Models	27
1.2	Expropriation Risk – Additional Models	32
2.1	Comparing Prediction Error	73
2.2	Description of Features	76
2.3	Average Marginal Effects on Foreign Fighter Supply	88
2.4	Average Marginal Effects on Foreign Fighter Supply in Listwise-Deleted Dataset	89
3.1	Degree Summary Statistics	102
3.2	Centrality Scores by Group	105
3.3	ANOVA of ERGMs	106
3.4	Best-Fit Models	108

ACKNOWLEDGMENTS

This dissertation is a product of the guidance and support – professional and emotional – of numerous people. On the professional side, first and foremost, I thank my advisor, Barbara Geddes, for her feedback and advice. As her student, coauthor, and mentee, I learned to ask tough questions and that “there’s no substitute for hard thinking”. I also thank her for the understanding she showed when, during the course of this dissertation, my professional aspirations changed. Furthermore, I thank my other committee members, Mark Handcock, Ron Rogowski, and Dan Treisman, for their input as teachers and advisors.

With respect to specific chapters, for Chapter 1 I am grateful for comments from Zachary Steinert-Threlkeld, Matthew Wilson, and participants at the 2017 Midwestern Political Science Association Annual Meeting and the 2017 UCLA Comparative Politics Graduate Student Conference, and I thank Fang Yeng for sharing his Forbes data. Chapter 2 is based on unpublished, coauthored work with Luke Sonnet, who grants me permission to include our work here. Luke and I contributed equally to that work. For Chapter 2, I thank Bryce Dietrich, Chad Hazlett, Jeff Lewis, David Rapoport, Art Stein, Barbara Walter, participants at the 2016 Midwestern Political Science Association Annual Meeting, and members of the UCLA Comparative Politics Reading Group for their feedback. For Chapter 3, I thank Amanda Rizkallah for her suggestions.

Additionally, I thank Joseph Brown and Esther Blair for helping with graduate school’s administrative hurdles. Finally, I acknowledge financial support from the UCLA Graduate Division and Political Science Department.

On the personal side, I would like to thank four colleagues-turned-friends. I am grateful to Sarah Brierley and Andrea Vilán for their kindness, support, and the countless social events they organized. Imil Nurutdinov and I shared the most memorable intellectual conversations I had during my time here; I will treasure all of them. Luke Sonnet – in addition to an exemplary coauthor – was an endless source of energy, humor, and general knowledge.

Outside the department, I would like to thank five friends: Richard Domurat and Ben Smith for their camaraderie, especially during the last and most important stage of graduate

school; Gianni Nicolò for being a breath of Mediterranean air and my roommate for three years; Andrés Schneider for his insights on everything from the markets and backgammon to Freud and Peronism. A special mention is reserved for Ioannis Kospentaris, a beacon of positivity and my best friend and brother in Los Angeles. I will miss him greatly.

I now turn to my family, to whom I owe so much. They provided me with all of the opportunities to accomplish what I've accomplished and never doubted me, even when my hobbies and interests were not typical for a future PhD. I also thank them for the impact their own experiences had on me. Dad's PhD experience in Berkeley is probably the main reason I pursued graduate school and chose LA over NY. Mom's determination in writing her book – at a difficult point in her life – strengthened my own determination when it was faltering. Alex's success in overcoming his obstacles and pursuing his veterinary studies was a constant source of inspiration. Pappou Kim's tough decision to halt his PhD in order to support his family strengthened my resolve to finish my PhD. My memory of Yiayia Kalliopi's smile when I told her I was admitted to Cambridge – my last words to her – served as fuel for pushing on. Finally, Yiayia Kaiti's regrets of not pursuing her dreams, including becoming a doctor, taught me to never give up. At the same time, her beginning an artistic career at 88 also taught me that it's never too late to start a new chapter in life.

Finally, I thank my better half, Marianna, for always being there for me and for believing in me more than I believed in myself. She is, in equal parts, an amazing partner, confidant, and friend, and an unending source of courage, inspiration, and love. I am excited beyond words about the new adventure we are embarking on, and I smile when I think of what the future has in store for us.

VITA

- 2010 B.Sc. (Economics and Politics), University of Bristol.
- 2011 M.Phil. (Economics), University of Cambridge.
- 2018 M.S. (Statistics), University of California, Los Angeles.

PUBLICATIONS

Derpanopoulos, George. 2018. “Public Choice and the Study of Autocratic Survival.” In Paul Dragos Aligica, Christopher J. Coyne and Stephanie Haeffele (eds.), *Exploring the Political Economy and Social Philosophy of James M. Buchanan*. London: Rowman & Littlefield.

Derpanopoulos, George, Barbara Geddes, Erica Frantz and Joseph Wright. 2017. “Are Coups Good for Democracy? A Response to Miller (2016).” *Research and Politics* 4(2):1–4.

Derpanopoulos, George, Barbara Geddes, Erica Frantz and Joseph Wright. 2016. “Are Coups Good for Democracy?” *Research and Politics* 3(1):1–7.

CHAPTER 1

Elites, Financial Networks, and Commitment in Dictatorships: Evidence from the Panama Papers

1.1 Introduction

Research Question The importance of secure property rights to economic growth is one of the few consensus facts in economics ([Barro, 1989](#); [Acemoglu, Johnson and Robinson, 2001](#)). A related consensus fact is that property rights protection requires limited government ([North and Weingast, 1989](#); [North, 1990](#)). Most theories of limited government view it as product of elites’ efforts to punish predatory behavior by the sovereign. Thus, at the core of the literature on economic growth, property rights, and limited government lies the question: how do elites constrain dictators?

Literature Scholars of authoritarian regimes¹ have focused on the role of political institutions.² For example, [Boix and Svolik \(2013\)](#) argue that interaction within legislatures enables elite monitoring of the dictator, hence deterring him from defecting on rent-sharing pacts. Similarly, [Gehlbach and Keefer \(2012\)](#) claim that authoritarian parties increase the observability of expropriation against party supporters, thereby tying the dictator’s hands. In short, the literature sees political institutions as an antidote to dictator-elite information asymmetries and a catalyst for elite coordination.

¹I use the terms “authoritarian regime”, “dictatorship”, “autocracy”, and “non-democracy” interchangeably. The same holds for the terms “dictator” and “autocrat”.

²See [Pepinsky \(2014\)](#) for a general critique of the institutionalist approach to the study of dictatorships.

Puzzle However, formal political institutions are only one medium through which elites can acquire, share, and act on private information regarding the dictator’s predation. As Boix and Svolik (2013) acknowledge, “several institutions may serve to reduce asymmetries of information between the ruler and his allies [...] less formal, idiosyncratic, or traditional institutions may perform this function” (p. 307). Informal institutions may matter most where formal ones are ineffective or controlled by the dictator; in “personalist” or “established” regimes (Geddes, 1999; Svolik, 2009).³ Therefore, by studying elites’ interaction within informal institutions we can advance our understanding of how elites constrain dictators.

Argument In this article, I explore the role of one informal institution, financial networks. I argue that the structure of elites’ financial networks affects the diffusion of private information on the true size of the regime’s rents. This information allows some elites to infer whether they receive low rents because the dictator reneges on his commitment to share rents or because of a negative economic shock. Informed elites do not have to threaten to punish the dictator whenever he delivers low rents—a threat that is non-credible, as elites incur a cost for reallocating their capital. Instead, informed elites can credibly threaten to punish the dictator *only* when rents are high but he does not share them. Crucially, when more elites learn the true state of the regime’s rents—through their financial ties to other elites—the credible threat posed to the dictator increases, and he is further deterred from predation. Hence, I hypothesize that, in countries where elites’ financial network enables larger diffusion of private information about the economy, economic predation by the dictator is lower.

Data To test this hypothesis, I operationalize the diffusion of information within a network as the network’s density: the number of observed ties divided by the number of possible ties (the latter is a function of the network’s size). To measure network density, I tap into the largest public source of private financial information to date, the leaked Panama Papers (International Consortium of Investigative Journalists, 2016). This unique, previously

³Crucially, Geddes, Wright and Frantz (2014) show that the share of all dictatorships that are personalist has been increasing near-steadily since 1950.

unused data includes information on offshore companies belonging to tens of thousands of individuals from all countries during the period 1990-2015.⁴ Since this information was leaked from a firm that charges large fees for its services and holds “special expertise in creating tax shelters for wealthy global elite”, I assume that the individuals in its records are, indeed, economic elites (Graham, 2016). Thus, for each dictatorial country-year since 1990, I treat the names associated with country k in year t as its elites, and I code elites i and j from k, t as financially tied if they are both associated with one or more of the same offshore companies. Through this method, I construct the elite financial networks needed to measure network density for a large sample of dictatorships and years.

Empirics To test my hypothesis, I estimate the effect of network density on the price of insurance against expropriation. I obtain the latter from the annual country ratings of the leading political risk insurance agency. My sample includes 51 dictatorships in the period 2002–2013. My analysis shows a statistically significant and substantively strong association in the hypothesized direction: all else equal, a one standard-deviation increase in elite financial network density is associated with as much as a half standard-deviation decrease in expropriation risk—equivalent to the difference between capital-controlling China and the capital-friendly United Arab Emirates. Other variables mostly have an inconsistent or statistically insignificant effect. My main finding is robust to a barrage of controls, unobservable effects, temporal dynamics, and statistical irregularities, and to using alternative dependent variables, independent variables, and rules for coding the underlying networks.

Contributions This study contributes to several literatures. The first literature is that from political economy and economic history, on the emergence of property rights and contract enforcement vis-à-vis limited government (North and Weingast, 1989; North, 1990). The economic interdependence of elites and sovereign has featured heavily in that literature; I expose how elites’ economic interdependence affects their interaction with the sovereign. A second literature this study builds on is that from comparative politics, on power-sharing and elite politics in dictatorships (Bueno De Mesquita et al., 2005; Gandhi, 2008; Svobik,

⁴To my knowledge, only one other research paper has used this information to date, to study the effects of offshoring on firm valuation (O’Donovan, Wagner and Zeume, 2016).

2012; Geddes, Wright and Frantz, 2018). Though elites’ bargains with dictators have been thoroughly examined—along with the role of private information in those bargains—this study explores how networks condition the effect of private information on elites’ position. A third literature related to this article is the international political economy literature on foreign direct investment and expropriation (Jensen, 2008; Jensen, Malesky and Weymouth, 2014). That literature looks at how political institutions—dictatorial or democratic—protect investors from economic predation. Instead, I focus on a non-institutional safeguard against predation, financial networks. A related body of work is the literature on *de facto* property rights that arise through the action of guilds, business associations, and supply chains (Greif, Milgrom and Weingast, 1994; Doner and Schneider, 2000; Johns and Wellhausen, 2016). These studies take the role of informal institutions in creating property rights seriously, but they do not examine networks.⁵ Methodologically, my approach belongs to a large literature in sociology, political science, and economics, which studies the effects of various types of networks on individual-level and network-level outcomes (see Jackson (2008) for a review). In terms of subject matter, a relevant literature is that on the causes and effects of tax haven and shell company usage (Johannessen and Zucman, 2014; Findley, Nielson and Sharman, 2014). Finally, from a policy perspective my findings suggest that, though dense financial ties between elites have negative normative connotations, they counter a more negative force, economic predation by the dictator.

Roadmap The remainder of this study proceeds as follows. Section 1.2 briefly reviews the literatures relevant to my research question. Section 1.3 lays out my argument in the form of a model sketch and derives its main empirical implication. Section 1.4 introduces my financial network data. Section 1.5 presents my empirical approach, the rest of my data, the results of my analysis, and robustness checks. Section 1.6 discusses alternative interpretations of my findings and limitations of my analysis. Section 1.7 summarizes and points to directions for future research.

⁵One exception is Razo (2009) and related pieces.

1.2 Literature

A number of literatures relate to this study: the literature on property rights, contract enforcement, and limited government; the literature on foreign investment and expropriation; the literature on business associations and lobbies; the literature on power-sharing and elite politics in dictatorships.

Common Structure Reviewing these literatures is a gargantuan task. Fortunately, all share a common theoretical structure: an agent in a position of authority, A, interacts with a group of agents in a position of lesser authority, B, in a sometimes competitive, sometimes co-operative manner. A is usually the state, government, ruler, elected leader, or dictator, and members of B are elites, domestic or international firms, or investors. When cooperative, their relationship involves A protecting B's property rights, committing to announced policies, or sharing rents and power. A cooperative relationship also involves B supporting A's rule, by lending capital to A, not aiding coup/rebellions against A, or not defecting to a challenger. When their relationship becomes competitive, A and B engage in the opposite actions from the above.

Examples For example, the property rights theory of [North and Weingast \(1989\)](#) features the English crown and the landed gentry. The crown first expropriates the gentry, then commits to protecting its property—by increasing the powers of parliament and judiciary—in order to secure loans. In [Svolik \(2009\)](#) and [Boix and Svolik \(2013\)](#), the dictator interacts with a coalition of allies; the dictator shares or withholds rents from allies, and allies support or rebel against the dictator. Similarly, in [Guriev and Sonin \(2009\)](#), the players are a ruler and oligarchs, with the former having the power to expropriate the latter and the latter having the power to replace the former. Examining a very different setting, [Johns and Wellhausen \(2016\)](#) study the interaction of a government with foreign firms, where the government can honor or break contracts, and firms invest in protecting themselves from expropriation. A number of other studies from international political economy adopt a similar theoretical framework ([Jensen, 2008](#); [Jensen, Malesky and Weymouth, 2014](#); [Wilson and Wright, 2017](#); [Graham, Johnston and Kingsley, 2017](#)).

Intervening Factor Within the above unifying framework, there is often a role for an intervening force that conditions the relationship between A and B: for example, an economic shock, conflict, or technological change. However, this force might also operate through changing the way actors *within* B interact, which, in turn, affects B’s interaction with A. A prominent example in the literature is political institutions. In [Boix and Svolik \(2013\)](#), institutionalized power-sharing allows allies in the dictator’s coalition to monitor his compliance with their rent-sharing pact. Similarly, in [Gehlbach and Keefer \(2012\)](#), the creation of a ruling party enables the dictator’s supporters to observe expropriations against party members. In both theories, institutions facilitate coordination among the actors in B to change the behavior of A—in a welfare-improving manner. However, institutions are only one among a myriad of factors that can alter the interaction of the actors in B.

Networks An understudied factor that affects the way firms, investors, or elites interact are *networks*—the ties that bind these actors. A large literature in sociology and economics shows that networks impact how rational actors coordinate joint actions, commit to reciprocate in certain actions, and build social capital ([Jackson, 2008](#)). Actors can have different ties (e.g. ethnicity, religion, nationality, education), both cooperative and competitive. For firms, investors, and elites, arguably the most important are *financial ties*—after all, these are primarily economic actors. A key function that financial ties serve is to diffuse valuable private information on the economy. This is particularly the case in dictatorships, where policy-making is more arbitrary and political connections especially valuable ([Fisman, 2001](#)). As such, one way to advance our understanding of how firms, investors, or elites coordinate with each other in interacting with dictators, is to study financial networks among these actors.

1.3 Theory

Elites I focus on the network of actors most likely to hold, share, and act on private information about the economy, elites. Elites are also the group most likely to benefit from constraining the dictator—they are the main target of predation.

Setup I present my argument in the form of a sketch model; an infinitely repeated game between a dictator and a set of N elites.⁶ Elites form a given financial network—a set of bilateral financial ties—that determine the communication of information (more below). In the initial time period, dictator and elites form an agreement to share a fraction $0 < \beta < 1$ of total rents from their joint rule in each subsequent period. In all subsequent periods, the following sequence of plays takes place. First, the total value of rents, r , is determined by stochastic economic forces: $r = 1$ with probability π (good times), and $r = 0$ with probability $1 - \pi$ (bad times). Crucially, only the dictator observes r , along with a randomly chosen elite—label her i .⁷ In the next step, informed elite i decides to which uninformed elites, if any, to reveal r (costlessly). Then, the dictator chooses how much to share with each elite (including i): β/N or 0. Subsequently, elites observe their share of rents and decide whether or not to punish the dictator.

Punishment & Cost Punishment can take the form of capital flight or any type of capital reallocation that might hurt the economy and, hence, create negative externalities for the dictator.⁸ However, elites incur a cost c for reallocating their capital.⁹ Therefore, elites prefer not to punish the dictator unless the expected gain from punishment—the rents they will earn discounted by the probability that the dictator will award them rents due to the threat of punishment—exceeds c . Unfortunately, for uninformed elites that probability is zero because the dictator knows that their threat is non-credible.

Predation Equilibrium To see how the dictator preys on uninformed elites, ignore informed elite i for now. Because uninformed elites do not observe the true size of rents, they can only adopt one strategy to deter the dictator from predation: punish him whenever they receive no rents. However, to execute this strategy uninformed elites have to punish

⁶I borrow some non-network-related elements from the formal models in [Svolik \(2009\)](#) and [Boix and Svolik \(2013\)](#).

⁷The theory's predictions are substantively similar if we allow more than one elite to observe r .

⁸Most theories of authoritarian survival argue that a strong economy bolsters the dictator's tenure. This effect can operate through rent-sharing with elites and the military or distributive politics to alleviate popular threats like revolutions and civil conflict.

⁹I think of c as the sum of brokerage fees, capital gains taxes, and regulatory compliance costs elites incur for reallocating capital, especially if moving it abroad.

the dictator *even when he has not reneged on their pact and times are simply bad*. The dictator knows this, and so the strategy contains a non-credible threat—uninformed elites are better off not punishing the dictator and foregoing cost c .¹⁰ Thus, in equilibrium, the dictator is not deterred from predation, keeps all rents to himself (when there are any), and uninformed elites always get 0. In short, *when there is a complete information asymmetry between dictator and elites regarding the true size of rents (i.e. all elites are uninformed), the unique equilibrium involves complete predation*.¹¹

Information The presence of informed elite i alters this equilibrium. Because i can condition her response on the economy’s state, she avoids costly, unnecessary punishments—when times are bad i attributes low rents to stochastic forces and does not blame the dictator. Conversely, when times are good yet she receives low rents, i infers defection and seeks to punish the dictator. Thus, i ’s threat to punish the dictator if he defects is credible and incentivizes the dictator to share the agreed-upon rents with i (β/N). In other words, i ’s private information and the threat she conditions on that information secures her a *selective commitment* by the dictator (Haber, Maurer and Razo, 2003; Razo, 2009).¹² For the dictator to make additional selective commitments, though, more elites need to join i in making credible threats. This requires that i communicates her private information about the economy’s true state, enabling additional elites to infer and punish defection.

Communication How does communication between elites occur? Recall that, after observing the true size of rents r , i can costlessly reveal r to other elites. However, i ’s incentive to *truthfully* reveal r to another elite will depend on their financial relationship—whether they are competitors or partners. Truthful communication between competitors is not possible. In particular, i has an incentive to lie to a competitor when times are bad and claim that the dictator defected; the competitor will then reallocate her capital, thinking

¹⁰That is, uninformed elites prefer 0 (no rents) to $-c$ (no rents and costly punishment).

¹¹This is similar to the no-power-sharing equilibrium in Boix and Svolik (2013).

¹²A richer implication would involve the dictator sharing rents equal to the punishment i can inflict on him—a form of “rent discrimination” by the dictator towards elites. In this manner, informed elites would be able to extract larger rent shares from the dictator the more capital they have. This feature can be incorporated without changing the core of the model but would needlessly complicate the analysis.

that she is punishing the dictator, which can hurt her and benefit i . On the contrary, i has no incentive to lie to a financial partner, say j , when times are bad. If i lies, j 's capital reallocation can adversely impact i ; for example, capital might be moved out of a joint venture.¹³ Thus, I limit my attention to communication between informed elites and their financial partners (if any).¹⁴

Network Structure Communication between informed elites and their partners depends on the financial network's structure: how many partners i has, how many partners they have, and so on and so forth. To illustrate the importance of network structure, I analyze two toy examples, depicted in Figure 1.1. Both networks have 7 elites, but vary significantly in how interconnected elites are.¹⁵ The left network has 5 elites with no ties and 2 elites that are only tied to each other, while in the right network every elite is tied to at least 3 others. Given my assumptions about communication between elites, in the left network 1 or 2 elites learn the true size of rents, versus 4 or 7 elites in the right network (depending on which elite is endowed with the private information). Recalling that, in every period, one elite i randomly observes the true size of total rents r , we can derive the *average* number of elites to which i communicates r . This statistic is simply the sum of each elite's probability of acquiring information ($1/N$) times the number of elites she can communicate that information to (i.e. her number of ties). This equals $5 * \frac{1}{7} * 1 + 2 * \frac{1}{7} * 2 = \frac{9}{7} = 1.28$ elites in the left network and $6 * \frac{1}{7} * 4 + 1 * \frac{1}{7} * 7 = \frac{31}{7} = 4.43$ elites in the right network. This means that, on average, more than 3 additional elites learn r in the right network, and will infer and threaten to

¹³In addition, if i lies, j will soon infer the lie, when i fails to punish the dictator herself. Note that, even though j will only observe i 's lie in the next time period, j might punish i then, by terminating their financial tie—a tie that i presumably derives value from. And because financial partners play a repeated game, i should be deterred from lying to j . Another reason, though more difficult to insert in a rational choice model, might be trust between i and j . Indeed, much sociological research records higher levels of trust between connected nodes in different kinds of social networks (Cook and Hardin, 2001).

¹⁴An altogether different motivation for the assumption that only financial partners can communicate is informational: elites that are not financially tied might not even be aware of each other's existence. This assumption is more realistic for large, populous, and/or ethnically diverse countries, where there are many, heterogenous elites. Also, note that if i has no ties, the unique equilibrium is the same as with no communication: complete predation.

¹⁵Throughout, I model and measure ties as undirectional; that is, for any i and j , if i is tied to j , then j is tied to i .

punish any defection by the dictator, thus securing selective rent-sharing commitments from the dictator.¹⁶

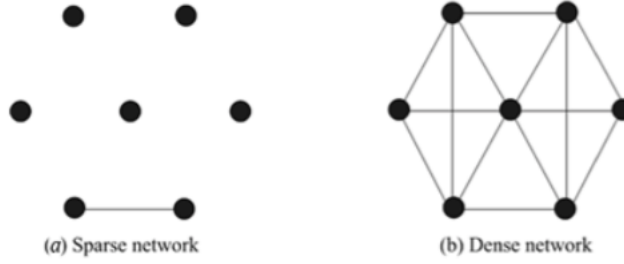


Figure 1.1: Two Examples of Elite Networks

Network Density A well-known and readily operationalizable variable that reflects how network structure affects information diffusion is network density. Indeed, a large literature in sociology and economics links network density to important behaviors like cooperation, exchange, and conflict (see Jackson (2008) for a partial review). For an undirected network k in period t , density is calculated as $d_{k,t} = \frac{\sum_i^{N_{k,t}} x_j}{N_{k,t}(N_{k,t}-1)}$, where x_j denotes the number of financial ties of elite j and $N_{k,t}$ the number of elites in network k in period t (Kolaczyk, 2009).

Hypothesis Coupled with the concept of network density, the above stylized example brings us to the core of my theory and its key prediction: *where elites' financial network is more densely interconnected, the dictator is less likely to expropriate elites*. A dense financial network diffuses information about the true size of the regime's rents more widely, allowing more elites to infer whether the dictator reneged on his rent-sharing agreement and threaten

¹⁶A richer version of my sketch model would feature the dictator sharing all of β with only those elites that secure selective commitments from him; say, n elites, where $n < N$. Under this modification, an informed elite might have an incentive to withhold information from her uninformed connections, in order to keep n low and reap a larger β/n from the dictator (along with other informed elites that secure commitments). However, to dispel this incentive we must simply assume that an informed elite gains more from sharing information with a financial partner—through the various channels outlined above—than she does by increasing her share of rents from $\beta/(n+1)$ to β/n , especially when n is large, i.e. there are many informed elites. Moreover, we might imagine that informed elites face a public goods provision problem: every informed elite has an incentive to share information with her connections, thereby gaining from aligning her strategy with her financial partners, while free-riding on other informed elites' efforts to keep n low by withhold information from their own partners. This would create an equilibrium whereby no informed elite withholds information from her partners.

to punish him. Credible threats of punishment by more elites, in turn, should act as a larger deterrent against predation, thereby producing selective commitments to protect more elites' property. In short, I hypothesize that, all else equal, *dictatorships with dense financial elite networks enjoy stronger property rights*.

Selective vs Collective Commitments Before proceeding to the study's empirical component, I note an important distinction between my theory and the literature. Most theories of property rights formation in dictatorships focus on elites' collective action problem in constraining the dictator (North and Weingast, 1989; Svolik, 2009; Boix and Svolik, 2013). A common feature of these theories is that elites – the group of agents attempting cooperation – secure a *collective* commitment by the dictator to protect their property. Though collective commitments are undoubtedly a feature of dictator-elite interactions, so are *selective* commitments (Razo, 2009). Indeed, since authoritarian politics are more discretionary and relation-based than democratic politics, especially in less institutionalized regimes, the literature's near-exclusive focus on collective commitments seems unwarranted. Moreover, the study of selective commitment seems like a natural progression, given the literature's recognition of the fragility of formal institutions in dictatorships and the shifting focus to the study of informal institutions (Keller, 2014). In short, in terms of the literature, my theory is situated as one about how an informal institution enables elites to secure selective commitments from the dictator.

1.4 Data: Elites & Financial Networks

Network Data Unlike other independent variables used in statistical analyses, we cannot obtain network density from existing data; we must calculate it. To do so for a sample of K countries in T years, we first need to construct an equal number of networks of elites' financial ties. Collecting network data on elites' financial ties is far from straightforward, especially in dictatorships. It requires defining who the elites are and recording every elite's ties to every other elite. For a large country like China, this could mean hundreds of thousands of elites (nodes) and millions of potential financial ties (edges). One approach would be

to use primary and secondary sources that identify economic elites and their financial ties in a sample of dictatorship-years. However, this would introduce bias, stemming from the disproportionate media coverage of prominent elites. Less prominent elites and ties between them are more likely to be omitted from news reports, which would result in networks that are smaller (fewer elites) and, possibly, sparser (fewer ties per elite) than reality. In addition, if we want to compare networks across countries and years, we need data on elites’ financial ties that is recorded in the same way for all units. To see why this is crucial, imagine that elites’ financial networks are identical in two dictatorships, k and k' , yet k receives more coverage than k' due to its alliance with the US. In addition, imagine that the US’s alliance with k also causes its dictator to expropriate less. The above data-collection approach will produce a network that is denser for k than k' , which will lead us to spuriously attribute k ’s lower expropriation to its higher network density instead of its alliance to the US. Clearly, traditional techniques of network data-collection—observing or surveying the network’s nodes (elites)—are of no use here (Wasserman and Faust, 1994). Equally blunt are the methods used to study other networks of political actors, such as co-sponsorship networks in Congress (Fowler, 2006).¹⁷

Panama Papers To resolve obstacles in network data-collection, I exploit a new, unique, and untapped source of information on hidden financial ties. Constituting the largest leak of private financial information to date—11.5 million files adding up to 2.6TB of data—the so-called Panama Papers are the full body of documents of one of the largest provider of offshore legal services, Mossack Fonseca (Mossfon) (International Consortium of Investigative Journalists, 2016). These documents were leaked by an anonymous source to journalists in 2015. After more than a year of preparatory work by a team of 400 individuals from 100 news organizations in 80 countries, a reduced version of the documents were made available for public download.¹⁸ In the words of the team behind the leak: “the real value of the database

¹⁷Mahdavi (2017) proposes a method to construct affiliation networks of political elites via joint appearances in gala events, while Mahdavi and Ishiyama (2018) construct an affiliation network of N. Korean party elites via joint appearances in state-sponsored factory inspections. However, these methods are not scalable to a large number of countries.

¹⁸See <https://offshoreleaks.icij.org/pages/about> for further information on the data. The vast majority of Mossfon’s leaked documents are excluded from the public dataset—email exchanges, bank account

is that it strips away the secrecy that cloaks companies and trusts incorporated in tax havens and exposes the people behind them” ([International Consortium of Investigative Journalists, 2016](#)). This is the data I use to construct elites’ financial networks in all dictatorships since 1990.¹⁹

Elites To use this data to record elites’ financial ties, we must assume Mossfon’s clients are elites. According to reports on the market for offshore services, Mossfon has “special expertise in creating tax shelters for the *wealthy global elite* [emphasis added]” ([Graham, 2016](#)). This is evident in the company’s reported fees: depending on the service required, charges range from \$1,500 per year for setting up an offshore company in a not-so-costly jurisdiction to \$17,500 per year for providing a nominee director that acts on the beneficiary’s behalf ([Harding, 2016a](#)). Though the data does not indicate which services were purchased by each elite, some of Mossfon’s fees exceed the median per capita income of the vast majority of post-1990 dictatorships.²⁰ In other words, merely purchasing Mossfon’s services is an expense that only the very wealth in dictatorships can incur.²¹ Another piece of evidence on the elite status of Mossfon clients is offered by [Alstadsæter, Johannesen and Zucman](#)

numbers, and financial transactions. However, for the purposes of this study, the excluded information is not necessary to record elites’ financial ties (see Financial Ties paragraph). The public version of the data includes all clients’ information on: their nationality, address, the companies they are associated with and their role in those companies, intermediaries used to establish the account (if any), and various dates relevant to the companies’ status. In some cases, some of this information is missing, though there is no obvious pattern to this missingness.

¹⁹In 11/2017, a related dataset, called the Paradise Papers, was leaked to the same media that made the Panama Papers publicly available. The Paradise Papers consist of files leaked from Appleby, a competitor of Mossfon. The Paradise Papers’ size is 1.4TB of data, a little more than half that of the Panama Papers. Though the Paradise Papers have not been made publicly available as of 11/14/2017, media coverage suggests that Appleby’s clients are mostly multinational corporations and wealthy Westerners. As such, it is unlikely that the Paradise Papers will illuminate financial ties among elites in dictatorships to the extent that the Panama Papers do.

²⁰Based on author’s calculations using data from the Quality of Government (QoG) dataset ([Teorell et al., 2013](#)).

²¹A simple back-of-the-envelope calculation adds further weight to my assumption—particularly for poorer dictatorships. Assume there is a 1% probability that, in any year, client i will lose a particular asset—through expropriation, imprisonment, etc—if she does not register it in an offshore company. Further assume that the annual fee Mossfon charges for incorporating i ’s asset in offshore company is \$2,000. This implies that, for i to be making a rational decision in seeking Mossfon’s service, her asset must be worth at least \$200,000. And, given that i is highly unlikely to have a single asset and no source of income, i ’s net worth must arguably be in the top percentile of the distribution in most post-1990 dictatorships.

(2017a), who are able to match Norwegian and Swedish account holders to their tax returns. The authors find that only the top 0.1% household earners from two of the world’s richest countries own Mossfon accounts. Thus, I safely assume that Mossfon’s clients are, indeed, economic elites, and I use the data leaked from Mossfon to record elites’ financial ties. Note a direct benefit of using the Panama Papers to identify elites: we do not have to identify elites individually, for a large number of countries. Their wealth, as implicitly revealed through purchasing Mossfon’s services, renders them elites, and no researcher discretion is needed to classify them as such.²²

Financial Ties To lend itself to testing my hypothesis, the Panama Papers must also measure elites’ financial *ties*; that is, meaningful and cooperative financial relations. The data includes information on the companies associated with each client—the companies’ names, jurisdictions (which countries they are registered in), incorporation dates, and inactivation dates (if applicable)—and clients’ positions in those companies (e.g. director, beneficiary, shareholder, secretary, etc).²³ I use this information to record financial ties between clients (elites). Namely, I code elites i and j from country k ²⁴ as tied in year t if they are both associated with the same company, m , and m is active in t .²⁵ Given that clients associated with a company have veto power over its activity, I assume that i and j have a strong, cooperative financial tie if they both have a substantive position in the same

²²Due to the desire of Mossfon’s clients’ for anonymity, in a large fraction of companies it is not possible to discern the ultimate beneficiaries. Anonymity is usually achieved through the use of bearer shares, which award beneficiary status to their physical holder. Accounts that use bearer shares show “The Bearer”—or some variation of that title—as the account name. I do not use information from such accounts in constructing my networks, but there does not seem to be a pattern in the use of bearer shares.

²³Though a key feature of offshore companies is their jurisdiction, it does not play a role in constructing my elite networks. I code elites’ ties based on their own nationality or country of residence, not the country where their companies are registered. In this manner, elites i and j that are nationals of country k will appear in k ’s network, independent of where their companies are registered.

²⁴Note that there can be other elites that are associated with company m but are not nationals of country k —indeed, this occurs frequently in the data. Though, these elites are financially tied to elites i and j —via their joint association with company m —I do not include them in the network for country k , but in the network for their own country.

²⁵For a company to be active in year t , its incorporation date must precede t and its inactivation date (if applicable) must be later than t . Also, note that two elites can be tied through more than one company. I plan to account for this feature of the data in future versions of the analysis, by allowing the strength of elites’ ties to vary: the more companies elites i and j are jointly associated with, the stronger their tie.

company.²⁶ This allows me to construct a financial network through elites’ joint association with one or more offshore companies. Again, note that no researcher discretion is needed to code elites’ financial ties—they are revealed to us through elites’ own actions, and these actions are recorded by an agent with no incentive to misreport them (Mossfon). Another advantage of this approach is that elites’ ties are recorded in a uniform way across countries and years, because a single agent records these ties (Mossfon), using the same standard across countries.

Ties Example The following is an example of a typical entry in the raw data: New Russia Venture Partners Inc is an entity incorporated in the British Virgin Islands on 5/23/2003, and whose “officers” (i.e. the individuals associated the company) are Russian nationals Anna Baskakova, Igor Kubanov, and Sergey Vykhodtsev. I code these names as three of the nodes (elites) in the Russia 2003 network, and I add 3 edges (financial ties) between them. To complete the Russia 2003 network, I repeat this process for all other Russian nationals in the data that are associated with companies incorporated in 2003 or earlier. To complete the Russian network for my whole period of study, I repeat this process for all other years during 1990-2015 that Russia was a dictatorship. The same process is applied for all other dictatorship-year observations.²⁷

Networks Example Figure 1.2 shows the financial networks for two countries in the year 2015, Chad (left) and Botswana (right). Note the richer structure of Botswana’s network. At the graph’s 9 o’clock position there is a very large clique—a group of nodes that are all tied to each other directly—formed by a single company of many officers. Also, in the graph’s center there is a very large component—a group of nodes that are all tied (in)directly—formed by several companies with overlapping members. On the contrary, roughly 80% of Chad’s elites

²⁶There is a large number of positions a client can have in an offshore company, and some positions do not grant real power over the company. For example, if a name appears as the company’s “nominee director” (or a variation of that title), that name is not the true beneficiary of the company, but usually an employee of Mossfon that acts on the beneficiary’s account. As such, there is no reason to code that name as an elite and code its ties to other elites. To ensure that I record only substantive ties between Mossfon’s clients (elites), I researched the legal powers of every position, and I retain only the positions that grant real powers over companies.

²⁷I discuss the use of alternative coding rules for constructing my elite networks in Section 1.5.3.

have no ties, while connected elites only have 1 or 3 ties. These differences are reflected in a range of statistics, aside from the network size (48 nodes in Chad vs. 96 in Botswana). Most relevant to testing my hypothesis is their difference in density: Chad’s elite network has a density of 0.008, a tenth of the respective density of Botswana (0.080).²⁸ Moreover, density is associated with property rights protection in Chad and Botswana as predicted by my theory—Botswana is hailed as an institutional miracle, while Chad’s property rights are among the world’s worst (Acemoglu, Johnson and Robinson, 2002).

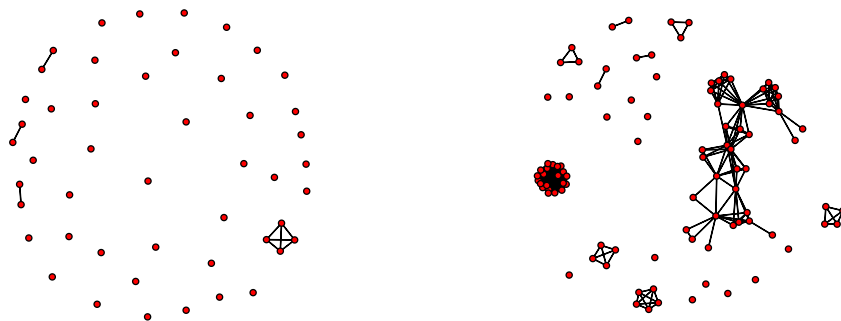


Figure 1.2: Elite Financial Network in Chad 2015 (L) vs. Botswana 2015 (R)

Notes: Each node represents an elite; each edge represents at least one financial tie between the corresponding nodes. See text for data methodology.

Mossfon Coverage How encompassing are the offshore networks in the Panama Papers? Obtaining estimates of Mossfon’s market share in offshore legal services is difficult, due to the obscure nature of that market. The only available estimates place the firm’s share at 5 – 10% (The Economist, 2016b), with one report labeling Mossfon the “industry leader” (The Economist, 2012), and others calling it the “fourth largest” provider (Harding, 2016b). Clearly, the Panama Papers by no means capture the universe of elites’ offshore financial

²⁸The difference between the two countries becomes much larger if we allow for indirect diffusion of information, due to the large component in the Botswanan network. That is, if elite i can communicate with j via their mutual tie with l , whenever an elite within the component acquires private information, it will diffuse to the whole component, which constitutes half the Botswanan network.

ties—there are elites that do not use Mossfon for their offshore financial activities, as there are elites that do use Mossfon but might have other offshore financial ties that are not established through Mossfon. That said, the data includes information on roughly 136,000 individuals and 310,000 offshore companies from all 101 countries that had one or more dictatorial spells during the period 1990-2015.²⁹ Moreover, there is no reason to expect that the pattern of elites’ non-Mossfon-established offshore ties is systematically different from that appearing in the Panama Papers.

Offshore vs Onshore Ties Another concern with using the Panama Papers to capture financial ties between elites is whether offshore ties are meaningful. One might argue that *onshore* ties are more important, since most individuals hold a larger fraction of their assets onshore and onshore ties are more transparent.³⁰ In the case of elites in dictatorships, this argument is problematic for three reasons. First, elites in dictatorships most likely hold the majority of their wealth offshore. Using estimates from [Alstadsæter, Johannesen and Zucman \(2017b\)](#), I find that the average country in my sample holds around 17% of its GDP offshore, with that figure reaching over 50% for Russia, Saudi Arabia, Venezuela, and the UAE. And given the progressive nature of taxation in most countries coupled with the transaction costs of holding assets offshore, we can assume that elites in these countries hold a higher percentage of their wealth offshore than the average income-earner. Indeed, simple calculations show that even in countries with a relatively low share of their GDP offshore, elites are more likely than not to hold more than half their wealth offshore.³¹

²⁹These statistics are produced after cleaning the original data to include only individuals and their offshore companies from countries that had dictatorial spells during the period 1990-2015. The original data includes many more individuals and their companies from democracies and/or from earlier years, as well as many observations that cannot be attributed to specific countries. I use an extended version of the Geddes-Wright-Frantz data to identify which countries are dictatorships during 1990-2015 ([Geddes, Wright and Frantz, 2014](#)).

³⁰I use “onshore ties” to refer to shared financial interests between elites in one or more assets that are located in elites’ country of residence. I use “offshore ties” to refer to shared financial interests in assets that are either (i) located in a country different from elites’ country of residence, or (ii) located in elites’ country of residence but incorporated in a company that is registered in a *different* country (e.g. an apartment in Russia owned by Russian elites but incorporated in a Bahamas-registered company).

³¹To verify this claim, recall that the average country holds 17% of its GDP offshore. Assume that the top 1% of income earners, call them elites, own 20% of GDP and the remainder is owned by the bottom 99%—a relatively unequal distribution for the countries in my sample. If the bottom 99% hold all of their

The second advantage of using offshore ties to measure elites’ financial connections is that they are unobservable to the dictator.³² On the contrary, the dictator can observe onshore ties, thereby possibly deterring connected elites that want to conceal their connection from establishing onshore ties. Thus, because of their legal obscurity offshore ties should be more accurate than onshore ties in capturing connections between elites in dictatorships. A third, related argument in favor of using offshore ties is that they limit spillovers for elites. Due to the threat of predation, financial ties formed through onshore assets expose connected elites to negative externalities from their connection. For example, if elites i and j form a joint venture through an onshore company that gets expropriated because the dictator wants to prey on i , j will incur a negative spillover. Crucially, the dictator cannot prey on offshore assets, while onshore assets that are incorporated in offshore companies are harder to prey on than those incorporated in onshore companies. Therefore, offshore ties provide elites in dictatorships greater protection against indirect predation than onshore ties. For these reasons, using data on offshore ties to capture elite financial networks actually strengthens my research design.

Network Representativeness Putting aside the arguments in favor of using offshore ties, it is useful to assess the representativeness of the offshore networks in my data. In particular, we might want to know whether my networks are atypically dense/sparse. Ideally, to make that assessment I would use information on the observed elites that predicts elites’ propensity to form financial ties. One feature that might predict tie formation is wealth. If the distribution of wealth among elites in the data is similar to that among elites in dictatorships at large, we would have one less reason to doubt the data’s validity. Unfortunately,

assets onshore, this implies that elites hold 85% of their assets offshore. Note that the percentage of elites’ assets held offshore increases the larger the percentage of the country’s GDP held offshore and the lower the percentage of GDP held by elites (i.e. the more equitable the income distribution). For the average country in my sample, as long as elites own less than 34% of GDP—roughly the figure of the US, one of the most unequal countries as measured by Gini—my calculations imply that elites hold more than half of their assets offshore. In short, for the majority of the countries in my sample, it is likely that the bulk of elites’ assets are offshore.

³²Unobservability does not only hold for ties formed through assets that are located offshore, but for ties formed through assets located onshore but incorporated in offshore companies. In fact, because of this unobservability, offshore companies are labeled “shells”—they are often used to obscure the physical person(s) that ultimately own the assets.

the Panama Papers provide no substantive information on the observed elites, while the sheer number of elites in the data renders an exhaustive merge with external data sources unfeasible. Thus, I restrict my attention to the most wealthy and public elites, billionaires, and calculate what percentage of the world’s known billionaires from dictatorships are observed in my data. I use the most comprehensive source on billionaires, Forbes’ *The World’s Billionaires* list (Kroll, Miller and Serafin, 2010). For 2010, 11.5% of the list’s billionaires in dictatorships appear in my data.³³ Recalling that Mossfon’s share of the whole market for offshore legal services could not have exceeded 15% that year, it seems that the ultra-wealthy are neither over- nor under-represented in the Panama Papers. Though this is clearly a limited assessment of the representativeness of observed elites’ wealth, let alone of other features that might affect the structure of the observed financial networks, it is reassuring that one test of the data’s comprehensiveness produces no concerns.

Network Cross-Checks Another way to assess whether the networks in the Panama Papers accurately reflect the broader set of ties between elites is to compare my data to other sources. As hinted earlier, there are no other cross-national datasets of elite networks that cover *all* dictatorships. Hence, I resort to the assessments of two region specialists. Haddad (2011), an authoritative source on Middle Eastern elite networks, argues that Syria’s elite network is particularly dense. Additionally, Cárdenas (2015) conducts an analysis of interlocking corporate directorates in Latin America and finds that the Chilean business network is denser than the Argentine one. And using a similar method, Cárdenas and Robles-Rivera (2017) find that Panama’s elites are more densely interconnected than Costa Rica’s, which, in turn, are more densely interconnected than El Salvador’s. Indeed, in my data the Syrian network is the densest among all Middle Eastern ones, Chile’s network is more dense than Argentina’s, and Panama’s is more dense than Costa Rica’s and El Salvador’s (though the latter is denser than the former).³⁴ Again, although these cross-checks do not provide

³³The Forbes list includes all billionaires with estimated net wealth above 2 billion US dollars. The methodology for estimating billionaires’ wealth is described in Kroll, Miller and Serafin (2010). I use the year 2010, as it was the latest year for which I could acquire comprehensive data.

³⁴Furthermore, Haddad (2011) and many others have emphasized the importance of Rami Mahklouf, a cousin of Bashar al-Assad, among Syrian elites; he is one of the wealthiest Syrian businessmen and controls several state-owned companies. In line with his perceived centrality in the business world, in my data

conclusive proof of my data’s validity, they do not raise any red flags for my data. Therefore, I proceed to use the Panama Papers as a unique source of information on elites’ financial ties, with respectable coverage and no documented concerns of non-representativeness.

1.5 Analysis

In this section, I describe the approach used in my statistical analysis and present and discuss my baseline results and robustness checks.

1.5.1 Approach

Statistical Model My data is structured as a cross-sectional time-series (maximum of 101 countries and 26 years). The regressions I estimate for country k and year t are generally of the form

$$y_{k,t} = \alpha + \beta d_{k,t-1} + \mathbf{x}_{k,t-1}\boldsymbol{\gamma} + \epsilon_{kt} \quad (1.1)$$

where d is network density, as described in Section 1.3, and $\mathbf{x}$ a row-vector of controls. I begin by estimating a simple model with few control variables, then progress to more sophisticated models and add more controls. In all of my models, I lag time-varying predictors by one year to mitigate some concerns of endogeneity, and I log predictors when it transforms their distribution to one that better approximates the Normal.³⁵ All of my models use country-clustered standard errors to account for within-country cross-year correlation of unobservables.

Dependent Variable Features My theory predicts a negative effect of elite network density on dictators’ predation of elites. Unfortunately, predatory economic policies are hard to observe, and any proxy variable will be imperfect. That said, a good proxy variable should exhibit at least the four features. First, it should cover an adequately large sample

Mahklouf appears as the most central node in the Syrian network. Note that the Latin American countries I mention here were overwhelmingly democratic during the period I study, so they are not included in my later analysis.

³⁵I address the issue of endogeneity in Section 1.6.

of dictatorships and years, in order to provide my regressions statistical power and external validity; second, it should be measured in a consistent way across dictatorships and years, to reduce the risk of non-random measurement error; third, it should be non-binary (ordinal, discrete, or continuous), so as to allow statistical models to discriminate between varying levels of predation in different dictatorship-years; fourth, it should capture *actual*, not *perceived*, property rights. Unfortunately, most variables employed in the property rights literature fail the last criterion, as they are indexes based on expert opinion surveys.³⁶ These indexes face well-known critiques (e.g. Treisman (2007)): experts base their answers to property rights survey questions on proxies for property rights (e.g. capital flows), thereby creating spurious correlations between these proxies and property rights.

Dependent Variable A variable that is not based on perceptions and features in several political science studies is the ratings scale of political risk insurance agency Credendo Group (formerly ONDD) (Credendo Group, 2017).³⁷ Credendo is a leader in the political risk insurance market, and insures companies operating in any country against various political risks. One of Credendo’s insurance products is closely related to the behavior I am studying: the *Expropriation Risk* product protects companies against direct expropriation or breach of contract by the government. Crucially, insurance agencies and their clients are rational profit-maximizers, thus we can be relatively certain that Credendo’s premia accurately reflect the risks it insures against.³⁸ Moreover, as documented in interviews conducted by Jensen (2008), Credendo is a price-maker in the political risk insurance market, and insurance brokers exhibit price convergence the premia they charge. Therefore, Credendo’s premia can also be considered representative.³⁹ The premia charged to each client are confidential, and the

³⁶In Section 1.5.3 I discuss the use of some of these indexes in my analysis.

³⁷Credendo Group insurance premia were first used by Jensen (2008) and have since been used by Jensen, Malesky and Weymouth (2014), Wilson and Wright (2017), and Graham, Johnston and Kingsley (2017), among others. The version of the data I use is from the replication files of Wilson and Wright (2017).

³⁸In addition, expropriation risk insurance seems to be the main revenue source for political risk insurance agencies. From 1991–2004, 84% of settlement claims received by the Overseas Private Investment Corporation, a major US government agency that insures against political risk, were for expropriations (Jensen, 2008). This further emphasizes Credendo’s incentive to price expropriation risk correctly.

³⁹To verify that Credendo ratings capture actual expropriation and other political risks affecting business, Graham, Johnston and Kingsley (2017) use the ratings to predict *de facto* measures of these risks, like

methodology for pricing the premia is proprietary. Fortunately, Credendo publicly releases an annual rating for each country, which is based on the underlying insurance contracts it sells to clients operating in that country. Like all insurance products, larger risks command higher premia, and so the higher the rating a country receives on Credendo’s scale, the greater the expropriation risk in that country. That scale ranges from 1–7 and is comparable across countries and years.⁴⁰ Though Credendo’s ratings for expropriation risk only cover the period 2002–2013, they include 70 out of 75 GWF dictatorships observed in that period, resulting in 666 data points. In short, the Credendo expropriation risk ratings satisfy all four criteria for my dependent variable: they are based on actual expropriation, are reasonably fine-grained, comparable across observations, and provide adequate coverage.⁴¹

Independent Variable As described in Section 1.4, my independent variable is the density of elite financial networks, as observed in the Panama Papers.⁴² That density exhibits a strong right skew, with a of 0.09, median of 0.045, and a standard deviation of 0.12. Figure 1.3 shows the geographic distribution of density for the sample used in my regressions (2002–2013), as a map of dictatorships’ average densities in that period. Like in many other spatial distributions, there is no unique interpretation. Yet, two patterns do stand out. The first is the low network density of dictatorships with former communist regimes. This might be owed to the relatively young nature of the economic elite in those countries, which were born through the haphazard transition from command to market economies. The second visible pattern is the high network density of countries in Southern Africa, which could be attributed to the ethnically cohesive nature of (clusters of) elites from the colonial era (e.g. British in Botswana and Zimbabwe, British and Germans in Namibia, and Arabs in Mozambique).

restrictions on capital flows. Using a series of statistical models, they find that the Credendo ratings closely predict actual measures of political risks to business.

⁴⁰Expropriation risk ratings (for GWF dictatorships) are distributed relatively symmetrically, with a mean of 4.12, median of 4, and standard deviation of 1.55. In my regressions, I standardize the variable for interpretability.

⁴¹In Section 1.5.3 I discuss the use of alternative measures of expropriation and property rights in my analysis.

⁴²I discuss the use of alternative independent variables in Section 1.5.3.

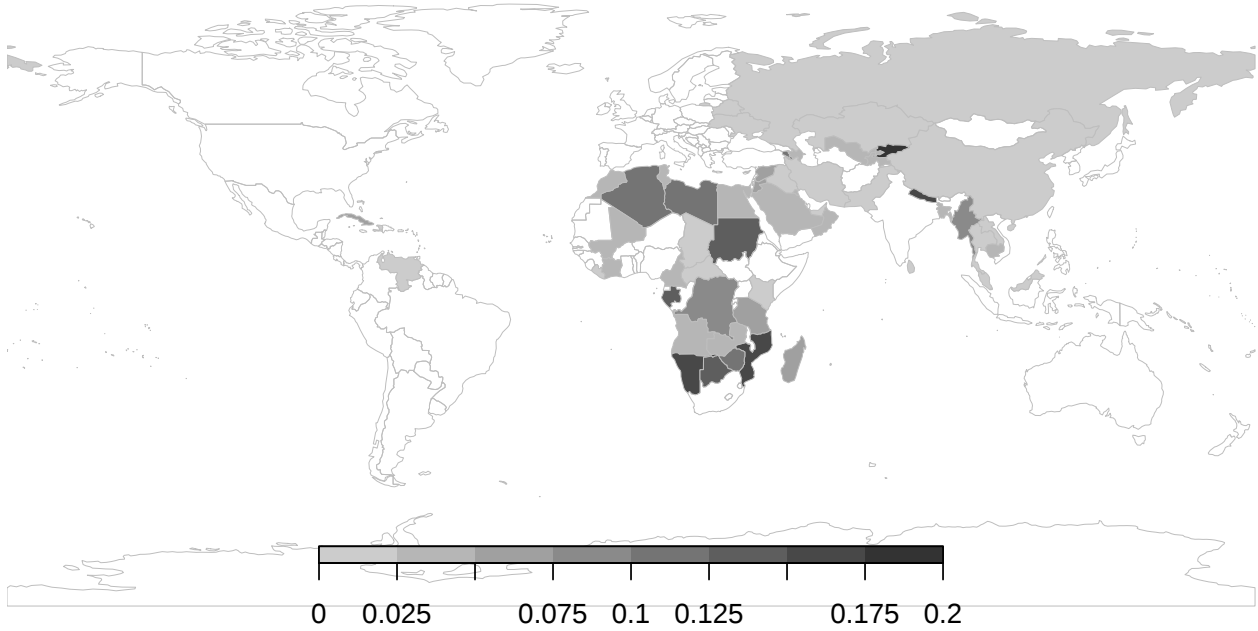


Figure 1.3: Global Distribution of Elite Financial Network Density

Notes: Average country density for 2002–2013 (years with non-missing dependent variable). Only 54 GWF dictatorships graphed (sample of interest); excludes 16 outliers (values 0/1). See text for data methodology.

IV & DV Since my data has a temporal dimension, it is informative to visualize the dependent and independent variables over time. Figure 1.4 graphs the mean value of both variables across all years for which data is available.⁴³ There are two main takeaways from the plot. The first is that average financial network density steadily decreases after 1995, from a high of 0.21 to a low of 0.03.⁴⁴ This drop might seem to counter the global trend of increasing financial ties between countries during this period. Yet, recall that the data captures financial ties among elites *within countries*. Moreover, there might be a tradeoff in elite ties, such that stronger between-country ties are associated with weaker within-country ties. Indeed, one of the conventional wisdoms on global financial openness is that it disrupts

⁴³Means are calculated using a different number and list of countries in each year, since countries phase in and out of my sample (transitions to and from dictatorship to democracy). I do not include confidence intervals for the two series, since they vastly reduce the graph's interpretability.

⁴⁴Before 1995, a maximum of 16 countries appear in the Panama Papers, so the mean density calculated for years up to 1995 is very noisy.

entrenched elite networks within countries (The Economist, 2016a). The second takeaway is that expropriation risk (standardized) increases near-steadily after 2005, from 0.17 standard deviations below its mean, to 0.24 standard deviations above it. A corollary of this pattern is that, as elite network density decreases during 2005–2013, expropriation risk increases. In fact, it is this trend that underpins the negative association between network density and expropriation risk reported in my regressions.

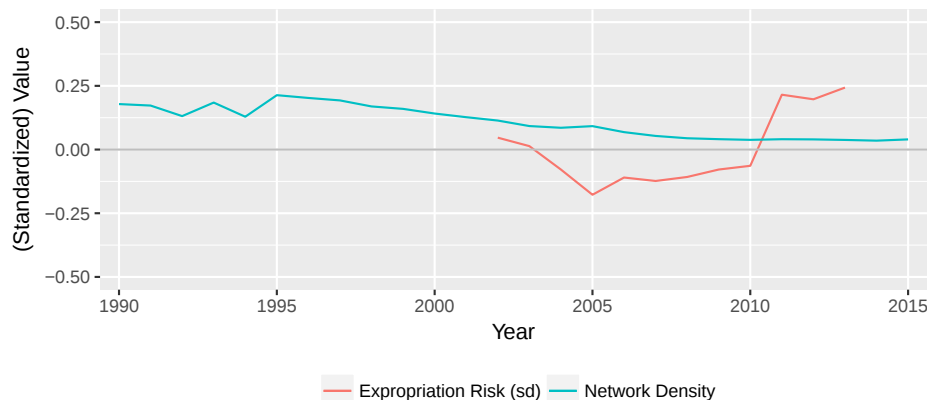


Figure 1.4: Dependent and Independent Variable Means Across Time

Notes: Expropriation risk values standardized for interpretability; data only exists for 2002–2013; mean for each year calculated using 36–42 countries. Density values in original scale (0–1); mean for each year calculated using 8–43 countries. See text for data sources and methodology.

Control Variables To control for spurious correlations between expropriation risk and density of elite financial ties I include several controls in my regressions:⁴⁵ first, the logged number of *nodes* (elites) in each network (country-year), since the number of nodes enters the formula for calculating density, and it might also affect the dictator’s predation directly (in larger networks there are more elites to constrain the dictator);⁴⁶ second, *GDP per capita* (logged, in current US dollars), as it might strengthen elite ties through increasing financial transactions, while decreasing dictators’ incentive to prey due to collecting more in taxes; third, the *growth* rate (of GDP per capita in current US dollars), because a faster growing

⁴⁵I include essentially all of the control variables that appear in previous regressions of the Credendo expropriation risk ratings (e.g. Jensen (2008)), while adding some further controls. I discuss the use of even more control variables in Section 1.5.3.

⁴⁶Also, note that nodes directly enter the formula for density in the denominator and the upper limit of the summation.

economy might also propel tie-generation, while making the dictator hesitant to disrupt a well-performing economic system; fourth, *inflation* (annual percentage change in consumer price index), as it might raise transaction costs for elites, hence reducing their financial ties, while tempting the dictator to seize control of companies in order to halt rising prices; fifth, *government net lending* (as a percentage of GDP), seeing as in countries where the government is a net lender to the private sector elites might have more resources to transact with, and the government will have less need to seize assets from the private sector; sixth, *trade* (import and export value as a percentage of GDP, logged), because elites from countries with more international transactions might also transact more domestically, while dictators in such countries might be disciplined by trade flows if they expropriate; seventh, *oil rents* (production value in current US dollars, logged), since oil wealth typically crowds-out the private economy, thereby reducing opportunities for elite transactions, while dictators often expropriate oil assets (Mahdavi, 2014); eighth, whether there is a *legislature* (dummy variable), given that elites often belong to and interact within legislatures, and legislatures are thought to constrain dictators (Wright, 2008); ninth, whether the country was a *British colony* (dummy variable), because the British colonial elite was particularly cohesive, while institutions endowed by the British are linked to stronger property rights (Acemoglu, Johnson and Robinson, 2001); tenth, *ethnolinguistic fractionalization* (0–1 index), since diversity might fractionalize elite networks, while also preventing elites from coordinating to discipline the dictator; eleventh, *regime duration* (years), because elites might become more integrated as time progresses, while regime survival might undermine the dictator’s urge to prey on the economy;⁴⁷ twelfth, *authoritarian regime-type* (GWF typology, 4 dummy variables), because the organizational structures associated with different regime types can affect elites’ interaction as well as the constraints placed on the dictator (Geddes, Wright and Frantz, 2018).⁴⁸

⁴⁷Following a similar rationale, instead of the regime’s duration, in alternative specifications I control for the leader’s tenure or the leader’s age.

⁴⁸GDP, growth, and trade data are from the World Bank’s World Development Indicators (World Bank, 2016). Inflation and government net lending data are from the IMF’s World Economic Outlook (International Monetary Fund, 2014). Oil data are from Ross (2013). Legislature data are from the World Bank’s Database of Political Institutions (Beck et al., 2001). British colony data and regional indicators (Asia, Latin America, Middle East and North Africa, Sub-Saharan Africa, and ex-USSR) are from Hadenius and Teorell (2007). Ethnolinguistic fractionalization data are from Fearon and Laitin (2003). Regime duration data and regime

Furthermore, I include 5 regional indicators, to control for time-invariant unobservables that operate at the region-level and might affect both network density and expropriation risk (e.g. culture, geography). Finally, I include a lagged value of the dependent variable (LDV), due to the strong autocorrelation revealed by statistical tests.⁴⁹

Sample Though the number of observations in my regressions varies depending on which controls I include, my sample generally consists of 293–476 observations from 47–51 countries and 11–12 years of data (2002–2013).

1.5.2 Baseline Results

Density Table 1.1 displays the results from several variations of my baseline regression (Equation 1.1). In line with my hypothesis, across all specifications there is a statistically significant negative association between network density and expropriation risk. Density’s significance ranges from above the 1.5% level to below the 0.01% level, and its coefficient ranges from roughly -0.035 (Model 2) to -0.160 (Model 1). I describe substantive effects for density using Model 1’s coefficient because, though Model 1 is unsophisticated and lacks many controls, the coefficient it produces is actually closer to that of more robust models I present later (Table 1.2). Recalling that the dependent variable is standardized, density’s substantive effect in Model 1 is moderate: a doubling of network density (roughly an increase of 1 log unit) is associated with a decrease of 0.16 standard deviations in expropriation risk the next year.⁵⁰ Alternatively, a 1 standard deviation increase in density (1.61 log units) is associated with a 0.26 standard deviation decrease in expropriation risk. A difference

type indicators are from [Geddes, Wright and Frantz \(2014\)](#). All preceding variables were obtained – though some were transformed – through the QoG dataset ([Teorell et al., 2013](#)).

⁴⁹Durbin-Watson, Breusch-Godfrey, and Wooldridge tests on the residuals of regressions without LDVs reveal substantial serial correlation, thereby calling for the inclusion of an LDV.

⁵⁰All right-hand-side variables are lagged by one year, but I refrain from giving a temporal interpretation to the reported effects in order to conserve space. A doubling of network density is not a particularly large increase, given that mean density in the sample used in the regressions is 0.06. Such an increase can be achieved by holding the number of nodes (elites) fixed and doubling the number of ties between them, or by holding the number of ties fixed and decreasing the number of nodes by roughly 28%. The exact decrease in nodes needed to double density depends on the initial number of nodes and initial density, due to the non-linear relationship between nodes and density.

of 0.26 standard deviations in expropriation risk is substantial, as it is similar to that between Mozambique and the United Arab Emirates. Yet another way to describe density's substantive effect is the following: an increase in density from its observed minimum to its observed maximum (-8.67 to 0.50 in log units) is associated with an decrease of 1.47 standard deviations in expropriation risk, similar to the difference between Azerbaijan and Singapore.⁵¹

Nodes The only controls that are statistically significant are nodes, oil rents, and two regime type indicators.⁵² Interestingly, the number of elites is negatively associated with expropriation risk. This implies that, all else equal, dictatorships with a larger number of elites expropriate less. However, I hesitate to give weight to this finding, as it is not particularly robust: the coefficient on nodes is insignificant in Model 2, only significant at the 10% level in Models 3 and 4, and insignificant in most robustness checks I discuss in Section 1.5.3. Moreover, including an interaction term between density and nodes renders the coefficients on both nodes and the interaction positive and/or insignificant, while preserving the sign and significance of density's coefficient.⁵³ At the risk of over-interpreting the results from the regression with the interaction term, they suggest that, all else equal, a larger elite exacerbates expropriation (positive coefficient on nodes), especially if densely connected (positive coefficient on interaction). Alternatively, we can interpret the negative coefficient on density coupled with the positive interaction term as suggesting that, only when they are small do dense elite networks constrains the dictator. In any case, the inconsistent sign and

⁵¹I also rerun my regressions after adding density's squared term, in order to capture any curvilinear effect density might have on expropriation risk. Density remains negative and statistically significant and its coefficient strengthens, while the squared term is also negative and not always significant. This suggests density does not have a U-shaped effect on expropriation. Thus, I present the regressions without the squared term, which are also easier to interpret.

⁵²Model 4 corroborates the deleterious effect of oil wealth on property rights documented in the literature. The dummy coefficient for Latin America (not reported) is also statistically significant, and its positive sign reflects the region's poor record of property protection. Unsurprisingly, the coefficient on the LDV is also positive and significant, as in every regression I will present.

⁵³I do not present the results of these regressions because an interaction with two logged variables (density and nodes) is particularly hard to interpret. I also rerun my regressions after adding the squared term of nodes, in order to capture any curvilinear effect nodes might have on expropriation risk. The subsequent regressions produce insignificant coefficients for both nodes and nodes squared.

Table 1.1: Expropriation Risk – Baseline Models

	(1)	(2)	(3)	(4)
Log Network density $_{t-1}$	−0.159** (0.063)	−0.035** (0.015)	−0.037*** (0.012)	−0.046*** (0.012)
Log Network size $_{t-1}$	−0.244*** (0.053)	−0.022 (0.016)	−0.023* (0.013)	−0.025* (0.014)
Log GDP/capita $_{t-1}$		−0.009 (0.015)	−0.009 (0.016)	−0.016 (0.014)
GDP/capita growth $_{t-1}$		−0.004 (0.003)	−0.004 (0.003)	−0.005 (0.003)
Inflation $_{t-1}$		0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Gov. net lending $_{t-1}$		−0.002 (0.002)	−0.002 (0.002)	−0.001 (0.002)
Log Trade/GDP $_{t-1}$		−0.024 (0.031)	−0.021 (0.030)	−0.002 (0.029)
Log Oil production $_{t-1}$		0.002 (0.001)	0.002 (0.001)	0.003** (0.001)
Legislature $_{t-1}$			0.036 (0.041)	0.012 (0.041)
British colony			−0.012 (0.021)	−0.029 (0.019)
Ethnolinguistic frac.			−0.054 (0.053)	−0.055 (0.055)
Regime duration $_t$			0.000 (0.000)	0.000 (0.000)
Military $_{t-1}$				0.088** (0.036)
Monarchy $_{t-1}$				−0.021 (0.040)
Personalist $_{t-1}$				−0.098*** (0.025)
Expropriation risk $_{t-1}$		0.919*** (0.023)	0.920*** (0.024)	0.919*** (0.022)
Constant	0.368*** (0.110)	0.106 (0.153)	0.072 (0.153)	0.112 (0.143)
Region Dummies	N	Y	Y	Y
N	476	355	355	355
n	51	47	47	47
T	12	11	11	11
Adj-R ²	0.07	0.91	0.91	0.91

Notes: *p < .1; **p < .05; ***p < .01. All models are OLS with country-clustered standard errors. Dependent variable is standardized. Variables without time indexes are time-invariant. Regime-type reference category is Party. Coefficients on region dummies suppressed to save space. See text for data sources.

statistical significance of the coefficient on nodes renders further discussion fruitless.

Regime Type A more puzzling finding than the (inconsistent) negative effect of nodes on expropriation risk is that personalist regimes have the lowest expropriation risk among GWF regime types. This contradicts the literature, which argues that personalist regimes score the highest on expropriation risk (Wright, 2008). Curiously, in one of their models Wilson and Wright (2017) also find that personalist regimes are associated with lower expropriation risk. However, when they probe that result further, the authors find that it is not robust. Similarly, if I interact the personalist regime-type dummy with density, the coefficient on personalism becomes insignificant, while the interaction term becomes positive, though not always significant. Again, at the risk of over-interpreting a non-robust finding, it suggests that, all else equal, dense elite networks constrain dictators' predation, but less so against personalists. The weaker effect of elite network density in personalist regimes might be owed to personalist dictators' concentration of power; already facing few constraints, powerful dictators might be able to prey on subjugated elites more easily when these elites are inter-connected (Svolik, 2009). Where in other regime types financial ties between elites foster information-sharing and monitoring, against a personalist dictator the same ties might exacerbate the dictator's predation through elites' financial contagion. Another way to interpret the estimates from the regression with the density–personalism interaction is the following: the adverse effect of personalism on property rights only holds in dictatorships where the elite is more densely connected. Regardless of the interpretation, the frailty of this finding limits the returns from discussing it further. Moreover, other coefficients on regime types indicators show the expected signs. Monarchies face statistically similar expropriation risk as single-party regimes (reference category). Finally, military regimes pose the largest risk for expropriation. This could be attributed to the military's generally myopic policy outlook and attempts to seize control of assets deemed strategically important.

Substantive Effects Comparison To better grasp the substantive effect of density on expropriation, it is useful to benchmark it against the effect of regime type. Figure 1.5 shows the mean predicted value of expropriation risk for all levels of density across different regime types. The maximum effect of regime-type – the mean predicted difference in (standardized)

expropriation risk between personalist and military regimes – is small; less than 0.19 standard deviations. On the other hand, the maximum effect of density – resulting from a change from a network consisting of isolated elites to one where all elites are connected to each other (0 to 1) – is almost double (0.38 standard deviations).⁵⁴ In short, elite network density seems to have a statistically significant and sizable effect on expropriation, much more so than other variables featured in the literature.

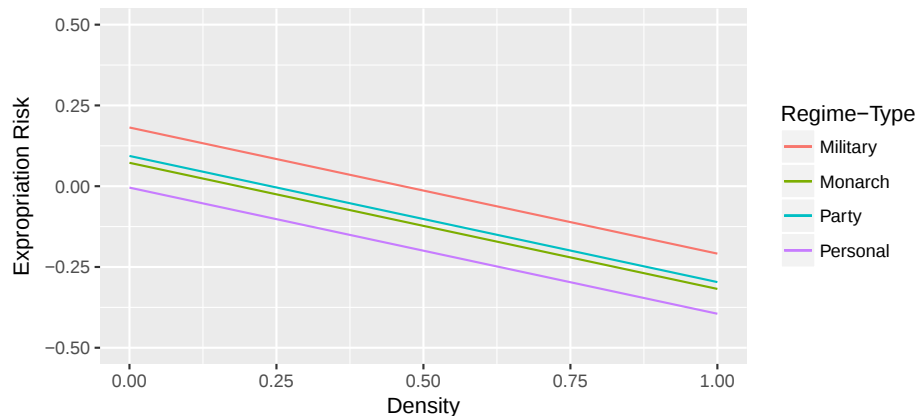


Figure 1.5: Marginal Effect of Elite Network Density on Expropriation Risk by Regime-Type.

Notes: Regression lines based on coefficients from Table 1.1, Model 4. Expropriation risk standardized. Density logged in regression model but graphed as unlogged for interpretability. See text for data sources.

1.5.3 Robustness Checks

Though the models presented in Table 1.1 include a battery of controls, regional indicators, and an LDV, there might still exist unobserved confounders that bias the estimated relationship between network density and expropriation risk.

First-Differences & Fixed Effects The first threat to inference I try to guard against are country-specific, time-invariant unobservables, such as culture or geography. Indeed, Wooldridge, Lagrange Multiplier, and F- tests all reject the null hypothesis of no country effects. Hence, in Table 1.2, Models 1–2, I estimate first-difference (FD) and country-fixed effects (FE) regressions. Both models confirm the statistically significant negative association

⁵⁴Note that this effect is almost a lower bound estimate, since the associated coefficient (Table 1.1, Model 4) is significantly smaller than that of other models.

between density and expropriation risk found in my baseline regressions. In addition, the coefficient on density is now stronger than the coefficients reported in Table 1.1. Comparing FD and FE models, the only differences are that the FD model finds a stronger effect for density than the FE model and a significant negative effect for nodes (null effect in FE model).

Two-Ways Effects Another unobservable confounder to guard against are year-specific shocks that affect countries homogeneously (e.g. global recessions). Indeed, LM- and F- tests both reject the null hypothesis of no two-ways effects (country and year). Thus, Model 3 estimates a two-ways FE model. Again, the coefficient on density is negative, significant, and stronger than the baseline regression coefficients, while the coefficient on nodes loses significance.⁵⁵

Arellano–Bond/GMM Including an LDV in either FD or FE model would make my estimates inconsistent, thus Models 1–3 omit the LDV.⁵⁶ However, as noted above, statistical tests suggest that an LDV might be necessary to remove the strong autocorrelation in expropriation risk. Hence, I employ an Arellano–Bond (A–B) estimator, which uses Generalized Methods of Moments (GMM) estimation within a country-FE model to instrument the lags of the dependent variable with higher-order lags.⁵⁷ Crucially, as evident in Model 4, the coefficient on density is still negative and highly significant, while its (absolute) size

⁵⁵One curious finding in Model 3 is the negative coefficient on inflation, suggesting that dictatorships where prices increase faster experience *less* expropriation. However, I hesitate to interpret this result further, due to the coefficient’s minute size and significance at only the 10% level.

⁵⁶In particular, an LDV makes the FD estimator inconsistent because the first-differenced LDV predictor $\Delta y_{i,t-1}$ is correlated with the first-differenced error $\Delta \epsilon_{i,t}$. Similarly, an LDV makes FE estimates inconsistent because the demeaned LDV predictor $y_{i,t-1} - \bar{y}_i$ is correlated with the demeaned error $\epsilon_{i,t} - \bar{\epsilon}_i$. The latter bias (Nickell bias) is a concern in “short” panels (low T/n), like my own. Note that, even though a *maximum* of 12 years of data are used in my regressions, the average T is around 9, due to the unbalanced structure of my panel and countries transitioning in/out of dictatorship at different durations.

⁵⁷I instrument the first lag of the dependent variable with its second lag. The results are substantively unchanged if I use both second and third lag as instruments, while the model passes the Sargan test of valid instruments. Using even more instruments is possible but results in (further) reductions of degrees of freedom. I also fit an A–B/GMM model with two-ways effects (Model 4 plus year effects), and the results are substantively unchanged. The only differences I observe versus Model 4 is that the coefficient on density is 0.25, the coefficient on government lending becomes insignificant, while the coefficient on oil production becomes significant.

is almost double that of other models.⁵⁸

Model Choice Choosing which model to rely on for inference in the absence of a tight mapping between theory and empirics or a well-defined body of empirical work is, to some extent, a matter of preferences.⁵⁹ On the one hand, the FD and A–B estimators are better suited to removing the strong autocorrelation in expropriation risk by removing medium- and long-term trends in the data, but only identify coefficients using variation from short-term trends.⁶⁰ On the other hand, the FE models allow for trends around the mean to inform their estimates, as they merely de-mean the data, but might provide inconsistent estimates in the face of strong autocorrelation. Indeed, autocorrelation is a serious concern here, and the FE models’ residuals do not pass the Durbin-Watson test of first-order autocorrelation, whereas the FD model’s (differenced) residuals do. This suggests that first-differencing adequately removes autocorrelation. However, as shown in the last line of Table 1.2, even the FD model fails the Breusch–Godfrey test of no serial correlation, while the A–B model is the only one that passes it. Moreover, Model 4 exhibits stronger fit than Models 1–3 in terms of R^2 , though I should add the usual caveats about assessing fit through R^2 .⁶¹ Overall, the tests and diagnostics favor the A–B model. This verdict is in line with practitioners’ advice that the A–B estimator is best applied to data with a large number of observations, few time periods, heteroskedasticity, serial correlation, a dynamic dependent variable, and dynamic and potentially endogenous predictors—all features of my data (Roodman et al., 2009).

Substantive Effects Revisited Table 1.2 shows that, apart from Model 1, the baseline regressions significantly underestimate the coefficient on density. The coefficient from the

⁵⁸Model 4 also reports significant negative coefficients for nodes and net government lending (the latter only significant at the 10% level).

⁵⁹Note that in all Table 1.2 models I am forced to drop institutional and demographic controls (legislature, British colony, regime duration, ethnolinguistic fractionalization), as well as regime-type and regional dummies, as they are either time-invariant or exhibit so little within-country variation over time (e.g. legislature, regime type) that their coefficients are non-identified.

⁶⁰I estimate the FD model both with and without a linear time trend (intercept). The results are substantively identical, and the coefficient on density is statistically identical.

⁶¹ R^2 and other measures of goodness of fit are not usually reported for A–B/GMM models, so I manually calculate R^2 .

Table 1.2: Expropriation Risk – Additional Models

	FD	FE	2-way FE	A-B/GMM
	(1)	(2)	(3)	(4)
Log Network density _{t-1}	-0.192*** (0.065)	-0.168** (0.076)	-0.175** (0.072)	-0.329*** (0.109)
Log Network size _{t-1}	-0.250** (0.100)	-0.004 (0.108)	0.077 (0.128)	-0.372** (0.157)
Log GDP/capita _{t-1}	-0.006 (0.125)	-0.083 (0.145)	-0.129 (0.164)	-0.046 (0.185)
GDP/capita growth _{t-1}	-0.002 (0.005)	-0.001 (0.006)	0.000 (0.007)	0.004 (0.008)
Inflation _{t-1}	0.000 (0.000)	0.000 (0.000)	-0.001* (0.000)	0.000 (0.001)
Gov. net lending _{t-1}	-0.003 (0.004)	-0.010 (0.006)	-0.007 (0.006)	-0.007* (0.004)
Log Trade/GDP _{t-1}	0.002 (0.122)	-0.088 (0.160)	-0.028 (0.188)	0.144 (0.146)
Log Oil production _{t-1}	0.060 (0.095)	-0.090 (0.100)	-0.102 (0.114)	0.186 (0.149)
Expropriation risk _{t-1}				0.591*** (0.222)
N	339	387	387	293
n	48	48	48	48
T	12	12	12	12
R ²	0.03	0.09	0.07	0.17
Robust Wald-test p-val	0.02	0.00	0.01	0.00
Autocorr. test p-val	0.03	0.00	0.00	0.88

Notes: *p < .1; **p < .05; ***p < .01. All models linear and include country-clustered standard errors. Dependent variable is standardized. Model 1 estimated in first differences. Model 2 includes country fixed effects. Model 3 includes country and year fixed effects. Model 4 includes country fixed effects and the first lag of the dependent variable, instrumented by its second lag. The model is estimated through Generalized Method of Moments (Arellano–Bond estimator). See text for data sources.

most appropriate model (A–B) suggests that, all else equal, a doubling of network density in country i in year $t - 1$ (roughly an increase of 1 log unit) is associated with a decrease of 0.33 standard deviations in expropriation risk in year t . Alternatively, a 1 standard deviation increase in log density (1.61 log units) predicts a 0.53 standard deviation increase in expropriation risk, equivalent to the difference between China and the United Arab Emirates. It should be noted that “all else equal” in Model 4 involves controlling for network size and several economic variables, unobservable country-specific factors (country FE), and only leaving variation in the dependent variable between t and $t - 1$ (due to the LDV). In other words, density has to clear a high bar in Model 4 to have such a significant and sizable effect on expropriation risk. Overall, after using a series of more sophisticated and robust models than those in Table 1.1, I find that network density retains its statistical significance, while displaying an even stronger negative effect on expropriation risk.

Additional Models To further probe the robustness of my findings, I estimate a series of additional specifications. Since expropriation risk ratings are measured on an 7-point scale, one could argue that an ordered choice model best describes the data-generating process behind my dependent variable and might also make my estimates more precise. However, the consensus in the literature is that the advantages of ordered choice over linear models are minimal, especially when the dependent variable has several values and does not deviate much from the normal distribution (Riedl and Geishecker, 2014).⁶² Nevertheless, I rerun my baseline regressions using an ordered logit and find substantively similar results. Another robustness check I conduct is fitting my baseline regressions to cross-sectional data, which I create by calculating country-level means for all variables (dependent and predictor).⁶³ Again, my results remain substantively unchanged. Furthermore, I repeat my analysis using unlagged values of all predictors, and find that my main results hold. Similar findings emerge from lagging density by 1 year and other predictors by 2 years—a specification that might increase the odds that the former is exogenous to the latter. Furthermore, my results are

⁶²Additionally, ordered choice models are inconsistent when combined with fixed effects, unstable when applied to panels with complex dynamics, and produce output that is significantly harder to interpret.

⁶³I do not fit the models of Table 1.2 to the cross-sectional data, since those models require data with a temporal dimension.

robust to using alternative types of robust standard errors, like Beck-Katz, Huber-White, and Newey-West, as well as regular standard errors. I also substitute fixed with random effects—though the latter problematically assume that all predictors are uncorrelated with the error term—and obtain very similar results (compared to Table 1.2, Models 3–4).⁶⁴ Finally, I substitute the first with the second lag of the dependent variable in my baseline regressions and also rerun those regressions with both lags, which removes further autocorrelation but reduces my sample size. Overall, the findings from both Tables 1.1–1.2 are not sensitive to variations of the estimated models.

Alternative DVs: Expropriation Incidents As argued in Section 1.5.1, my dependent variable exhibits several desirable traits in testing my argument. That said, it is interesting to explore whether my findings hold when using three sets of alternative dependent variables. First, I explore two datasets of *actual incidents of expropriation*. [Albertus \(2015\)](#) codes all international instances of land expropriation in the period 1900–2009, while [Hajzler \(2012\)](#) records all international expropriation events against agriculture, mining, petroleum, and utility companies from 1960 to 2006.⁶⁵ Unfortunately, the rare frequency of expropriation (roughly 3% of cases for the country-years in my sample), coupled with missing data, leaves too little variation in the dependent variable to allow for robust estimation using these datasets. Another weakness of these datasets is that, though the incidents of expropriation that they record are clearly cases of the dictator’s predation on elites, they ignore “softer” predatory behavior like targeted regulation, extortionary taxation, and transfer restrictions. Putting these limitations aside, I adapt my baseline specifications to logit models and find a significant negative effect for elite network density on the odds of expropriation, as measured by both datasets—though the effect is stronger for land expropriation.

Alternative DVs: Property Rights Indexes The second set of alternative dependent variables I use is *indexes of property rights*, whose limitations were touched upon in

⁶⁴A Hausman test rejects the null hypothesis that fixed and random effects estimates are similar. I cannot estimate two-ways random effects (country and year), due to the unbalanced nature of my panel.

⁶⁵[Albertus and Menaldo \(2012\)](#) compile a more detailed record of expropriation – of not just land – in Latin America. Unfortunately, their dataset overlaps with my network data only in the period 1990–2002, during which there were too few years of dictatorship in Latin America to allow for a meaningful analysis.

Section 1.5.1. I briefly describe some of these indexes.⁶⁶ The PRS Group International Country Risk Guide produces a 12-point “investment profile” rating, which is an aggregate of three components: contract viability/expropriation, profits repatriation, and payment delays (Howell, 2012). The Heritage Foundation compiles the Index of Economic Freedom, which includes a property rights component scoring from 0 to 100 “the degree to which a country’s laws protect private property rights and the degree to which its government enforces those laws” (Miller, Kim and Holmes, 2015, p. 353). The World Economic Forum releases the Global Competitiveness Report, which subsumes an 8-point rating of property rights protection (Schwab and Sala-i Martin, 2016). The World Bank issues the World Development Indicators, which provide a 7-point variable reflecting property and contract rights legislation and enforcement (World Bank, 2016). Finally, the Fraser Institute produces the Economic Freedom of the World dataset, which includes a 0 to 10 rating of “legal structure and security of property rights” (Gwartney et al., 2016).⁶⁷ Though the above indexes differ in their methodology, granularity, and geographic and temporal coverage, they all yield roughly the same conclusion in my regressions: dictatorships with more dense elite financial networks are associated with stronger property rights.

Alternative DVs: Property Rights Proxies The last group of alternative dependent variables that I analyze are three *proxies for property rights*. First, I follow Weymouth (2011) in using financial dollarization, the proportion of a country’s bank deposits denominated in foreign currency, as a proxy for “indirect expropriation” and property rights (Yeyati, 2006). One concern with this proxy is that dollarization mostly reflects the average household’s hedge against inflation and currency devaluation, not elites’ response to predation by the dictator. The second proxy I use is net foreign direct investment (FDI) outflows as a percentage of GDP (World Bank, 2016). Though FDI certainly captures the decisions of holders

⁶⁶All indexes are annual and cross-national, with the exception of the monthly PRS Group ratings (I convert them to annual averages). Most of these variables are obtained through the QoG dataset (Teorell et al., 2013).

⁶⁷There are also several datasets with rule of law indexes, such as Freedom House’s Freedom in the World report and the World Bank’s Worldwide Governance Indicators (Freedom House, 2016; World Bank, 2014). I omit these indexes from my analysis because rule of law relates to the enforcement of a much broader set of rights than property rights.

of large amounts of capital (i.e. elites), using it as a property rights proxy raises a different concern: FDI decisions depend on many more factors than economic predation, some of which I cannot control for (e.g. relative return on foreign vs domestic assets). Thus, FDI is arguably a weak proxy for property rights. The final proxy I employ is an estimate of illicit financial flows, which measures the value of outflows “illegally earned, transferred, and/or utilized” (Solomon and Spanjers, 2017, p. 1). A concern with this variable is that illicit outflows might be funds the dictator *rewards* elites with, just as much as they might be funds elites try to hide from the dictator. Additionally, illicit outflows are often proceeds of crime, and crime’s relationship to economic predation in dictatorships is unclear. Perhaps due to the above limitations, I find that the effect of elite network density is insignificant and/or has an inconsistent sign in regressions of all three property rights proxies.

Alternative Independent Variables Another robustness check I conduct involves using different network statistics as independent variables. Though I formulate my theory in terms of network *density*, I cannot rule out other measures of network structure affecting elite’s coordination in constraining the dictator. One network statistic that might capture the dynamics described in my theory is the fraction of the network that is non-isolated (non-isolate fraction); that is, has at least one tie. This variable reflects a weak form of density, in that it does not discriminate between elites with one financial tie and elites with dozens of ties. This “loss of information” in the variable might be the reason why, despite displaying a negative and significant effect on expropriation risk in all of my models, non-isolate fraction’s coefficient is roughly half that of density’s. A different metric often used in network analysis is transitivity, otherwise known as clustering. A network’s (global) clustering coefficient captures the ratio of observed triads (e.g. i - j , j - k , k - i) to potential triads (Wasserman and Faust, 1994). Since we might expect that transitivity in financial ties fosters elite coordination, I rerun my regressions with networks’ clustering coefficient as the independent variable.⁶⁸ I

⁶⁸To illustrate the potential role of network transitivity in elite coordination, imagine a scenario where elites i , j , and k gain by punishing D , but only succeed if all three coordinate. Assume an elite only knows the participation decision of elites she is tied to, and that i is tied to j and k but j and k are not tied to each other. Even though i knows that j and k want to punish D , and the vice-versa also holds, j (k) does not know that k (j) wants to punish D ; thus the participation threshold of 3 is not met and coordination fails. In short, transitive financial ties among elites might foster coordination in constraining the dictator, and

find substantively similar results as in the tables presented above, with the exception of the fixed-effects regressions (null effect). Finally, I substitute density with another prominent network statistic, degree centralization. Loosely speaking, a network’s degree centralization measures the influence – in terms of network ties – of the most centrally located node(s) (Freeman, 1979). One might imagine that, if one elite has significantly more ties than any other elite, she might be able to control the flow of resources and information through the network and act as a focal point for coordination to constrain the dictator. As such, dictatorships with elite networks with high degree centralization might experience stronger property rights. Indeed, I find that degree centralization has a significant negative effect on expropriation risk – and stronger than density – in all of my regressions other than the A-B model (null effect).⁶⁹

Additional Controls: Economic A further robustness check I conduct entails using additional control variables. Due to the myriad of determinants of expropriation risk (DV) that might also affect elite network density (IV), coupled with the absence of a literature on my IV, I try to control for more potential confounders than is customary. In particular, I experiment with additional controls from three categories: economic, political, and demographic. Among economic variables, I begin by controlling for exchange rate restrictions with data from Ilzetzki, Reinhart and Rogoff (2017); it is possible that in dictatorships with such restrictions elites interact more to overcome them, while economic mismanagement stemming from the restrictions might lead the dictator to expropriate elites in order to gain control of key companies (e.g. Venezuela recently). A related variable I account for is capital controls, using data from Fernández et al. (2016). As with exchange rate restrictions, elites might interact more to overcome capital controls, while the economic burden controls cause might incentivize dictators to scapegoat elites and seize their assets (e.g. Venezuela again). I also control for government expenditure as a share of GDP (with data from International Monetary Fund (2014)), since a large government sector might crowd out the private sector

dictatorships with elite networks that have higher clustering coefficients might experience less predation.

⁶⁹Density is correlated 42% with the non-isolate fraction, -1% with transitivity, and 71% with degree centrality in my sample.

and elite interactions within it, while large governments are generally more likely to prey on private firms (owned by elites). The next variable I control for is income inequality (using data from [World Bank \(2016\)](#)), because concentration of income might lead to the concentration of financial interactions between few elites, while also tempting the dictator to redistribute elites' income through expropriation. Overall, my results are substantively unchanged, though using these controls substantially reduces my sample.⁷⁰

Additional Controls: Political The second group of additional controls I explore are political/institutional. First, I include an indicator of the presence of multiple parties from [Cheibub, Gandhi and Vreeland \(2010\)](#). Similar to the effect of legislatures, it is possible that the ties observed in my data are the result of elite interactions within parties, which are argued to constrain dictators ([Gandhi, 2008](#)). Furthermore, I control for executive constraints, as measured by [Marshall, Jaggers and Gurr \(2016\)](#), because an unconstrained executive might both limit elites' interactions and prey on them. I also control for corruption, using the Bayesian index of [Standaert \(2015\)](#); this is because corruption often involves elite cronyism, which should foster elite interactions, but also the misuse of public office for private gain, which result in economic predation. Finally, I control for the origins of countries' commercial law, due to their potentially joint effect on elites' financial transactions and property rights ([La Porta et al., 1999](#)). None of these four controls nullifies the effect of density on expropriation, though some weaken its coefficient and/or limit my sample.

Additional Controls: Demographic The last group of controls I account for are three demographic variables.⁷¹ The first is (log) population, because it might be harder for elites to form ties within a large population, while populous countries imply larger, more valuable firms for dictators to expropriate. The second demographic control I include is (log) population density, since spatial clustering might imply network clustering for elites and also larger elite influence on the dictator due to proximity. For the same reason, I control for urbanization (% of total population). Overall, density's effect on expropriation risk remains

⁷⁰One change in my results is that density is insignificant in the fixed effects regressions that control for exchange rate restrictions.

⁷¹All demographic variables are from [World Bank \(2016\)](#).

negative and statistically significant in all models, and in some models (urban) population has a significant (negative) positive effect.

Alternative Network Codings The final set of robustness checks I carry out involve changing the way I construct my elite networks from the Panama Papers. I experiment with four variations of my coding rules; two affect the ties coded, another two the nodes coded. First, I drop ties formed through offshore companies that are dissolved. In other words, I assume that elites i and j are no longer tied in year t if company k , through which the tie was formed in the data, dissolves. This coding rule assumes – perhaps naively – that dissolving an offshore company has substantive bearing on the connections of the respective elites; that the dissolution is not due to accounting, legal, or financial reasons. The second coding rule I change relates to offshore companies with the same name. Since some company names appear many times in the data (e.g. New Ventures Ltd.), I merge companies with the same name when they are incorporated in the same country (e.g. Bahamas).⁷² The third change I make to my coding scheme is analogous to the second one, but applied to officers (i.e. the individuals associated with each company); in particular, I merge officers with the same name and nationality.⁷³ Note that merging two officers reduces the number of nodes in the network by 1 and aggregates their ties, thereby increasing network density. Finally, I use different techniques to identify and drop “nominee” directors from my networks—officers contracted by the real owners of offshore companies to act as frontpeople. Namely, I purge my networks from officers that are associated with a large number of companies (I vary that threshold) or list law firms as their address. None of the above alterations to my networks substantively changes my regression results, while some merges strengthen the coefficient on density.⁷⁴

⁷²Though every company has a unique identifying number in the data, it is not clear that every number identifies a unique company. I also experiment with “fuzzy” matches of company names (e.g. merging New Ventures Ltd. with New Ventures).

⁷³As with company merges, I experiment with “fuzzy” matches of officer names. I also use information on officers’ listed addresses, though noisy and incomplete, to match officers with the same name, nationality, and address.

⁷⁴Officer merges by name have a larger effect on the coding of networks from East Asian dictatorships (e.g. China, Singapore) because these countries have lower-than-average variation in names. Crucially, these countries also have above-average property rights protection, thus merging officers reinforces the negative

1.6 Discussion

In the previous section I documented a robust statistical association between my measure of elite financial network density and various measures of expropriation and property rights in dictatorships. However, the interpretation I have given to my findings might not be unique. In this section I explore five alternative interpretations of my findings, as well as four limitations of my empirical approach.

Elite Collective Action Several prominent theories of property rights formation and power-sharing in dictatorships focus on elites' collective action problem in constraining the sovereign / dictator (North and Weingast, 1989; Boix and Svolik, 2013). As such, it is important to evaluate my findings vis-à-vis collective action theory. According to the latter, all else equal, collective action is more likely to be achieved the smaller the group in question, the greater its ability to provide selective rewards and punishments to its members, and the more unequal the distribution of expected gains within the group.⁷⁵ Interestingly, my findings are not consistent with any of these predictions. First, as noted in Section 1.5.2, the number of nodes (elites) in the network has a negative effect on expropriation risk—though the effect is inconsistent in sign and statistical significance, especially when I interact nodes with density in my regressions. This result contradicts a collective action interpretation, whereby a larger network of elites should be less likely to constrain the dictator and should thus experience more predation (positive coefficient). The second piece of evidence against a collective action view is that the legislature dummy in my regressions has a positive (insignificant) effect on expropriation. Given that legislatures institutionalize the selective allocation of rewards among elites, we should expect them to facilitate collective action, thereby producing a negative effect on expropriation, instead.⁷⁶ Finally, assuming income inequality is a good proxy for elites' unequal gains from curtailing expropriation – with the wealthiest

relationship between network density and expropriation risk in my regressions.

⁷⁵Ideally, one agent should value the public good to be produced under collective action enough to find it individually rational to provide it herself.

⁷⁶Note that my finding is consistent with Jensen, Malesky and Weymouth (2014) and Wilson and Wright (2017), who find that legislatures have a null effect on expropriation when pooling different regime types.

elites gaining more from constraining the dictator – we should expect the Gini coefficient to have a negative effect on expropriation.⁷⁷ Once again, my regressions are inconsistent with a collective action mechanism underpinning the effect of elite networks, as income inequality has a positive insignificant effect on expropriation.

Dictator’s Encompassing Interest My theory assumes that only financially connected elites share private information on the economy, in order to protect their shared economic interests from the dictator’s predation—elites that are not financially tied might be competitors with an incentive to misinform each other. Under this view, a dense financial network reflects overlapping economic interests for elites. However, it is instead possible that a dense financial network reflects the dictator’s encompassing interest in the economy (Olson, 1993). That is, financial ties among elites might be the result of the dictator distributing economic resources across elites in a manner that maximizes allocative efficiency. Under the Olsonian view of the dictator as a stationary bandit, when the dictator has a larger encompassing interest (stake) in the economy he chooses policies that maximize economic output and preys less. Though this theory might seem consistent with my analysis, it faces two obstacles. First, to argue that the dictator can foster ties between elites and that *those* ties increase his encompassing interest in the economy and reduce his predation, I would have to use data on ties that the dictator controls, not ties formed through offshore companies that the dictator is not even aware of. In other words, unless onshore financial ties are identical to offshore ones, we cannot assume that the dictator shapes the ties used in my analysis.⁷⁸ Second, if the dictator can shape (offshore) ties between elites, we should expect that personalist regimes, where the dictator yields more influence, to have denser elite networks. Interestingly, personalist regimes are robustly associated with lower network density. Coupled with my other finding that personalist regimes – surprisingly – experience lower expropriation risk, it is

⁷⁷Admittedly, this assumption is problematic when using the Gini coefficient as a measure of income inequality, because top income earners (elites) might have the same income and still create a high Gini—if their income is a large multiple of lower income earners.

⁷⁸Another, econometric piece of evidence against the endogeneity of my estimate of density’s effect is that, if density were endogenous, its coefficient should be similar across FD and FE estimates. In fact, the coefficient for (log) density has the same sign and sufficiently similar point estimates and standard errors in Table 1.2, Models 1–2.

hard to treat dense offshore elite networks as a result of the dictator’s encompassing interest in the economy. Similarly, I cannot attribute lower predation in countries with dense elite networks to the dictator’s encompassing interest.

Intra-Elite Incentive Alignment Another potential interpretation of my findings is that the negative effect of dense elite networks on expropriation operates mostly through the incentive alignment that financial ties create between elites. Though I naturally assume that connected elites have shared economic interests – which allow information-sharing – aligned preferences should mostly affect *horizontal* expropriation; that is, elites’ predation on each other. The more elites’ economic interests overlap, the larger the cost for one elite to undermine another elite’s interests, so the lower the horizontal expropriation. However, it is unclear why shared interests *per se* should limit *vertical* expropriation—by the chief executive (dictator)—measured by my dependent variable. If aligned elite incentives reduce vertical expropriation, other determinants of shared elite interests should also have a negative effect on expropriation risk. Two potential proxies of these determinants are ethnolinguistic fractionalization (ELF index) and how concentrated industries are (Herfindahl-Hirschman index)—ethnolinguistically non-diverse elites might have shared interests due to shared culture, while elites in an economy with concentrated industries will experience less market competition. Nevertheless, my regressions show a null effect for both proxy variables, giving no support to a mechanism where dense elite networks limit vertical expropriation only through aligning elites’ incentives.

Vertical vs Horizontal Expropriation At this stage, it is fruitful to take the distinction between vertical and horizontal expropriation more seriously. [Jensen, Malesky and Weymouth \(2014\)](#) argue that vertical expropriation is limited by stronger property rights, which govern the relationship between state (dictator) and private actors (elites), while horizontal expropriation is limited by contracting institutions, which govern the relationships among private actors (elites). How do dense elite networks affect each type of expropriation? My analysis shows a robust negative effect of dense elite ties on vertical expropriation. However, I find a negative effect for network density on indexes of investor protection and rule of law, like those used by [Jensen, Malesky and Weymouth \(2014\)](#) to measure horizontal expro-

priation. These findings suggest that it is possible to extend my theory to produce a dual effect of elite networks on expropriation. On the one hand, dense elite ties reduce vertical expropriation by diffusing private information that elites use to secure selective commitments from the dictator. On the other hand, financial ties reduce horizontal expropriation by aligning elites' interests and reducing their incentive to prey on each other. I leave such an extension of my theory to future work.

Regime Insiders vs Outsiders Another competing explanation of my findings rests on the observation that the dictator does not make decisions about which elites to prey on alone; members of his regime, inner circle, or ruling/winning coalition – i.e. insiders – also participate in economic decision-making. Crucially, insiders are relatively secure from predation, compared to elites excluded from decision-making, i.e. outsiders. As such, a less simplistic and more important distinction than dictator-vs-elites might be insiders-vs-outsiders. Moreover, an insiders-vs-outsiders approach might change the interpretation of my data, since some of the financial ties I observe might be between insiders and outsiders, not outsiders and other outsiders. In other words, it is possible that the financial networks I observe align the incentives of elites excluded from decision-making – potential prey – with elites included in decision-making – potential predators, and one could interpret the negative effect of elite network density on (vertical) expropriation against outsiders as the result of financial ties binding insiders with outsiders. If this mechanism underlies my findings, we might expect two implications to hold. First, density should have a weaker effect in constraining expropriation in personalist regimes. Given that insiders – other than the dictator – are fewer and less powerful in personalist regimes, we should expect insider-outsider ties to do little to limit outsiders' expropriation. Yet, to the extent that there is a consistent pattern, density actually has a stronger effect in the subsample of personalist regimes versus that of all other regime types. The second implication we might expect if insider-outsider ties underpin my findings is for density to matter more where there are fewer elites (after controlling for population). Since, all else equal, a smaller elite implies higher chances that denser ties bring outsiders closer to insiders, we might expect a negative effect from the interaction of density with nodes. However, as noted in Section 1.5.2, the nodes-

density interaction term has an inconsistent sign and significance across specifications. In short, there is no evidence that the negative effect of elite network density on expropriation operates through insider-outsider ties, as opposed to information diffusion among outsiders.

Reverse Causality Having explored several competing interpretations of my findings, I move on to consider four potential limitations of my empirical approach. First, I address the question of reverse causality. One might argue that lower expropriation risk cultivates trust between dictator and elites, thereby encouraging stronger financial ties between elites. There are two issues with this criticism. First, it is unclear why trust between dictator and elites would foster elite-to-elite ties, instead of dictator-to-elite ties. In other words, reverse causality would be a natural concern if my independent variable measured the density of ties in a dictator-centric network, not in a network with only elites. The second problem with the reverse causality claim is that, as argued in Section 1.4, the elite ties in my data are formed through offshore companies, which are unobservable to the dictator and hold assets that are harder to prey on. Thus, even if lower expropriation by the dictator cultivates trust between elites, that should be reflected in denser *onshore* ties, due to onshore assets' visibility and their vulnerability to predation. Since my independent variable does not measure onshore ties, though, it is not obvious how reverse causality can account for the negative association between dense offshore elite ties and expropriation.

Network Measurement Error The second criticism of my approach that I address is non-random measurement error in my independent variable. Given the incentives for substituting onshore with offshore financial ties outlined in Section 1.4, it follows that the worse the property rights are in a dictatorship the larger the share of elites' ties that will be offshore. And since my data only captures offshore ties, this implies that my networks are more representative of elite ties in dictatorships with higher expropriation risk. In other words, if the "latent" independent variable I am trying to capture is density of *aggregate* elite ties, it is possible that my "proxy" independent variable, density of *offshore* ties, is more accurately measured in countries with high values of the dependent variable, expropriation risk. Can this non-random noise account for my findings? Recall that my analysis associates higher expropriation risk with sparser elite networks. For this association to be the result

of non-random noise, it must be the case that more accurate measurements of the latent independent variable produce *sparser* networks; that is, getting a clearer picture of all elite ties must produce a network with fewer ties per node. However, if anything, the opposite is likely to be true: more accurately measured elite ties should produce *denser* networks. Thus, to the extent that removing measurement error would alter my analysis, it would probably be in the direction of a stronger negative effect of elite network density on expropriation.

Offshore Assets & Ability to Punish The third potential issue with my empirical approach also relates to the use of data on offshore ties. Since the elites in the Panama Papers have offshore company accounts, my data might be unsuitable for testing my theory’s main implication. In particular, because I argue that elites extract selective commitments from the dictator by threatening to reallocate their capital, it might seem that my theory rests on elites holding most of their assets onshore—otherwise their potential economic harm to the dictator would not be sufficient to deter him from predation. This criticism has two limitations. First, as noted in Section 1.4, having an offshore company account does not mean that one’s assets are located offshore; for example, one can own real estate in his country of residence through a company registered offshore.⁷⁹ Relatedly, even if an elite’s offshore company owns assets located offshore, this does not mean that said elite does not also own assets onshore. In short, most elites in the Panama Papers are likely to have some onshore assets that they can use to threaten the dictator through punitive capital reallocation. The second issue with this criticism is that it ignores the two-way nature of my theory: elites can also extract selective commitments by promising to *reward* the dictator through their *offshore* capital. Elites can deter the dictator from preying on them by committing themselves to repatriating their offshore capital and using it to boost the economy and/or aid the dictator: investments to increase employment, public-private partnerships to develop infrastructure, participation in distributive politics, or simply transfers to the dictator. Moreover, combining the two issues raised above, if I were to use data on elites that have only onshore assets, I would risk producing a weaker test of my theory. Since they do not have offshore assets, these elites

⁷⁹Indeed, this seems to be a common use for offshore companies, according to anecdotes from classified reports on some of the Panama Papers accounts.

would have to incur a cost to set up offshore companies to reallocate their onshore assets, and would thus have reduced capacity to punish the dictator. Furthermore, without having offshore capital, these elites would have limited capacity to reward the dictator through capital repatriation. Overall, it is clear that my theory does not rest on elites having solely onshore assets. On the contrary, my analysis becomes more powerful vis-à-vis my theory when using data on elites that have a mix of onshore and offshore assets, like the ones in the Panama Papers.

Predation: Elites vs Foreign Firms The final potential limitation of my analysis concerns my dependent variables; namely, whether the measures I use capture predation against domestic elites or foreign companies. My main dependent variable, the price of insurance against expropriation, factors in the risk of expropriation against foreign firms, not just domestic ones. Domestic elites might have no stake in these foreign firms, and so my dependent variable might not be measuring predation against the actors I am interested in. Furthermore, as noted in Section 1.5.3, the property rights indexes I use are based on the judgement of mostly foreign experts, while the incidents of expropriation variables also include incidents against foreign firms. Though none of my dependent variables perfectly captures predation against domestic elites, it is reassuring that my main finding holds across all of these variables. Each variable is likely to capture a different degree and type of predation against domestic elites. Interpreting the negative effect of network density on all of these variables *collectively*, it seems reasonable to take the beneficial effect of dense networks on domestic elites seriously.

1.7 Conclusion

Recap I have argued that an informal institution—financial networks—aids elites in constraining dictators, much like formal political institutions do. The strong, cooperative ties elites form within financial networks allow them to share private information on the true state of the economy and monitor the dictator’s compliance with their rent-sharing agreement. Where financial networks are denser, private information diffuses to more elites, which

present the dictator with a larger credible threat if he preys on their shared rents. As such, dense elite networks deter economic predation and constrain dictators, thus strengthening property rights. To test my argument, I uncovered the structure of one type of elite financial network, ties in offshore companies, in all dictatorships post 1990, using the largest leak of information on offshore finance, the Panama Papers. After constructing these networks, I derived measures of elite density, which I used as the independent variable in my regression analysis. Controlling for a host of confounders, country- and year-level effects, temporal dynamics, and statistical irregularities, as well as using alternative dependent variables, independent variables, and rules for constructing the underlying networks, I found that dictatorships with denser elite offshore financial networks are associated with significantly lower expropriation risk. I interpret these results as evidence that, though dense financial ties between elites are usually associated with corruption, nepotism, and patronage, in dictatorships, where the main threat to property rights are unchecked dictators, a densely financial connected elite is the lesser of two evils—it counters dictators’ predation.

Future Research Future research could revolve around the broader role of networks in dictatorships. Since most networks of political importance are endogenous to politics, one could develop an integrated theory on how networks’ role varies, depending on such factors as the dictator’s strength, institutional environment, and level of development. Furthermore, in addition to financial ties among economic elites, one could study the role of familial, ethnic, religious, or other ties. Similarly, one could extend the focus to networks among other key players in dictatorships, such as party elites in regimes with dominant parties, ruling families in monarchies, and officers in military regimes. In all of these cases, the main empirical challenge is capturing ties among the actors of interest—in a comparable way within and across countries and/or time. Thus, to test the role of networks in dictatorships researchers will have to combine innovative data collection and contextual knowledge of the cases under study with careful research design. I leave these tasks to other scholars.

CHAPTER 2

Predicting Foreign Fighter Flows to Syria Using Machine Learning: An Introduction to Kernel Regularized Hurdle Negative Binomial

2.1 Introduction

The ongoing Syrian Civil War has been named “the world’s largest humanitarian crisis since WWII” ([Commission, 2015](#), p. 1). 230,000 dead, 850,000 injured, 4 million refugees, and 7.5 million internally displaced; these are some conservative estimates of the conflict’s cost.¹ An integral component of the conflict are the multiple rebel groups involved, like Islamic State, Jabhat al-Nusra, and the Free Syrian Army. Although rebel manpower is hard to estimate, there is consensus that a significant portion of it comes from foreign fighters ([Byman and Shapiro, 2014](#)), “non-state actors involved in military activity in a foreign country” ([Hegghammer, 2013b](#), p. 1). Moreover, with an estimated 21,000 foreigners from 50 countries fighting in Syria ([Neumann, 2015](#)), foreign fighter supply has reached a historical high ([Hegghammer, 2013a](#)).

Naturally, policy-makers are asking themselves what draws their voters to participate in a foreign conflict. Common concerns were best expressed by Britain’s Prime Minister, David Cameron: “one of the most disturbing aspects is how this conflict is sucking in our own young people, from modern, prosperous societies”. So far, public discourse has addressed

¹All figures are the most recent estimates that could be found as of June 2015 ([UNOCHA, 2015](#)).

these concerns by focusing on individual fighters’ motivations. This has been complemented by an emerging case-study literature based on returning fighter interviews (Stenersen, 2011; Weggemans, Bakker and Grol, 2014; Nilsson, 2015). As a result, a popular conception of foreign fighters has emerged; that of young Muslims from poor urban environments, with “deep-seated feelings of marginalization and exclusion” (Noor and Dorsey, 2014, p. 2).

Whether the inferences of this literature are accurate and generalizable is an open question. Before settling it, though, policy-makers might want to know what the role of *country-level* features is in individual fighters’ calculus. In particular, we might ask to what extent can foreign fighter supply be attributed to *policy*, which is under the control of government, versus countries’ *structural features*, which are sticky. To date, there is no study addressing these questions. However, the ongoing nature of the conflict, its spillovers into the region, and the continuing flow of fighters throughout the globe call for a systematic analysis of the evidence. We take a first step in that direction, thereby responding to the recommendations of policy reports, that “strategies would benefit immensely from more evidence-based research” (GCCS, 2014, p. 17).

An analysis of the predictors of foreign fighter supply is hindered by two classes of problems: data availability and modeling challenges. In this study, we focus on overcoming the difficulties in modeling this process. Namely, our approach addresses four sets of issues, relating to our priors about the country-level mechanism generating foreign fighter supply, the nascent state of the literature, and our dual interest in prediction as well as inference and interpretability.

First, our theory-motivated prior that foreign fighter supply is a two-component process—some function of countries’ features should predict *whether* they supply any foreign fighters and, if so, another function should predict *how many* fighters they supply. This is motivated by qualitative evidence that supplying at least one fighter alters supply dynamics within a country; radicalization and recruitment networks form, along with policies to contain them, and a different mechanism takes over the “scaling-up” of supply. Such a theoretical structure suggests that a two-component mixture model is appropriate, with one component predicting assignment to supplier countries (binary response), and a second component predicting the

number of fighters supplied (count response left-truncated at 1). This directs us to the familiar hurdle model (Mullahy, 1986). In addition, the count nature of the response and its large variation urge us to further refine our specification to the hurdle negative binomial model.

As with any mixture model, we can allow different features, or different functions of the same features, to enter each component. However, this brings us to a second modeling challenge: there is little theory on how a country’s features combine to affect foreign fighter supply, let alone how this should vary between the model’s components. The assumption behind most regression models is that the systematic component of a unit’s response is a linear function of its features (King, 1989b). When models are specified using vague theoretical priors, though, parametric assumptions are hard to justify (Ho et al., 2007). Instead, model-fitting could greatly benefit from a flexible semi-parametric approach. Semi-parametric and non-parametric models are not uncommon, yet existing models do not allow us to impose the theoretically-motivated structure of the hurdle model.

Another downside of many non-/semi-parametric models is that they overfit the data, thereby yielding poor out-of-sample predictions – a third challenge to address. This can be addressed by machine learning algorithms that penalize complexity, like LASSO and Random Forest. Regularizations based methods like the LASSO or Ridge regression use k -fold cross-validation to tune penalty, or regularization parameters in order to strike a better balance on the bias-variance tradeoff than classical estimators. Unfortunately, though, semi-parametric regularized models give way to a fourth challenge: they do not produce quantities of interest familiar to social scientists, like marginal effects and confidence intervals. In fact, demanding such quantities of interest requires sacrificing flexible model-fitting—for example, output from ridge regression is interpreted like that of GLMs, but the model assumes that the features affect the response linearly.

In short, we seek an algorithm that combines attractive features from multiple approaches: the intuitive structure of the hurdle model, the agnosticism of non-parametric models on how the features combine to affect the response, the good out-of-sample performance of regularized algorithms, and the interpretability of standard regression models. We

bridge the generalized linear model and machine learning literatures to arrive at such an algorithm, the Kernel Regularized Hurdle Negative Binomial (KRHNB). Specifically, we derive our estimator by applying kernel expansion and regularization to a hurdle negative binomial, then develop companion software to compute parameter estimates that minimize prediction error, produce pointwise marginal effects, and get bootstrapped estimates of uncertainty.

Our procedure consists of several simple steps. We begin by assuming the data is generated by a two-component mixture model: a logit predicting whether a positive count is observed and a negative binomial (left-truncated at 1) predicting the conditional count. Then we form the sample log-likelihood function and add a regularization term (using an L_2 norm), thereby arriving at our target function—the penalized log-likelihood. To move away from a fully parametric form, we expand our feature matrix into a higher-dimensional space; in our application we use an infinite-dimensional space corresponding to all possible expansions of the data (e.g. polynomial, logarithmic, exponential, multiplicative). This makes our target function linear in the *mapping* of the features, instead of the features themselves. We then show that the features enter the minimum of the target function solely through inner products and thus can be substituted by positive semi-definite kernels, like the Gaussian kernel. This is known as Mercer’s Theorem and enables numerical optimization of the target function.

Predicted responses are derived by substituting our estimates into the hurdle negative binomial’s conditional expectation function (CEF). Pointwise marginal effects for feature j are computed using numerical derivation of the CEF with respect to feature j . Averaging these effects for each feature gives us its average pointwise marginal effect. Finally, redrawing 1,000 samples from our data with replacement and repeating the above computations gives us non-parametric bootstrapped estimates of uncertainty—a distribution of average pointwise marginal effects for each feature. All relevant computations are performed with our companion software, written in the R language.

KRHNB has two attractive properties over the standard hurdle negative binomial. First, regularization penalizes complexity, thereby striking a better balance on the bias-variance tradeoff. Thus, our procedure is less prone to overfitting the training data and better predicts

test data. Second, feature expansion implicitly allows our features to affect the systematic component of the response through any functional form. This is particularly valuable in this context; the absence of theories of country-level foreign fighter supply should deter us from making strong parametric assumptions. Nevertheless, one might ask: does our method improve over other machine learning models? After all, there are other algorithms that penalize complexity, some of which impose less structure on the data.

To answer this question, we compare our method to three popular machine learning algorithms and find that it performs better or comparably to all of them. Using a cross-section of 163 countries, and an array of 27 demographic, economic, geographic, and political features, we compute the leave-one-out cross-validation root mean square error and the mean absolute error. Higher mean absolute error rates are produced by all 3 of our benchmark models: Kernel Regularized Least Squares (KRLS) ([Hainmueller and Hazlett, 2014](#)), KRLS truncated at 0, and a Random Forest. Only Random Forest outperforms KRHNB with respect to root mean squared error and this is largely due to its relative success in predicting a few large outliers.

Having established the merits of our approach – theoretical motivation, flexibility, predictive power, and interpretability – we apply it to our data, in an effort to contribute to the literature on foreign fighters. Substantively, we find that structural features dominate in predicting foreign fighter supply: populous, developed countries, that are close to Syria, and have a high concentration of Sunnis supply more fighters. Some features that respond to government policy more easily also matter: internet usage, refugee intake, government regulation of religion, and government favoritism of a particular religion positively predict foreign fighter supply. Our method also lends itself to the discovery of interesting heterogeneity in the predictive effect of our features. We find that the positive relationships of refugee intake and government regulation of religion with foreign fighter supply stem from European and Muslim-majority countries, respectively. Throughout this paper, we refer to effects in the sense of partial derivatives; we do not make causal claims.

To the limited extent that our research design allows us to inform policy, the implications of our findings for constraining foreign fighter flows are grim: little can be done and what

can be done is costly. This is because the strongest and most accurate predictors of foreign fighter supply are structural features, while the cost of altering policies that predict supply might exceed the benefit of curbing it; in European countries, reducing refugee intake would arguably alienate the median voter, while in Muslim-majority countries, liberal government policies towards religion might be resisted. Furthermore, the relationships uncovered by this analysis say little about the general equilibrium effects of a large shift in government policy.

The remainder of this study proceeds as follows. Section 2.2 provides background information on the Syrian Civil War and foreign fighters. Section 2.3 reviews the emerging literature on foreign fighter supply and synthesizes a theoretical framework to analyze the data. Section 2.4 outlines some obstacles to empirical modeling in this setting, and motivates our method. Section 2.5 presents the theory and mechanics of KRHNB. Section 2.6 covers issues related to modeling choices and computation. Section 2.7 compares our predictions to those of other models. Section 2.8 displays the substantive findings of our analysis and provides possible substantive interpretations. Finally, Section 2.9 summarizes, underlines the limitations of our approach, and suggests directions for future research.

2.2 Foreign Fighters in Syria

Foreign fighters are not a novel fighting technology. They have existed since at least the Greek War of Independence in the 1820s, have participated in different types of conflicts (ethnic, political, religious), and in several regions of the world (Europe, Asia, Latin America, Africa) (Malet, 2010). However, the scale of foreign fighter presence in Syria is unprecedented. Conservative estimates place it around 21,000, far exceeding supply to conflicts in the past 40 years, like those of Afghanistan, Bosnia, or Somalia (Hegghammer, 2011). Already, confirmed cases of foreign fighters have been noted in 50 countries (Neumann, 2015) – their supply is mapped in Figure 2.1.

The scale of the foreign fighter phenomenon has raised a number of concerns in both Syria and supplier countries. Regarding the conflict itself, there is concern that foreign fighters may swing outcomes. Indeed, there is anecdotal evidence that insurgencies with

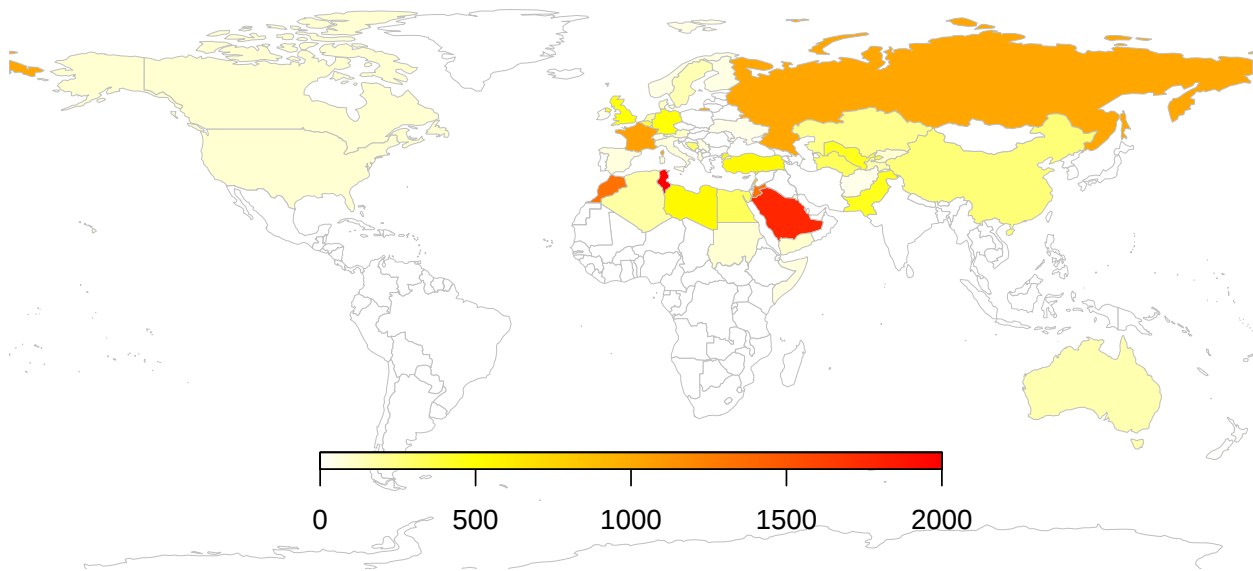


Figure 2.1: Global Distribution of Foreign Fighter Supply, Through 2014

foreign fighters are more successful (Hegghammer, 2011; Malet, 2010). Moreover, the Syrian case is complicated by the presence of multiple organizations with competing ideologies. On the one hand, groups like the Free Syrian Army (FSA) promise a secular, democratic and pro-Western government upon toppling Assad. On the contrary, groups like Islamic State (IS) strive to replace the Ba’athist regime with an Islamic caliphate. To the extent that both of these opposing camps are continuously strengthened by foreign fighters, it is unlikely that either will prevail. Hence, even if insurgents succeed in defeating the state, factional conflict will make a ceasefire unlikely.²

Just as many concerns have been raised in the countries where foreign fighters originate. Despite some governments taking an active stance on the conflict, they do not condone their citizens’ participation, even if it is in support of an organization allied with their government (e.g. FSA).³ More problematic is the case of foreign jihadis, since a coalition of numerous countries has declared war on groups like IS, yet some jihadis are citizens of these countries.

²The dynamics are complicated further if we consider Hezbollah and other Shia or pro-government militias’ support of Assad.

³In order to deter their citizens from joining the conflict, some governments have passed legislation to revoke returning fighters’ citizenship (e.g. Australia, Britain, Canada, USA). See <http://news.nationalpost.com/news/canada/canadian-government-revoking-passports-of-citizens-trying-to-join-extremist-groups>.

A prominent example is “Jihadi John”, a British citizen—and IS fighter—that was bombed by air strikes funded through his taxes. Cases like this have troubled Western governments, whose citizens comprise an estimated 20% of foreign fighters in the conflict.⁴

In addition to shaking the liberal-democratic foundations of these countries, the greatest concern relates to the “veteran effect”: the risk of returning fighters carrying-out terror attacks in their home country (Hegghammer, 2013b). Veterans’ training, experience in combat and access to networks of fighters makes them prime candidates for terrorist recruiting, with Hegghammer (2011) claiming that “most transnational jihadi groups today are by-products of foreign fighter mobilizations” (p. 53). Naturally, policy-makers are worried that their citizens can participate in foreign conflicts and then import violence to their home country.

A similar pattern holds in the literature. Few studies ask what causes one to become a foreign fighter, while virtually none focuses on country-level features. Crucially, ignoring these features impedes our ability to prevent foreign fighter flows in future conflicts. Admittedly, the Syrian conflict is well underway, and thus the most efficient way to minimize its costs is to contain its spillovers. In future cases, though, given that prevention is cheaper than treatment, a more efficient strategy is to identify potential supplier countries. This will also benefit countries on the receiving end; negotiation is always preferable to conflict, but the potential inflow of fighters makes conflict more likely.

This study takes the first step towards predicting country-level foreign fighter flows. Before proceeding, though, in the next section we briefly review the emerging literature on foreign fighters and synthesize it into a broad theoretical framework. We do not put forth an argument for how foreign fighter flows take shape at the country level, let alone the individual level. We simply impose the minimum theoretical structure necessary to later enable a substantive interpretation of our findings.

⁴This estimate is based on the data we employ (Neumann, 2015). In this case, the West is defined as non-Muslim-majority countries.

2.3 Predictors of Foreign Fighter Supply

The literature on foreign fighters is still in its nascent stage, yet combines insights from a range of disciplines: political science, history, sociology, public policy, security, and counter-terrorism studies. A comprehensive review of this work is beyond our scope, and instead we focus on studies that touch on the determinants of foreign fighter supply. Unfortunately, we are aware of only one study that explores *country*-level predictors ([Hewitt and Kelley-Moore, 2009](#)); remaining studies and policy reports tackle the question at the individual level.

Crucially, although we extract the covariates that feature in this literature to synthesize a framework for country-level explanations of foreign fighter supply, we do not claim to *test* the theories posited in the literature. To make that claim would be to commit an ecological fallacy, since existing theories are about individuals, while our data covers countries. We merely borrow from individual-level theories to motivate our search for predictors at the country level. Similarly, we do not present every possible channel through which country features might operate on foreign fighter supply, as our research design does not have the power to adjudicate between competing mechanisms.⁵

We divide country-level explanations of foreign fighter supply into two broad camps: structural vs. policy-related. The former involve features of a country that are “sticky”, and respond very slowly to policy or shocks (e.g. geography, demography, development level).⁶ Policy-related explanations, on the other hand, point to country features that are either the direct output of government decisions (e.g. discrimination laws), or an outcome that is significantly influenced by government (e.g. respect for human rights). Although, policy and structure are interrelated—structure constrains policy, while policy can alter structure in the long-run—we adhere to this simplistic yet powerful classification for the sake of conceptual clarity.

⁵Many of the potential mechanisms we review come from the transnational terrorism literature. For a discussion of the distinction between foreign fighting and other forms of violence, see [Hegghammer \(2011\)](#).

⁶[Jackson and Kollman \(2012\)](#) define a “sticky” variable as “[a] highly autoregressive process that is slow to adapt to changes” (p. 163). They add that “[in] many policy areas, the influence of interest groups, the security of incumbents, the number of veto players, and bureaucratic inertia are likely to produce high values of ρ [the autocorrelation index] in any model of the policy processes” (p. 163).

Numerous structural variables have figured in popular conceptions of foreign fighter-prone countries. Proximity to the conflict zone, because it decreases the travel cost incurred by fighters, along with spillovers of the conflict.⁷ Population, since a positive (unconditional) probability of any individual becoming a foreign fighter implies that larger countries should have a higher supply. Urbanicity and population density, via the notion that radicalization networks operate more easily in urban centers and densely inhabited areas (Gibbs, 1989). A right-skewed age distribution and large male population share, on account of rebel groups' preference for young male recruits.⁸ The share of the country's population that identifies with a particular party in the conflict, by virtue of common nationality, ethnicity, or religion. This operates both on the supply side, with members of the diaspora having a stake in the conflict on account of their shared identity with some rebel group, and on the demand side, with recruiters manipulating the salience of identities to construct a sense of moral obligation to fight (Malet, 2010).⁹ Development, via the findings of Hewitt and Kelley-Moore (2009) that more developed countries supplied more foreign fighters to Iraq.¹⁰ Unemployment, particularly among the young and males, because they decrease the opportunity cost of participating in conflict vis-à-vis recruiters' preference for that target group (RAN, 2014). Homicide rates, given the notion that individuals raised in violent environments might export that violence to different theaters, especially if different forms of violence are substitutable.

An equally large number of policy variables has been associated with foreign fighter supply. Poor human and civil rights, repression, and censorship, due to their association with authoritarianism. Scholars have argued that aspects of liberal democracy alleviate grievances

⁷Distance has also been used to explain the unprecedented *aggregate* flow of fighters to Syria, due to the conflict's proximity to Europe (Hegghammer, 2011).

⁸Nevertheless, it is worth noting that the Syrian foreign fighter movement is the first with such a high participation of females (see <http://www.theguardian.com/world/2014/sep/29/schoolgirl-jihadis-female-islamists-leaving-home-join-isis-iraq-syria>).

⁹Malet (2015) qualifies this by arguing that shared ethnicity is not as strong of a motivating factor as shared religion.

¹⁰The authors do not offer an interpretation of their findings. However, they do review the more conventional view expressed by Lewis (2004) and Gerges (2009); that Islamic radicalism is owed to a lack of modernization in Muslim-majority countries. Clearly, this explanation is problematic when applied to the case of Syria, as it cannot account for the large supply share held by the developed world.

that lead to terrorism (Crenshaw, 1981), but also that they decrease the cost of organizing terrorist acts and the punishment for carrying them out (Schmid, 1992).¹¹ Discriminative policies against minorities that share an identity with some party in the conflict, as they create grievances that have been argued to fuel terrorism (Piazza, 2011; Sageman, 2008). The number of migrants and refugees, via two channels; first, their own radicalization, if they originate from countries that have a stake in the conflict (e.g. Muslim refugees and migrants in France), and second, their effect on citizens of the host country that are prone to radicalization (e.g. French citizens of Moroccan origin). Lastly, internet penetration, as most foreign fighter recruitment occurs online (Hegghammer, 2011).

2.4 Hurdles to Predicting Foreign Fighter Supply

Any empirical investigation of the correlates of foreign fighter supply encounters two central difficulties. The first relates to data availability, as foreign fighter supply is poorly measured. Naturally, foreign fighters do not report their participation in conflict, especially in supplier countries where returning fighters face punishment.¹² As such, any data on foreign fighter supply remains an estimate and is subject to non-random missingness and measurement error. Nevertheless, it is our conviction that the (nascent) literature on this question can gain enough from analyzing this data to warrant a full investigation. Thus, we will treat these estimates as our response variable (Section 2.8) and delegate improved data-collection to future work.

We instead focus on the second difficulty, which relates to prediction, inference, and interpretability. In particular, our approach is motivated by a desire to address four issues. First, we believe that foreign fighter supply at the country level can be theorized as a two-component process. *Whether* a country supplies any foreign fighters can be thought of as one component of the process; *how many* fighters it supplies can be thought of as

¹¹For a review of a similar debate on the role of regime-type in terrorism, see Chenoweth (2013).

¹²Increasingly, fighters are using social media to publicize their activities. This has allowed more fine-grained measures of foreign fighter supply to develop, but not to the extent where we can use them in our analysis.

another component. In other words, using an analogy of countries as firms and foreign fighter supply as a firm choice, one mechanism should determine which countries become suppliers of fighters and another mechanism should determine how much they scale-up their supply by. This is justified if different supply dynamics take over once a country “decides” to become a supplier.¹³ Indeed, anecdotal evidence suggests that after radicalization and recruitment networks form, along with the policies to contain them, the mechanism underlying foreign fighter supply changes; incentives and costs for prospective fighters are fundamentally altered by the existence of previous fighters (Felter and Fishman, 2007; Klausen, 2015).

As such, we require a model that allows us to make a conceptual distinction between the two components of foreign fighter supply. The model should allow a country’s features to differentially affect whether it becomes a supplier or not and how many fighters it supplies. More precisely, we allow for the possibility that there are different functional forms that govern the relationship between the features and the two components of the model. For example, security policies might have a small effect on whether aggrieved citizens decide to form a foreign fighter movement, but they may have a large effect on whether additional citizens join that movement—it may deter them from undertaking the potential legal costs of becoming a fighter. It is also possible that some features influence whether any foreign fighters are supplied, while having no effect on how many are supplied.

The above points us to the familiar hurdle model of Mullahy (1986), a two-part mixture model combining a binary component with a truncated count component.¹⁴ Applied to the question of foreign fighter supply, the former predicts which countries are suppliers and the

¹³The use of “once” here does not imply a temporal dimension, since our model does not take account of time—in the statistical setup, the two decisions (whether to supply, how much to supply) are simultaneous.

¹⁴The hurdle model was introduced to political science by King (1989a) and King (1989b). However, to the best of the authors’ knowledge, its only application to a political science question is Marschall, Ruhil and Shah (2010), which fits a Poisson process to the count component. For a comparison and application of the hurdle versus its more popular counterpart—the zero-inflated model—directed at a political science audience, see Zorn (1998). Virtually all political science studies that employ the zero-inflated negative binomial do so to model terrorism (e.g. Li (2005), Burgoon (2006), and Wilson and Piazza (2013)), or politically violent events more generally (e.g. Bagozzi (2015)). Note that zero-inflation is also used in other processes of discrete random variables, generating models like the zero-inflated ordered probit (Bagozzi et al., 2015), the middle-inflated ordered probit (Bagozzi and Mukherjee, 2012), and the baseline-inflated multinomial logit (Bagozzi and Marchetti, 2017; Bagozzi, 2016).

latter the number of fighters supplied by each country. Crucially, different features can enter each component and even if the same features are entered different marginal effects are returned. Thus, the hurdle model can handle both conceptual distinctions drawn above. Moreover, it closely follows our prior about the two-component structure of foreign fighter supply at the country level.

To adapt the hurdle model to the question of foreign fighter supply, appropriate statistical processes must be chosen for each component. For the binary component, the choice does not matter greatly, with the logit being the standard.¹⁵ As for the count component, the choice of process should depend on the distribution of foreign fighter supply across supplier countries. A common feature of count data is over-dispersion—the variance of the distribution exceeding its mean.¹⁶ Indeed, as Figure 2.1 shows, this is the case with foreign fighter data—some countries supply only a handful of fighters (e.g. New Zealand), while others supply thousands (e.g. Tunisia). Since this pattern cannot arise under the canonical Poisson process, estimates from a Poisson hurdle will suffer from high variance. A popular fix to this problem is to instead fit a negative binomial distribution to the count component. This restores the good statistical properties that the Poisson process holds under no over-dispersion (Lawless, 1987).

An altogether different concern with our empirical setting is the absence of a theory to guide functional form selection. Still at its nascent stage, the literature is very far from specifying the correct function through which country features affect foreign fighter supply. That is, even if there was consensus that the count of foreign fighters from each country is a function of, say, its poverty and unemployment rates, there would be no consensus on whether that function is linear, exponential, logarithmic, or of any other form. This is made worse by the two-component nature of our model, as the danger of misspecification is doubled. The assumption behind most regression models, not just the linear one, is that the systematic component of the response is a linear function of its features King (1989b). However, when models are specified using vague theoretical priors, parametric assumptions

¹⁵For a theoretical overview of count models, see Cameron and Miller (2015). For an applied overview in the context of the R language, see Zeileis, Kleiber and Jackman (2008).

¹⁶See Cox (1983) for an early discussion of problems created by over-dispersion in count models.

may be unwarranted (Ho et al., 2007).

Instead, model-fitting could greatly benefit from a flexible nonparametric approach. Nonparametric models are not new to political science, but their application has been relatively limited.¹⁷ We argue that our discipline has a lot to gain from drawing a closer connection between the theories posited and the functions fit to the data. It is rarely the case that our theories are developed enough to accurately test them using assumptions as strong as those of GLMs. Therefore, to conduct fairer investigations of the validity of our hypotheses, we must relax the narrow confines of parametric models.

One form-free way to include a set of features in the systematic component of the response is to map them into a high-dimensional space, such as a high order polynomial expansion of the data or even more complicated expansions. Yet, such an expansion of the features into a regression model often will make computation unfeasible. Therefore, we need a way to *implicitly* consider higher-order expansions of the features, without actually computing them. This is possible if the expanded feature matrix enters our likelihood function only as an inner product; the inner product can be substituted with an appropriate kernel matrix of the features (see next section), via Mercer’s Theorem. As such, kernels allow us to consider a high-dimensional function space, thereby fitting the CEF of the hurdle negative binomial much more flexibly than its classical counterpart. In our case, we use the Gaussian kernel which corresponds to an infinite-dimensional mapping of the data, allowing for very flexible functions.

A third issue that our approach seeks to address is overfitting—the danger of producing predictions that generalize poorly to other samples. This danger is particularly grave in settings where the data-generating process is under-theorized, as is the case with foreign fighter supply. The researcher can embark on a quest to find the best-fitting model, trying numerous different specifications in order to minimize an appropriate metric, such as mean squared error. That model, in turn, may fit the available data very well, but may not fit other

¹⁷We were unable to find a single review of nonparametric methods in a political science journal or textbook. This stands in contrast to related disciplines, like economics, where nonparametric estimators are more widely employed.

samples well. Overfitting becomes even more worrisome in under-theorized settings when coupled with nonparametric methods. Algorithms such as local polynomial regression and kernel regression fit complex surfaces through the training data, but make large prediction errors when applied to test data (Mroz and Savage, 1999).¹⁸ In other words, their emphasis is on minimizing bias in the given sample, as opposed to generating predictions generalizable to withheld samples.

To strike a better balance on the bias-variance tradeoff, we employ regularization and cross-validation. These are standard tools in machine learning methods that penalize complexity to improve out-of-sample prediction. Regularized algorithms based on the L_2 penalty, such as ridge regression (Hoerl and Kennard, 1970), shrink the coefficients of features that do not significantly improve prediction, while others based on the L_1 penalty, like Lasso (Tibshirani, 1996), perform well in high-dimensional settings where feature selection is desired. Cross-validation, in turn, enables the optimal tuning of the regularization parameter, especially when there are few data points in the training set. Regularization and cross-validation have spurred the development of numerous algorithms. However, until recently these tools had not been used alongside kernel expansion.¹⁹ This might not seem important, given that nonparametric model fitting and regularization can both be accommodated without using kernels (e.g. Random Forest).

Avoiding kernels, though, gives rise to our fourth issue: interpretability and inference. Many machine learning methods produce output that social scientists are not accustomed to analyzing.²⁰ For example, familiar quantities of interest, such as marginal effects and confidence intervals, are absent from algorithms like Random Forest. Similarly, other regularized algorithms, like ridge regression, do produce familiar output, but make strong parametric

¹⁸Kernel regression does not necessarily involve regularization, and thus should not be confused with kernel *regularized* regression models like KRLS and KRHNB.

¹⁹Zhu and Hastie (2005) introduce kernel logistic regression, Shim and Hwang (2011) and Shim (2012) apply kernel regularization to Poisson regression, and Hainmueller and Hazlett (2014) present kernel regularized least squares.

²⁰For a relatively comprehensive review of interpretability issues in machine learning methods, see pp. 9-12 of Hainmueller and Hazlett (2014).

assumptions—namely, that the response is a linear function of the features, as in OLS.²¹ In short, kernel expansion vis-à-vis regularization and cross-validation aggregates the benefits of all aforementioned models: fitting a flexible solution surface, penalizing complexity, and communicating output to social scientists.

Before presenting the mechanics of our approach, we note that machine learning methods are not new to political science. Although already in Beck, Katz and Tucker (1998) and Beck, King and Zeng (2000) efforts were made to import some of these ideas to the discipline, limitations to computing power contained their expansion. Recently, though, this containment has ceased: Kenkel and Signorino (2013) introduce *polywog*, a model that fuses basis regression with regularization, cross-validation, and bootstrapping; Hill and Jones (2014) and Muchlinski et al. (2016) use Random Forests to model state repression and civil war onset, respectively; Wilson and Wright (2017) use KRLS as a robustness check in modeling expropriation risk in autocracies; Green and Kern (2012) and Montgomery et al. (2015) apply Bayesian Additive Regression Trees to survey experiment data and election fraud measures, respectively; Montgomery and Olivella (Forthcoming), Fariss and Jones (2018), and Jones and Fridolin (2015) discuss possible contributions of machine learning to political science.²² In the following section, we develop a method that we hope will contribute to the growing use of machine learning in political science.²³

²¹Hainmueller and Hazlett (2014) note that applying kernel expansion (with a Gaussian kernel) and regularization (with an L_2 norm) to a least squares problem produces an *infinite-dimensional* ridge regression model. This should be contrasted to standard ridge regression (without kernel expansion), which solves a P -dimensional linear problem and, hence, produces a parametric fit.

²²These studies by no means constitute the universe of political science papers employing machine learning tools.

²³For alternative ways of motivating and deriving kernel regularized models, we refer the reader to the simple and intuitive expositions in Hainmueller and Hazlett (2014) and their supplementary appendix. In what follows, we only take one possible approach to the derivation, in order to expose the mechanics of our method.

2.5 The Model

In this section we construct the likelihood for the hurdle negative binomial model, reparameterize our model in an infinite-dimensional space, demonstrate that our features only enter our penalized likelihood via inner products, and then use Mercer’s theorem to rewrite the problem using Gaussian kernels rendering optimization feasible.

Assume y_i is a count ($y_i \in \{0, 1, \dots, \infty\}$). We begin with the general formulation of the two-component density of the hurdle model (Mullahy, 1986), which combines a zero-hurdle component, right-censored at $y_i = 1$, with a positive count component, left-truncated at $y_i=1$.²⁴

$$p(y_i) = \begin{cases} p_0(y_i = 0) & \text{if } y_i = 0 \\ \frac{p_1(y_i)}{1-p_1(y_i=0)}(1 - p_0(y_i = 0)) & \text{if } y_i \geq 1 \end{cases} \quad (2.1)$$

Any binomial model (e.g. probit) or right-censored count model (e.g. Poisson) can be chosen for the zero hurdle component. We follow the literature in opting for the computationally simple logit. Similarly, any count model can be chosen for the positive count component. We opt for the negative binomial, because it allows us to account for over-dispersion—as we will see, a characteristic of our response variable. Before providing the likelihood for y_i , we note the following useful densities:

$$p_0(y_i = 0) = \frac{1}{1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})} \quad (2.2)$$

$$p_1(y_i) = \frac{\Gamma(\zeta + y_i) \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \left(\frac{\exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^{y_i}}{\Gamma(1 + y_i)\Gamma(\zeta)} \quad (2.3)$$

$$1 - p_1(y_i = 0) = 1 - \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \quad (2.4)$$

²⁴For the original formulation of the hurdle model—albeit not for count variables—see Cragg (1971). Throughout, positive counts refer to *strictly* positive counts.

The first density is the likelihood of observing a zero outcome, where $\mathbf{x}_i$ is a length- P vector of features for observation i , $\boldsymbol{\alpha}$ is the parameter vector for the binary component, α_0 is an intercept, and $\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}$ is the linear predictor.

The second density is that of the standard negative binomial, where $\boldsymbol{\beta}$ is the parameter vector for the count component, β_0 is an intercept for the count component, ζ is the overdispersion parameter²⁵, and $\Gamma(\cdot)$ is the gamma function.²⁶ The hurdle model allows us to specify each component as a function of different country features—say, $\mathbf{x}_i$ for the binary component and $\mathbf{z}_i$ for the count. This choice can be motivated from a substantive perspective; if theory dictates that different country features affect the likelihood of a country being a supplier (binary component), versus its likelihood of supplying a certain number of fighters conditional on being a supplier (truncated count component), including different features in each component is appropriate. However, given the nascent state of the literature on foreign fighters, we do not feel justified in making that choice, and assume that the same set of features ($\mathbf{x}_i$) affects both likelihoods. Furthermore, because we allow for very flexible functional forms in both components and penalize complexity, including irrelevant variables will not cause overfitting or induce misspecification bias.²⁷

The third density is simply the complement of the negative binomial density evaluated at zero. This term plays a crucial role in the hurdle model, since a zero count *cannot* arise from the count component, which is left-truncated at unity—hence the “hurdle”. As such, in the formula for the density of a positive count we scale the binomial density by $1 - p_1(y_i = 0)$. This is the key difference between the hurdle and the zero-inflated model; the latter allows for zero counts to arise from *both* components.²⁸ We opt for the hurdle model not just

²⁵For completeness, we note the following properties of the negative binomial density: $\mathbb{E}[y_i|\mathbf{x}_i, \boldsymbol{\beta}] = \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i) = \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) \exp(\epsilon_i) = \mu_i h_i$, where $h_i \sim \Gamma(\zeta, \zeta)$, $\mathbb{E}[h_i] = \zeta/\zeta = 1$, and $V[h_i] = 1/\zeta$. Thus, after conditioning on ζ , we obtain $\mathbb{E}[y_i|\mathbf{x}_i, \boldsymbol{\beta}, \zeta] = \mu_i$, as in the Poisson, and $Var[y_i|\mathbf{x}_i, \boldsymbol{\beta}, \zeta] = \mu_i(1 + \mu_i/\zeta)$, which tends to the Poisson’s variance as $\zeta \rightarrow \infty$.

²⁶Note that $\Gamma(n) = (n-1)!$, where n is a positive integer.

²⁷Of course, if a feature truly has no relationship to one component of the model, including it will reduce the efficiency of our estimator. Knowing this a priori is very difficult and a benefit to regularized, flexible methods such as KRHNB is the ability to include many features and allowing the estimator to learn the appropriate features and functional form.

²⁸A case for using the hurdle is also made by [Porter, White et al. \(2012\)](#), but with respect to terrorism

because it is theoretically appropriate, as explained in Section 2.4, but also because it is computationally more straightforward; the likelihood function is perfectly separable with respect to the two components and hence each parameter vector can be fit by independently maximizing the respective component.²⁹

Now we can form the likelihood for observation i :

$$L_i(\cdot) = [p_0(y_i = 0)]^{1-d_i} \left[\frac{p_1(y_i)}{1 - p_1(y_i = 0)} (1 - p_0(y_i = 0)) \right]^{d_i} \quad (2.5)$$

$$d_i = \begin{cases} 0 & \text{if } y_i = 0 \\ 1 & \text{if } y_i \geq 1 \end{cases}$$

We use the densities in Equations 2.2, 2.3, and 2.4 to form the joint (sample) likelihood for N observations. Taking the log, we arrive at the sample log-likelihood.³⁰ Where $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}^\top, \beta_0, \boldsymbol{\beta}^\top, \zeta)^\top$ and $\mathcal{D} = (\mathbf{Y}, \mathbf{X})$,

$$\begin{aligned} \ell_N(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^N & -\log(1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})) + d_i \left[\log \Gamma(\zeta + y_i) + \zeta \log \zeta \right. \\ & - (\zeta + y_i) \log(\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})) + y_i(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) + (\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}) \\ & \left. - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \right) \right] \end{aligned} \quad (2.6)$$

Note that α_0 and β_0 are intercept terms. At this stage, we reparameterize the log-likelihood, by substituting the linear predictor functions $\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}$ and $\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}$ with $\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi}$ and $\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega}$, respectively, functions linear in $\boldsymbol{\phi}(\mathbf{x}_i)^\top$, a mapping of the

data.

²⁹This is not the case with the zero-inflated model, whereby a mixing of zeros occurs under the two components. The computational advantage of the hurdle over the zero-inflated model becomes even larger when the *same* set of features is used in both components, as we do in our specification. This is because the mixing of zeros from the two components hinders identification of the two sets of coefficients for the features.

³⁰See Appendix 2.11 for the intermediate steps.

features. That is, the feature space $\mathbf{X} \in \mathbb{R}^P$ is expanded onto a higher-dimensional space $\mathbb{R}^{P'}$, where $P \ll P'$, and the parameter vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^P$ are accordingly substituted with $\boldsymbol{\psi}, \boldsymbol{\omega} \in \mathbb{R}^{P'}$. Where $\boldsymbol{\theta}_\phi = (\psi_0, \boldsymbol{\psi}^\top, \omega_0, \boldsymbol{\omega}^\top, \zeta)^\top$ and $\mathcal{D} = (\mathbf{Y}, \mathbf{X})$,

$$\begin{aligned} \ell_N(\boldsymbol{\theta}_\phi|\mathcal{D}) = \sum_{i=1}^N & -\log(1 + \exp(\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi})) + d_i \left[\log \Gamma(\zeta + y_i) + \zeta \log \zeta \right. \\ & - (\zeta + y_i) \log(\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})) + y_i(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega}) \\ & + (\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi}) - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) \\ & \left. - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} \right)^\zeta \right) \right] \end{aligned} \quad (2.7)$$

Again, ψ_0 and ω_0 are unregularized intercept terms; for example, $\boldsymbol{\psi}$ is defined as $\boldsymbol{\psi} = [\psi_1 \ \psi_2 \ \dots]^\top$ and does not include ψ_0 . Next, we take the negative of the log-likelihood, Equation 2.7, turning our exercise into a minimization problem. In addition, we add $\|\boldsymbol{\psi}\|^2$ and $\|\boldsymbol{\omega}\|^2$, each of which is the square of the L_2 norm in our expanded feature space.³¹ These norms are multiplied by $\lambda_\psi, \lambda_\omega \in \mathbb{R}^+$, tuning parameters that govern the tradeoff between fit and complexity for the coefficients on the features in each component. In sum, the norms and the regularization parameters are penalties that ensure that smoother functional forms are favored, thereby protecting against overfitting. The penalized log-likelihood arises as our target function:

$$R_N(\boldsymbol{\theta}_\phi, \lambda_\psi, \lambda_\omega|\mathcal{D}) = -\ell_N(\boldsymbol{\theta}_\phi|\mathcal{D}) + \lambda_\psi \|\boldsymbol{\psi}\|^2 + \lambda_\omega \|\boldsymbol{\omega}\|^2 \quad (2.8)$$

Next, we solve the First Order Condition (FOC) for each parameter vector, in order to

³¹The choice of the L_2 norm can be motivated from a Bayesian perspective. As in [Hainmueller and Hazlett \(2014\)](#), it can be shown that the parameter estimates that maximize our target function—the penalized log-likelihood with an L_2 norm (Equation 2.8)—are the Maximum a Posteriori estimates of the hurdle negative binomial posterior, when a Normal prior is chosen for the parameters of the features. For a general treatment of the correspondence between Bayesian inference and regularization, see [Kimeldorf and Wahba \(1970\)](#).

demonstrate our use of Mercer's Theorem to reduce our problem from a potentially infinite-dimensional one to a tractable function. The details of this derivation can be found in Appendix 2.11. In both FOCs, many of the terms reduce to a scalar, which we can label c_i^ψ and c_i^ω . As such, we rewrite our FOC solutions for $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ as:³²

$$\boldsymbol{\psi}^* = \sum_{i=1}^N c_i^\psi \boldsymbol{\phi}(\mathbf{x}_i) \quad (2.9)$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^N c_i^\omega \boldsymbol{\phi}(\mathbf{x}_i) \quad (2.10)$$

Now we substitute the solutions for $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ back into the target function. Where $\boldsymbol{\theta}_c = (c_0^\psi, \mathbf{c}^{\psi^\top}, c_0^\omega, \mathbf{c}^{\omega^\top}, \zeta)^\top$ and $\mathcal{D} = (\mathbf{Y}, \mathbf{X})$,

$$\begin{aligned} R_N(\boldsymbol{\theta}_c, \lambda_\psi, \lambda_\omega | \mathcal{D}) = & - \sum_{i=1}^N \left(- \log \left(1 + \exp(c_0^\psi + \boldsymbol{\phi}(\mathbf{x}_i)^\top \sum_{j=1}^N c_j^\psi \boldsymbol{\phi}(\mathbf{x}_j)) \right) \right. \\ & + d_i \left[\log \Gamma(\zeta + y_i) + \zeta \log \zeta - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) \right. \\ & - (\zeta + y_i) \log \left(\zeta + \exp(c_0^\omega + \boldsymbol{\phi}(\mathbf{x}_i)^\top \sum_{j=1}^N c_j^\omega \boldsymbol{\phi}(\mathbf{x}_j)) \right) \\ & + y_i \left(c_0^\omega + \boldsymbol{\phi}(\mathbf{x}_i)^\top \sum_{j=1}^N c_j^\psi \boldsymbol{\phi}(\mathbf{x}_j) \right) + \left(c_0^\psi + \boldsymbol{\phi}(\mathbf{x}_i)^\top \sum_{j=1}^N c_j^\psi \boldsymbol{\phi}(\mathbf{x}_j) \right) \\ & \left. \left. - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(c_0^\omega + \boldsymbol{\phi}(\mathbf{x}_i)^\top \sum_{j=1}^N c_j^\omega \boldsymbol{\phi}(\mathbf{x}_j))} \right)^\zeta \right) \right] \right) \\ & + \lambda_\psi \left\langle \sum_{i=1}^N c_i^\psi \boldsymbol{\phi}(\mathbf{x}_i), \sum_{i=1}^N c_i^\psi \boldsymbol{\phi}(\mathbf{x}_i) \right\rangle + \lambda_\omega \left\langle \sum_{i=1}^N c_i^\omega \boldsymbol{\phi}(\mathbf{x}_i), \sum_{i=1}^N c_i^\omega \boldsymbol{\phi}(\mathbf{x}_i) \right\rangle \end{aligned} \quad (2.11)$$

Noting that $\sum_{j=1}^N c_j^m \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\phi}(\mathbf{x}_j) = \sum_{j=1}^N c_j^m \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle$, $m \in \{\psi, \omega\}$, it becomes obvious that the expanded features enter our target function only as inner products. Mercer's Theorem, allows us to replace these inner products with any positive semi-definite kernel,

³²Alternatively, this can be directly shown by invoking the Representer Theorem (Kimeldorf and Wahba, 1971).

$k(\mathbf{x}_i, \mathbf{x}_j)$.³³ Crucially, this means that we do not actually have to expand our features onto the higher-dimensional space that they are allowed to span via $\phi(\cdot)$, but merely pass them through kernels. Although any positive semi-definite kernel suffices for performing kernel substitution, we opt for the Gaussian, due to its well-known properties.³⁴ Namely, $\phi(\cdot)$ will be infinite-dimensional. A way to think about this infinite-dimensional vector is to imagine that it contains all possible combinations and functions of the original features. For example, it contains $\mathbf{x}^{(1)}$, $\mathbf{x}^{(1)}\mathbf{x}^{(2)}$, $\sqrt{(|\mathbf{x}^{(3)}|)}$, $\mathbf{1}\{(\mathbf{x}^{(1)} > 0)\}$ and so on and so forth, where $\mathbf{x}^{(j)}$ is the j th feature.

Thus, letting $\mathbf{c}^m = \begin{bmatrix} c_1^m & c_2^m & \dots & c_N^m \end{bmatrix}$ where N is the number of observations and m represents either the first or second components, ψ or ω , letting $\mathbf{K}$ be the kernel matrix of our sample, and letting $\mathbf{k}_i$ be the i th column of $\mathbf{K}$, we can rewrite our target function as:

$$\begin{aligned}
R_N(\boldsymbol{\theta}_{\mathbf{c}}, \lambda_{\psi}, \lambda_{\omega} | \mathbf{Y}, \mathbf{K}) = & - \sum_{i=1}^N \left(-\log \left(1 + \exp(c_0^{\psi} + \mathbf{c}^{\psi \top} \mathbf{k}_i) \right) + d_i \left[\log \Gamma(\zeta + y_i) + \zeta \log \zeta \right. \right. \\
& - (\zeta + y_i) \log (\zeta + \exp(c_0^{\omega} + \mathbf{c}^{\omega \top} \mathbf{k}_i)) + y_i (c_0^{\omega} + \mathbf{c}^{\omega \top} \mathbf{k}_i) \\
& + (c_0^{\psi} + \mathbf{c}^{\psi \top} \mathbf{k}_i) - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) \\
& \left. \left. - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(\mathbf{c}^{\omega \top} \mathbf{k}_i)} \right)^{\zeta} \right) \right] \right) \\
& + \lambda_{\psi} \mathbf{c}^{\psi \top} \mathbf{K} \mathbf{c}^{\psi} + \lambda_{\omega} \mathbf{c}^{\omega \top} \mathbf{K} \mathbf{c}^{\omega}
\end{aligned} \tag{2.12}$$

The resulting optimization problem does not have a closed-form solution. Therefore, we

³³That is, Mercer's Theorem holds that, for any positive semi-definite kernel $k(\cdot, \cdot)$, there exists a mapping $\phi(\cdot)$ that projects $\mathbf{x}_i$ into a higher-dimensional vector $\phi(\mathbf{x}_i)$ such that $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, $\forall i, j$. Hence, this is also known as the "kernel trick", or kernel substitution.

³⁴For any two data points $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^P$, the Gaussian kernel-based distance is $k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2} \right)$, where σ^2 is the kernel bandwidth. We follow [Hainmueller and Hazlett \(2014\)](#) in setting $\sigma^2 = P$, where P is the number of features. This provides good performance and allows for differentiation among observations. For a review of kernel-based machine learning methods, see [Schölkopf and Smola \(2002\)](#). For a critique of the use of Gaussian kernels in KRLS with small samples, see [Braga and Monard \(2015\)](#).

minimize the above with respect to $\{c_0^\psi, \mathbf{c}^\psi, c_0^\omega, \mathbf{c}^\omega, \zeta\}$ through numerical optimization.

To summarize, we receive estimates for $\mathbf{c}^\psi$ and $\mathbf{c}^\omega$, which act as a kind of weight for each observation i in the two CEFs (describing the mean of the hurdle and the truncated count components, respectively). For example, our estimate of the probability of sending no foreign fighters, the probability in the logit component, is estimated as

$$p_0(y_i = 0) = \frac{1}{1 + \exp(\hat{c}_0^\psi + \hat{\mathbf{c}}^{\psi\top} \mathbf{k}_i)}$$

As a result, directly interpreting the estimated coefficients $\hat{\mathbf{c}}^\psi$ and $\hat{\mathbf{c}}^\omega$ can be very difficult for two reasons: they influence the outcome through non-linear transformations like the logistic function and are acting on $\mathbf{k}_i$ instead of $\mathbf{x}_i$, our features of interest. Therefore, in Section 2.6.2 we take the partial derivatives of the CEF for the outcome y_i with respect to our columns of $\mathbf{X}$ so that we can interpret our results using our input features.

2.6 Computation & Quantities of Interest

2.6.1 Optimization

There are three main difficulties with fitting this model. First, there is minor sensitivity to starting values in the numerical optimization. Using the BFGS algorithm limits this problem dramatically, and the sensitivity only arises when using implausible starting values.³⁵ However, this sensitivity could be guarded against by doing a grid search over some set of starting values.

This solution is difficult to implement because of the second problem with fitting this model numerically—speed. While we supply the analytic gradient of our target function to the BFGS algorithm, these functions are quite complicated and high-dimensional. We numerically optimize with starting values for all coefficients $\mathbf{c} = \mathbf{0}$ and $\zeta = 1$. From some

³⁵Implausible here means uniformly positive or negative starting values. They are implausible because the data have been scaled, thus the coefficients will generally be distributed around 0.

rudimentary grid searches, this starting value has succeeded in finding the best minimum using our data and makes intuitive sense; setting $\mathbf{c} = \mathbf{0}$ means our initial estimate of the CEF is simply the sample average over the entire feature space.³⁶

Third, selecting the appropriate regularization parameters (λ_ψ , λ_ω) can be difficult, as there are two parameters. (This also slows-down optimization.) The traditional approach is to use cross-validation (CV) (Stone, 1974) to select parameters that minimize the cross-validation RMSE (Friedman, Hastie and Tibshirani, 2001). We advocate and implement in our software k -fold cross validation and a grid search over the two regularization parameters. For the application to foreign fighter supply, we select regularization parameters after several manual grid searches.

2.6.2 Pointwise Marginal Effects

The CEF for test observation i is:

$$\begin{aligned}\mathbb{E}[y_i|\mathbf{k}_i] &= p_0(y_i = 0) * 0 + (1 - p_0(y_i = 0)) * \mathbb{E}[\hat{Y}_i|\hat{Y}_i > 0] \\ &= (1 - p_0(y_i = 0)) \frac{\mu_i}{1 - p_1(y_i = 0)} \\ \widehat{\mathbb{E}[y_i|\mathbf{k}_i]} &= \frac{\exp(\hat{c}_0^\psi + \hat{\mathbf{c}}^\psi^\top \mathbf{k}_i)}{1 + \exp(\hat{c}_0^\psi + \hat{\mathbf{c}}^\psi^\top \mathbf{k}_i)} \frac{\exp(\hat{c}_0^\omega + \hat{\mathbf{c}}^\omega^\top \mathbf{k}_i)}{\left(1 - \left(\frac{\hat{\zeta}}{\hat{\zeta} + \exp(\hat{c}_0^\omega + \hat{\mathbf{c}}^\omega^\top \mathbf{k}_i)}\right)^{\hat{\zeta}}\right)}\end{aligned}\quad (2.13)$$

where μ_i is the mean component of the negative binomial distribution, and $\{c_0^m, \mathbf{c}^m\}$, $m \in \{\psi, \omega\}$ are the intercepts and coefficient vectors along with $\hat{\zeta}$ that minimize our target function (Equation 2.12).

Using this formula, we obtain a quantity of interest analogous to marginal effects in GLMs. We use numerical differentiation to take the partial derivative of $\widehat{\mathbb{E}[y_i|\mathbf{k}_i]}$ with respect

³⁶In some instances, the algorithm gets stuck in local minima. In particular, it sometimes estimates $\zeta < 0.0001$, which results in estimates of y_i that are too large by 3 or 4 orders of magnitude. However, because the performance is so poor, it fails to find a way to improve it. For now, we forcibly prevent ζ from reaching such implausibly small numbers to avoid this shortcoming of the BFGS algorithm.

to each feature $\mathbf{x}^{(j)}$ and evaluate it at each observation.³⁷ Because our CEF is non-linear, these pointwise marginal effects will vary over the feature space and provide rich detail about the shape of the CEF. Indeed, wherever an observation lives in the feature space, we will have an estimate of the slope of the CEF with respect to each feature.

This allows for us to summarize average effects and estimate heterogeneous treatment effects. For example, we can display the range of pointwise marginal effects across the training values of $\mathbf{x}^{(j)}$ for feature j as a histogram (see Section 2.8). Alternatively, one can focus on the mean of this distribution, the sample-average pointwise marginal effect of $\mathbf{x}^{(j)}$, which would be analogous to the marginal effect produced by a linear model. Yet another alternative is to report the pointwise marginal effect for the “typical” training point—an observation with mean/median/modal values of the features. This approach is subject to the usual limitations.³⁸ Our companion software allows the user to choose the effects reported.

2.6.3 Estimating Uncertainty of Sample-Average Pointwise Marginal Effects

As noted in Section 2.5, there is no closed-form solution to our optimization problem (Equation 2.12). Consequently, obtaining an analytical estimate of the uncertainty of our predictions is not straightforward and we opt for a computational one.³⁹ Namely, we obtain non-parametric bootstrapped estimates of our pointwise marginal effects (Efron, 1979).⁴⁰ This involves treating our training set as the population and repeatedly resampling from it (with replacement) to calculate sample-average pointwise marginal effects. Crucially, we hold the regularization parameters fixed across the resampled training sets and only re-fit the remaining parameters ($\mathbf{c}$ and ζ). This economizes greatly on computational time, as we do

³⁷Remember that $\mathbf{k}_i$ is a function of the input features.

³⁸See Gelman and Hill (2006) for different approaches to summarizing predictive effects.

³⁹For one analytical approach in the context of a penalized zero-inflated negative binomial, see Wang, Ma and Wang (2014). It is based on the sandwich estimator, whose consistency when applied to non-concave penalized likelihood problems was demonstrated in Fan and Li (2001).

⁴⁰Kenkel and Signorino (2013) also follow a bootstrapping approach to obtain a measure of uncertainty for their pointwise marginal effects. For an application and extension of the bootstrap to the zero-inflated negative binomial, see Garay et al. (2011).

not have to execute the full optimization routine on every bootstrapped sample. However, it will result in smaller estimates of uncertainty as the variability in the regularization parameter is not incorporated (Tibshirani, 1996). The resulting estimates provide a distribution of marginal effects, which can be graphed as a histogram, or summarized numerically to create a bootstrapped percentile interval.⁴¹

2.7 Performance Relative to Other Models

Beyond getting purchase on the causes of foreign fighter supply, we are also interested in benchmarking the performance of our method against other machine learning techniques. In order to do this, we take a leave-one-out cross-validation approach: we train our method and four competing methods on all but one observation and then predict the foreign fighter supply to the withheld observation, repeating this for every observation. We take these out-of-sample predictions for each observation and their true values and calculate the root mean squared (RMSPE) and mean absolute prediction error (MAPE) of our estimated outcomes. The first method we benchmark against is KRLS. The rationale behind this choice is simple: if we cannot perform similarly or better to the method we are adding complexity to, then there is no reason for this extension beyond the ability to interpret the two components of the hurdle model. The second method is KRLS with predicted values truncated at 0, which will uniformly improve the performance of KRLS, but will completely impair its interpretability. The third method is a random forest (Breiman, 2001) with 500 trees, and the number of parameters available at each node set to 5 – roughly the square root of the total number of features (27). The last method is standard OLS, where the predictors enter as a simple linear function.

The results of this leave-one-out cross-validation approach are presented in the upper half of Table 2.1. KRHNB beats all other methods using MAPE and loses only to random forest on RMSPE. The source for this discrepancy can be seen easily if all of the predicted

⁴¹Our companion software provides a range of summary statistics for the bootstrapped estimates of these effects.

Table 2.1: Comparing Prediction Error

Data	Method	CV RMSPE	CV MAPE
Foreign Fighters	KRHNB	282.52	107.91
	KRLS	288.28	141.59
	KRLS (0-truncated)	285.36	129.79
	Random Forest	278.08	131.60
	OLS	299.88	182.86
Burgoon (2006)	KRHNB	5.81	2.67
	KRLS	5.71	2.72
	KRLS (0-truncated)	5.71	2.71
	Random Forest	5.60	2.68
	OLS	6.49	2.94

Notes: Bold denotes the lowest prediction error. We use leave-one-out CV error for the foreign fighters dataset and 5-fold CV error for the Burgoon dataset, as it is much larger. Therefore, while the prediction errors for the Foreign Fighters dataset are fixed, the 5-fold CV error for the Burgoon data might vary depending on how the folds are constructed. Other results produce the same rankings.

outcomes are plotted against the observed outcomes. Figure 2.2 plots the observed number of foreign fighters on the x -axis, and the predicted number on the y -axis. The key advantage that random forest has over KRHNB is that it far outperforms it when predicting the largest value, Tunisia. However, all of these methods struggle to predict countries with high levels of foreign fighter supply, presumably because the predictors that explain the difference between supplying 200 and 2,000 fighters are not in our data. KRHNB performs well with respect to MAPE, because for values of y between 0 and 200, KRHNB does quite well, while KRLS and Random Forests are more likely to predict values that are too large. This can be seen more clearly when we zoom in to the predictions for outcomes less than 150 – the right panel of Figure 2.2.

Therefore, KRHNB is good at predicting whether there are any supplied foreign fighters. This means the first component is making good predictions, while the second component has little power, and most values are predicted near the mean. MAPE is much larger for random forest than KRHNB, which exhibits better fit at lower levels of foreign fighter supply. Thus, it appears that the added structure in the form of the hurdle improves fit over models that do not assume much structure, such as KRLS or random forests. However, that lack

of structure does allow some improved prediction for high levels of foreign fighter supply. Nonetheless, all these methods do quite poorly in the tail of the data, indicating a lack of explanatory power for many variables, or the fundamental unpredictability that describes observations living in tails of distributions.

As an additional evaluation of our method, we replicate [Burgoon \(2006\)](#), who uses a zero-inflated negative binomial model to predict cross-national time variation in terrorist attacks. As shown in the lower half of [Table 2.1](#), we observe a similar pattern of performance on this data; KRHNB minimizes MAPE, but performs worse on RMPSE.

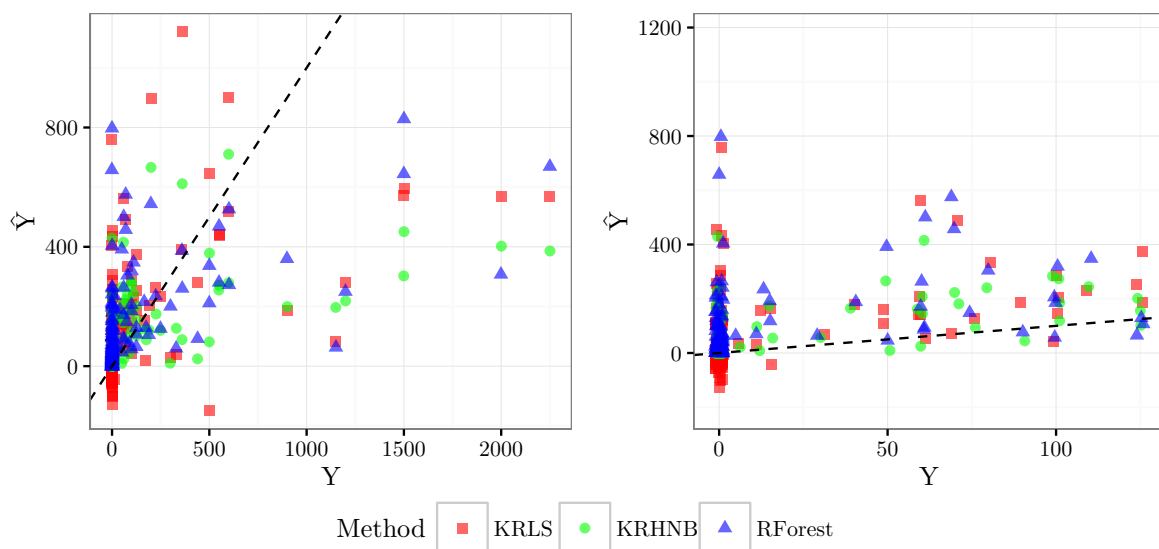


Figure 2.2: Leave-one-out CV Predictions

Notes: Black line is where perfect predictions would lie ($Y = \hat{Y}$). The left panel is the full dataset, while the right panel zooms in to where $Y < 150$. The values of Y are jittered because of the clustering at 0.

2.8 Results & Discussion

2.8.1 Data

To uncover predictors of foreign fighter supply, we build a design matrix of 27 geographic, demographic, economic, and political variables for 163 countries. All of these features have been measured prior to 2014, although issues of endogeneity are beyond the scope of this

paper. Furthermore, there is some (though very little) missingness, because we aggregate over several years (where applicable) to get as complete a dataset as possible. There is some remaining missingness, which we address by taking the mean over 1,000 datasets, imputed using Amelia II (Honaker et al., 2011).⁴² In the Appendix Table 2.4, we report full results with list-wise deletion, leaving us with 147 observations. The results are largely the same.

Table 2.2: Description of Features

	Mean	Std. Dev.	Min.	Median	Max.	Source
Foreign Fighter Supply	114.08	332.34	0.00	0.00	2250.00	ICSR
Contiguous	0.04	0.20	0.00	0.00	1.00	CEPII
Europe	0.21	0.41	0.00	0.00	1.00	Hand Coded
Distance (km)	5251.69	3743.59	85.94	4363.00	15630.00	CEPII
Log Population	16.23	1.51	13.20	16.15	21.03	World Bank WDI
Log Pop. Density	4.19	1.40	0.60	4.30	8.95	World Bank WDI
Youth Bulge (15-24 pct)	17.06	3.68	9.60	18.10	23.10	World Bank WDI
Sunni Pct	25.45	35.33	0.07	5.07	99.40	Pew
Shia Pct	2.61	10.82	0.00	0.06	92.22	Pew
Religious Frac.	0.43	0.24	0.00	0.44	0.86	Fearon & Laitin
Govt. Reg. Religion	2.78	3.24	0.00	1.39	10.00	ARDA
Govt. Fav. Religion	4.50	3.18	0.00	4.73	10.00	ARDA
Soc. Reg. Religion	4.22	3.58	0.00	3.67	10.00	Freedom House
Freedom House Pol. Rights	3.63	2.11	1.00	3.00	7.00	Freedom House
Migrants as Pct of Pop	7.87	12.06	0.05	2.94	74.61	World Bank WDI
Log Refugees in Country	8.50	3.04	0.00	8.85	14.30	UNHCR
Log Refugees from Country	7.35	2.92	0.00	7.18	14.75	World Bank WDI
Log GDP pc, PPP	9.10	1.24	6.37	9.26	11.76	World Bank WDI
Life Expectancy	69.89	9.48	45.55	72.27	83.83	World Bank WDI
Male LFP	79.42	7.03	48.40	79.80	95.90	World Bank WDI
Youth Unemployment	18.33	12.58	0.70	14.80	60.40	World Bank WDI
Internet User per 100	40.01	29.46	0.00	39.20	95.05	World Bank WDI
Log Homicides Per 100k	1.46	1.20	-1.61	1.57	4.50	World Bank WDI

Notes: ICSR is Neumann (2015). CEPII is the Centre d’Études Prospectives et d’Informations Internationales (Mayer and Zignago, 2011). Fearon & Laitin is Fearon and Laitin (2003). ARDA is the Association of Religion Data Archives (Finke, 2010). The Pew data can be found at in a report by Grim and Hackett (2006). The Freedom House data are from Teorell et al. (2013). The UNHCR data can be found at http://popstats.unhcr.org/en/time_series and are different from the World Bank summary of refugees in asylum because we exclude internally displaced persons.

The full list of features can be found in Table 2.2. We have data for all 163 countries⁴³ with

⁴²This algorithm relies on the assumption that the data are missing at random, which is almost certainly invalid. However, no feature has more than 4 missing observations, save for the measure of how many refugees are in a country (missing 10 observations). Furthermore, only 0.7% of the data are missing.

⁴³We exclude Macao and Puerto Rico as well. They had high rates of missingness due to many organizations not collecting data on these polities.

populations above 500,000 except for Syria and Iraq, where our outcome is not measured. Foreign fighter data come from the ICSR Report of 1/25/2015 (Neumann, 2015), which estimates the number of all foreign fighters through the end of 2014. The estimates are either a single value, or a range. If there is a range, we take its mean, and round to the nearest integer.⁴⁴

2.8.2 Main Results

We fit KRHNB on the full data set, with λ_ψ and λ_ω selected by a grid-search using leave-one-out cross validation. We use the Gaussian kernel to transform our data so that we are essentially working in an infinite-dimensional expansion of the design matrix and thus considering an infinite variety of complex functional forms. Following this optimization, we analyze the marginal effects of each feature on country-level foreign fighter supply in two ways. The first way is presented in Figure 2.3, which contains the sample-average marginal effect and bootstrapped confidence intervals, for both the binary and count components, as well as the two components jointly. Note that the confidence intervals are constructed using the empirical 2.5th percentile and 97.5th percentile of the bootstrapped distribution of marginal effects.⁴⁵ The exact results and confidence intervals for the sample-average marginal effects are presented in Table 2.3 of the Appendix. The bootstrapped intervals are skewed away from 0 for many sample-average marginal effects in the foreign fighters application, indicating that several outliers and the complex target function may prevent the bootstrapped percentile interval from achieving nominal coverage.

⁴⁴ICSR also lack data from the West Bank and Gaza. In the interest of maximizing our sample size, we split the count for Israel and assign half of those sent from Israel to the West Bank and Gaza. Table 2.4 presents the results were we instead drop the West Bank and Gaza from our analysis, as well as other countries that have missingness. The results are substantively similar.

⁴⁵This is sometimes known as the percentile interval, or the percentile bootstrap.

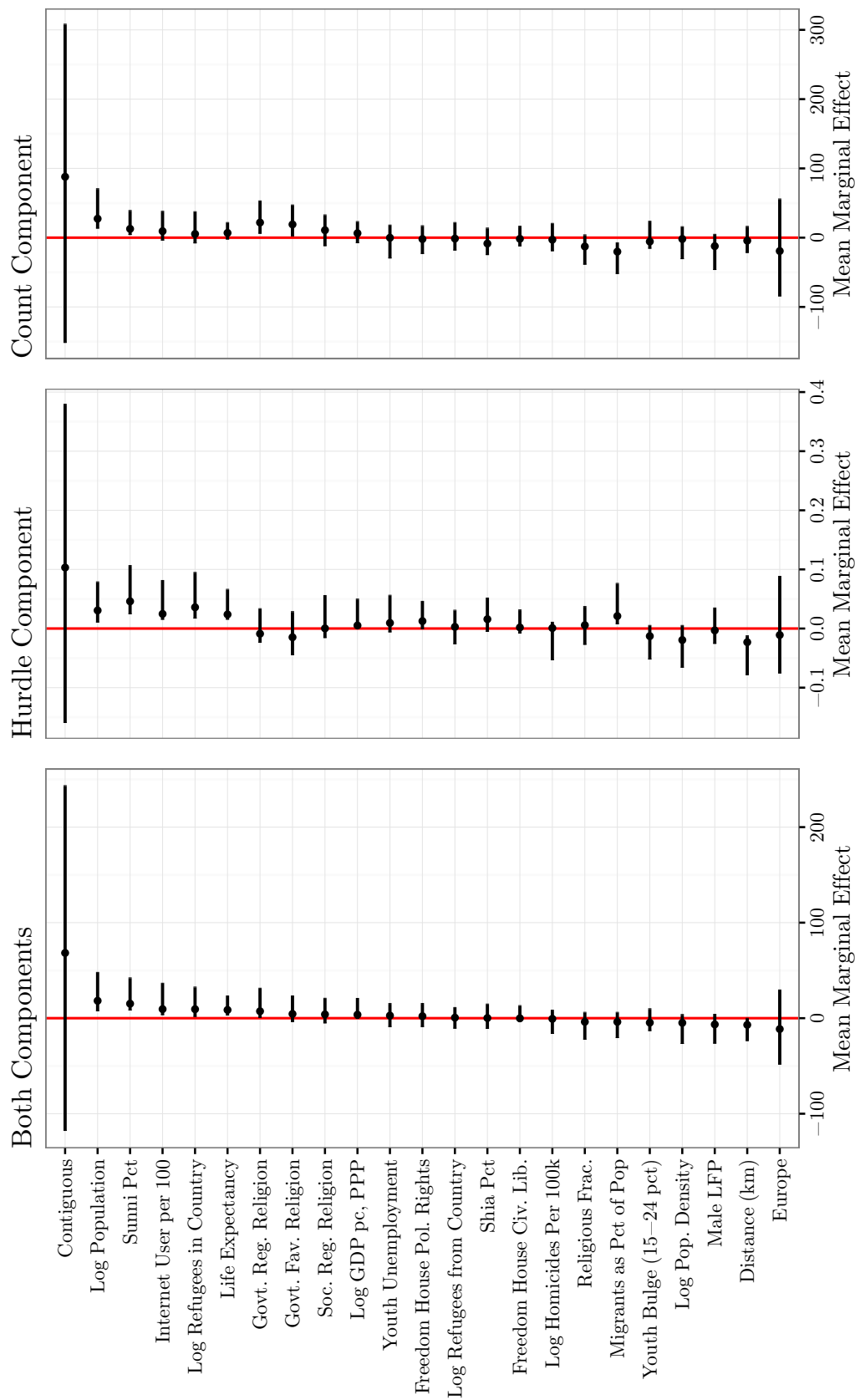


Figure 2.3: Sample-Average Pointwise Marginal Effects

Notes: Outcome variable is foreign fighter supply for the first column, probability of sending any fighters in the hurdle logit component in the second column, and mean of the truncated negative binomial component in the third column. As a result, note that the scaling of the x axes varies.

The leftmost panel of Figure 2.3 contains the estimated sample-average pointwise marginal effects for the combined model. Consistent with our usage throughout the paper, effects here mean the marginal effects our model produces, not causal effects. Furthermore, our features are scaled to have standard deviation of one so effects can be interpreted in standardized units. (Dichotomous variables for Europe and Contiguous are not pointwise marginal effects, but first differences between the two values.) Most effects are near zero, showing that there is little predictive power for many of the features. Nonetheless, the bootstrapped 95 percentile intervals around seven of the features do not include 0.

Substantively, it appears that the two strongest predictors are population, and the percent of the country’s population that is Sunni.⁴⁶ The effect of population is not surprising, as a positive (unconditional) probability of any individual becoming a foreign fighter implies that larger countries should have a higher supply. Similarly, Sunni population share intuitively predicts participation in the conflict, as it captures the strength of identification with the opposition; virtually all groups fighting the Assad regime are Sunni. Thus, more Sunnis in a given country means that more people feel they have a stake in the conflict and recruiters can more easily manipulate the salience of Sunnis’ religious identity to construct a sense of obligation to fight. Also unsurprising is the strong negative effect of distance to the conflict; increased travel costs should deter potential fighters at the margin, while dissociation from the conflict should increase with distance. Another noteworthy effect is that on life expectancy, which positively predicts foreign fighter supply. Although this seems troubling, it is in line with popular conceptions of supplier countries as developed, as well as the findings of Hewitt and Kelley-Moore (2009) that more developed countries supplied more foreign fighters during Iraq’s insurgency. Also consistent with conventional wisdom is the positive effect of internet usage; foreign fighters can be recruited online (Hegghammer, 2011), thus countries with deeper internet penetration allow recruiters to cast their net more widely. There also is a positive relationship between government regulation of religion and foreign fighter supply, something we discuss further below. Finally, a puzzling finding is the positive

⁴⁶The point estimate for contiguity is quite large, although the low number of foreign fighters from Iran, a contiguous state to Iraq, is a source of the large heterogeneity in these marginal effects.

effect of refugee presence. Below we show that this effect originates in European countries, hence we defer its interpretation for now.

More can be learned by disaggregating the effects into the hurdle and count components of our model. The central panel of Figure 2.3 contains the estimated sample-average pointwise marginal effects of standardized variables in the hurdle component. These effects can be interpreted as changes in the probability that a country passes the hurdle of no supply and sends some foreign fighters. The results are quite similar to the left panel, indicating that our features are better at predicting whether or not any foreign fighters are supplied (binary component), rather than how many are supplied (count component). Six of the seven predictors that are distinguishable from zero in the combined model remain so here, and their ranking in terms of substantive significance is also similar. Sunni population share, population, and refugee presence are still the three strongest predictors, again followed by internet penetration, distance to the conflict, and life expectancy. Government regulation of religion is not distinguishable from zero in this component. In addition, there are two new results worth noting. First, GDP per capita has a positive effect. This is consistent with the positive effect of life expectancy, popular conceptions regarding foreign fighters as originating in the developed world, and prior findings on Iraq (Hewitt and Kelley-Moore, 2009). Second, migrant population share also has a positive effect. Yet, we hesitate to interpret this, as the effect is reversed in the count component.

The rightmost panel contains the estimated sample-average pointwise marginal effects of standardized variables in the count component. Only five features are clearly driving prediction of the count of foreign fighter supply. As in the combined model and hurdle component, population and Sunni population share have a positive effect distinguishable from zero. However, in contrast to the hurdle component, the effect of migrant population share is negative. Clearly, the opposite direction of the migrant effect in the two components accounts for its null effect in the combined model. Another difference from previous results is that government regulation of religion and government favoritism of religion are positive predictors.⁴⁷ This means that, among supplier countries, those that do not respect freedom

⁴⁷These indices reflect how much the government respects freedom of religion, and whether it funds and

of religion and/or fund a particular religion send more foreign fighters. Further inspection of these effects reveals that they are due to Sunni-majority countries, hence we interpret them later.

2.8.3 Effect Heterogeneity

Our method allows for a rich exploration of marginal effect heterogeneity, by providing pointwise estimates of marginal effects for each feature. To demonstrate some of KRHNB’s strength in learning from the data, Figure 2.4 plots the distribution of pointwise marginal effects for four features. Most of the effects are near zero, indicating that much of the estimated CEF is flat across the support of the features. However, a clear difference can be seen between a feature that appears to have little to no relationship to foreign fighter supply (homicides, bottom left), and one that has a systematic relationship (Sunni population share, bottom right). Heterogeneity in the pointwise marginal effects can itself provide a lot of information about certain features of interest. Furthermore, by plotting or regressing the pointwise marginal effects on the features, we can see whether there appear to be interaction effects, or non-linear relationships (Hainmueller and Hazlett, 2014).

In Figure 2.5, we demonstrate how the pointwise marginal effects differ for countries within and outside of Europe. While internet usage has uniformly positive effects in Europe, in the rest of the world its marginal effect is clustered around zero. This could suggest that radicalization and recruiting happens mostly online in European countries, but through non-digital means in other countries (e.g. personal networks, mosques). The plot can also shed light on the puzzling result on refugee presence, as it suggests that refugee presence is associated with larger foreign fighter supply only in Western countries. A possible interpretation is that refugees’ predicament – persecution, poverty, loss of family members – exacerbates feelings of injustice among citizens of the host country that are already at high risk of becoming foreign fighters (e.g. young Sunni unemployed males in urban centers).⁴⁸

supports one religion in particular (Grim and Finke, 2006).

⁴⁸We believe it is unlikely that the positive effect of refugee presence means that refugees are more likely to fight in Syria, since there is no anecdotal evidence of refugees from Western countries becoming foreign

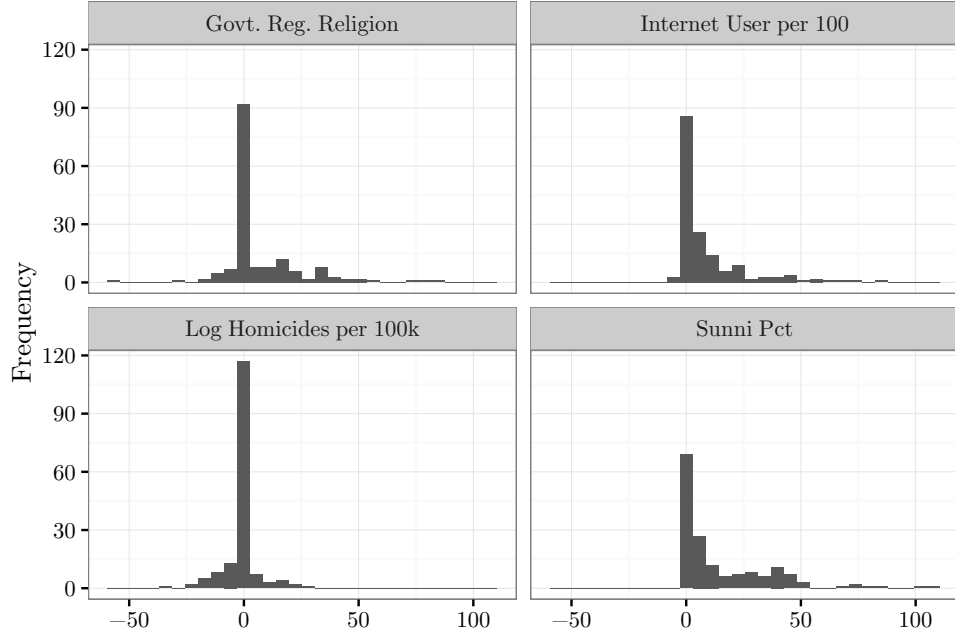


Figure 2.4: Distribution of Pointwise Marginal Effects

The latter might view their participation in the Syrian conflict as a chance to address the conditions that resulted in refugees' predicament. Radicalization might be especially strong for citizens of the host country with a shared religion or ethnicity with the refugees, the more so if the refugees fled countries with similar conditions to those that sparked the Syrian conflict. This applies to refugees from a number of countries that repressed Islam(ists) – for example, Algeria, Bosnia, Egypt, Iraq, Libya, Russia, and Tunisia – much like the Assad regime did in Syria.

This discussion provides an opportunity to return to the effect of government regulation of religion. As previewed above, intervention in religious affairs has a positive effect on foreign fighter supply in the model's count component. Figure 2.6 reveals an interesting heterogeneity in this effect, by plotting the pointwise marginal effect of government regulation of religion by countries' Sunni population share with a simple LOWESS smoother.⁴⁹ As the Sunni share of the population increases the effect of government regulation of religion

fighters. We also note that our refugee measure predates the large influx of refugees from Syria into Europe; this rules-out the explanation that Syrians are seeking refuge in Europe, only to return to Syria as fighters.

⁴⁹To see the full picture of possible interaction effects, see Appendix Figure 2.7 which presents an overview of heterogeneous effects.

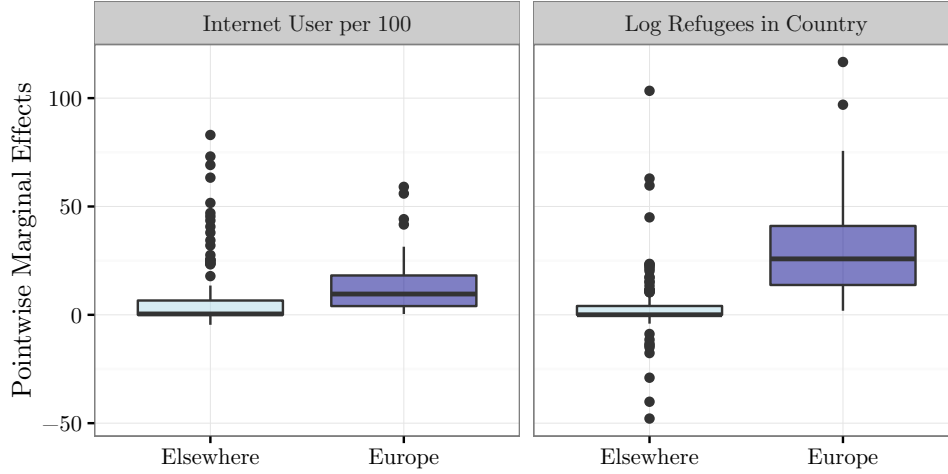


Figure 2.5: Effects Inside and Outside Europe

becomes stronger and more precise. That is, only in Sunni-majority supplier countries does religious intervention consistently predict a larger supply of foreign fighters. Given that virtually all foreign fighters are Sunni, this is consistent with a view of illiberal religious policies as radicalizing the population in Sunni-majority countries. This radicalization, in turn, might map onto foreign fighter supply through two possible mechanisms. One is that repressing (Sunni) Islam in countries where it is popular can create deep-seated grievances. Since the Assad regime also fought Islam, repressed Sunnis from the Muslim world might see Syria as a battleground for addressing their grievances. This could be the case in Tunisia, an overwhelmingly Sunni country that repressed Islamists until 2011, and has by far the largest per capita supply of foreign fighters. The second mechanism is that government repression of some religions or sects to benefit the official state religion might lead to exporting this attitude abroad. Thus, Sunnis from countries that enforce the dominance of Sunni Islam might consider it their duty to defend that religion wherever it is threatened, as in Syria. This could be the case of Saudi Arabia, which was founded on the radical Wahhabi (Sunni) interpretation of Islam, and has the second largest supply of foreign fighters.

These possible explanations are raised by our model because it is able to uncover non-linear and interactive relationships without relying on the a priori specification of complex functional forms and without overfitting. Therefore, our model provides fodder for future work that should causally identify the mechanisms that may lead from government interven-

tion in religion to the generation of foreign fighters.

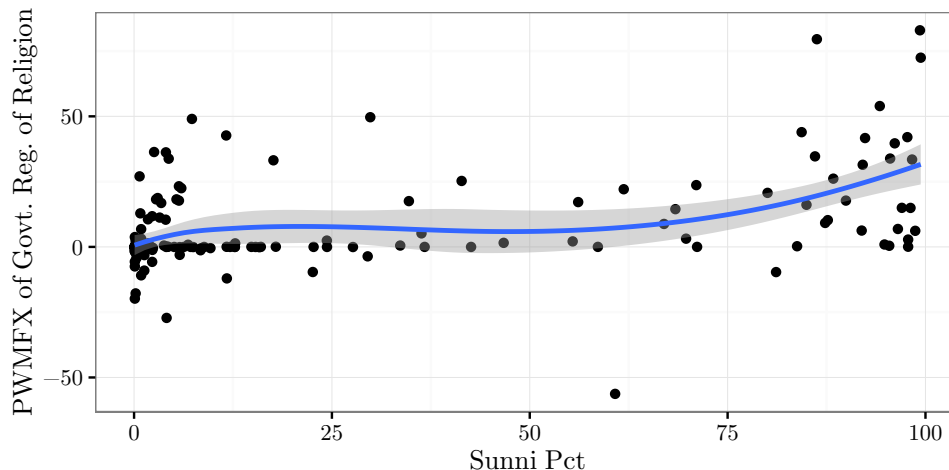


Figure 2.6: Effects of Government Regulation of Religion by Sunni Population Share

Overall, it seems that, while structural features dominate the hurdle component (distance, population, Sunni population share, life expectancy, GDP), policy variables mostly drive variation in the count component (government regulation of religion, government favoritism of a religion, migrant population share). Therefore, a tentative conclusion is that policy-makers cannot do much to prevent a foreign fighter network from emerging in their country, but, once it emerges, they can affect its size. Nevertheless, inspecting the effect of our features on the combined model, it is clear that structural features dominate in explaining foreign fighter supply, thereby limiting room for policy intervention.

2.9 Conclusion

“The best way to reduce a foreign fighter returnee problem is to never have them go in the first place” (Byman, 2015, p. 15). Our analysis helps identify country features that are associated with higher foreign fighter supply. Substantively, our results suggest that countries’ structural features play a larger role in shaping supply than their policies — four of the strongest and most accurate predictors of supply are population, the share of the population that is Sunni, distance to Syria, and life expectancy. Some of these features cannot be altered by policy-makers, while others can, albeit at a slow pace and high cost. Moreover, with

the exception of internet usage, all of the features that do respond to government policies – refugee intake, regulation of religion, and favoritism of a particular religion – are central to some countries’ political culture. For example, French voters consistently prefer liberal policies towards refugees, while most Saudi citizens are likely to value state protection of Sunni Islam, and regulation of other religions. Therefore, policy-makers in such countries will hesitate to change these policies, unless the cost of doing so is less than that imposed by their foreign fighter supply. This might be unlikely, even for countries facing a high supply, like France and Saudi Arabia. In short, to the extent that our research design allows us to offer any policy recommendations, it is unclear whether policy-makers can feasibly curtail foreign fighter supply.

Our approach is hindered by three limitations. On the empirical side, there is the inherent difficulty of measuring country-level foreign fighter supply. Fighters do not always publicize their participation in a conflict, not least because of fears of legal repercussions upon returning home. In addition, due to the nascent state of the literature, we lack a compass in our search for meaningful predictors to include in our specifications. Although we have carefully collected data on an array of economic, political, social, and demographic country features, there are many variables we have excluded from our specifications that may matter to foreign fighter supply. Future theoretical work will hopefully yield insights that can update our search for more informative predictors.

Our second limitation is that our algorithm is computationally expensive. This is owed to three factors. First, our complex optimization problem (Equation 2.12) involves a large number of parameters ($2n + 3$), and requires evaluating derivatives of the Gamma function. Second, our process for selecting the regularization parameters $(\lambda_\psi, \lambda_\omega)$, involves a grid search and leave-one-out cross validation. Third, the difficulties of deriving an analytical estimate of uncertainty for our quantities of interest force us to resort to the bootstrap, which involves repeatedly refitting the model to resampled datasets. One of the authors is also working on lowering the dimensionality of a related problem to improve speed and provide estimates of uncertainty that avoid the bootstrap.

On the theoretical side, a limitation of our framework is that it ignores psychological,

ideological, and organization factors that figure prominently in the terrorism literature. This is inevitable, given that such factors are only observable at the individual- or group-level, while our analysis is at the country-level. In theory, one could collect data on confirmed fighters and match them to non-fighters, in an effort to identify individual-level predictors of their decision to go to Syria. Indeed, several recent papers based on interviews of a small sample of returning fighters focus on individual features, albeit through a purely descriptive qualitative approach (Stenersen, 2011; Weggemans, Bakker and Grol, 2014; Nilsson, 2015). Expanding this dataset and applying classification models, such as Random Forest, is a promising avenue for future empirical research.

In future work, we would like to generalize the model to allow for different parameterizations of the two components. For example, if we use the procedure to model individual choice, one may want to fit a probit to the binary component, motivated by a theory that the stochastic term is normally distributed. Similarly, if there is no over-dispersion in the truncated count, we may want to make efficiency gains by fitting a Poisson.⁵⁰ Yet another extension is to allow for different features to enter each component. This is appropriate for questions where the theory guiding model specification is more developed. Nonetheless, with flexible estimators it is likely better to include as many relevant features as possible into the two components.

Broader extensions of our method would involve applying it to different political science questions. KRHNB can be used to predict any count processes that might have a hurdle structure, ranging from the number of terrorist attacks perpetrated in each country (e.g. Burgoon (2006)) to the number of bills passed by female legislatures in Congress (e.g. Volden, Wiseman and Wittmer (2013)). Moreover, by abandoning the hurdle structure and applying kernel regularization to different classes of GLMs (e.g. binary choice), the range of political science processes that can be modeled becomes infinite. We hope that our study will motivate political methodologists to take up this task.

⁵⁰Zorn (1998) notes that in the hurdle model over dispersion can be largely accounted for by the binary component, and thus a Poisson model may suffice for the count component.

2.10 Appendix 1: Results Table and Heterogeneous Effects

2.10.1 Full Results

This section contains two tables that have the sample-average pointwise marginal effects for two different datasets. Table 2.3 contains the same information as Figure 2.3 and represents the results from our main specification where we impute missing values and divide foreign fighter supply between Israel and the West Bank. The first column has the joint effect on the outcome of the full hurdle negative binomial model, while the second and third columns represent the effect on the probability of any fighters in the hurdle and the mean of the negative binomial distribution in the count component. In brackets are the 2.5th and 97.5th percentiles of the 1000 bootstrapped sample-average pointwise marginal effects.

Table 2.4 contains the same results, but instead on a dataset where listwise deletion is used to remove observations with missing values. Furthermore, the West Bank and Gaza observation has been dropped rather than manually imputed. The results are very similar. The only substantive difference is that the percentile interval now includes 0 for the number of refugees in the country, although the effect is still strictly positive in the hurdle component.

2.10.2 Heterogeneous Effects

Because we are able to compute pointwise marginal effects, it is possible to explore these effects to see how they vary in different parts of the feature space. One way to do this is simply to plot the pointwise marginal effects with respect to one variable by some other variable to see an interaction effect. A more robust way to explore these effects is to regress the pointwise marginal effects of some feature and all of the features in the data. This way, we can uncover the conditional interaction or non-linearity of our features. Of course, the pointwise marginal effects can themselves be modeled using flexible models or visualized using partial residual plots.

A way to summarize all of these interaction effects is to regress the pointwise marginal effects for each feature on the full set of features. To be precise, we estimate a regression of

Table 2.3: Average Marginal Effects on Foreign Fighter Supply

Variable	Both Components	Hurdle	Count
Life Expectancy	8.64 [2.88, 22.87]	0.02 [0.02, 0.07]	6.87 [-2.59, 21.3]
Contiguous	68.3 [-117.95, 243]	0.1 [-0.16, 0.38]	87.89 [-151.81, 307.67]
Distance (km)	-6.99 [-23.87, -0.51]	-0.02 [-0.08, -0.01]	-4.36 [-21.98, 15.74]
Internet User per 100	9.48 [3.23, 36.21]	0.02 [0.01, 0.08]	9.37 [-4.11, 37.57]
Migrants as Pct of Pop	-3.84 [-20.76, 5.71]	0.02 [0.01, 0.08]	-20.19 [-52.31, -7.87]
Male LFP	-6.52 [-26.61, 3.66]	0 [-0.03, 0.03]	-12.23 [-46.65, 4.27]
Youth Unemployment	2.73 [-9.21, 14.96]	0.01 [-0.01, 0.06]	-0.12 [-29.94, 17.65]
Freedom House Civ. Lib.	-0.12 [-4.01, 12.74]	0 [-0.01, 0.03]	-1.66 [-12.41, 15.97]
Freedom House Pol. Rights	2.06 [-9.17, 15]	0.01 [0, 0.05]	-2 [-23.41, 16.69]
Sunni Pct	15.17 [8.06, 41.84]	0.05 [0.02, 0.11]	12.7 [3.92, 38.71]
Shia Pct	0.19 [-11.1, 14.41]	0.02 [-0.01, 0.05]	-8.48 [-25.07, 13.32]
Govt. Reg. Religion	7.3 [0.34, 30.97]	-0.01 [-0.02, 0.03]	21.84 [5.75, 52.54]
Govt. Fav. Religion	4.45 [-3.94, 22.98]	-0.01 [-0.05, 0.03]	19.11 [2.04, 46.65]
Soc. Reg. Religion	3.96 [-5.36, 20.46]	0 [-0.02, 0.06]	10.8 [-12.12, 32.33]
Youth Bulge (15-24 pct)	-4.66 [-13.42, 9.51]	-0.01 [-0.05, 0]	-5.6 [-15.88, 23.18]
Religious Frac.	-3.8 [-22.42, 5.73]	0.01 [-0.03, 0.04]	-12.71 [-38.91, 3.54]
Log Homicides Per 100k	-0.69 [-16.45, 7.97]	0 [-0.05, 0.01]	-2.9 [-19.46, 20.01]
Log Pop. Density	-4.85 [-26.92, 3.47]	-0.02 [-0.07, 0]	-1.93 [-30.98, 15.26]
Log Population	18.24 [7.31, 47.54]	0.03 [0.01, 0.08]	27.37 [13.14, 70.14]
Log GDP pc, PPP	3.67 [-0.75, 20.31]	0.01 [0, 0.05]	6.38 [-7.99, 22.69]
Log Refugees in Country	9.42 [1.4, 32.18]	0.04 [0.02, 0.09]	5.35 [-8.18, 36.92]
Log Refugees from Country	0.55 [-10.9, 10.77]	0 [-0.03, 0.03]	-1.34 [-18.47, 21.29]
Europe	-11.37 [-48.52, 29.19]	-0.01 [-0.08, 0.09]	-19.29 [-84.71, 55.31]
N	163		

Notes: The point estimates are the sample mean of pointwise marginal effects over the full data. See text for details.

Table 2.4: Average Marginal Effects on Foreign Fighter Supply in Listwise-Deleted Dataset

Variable	Both Components	Hurdle	Count
Life Expectancy	15.24 [2.71, 35.62]	0.04 [0.01, 0.06]	12.01 [-9.94, 33.48]
Contiguous	78.05 [-91.57, 266.88]	0.14 [-0.15, 0.34]	82.55 [-138.82, 304.3]
Distance (km)	-8.6 [-27.97, 4.96]	-0.04 [-0.06, -0.01]	0.75 [-25.44, 35.02]
Internet User per 100	21.87 [4.38, 55.05]	0.04 [0.02, 0.07]	25.45 [-4.6, 65.7]
Migrants as Pct of Pop	-8.87 [-23.49, 8.55]	0.03 [0.01, 0.07]	-31.1 [-67.27, -2.6]
Male LFP	-16.44 [-44.22, 0.94]	0 [-0.02, 0.03]	-30.12 [-83.99, -1.16]
Youth Unemployment	1.53 [-18.73, 18.49]	0.01 [-0.02, 0.05]	-6.74 [-46.98, 25.33]
Freedom House Civ. Lib.	-2.27 [-7.19, 16.65]	0.01 [-0.01, 0.03]	-8.71 [-20.98, 18.74]
Freedom House Pol. Rights	0.65 [-17.51, 15.42]	0.03 [0, 0.04]	-11.18 [-41.82, 15.94]
Sunni Pct	26.57 [7.78, 52.81]	0.07 [0.03, 0.09]	23.85 [0.43, 54.28]
Shia Pct	-2.72 [-13.12, 15.71]	0.01 [-0.01, 0.04]	-12.77 [-31.06, 19.03]
Govt. Reg. Religion	14.75 [1.06, 43.85]	-0.01 [-0.02, 0.03]	34.68 [3.94, 80.33]
Govt. Fav. Religion	5.75 [-3.81, 31.51]	-0.02 [-0.03, 0.03]	23.54 [-5.12, 67.13]
Soc. Reg. Religion	8.3 [-4.82, 28.08]	0.01 [-0.01, 0.05]	13.15 [-13.39, 45.4]
Youth Bulge (15-24 pct)	-4.01 [-16.44, 15.04]	-0.02 [-0.04, 0.01]	-4.19 [-20.9, 34.81]
Religious Frac.	-6.56 [-32.4, 6.5]	0 [-0.03, 0.03]	-16.27 [-58.49, 6.34]
Log Homicides Per 100k	2.82 [-17.66, 16.3]	0 [-0.05, 0.01]	3.57 [-23.4, 40.09]
Log Pop. Density	-7.92 [-33.61, 6.05]	-0.02 [-0.05, 0.01]	-5.83 [-46.81, 20.93]
Log Population	31.27 [7.58, 71.29]	0.05 [0, 0.06]	46.34 [14.38, 109.16]
Log GDP pc, PPP	7.48 [-0.77, 30.97]	0.01 [0.01, 0.05]	11.16 [-11.26, 37.61]
Log Refugees in Country	13.62 [-0.21, 47.45]	0.04 [0.01, 0.08]	15.12 [-9.63, 63.35]
Log Refugees from Country	-0.22 [-11.83, 19.31]	-0.01 [-0.04, 0.01]	5.31 [-10.72, 47.96]
Europe	-11.1 [-71, 46.84]	0.01 [-0.08, 0.09]	-25.47 [-135.44, 97.85]
N	147		

Notes: The point estimates are the sample mean of pointwise marginal effects over the full data. See text for details.

the following form

$$\frac{\partial \mathbb{E}[y_i | \mathbf{x}_i]}{\partial x_i^{(j)}} = \boldsymbol{\gamma}^\top \mathbf{x}_i,$$

where $x_i^{(j)}$ is the i th observation of the j th feature. We do this for all features j . The coefficients in $\boldsymbol{\gamma}$ are analogous to interaction terms ($\gamma_i \forall i \neq j$) or quadratic terms (γ_j). Thus if we are predicting the pointwise marginal effects of life expectancy on foreign fighter supply and we estimate a large positive γ_i on, for example, the percent of the population that is Sunni, we then conclude that the marginal effect of life expectancy is much greater in heavily Sunni countries. Figure 2.7 contains the results of these regressions for all features, demonstrating how the variables interact. Large blue dots indicate the two have a positive interaction effect while large red dots indicate the two have a negative interaction effect.

2.11 Appendix 2: Target Function Derivation

2.11.1 Appendix 2.1: Sample Log-Likelihood

The likelihood for observation i in Equation 2.5 can be written more explicitly with respect to the densities in Equations 2.2, 2.3, and 2.4. Where $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}^\top, \beta_0, \boldsymbol{\beta}^\top, \zeta)^\top$,

$$L_i(\boldsymbol{\theta} | y_i, \mathbf{x}_i) = [p_0(y_i = 0)]^{1-d_i} \left[\frac{p_1(y_i)}{1 - p_1(y_i = 0)} (1 - p_0(y_i = 0)) \right]^{d_i} \quad ; \quad d_i = \begin{cases} 0 & \text{if } y_i = 0 \\ 1 & \text{if } y_i \geq 1 \end{cases}$$

$$= \left[\frac{1}{1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})} \right]^{1-d_i} \quad (2.14)$$

$$\times \left[\frac{\Gamma(\zeta + y_i) \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \left(\frac{\exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^{y_i}}{\Gamma(1 + y_i) \Gamma(\zeta) \left(1 - \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \right)} \frac{\exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})}{1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})} \right]^{d_i} \quad (2.15)$$

Next we derive the sample log-likelihood in Equation 2.6:

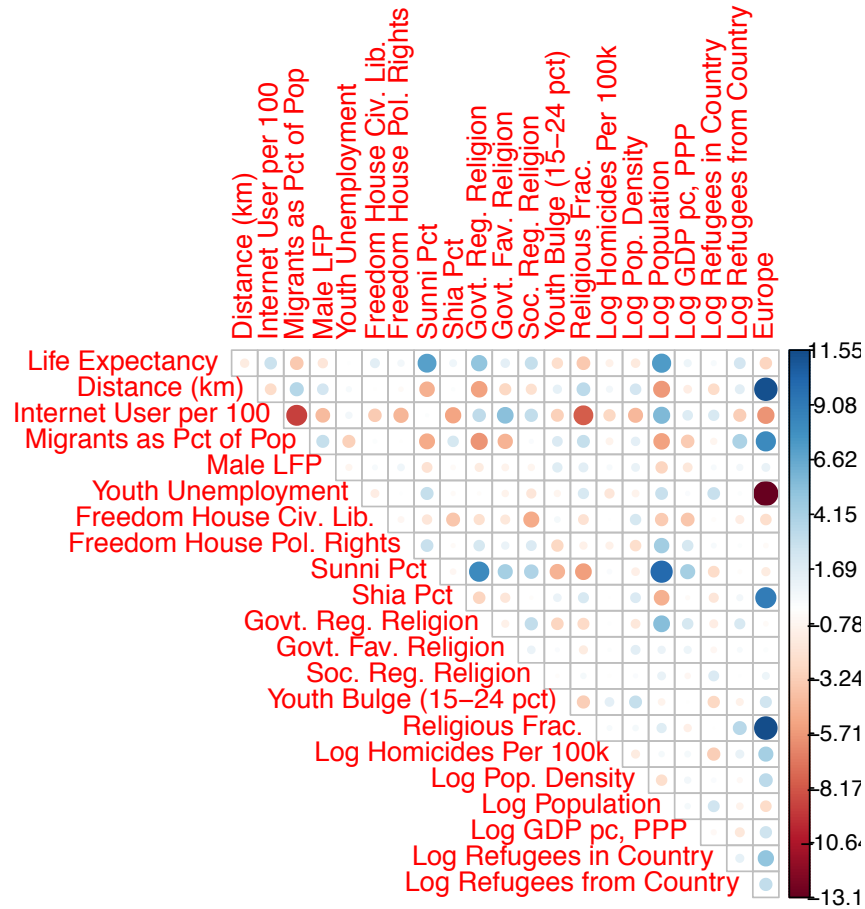


Figure 2.7: Interaction Effects

Notes: Each cell is the coefficient from of a regression of one set of pointwise marginal effects on the rest of the data. A blue cell means that the two have a positive interaction while a red cell means the two have a negative interaction.

$$\begin{aligned}
\ell_N(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) &= \sum_{i=1}^N (1 - d_i) [\log 1 - \log (1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}))] + d_i \left[\log \Gamma(\zeta + y_i) \right. \\
&\quad + \zeta \log \zeta - \zeta \log (\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})) + y_i(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) \\
&\quad - y_i \log (\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})) - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) \\
&\quad - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \right) + (\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}) \\
&\quad \left. - \log (1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})) \right] \\
&= \sum_{i=1}^N -\log (1 + \exp(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha})) + d_i \left[\log \Gamma(\zeta + y_i) + \zeta \log \zeta \right. \\
&\quad - (\zeta + y_i) \log (\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})) + y_i(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) + (\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}) \\
&\quad \left. - \log \Gamma(1 + y_i) - \log \Gamma(\zeta) - \log \left(1 - \left(\frac{\zeta}{\zeta + \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})} \right)^\zeta \right) \right] \tag{2.16}
\end{aligned}$$

2.11.2 Appendix 2.2: Using Mercer's Theorem

Given Equation 2.8, we solve the First Order Condition (FOC) for each parameter vector in order to demonstrate our ability to use Mercer's Theorem to reduce our problem from a potentially infinite-dimensional one to a more tractable function. Where $\boldsymbol{\theta}_\phi = (\psi_0, \boldsymbol{\psi}^\top, \omega_0, \boldsymbol{\omega}^\top, \zeta)^\top$,

$$\begin{aligned}
\frac{\partial R_N(\boldsymbol{\theta}_\phi, \lambda_\psi, \lambda_\omega | \mathbf{Y}, \mathbf{X})}{\partial \boldsymbol{\psi}} &= 0 \\
0 &= - \sum_{i=1}^N \left(- \frac{\boldsymbol{\phi}(\mathbf{x}_i)^\top \exp(\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi})}{1 + \exp(\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi})} + d_i \boldsymbol{\phi}(\mathbf{x}_i)^\top \right) + 2\lambda_\psi \boldsymbol{\psi} \\
\boldsymbol{\psi} &= \sum_{i=1}^N \left\{ \frac{1}{2\lambda_\psi} \left(- \frac{\exp(\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi})}{1 + \exp(\psi_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\psi})} + d_i \right) \right\} \boldsymbol{\phi}(\mathbf{x}_i) \tag{2.17}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial R_N(\boldsymbol{\theta}_\phi, \lambda_\psi, \lambda_\omega | \mathbf{Y}, \mathbf{X})}{\partial \boldsymbol{\omega}} &= 0 \\
0 &= - \sum_{i=1}^N d_i \left[- \frac{(\zeta + y_i) \boldsymbol{\phi}(\mathbf{x}_i)^\top \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} + y_i \boldsymbol{\phi}(\mathbf{x}_i)^\top \right. \\
&\quad \left. - \frac{-\zeta \left(\frac{\zeta}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} \right)^{\zeta-1} \left(\frac{-\zeta \boldsymbol{\phi}(\mathbf{x}_i)^\top \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})}{(\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega}))^2} \right)}{1 - \left(\frac{\zeta}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} \right)^\zeta} \right] \\
&\quad + 2\lambda_\omega \boldsymbol{\omega} \\
\boldsymbol{\omega} &= \sum_{i=1}^N \left\{ \frac{1}{2\lambda_\omega} d_i \left[- \frac{(\zeta + y_i) \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} + y_i \right. \right. \\
&\quad \left. \left. - \frac{\zeta \left(\frac{\zeta}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} \right)^{\zeta-1} \left(\frac{\zeta \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})}{(\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega}))^2} \right)}{1 - \left(\frac{\zeta}{\zeta + \exp(\omega_0 + \boldsymbol{\phi}(\mathbf{x}_i)^\top \boldsymbol{\omega})} \right)^\zeta} \right] \right\} \boldsymbol{\phi}(\mathbf{x}_i)
\end{aligned} \tag{2.18}$$

Note that in both FOCs, the terms inside $\{\}$ form a scalar, which we label c_i^ψ and c_i^ω . As such, we rewrite our solutions for $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ like we did in the main body of the paper:

$$\boldsymbol{\psi}^* = \sum_{i=1}^N c_i^\psi \boldsymbol{\phi}(\mathbf{x}_i) \tag{2.19}$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^N c_i^\omega \boldsymbol{\phi}(\mathbf{x}_i) \tag{2.20}$$

We use these solutions to rewrite Equation 2.8 as Equation 2.12, our final target function.

CHAPTER 3

Modeling Inter-Rebel Group Conflict with Network Analysis: The Case of Lebanon’s Civil War

3.1 Introduction

It is well-established that civil wars are more frequent, longer, more violent, and economically costlier than international wars (Blattman and Miguel, 2010). To end any conflict, we must first understand the complex dynamics between the major actors involved. Aside from the state, this often includes multiple opposition groups with heterogeneous goals, resources, and tactics. However, the literature usually treats civil war as conflict between two *unitary* actors—state and opposition (Collier and Hoeffler, 1998; Fearon and Laitin, 2003). This framework goes a long way towards explaining some conflicts, and it provides a useful starting point for thinking about all conflicts. Yet, as our understanding of civil war grows, data availability and computing power increase, and conflicts become more multidimensional, it becomes both possible and necessary to advance the empirical study of civil war.

One topic that has recently attracted the attention of civil war scholars is that on violence *between* rebel groups. Reflecting the state of the literature, Pearlman and Cunningham (2012) note that “[the] norm in more recent civil conflicts is not coherent antagonists as much as shifting coalitions of groups with malleable allegiances and at times divergent interests, only some of whom actually engage in violence at any given point in time” (p. 4). This evolving research agenda explores two main questions: how is conflict among rebel groups structured, and what explains the presence or absence of hostilities among different rebel

groups. Applying this agenda to the ongoing Syrian conflict, one might ask: why is the Islamic State in war with other rebel groups fighting the Syrian government, like the al-Nusra Front? Why, in turn, do both of these groups fight the Free Syrian Army? These questions are important, as anecdotal evidence suggests that the Syrian war is not unique (see below). Additionally, as power becomes more balanced within the international system, we might expect internationalized civil wars like the Syrian one to involve more rebel factions and inter-rebel hostilities, reflecting the diverging agendas of each faction’s foreign sponsors (Jenne and Popovic, 2016).

Setting-out to uncover the correlates of inter-rebel violence, recent studies sketch the profile of the groups most likely to attack other rebels: they are significantly stronger or weaker than the average group, face greater competition from groups sharing the same ethnic identity, control valuable resources or are in conflict over such resources, are located in territory beyond the state’s reach, face a weak state, or are in negotiation with the state (Cunningham, Bakke and Seymour, 2012; Eck, 2010; Fjelde and Nilsson, 2012; Pischedda, 2015). Undoubtedly, these findings advance our understanding of inter-rebel conflict. That said, the research design employed by these studies suffers from a key limitation: it ignores the *relational dependence* in conflict data. Given that warring groups’ decisions are interdependent, studying the determinants of inter-rebel hostilities requires a statistical approach that is sensitive to the *joint, non-independently distributed* nature of hostilities (Cranmer and Desmarais, 2011).

In this study, I follow such an approach. Namely, I treat rebel groups fighting the same civil war as nodes in a network, and model hostilities between rebels as directed edges.¹ A network approach has three advantages over competing approaches. First, it allows us to summarize useful information on the *structure* of inter-rebel group violence, through graphs, and descriptive statistics on dyad-level and network-level variables. Though the quantitative information that network graphs and statistics provide are not intended to replace qualitative

¹Where it does not create confusion, I use the terms node, vertex, rebel group, faction, and militia interchangeably. The same applies to the terms edges, ties, and hostilities, and I also interchange the terms network and graph.

research on conflict, they provide a more efficient overview of the dynamics between groups. More importantly, it is difficult for qualitative studies to provide equal and impartial coverage of opposing factions in a conflict. Indeed, historical accounts are often labeled as biased, a criticism that statistics are less vulnerable to.

The second advantage a network approach enjoys is that it allows us to unbiasedly and consistently estimate the effect of node-, dyad-, and higher-level variables on hostilities between rebel groups.² This is done through the Exponential-Family Random Graph Model (ERGM) (Besag, 1975). Crucially, the standard approach for modeling conflict, regression on dyadic data, with each observation modeling the likelihood of conflict between groups i and j , is problematic. Even if our goal is only to estimate the effect of node-level covariates on the outcome variable (e.g. how do group i 's resources affect its likelihood of attacking group j), the inability of classical regression models to *include* triad-level effects (e.g. group i more likely to attack group j , all else equal, if group j allied to an enemy of group i) or higher-level effects makes regression estimates biased and their standard errors inconsistent. Moreover, if we are want to *estimate* the effect of triad-level or higher-level terms on the outcome variable, standard regression on dyadic data is inapplicable. The ERGM offers an alternative to the standard approach that overcomes both of its limitations.

The third advantage that a network approach to conflict offers is that it can effectively uncover clusters among groups. Clusters in a conflict network might emerge due to homophily based on observed attributes (e.g. shared political, regional, religious, ethnic, or other identities), transitivity (e.g. the enemy of my enemy is my friend, the enemy of my friend is my enemy), or coordination dynamics (e.g. attack the group most groups are attacking, ally with the group most groups are allying). Overall, we might expect clusters in a network to form around alliances or, alternatively, sub-conflicts within the larger conflict. Though detecting patterns like clusters is often done through visual inspection, that approach is imperfect and misleading, as confirmation bias causes researchers to project clusters onto the network that

²Node-level variables measure features of rebel groups, dyad-level variables capture features of pairs of rebel groups, and higher-level variables describe features of network terms like triads (triplets of rebel groups) and components (sub-networks whose nodes are all connected to each other), etc.

are consistent with their prior beliefs or theory. A more precise and data-driven alternative is offered by the Latent Position Cluster Model (LPCM). The latter places nodes on a latent space, based on how similar nodes are in terms of their covariates, then it detects the number of clusters and assigns nodes to them.

The conflict I apply these tools to is the Lebanese Civil War; in particular, the years 1980 – 1991. This choice is made for two reasons, each allowing this study to make a separate contribution. First, though the Lebanese conflict lasted long, claimed many lives, shaped future regional politics, and drew-in many countries, consensus is lacking on many of the conflict’s dimensions. Through examining one dimension – hostilities between rebel groups – I aim to shed light on the conflict’s complex dynamics. Specifically, I contribute network graphs, descriptive statistics, and a clustering model illuminating the structure of the network of inter-rebel violence. This quantitative information can be used to complement the rich qualitative accounts of the conflict. In addition, I contribute predictive models (ERGMs) of hostilities among rebel groups that are relatively accurate. Although they are based on observational data, in the future these models can be trained for forecasting purposes, in order to yield early warnings of rebel hostilities. In turn, accurate conflict forecasting can allow the international community to intervene – via diplomacy or force – so as to minimize further violence.

The second distinctive feature of the Lebanese Civil War is the number of groups involved, their cross-cutting religious, ethnic, and political identities, and the variation in their capacity, objectives, and strategies. This is convenient from a statistical perspective: the presence of multiple groups enlarges the sample, thereby allowing for consistent and efficient estimates of quantities of interest. Similarly, the frequent hostilities between groups with different features makes for a sufficiently dense network and covariates with common support, thus enabling identification of covariate effects.³ As such, the Lebanese Civil War is an appropriate testing-ground for introducing network models to *civil* conflict between rebels.

³If hostilities are rare, thereby producing a sparse conflict network, it becomes more likely that hostile groups exhibit different covariate values from non-hostile groups, particularly for binary covariates (no overlap). This makes the coefficients on these covariates non-identified or, at best, noisily estimated.

This points to the first contribution of this study – as, to the best of the author’s knowledge – all previous applications of network analysis are to *international* conflict between states.

Using data on a network of 22 rebel groups and their hostilities during the 1980 – 1991 period of the Lebanese Civil War, I showcase the strengths of the network approach to studying inter-rebel conflict. I begin by graphing the network and displaying descriptive statistics at the node-, dyad-, and network-level. These reveal patterns in line with historical accounts of the conflict: a relatively dense network of hostilities, high reciprocity in hostilities (i attacks $j \Leftrightarrow j$ attacks i), low transitivity in hostilities (i.e. the enemy of my enemy is my friend), significant infighting within sects, and the presence of 3 central groups (Amal, Fatah/PLO, South Lebanon Army) belonging to the three largest sects (Shia, Palestinians, Maronites). Then, I estimate a series of ERGMs, uncovering several correlations that speak to the literature. Like previous research, I find that groups that command support from the ethnic community they belong to, as well as groups that control valuable natural resources and/or territory, are, all else equal, more likely to initiate hostilities against other rebels. On the other hand, and contrary to some of the literature, I find that groups that are able to strike an agreement with the state are less likely to attack other groups. Furthermore, I make two novel findings: groups that use terrorist tactics attack other groups with a higher probability, while the opposite holds for groups using ethnic cleansing tactics. Finally, I estimate an LPCM, which uncovers 2 sub-conflicts in the network: a narrow cluster that includes the infighting among Palestinian groups and their Sunni allies and a broader cluster that also includes the hostilities between the two rival Shi’ite groups (Amal and Hezbollah). In sum, for the most part my findings confirm previous qualitative and quantitative research – comparative and case-specific – on inter-rebel violence. This suggests that a network approach to the study of civil conflict complements other approaches—by formalizing and quantifying their insights.

The rest of this study is structured as follows. Section 3.2.1 introduces the data, graphs the network, and presents descriptive statistics at the node, dyad, triad, and network level. Section 3.2.2 displays the results of the ERGMs predicting inter-rebel group hostilities. Section 3.2.3 presents the output of the LPCM. Section 3.3 discusses the significance of my

results vis-à-vis the literature and the history of the Lebanese Civil War. Section 3.4 summarizes and offers policy implications and directions for future research.

3.2 Analysis

This section introduces the data and presents the network graph, descriptive statistics, output from ERGMs, and output from the LPCM.

3.2.1 Data & Descriptive Statistics

The network I analyze is constructed using the Minorities at Risk Organizational Behavior (MAROB) dataset (Asal, Pate and Wilkenfeld, 2008). MAROB restricts its attention to the Middle East and North Africa in the period 1980 – 2004, and codes “the characteristics of those ethnopolitical organizations most likely to employ violence and terrorism in the pursuit of their perceived grievances” (Asal, Pate and Wilkenfeld, 2008, p. 1). Subsetting the observations for Lebanon from 1980 to 1991, I am able to capture all but 5 years of the Lebanese Civil War (1975 – 1979).

Because the non-state (rebel) actors involved are groups representing different ethnic, religious and political goals, they are all observed in MAROB. These 22 groups constitute the nodes in my network. The (directed) edges in the network are indicators of hostilities between groups, coded using MAROB variables on “inter-organization conflict” (Asal, Pate and Wilkenfeld, 2008, p. 30).⁴ Note that my edges are *binary* indicators of hostility by group i towards group j , not counts of hostilities. Similarly, there is no temporal dimension to the edges; they merely capture whether *at least one* hostility by group i towards group j took place between 1980 – 1991, not whether a hostility was observed each year.

The network is graphed in Figure 3.1.⁵ Four features are worth noting. First, the network

⁴These variables are INTERSEV1DES, INTERSEV2DES and INTERSEV3DES, which record the “organization with [the] highest level of inter-organizational conflict” (Asal, Pate and Wilkenfeld, 2008, p. 30-31).

⁵I cross-check the MAROB data with the UCDP Non-State Actors Dataset (NSA) (Sundberg, Eck and Kreutz, 2012) and the historical accounts in O’Ballance (1998). I find general agreement across the different sources regarding the pattern of hostilities. However, note that the NSA data only covers the last 3 years of

is neither overly sparse, nor dense. This is confirmed by the network’s density score: for two randomly chosen rebel groups i and j , there is an 8% chance that i attacked j at least once during the period in question.⁶ Though this figure might seem low, for a conflict network it is relatively dense—for example, the international conflict network’s density during 1990–2000 is 1%. One factor reducing the network’s density is that 4/22 nodes are isolates, i.e. they have no edges. Interestingly, 3/4 of isolates are Palestinian – the ethno-religious group with the largest number of factions in the conflict – but non-isolate Palestinian groups are relatively hostile (e.g. Fatah/PLO). This is consistent with perceptions of Palestinians as the most strategically diverse ethno-religious group in the War. Indeed, it will not surprise Lebanon scholars that Palestinians’ wide range of preferences and tactics maps into significant within-Palestinian variation in hostilities.

A second feature to note in Figure 3.1 is the high degree of mutuality in ties (if i attacks j , most likely j attacks i , and vice-versa). This is also confirmed by the network’s relatively high edgewise reciprocity score (fraction of ties that are mutual): 0.54.⁷ High mutuality/reciprocity is intuitive for a civil conflict network—especially when the state collapses, as in Lebanon. With few constraints on rebels’ strategies other than resources we should expect violence to be met with violence.⁸ This stands in contrast to international conflict, where domestic political constraints (laws, elections) and international political constraints (treaties, sanctions) limit states’ abilities to attack each other.

The third interesting feature of the network relates to hostilities within triads; in particular, whether hostilities are transitive (i.e. if i attacks j and j attacks k , how likely is it that i also attacks k). Transitivity is often recorded when edges denote cooperative behavior

the period I study, and thus a complete cross-check is unfeasible.

⁶The density score can be derived simply by dividing the number of edges (37) by the number of dyads (462); the latter is also the maximum feasible number of edges.

⁷Note that edgewise reciprocity is probably suppressed by the fact that some rebel groups in the network were actually eliminated during this period (e.g. National Liberal Party). If they were eliminated by a group attacking them for the first time and, thus, were unable to reciprocate, then the respective edge will necessarily be non-reciprocal.

⁸In civil wars where the state survives, we might imagine that the state limits rebel groups’ ability to reciprocate hostilities.

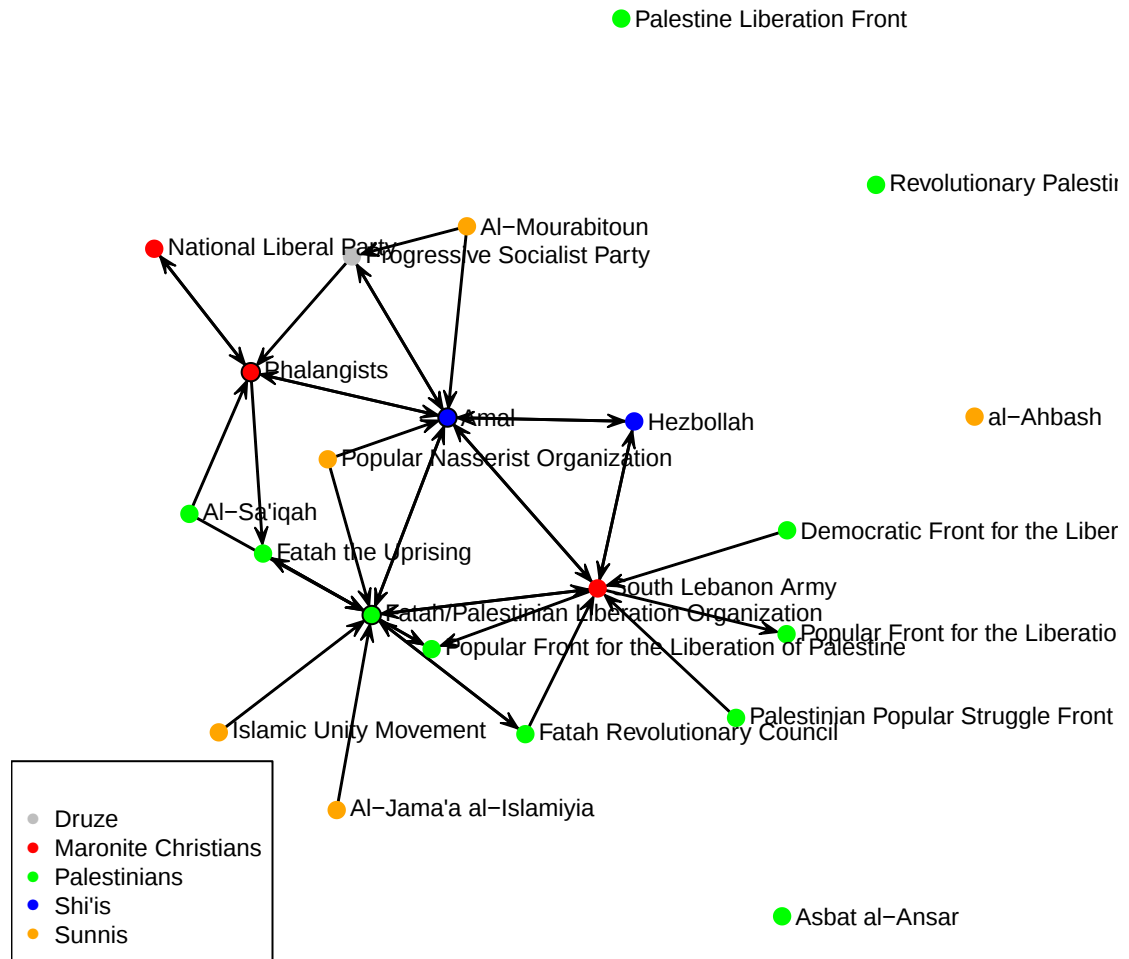


Figure 3.1: Inter-Rebel Hostilities in Lebanon's Civil War, 1980 – 1991

Notes: An edge from i to j represents at least one hostility directed by i to j during the period in question. Groups are labeled as in the MAROB data and might be labeled differently in other sources.

or positive preferences, as in business or friendship networks (Wasserman and Faust, 1994). However, in conflict networks we should expect edges to be intransitive, since i might gain from coalescing with k to defeat j —per the proverb “the enemy of my enemy is my friend”.

Alternatively, i might have no incentive to attack k , as i can free-ride on j 's hostilities towards k . In any case, all else equal, the strategic logic behind i attacking k just because j attacks k is weak. Perhaps for this reason, the Lebanese network has a transitivity score of just 0.23 (fraction of triads with transitive edges).

The fourth noteworthy feature in Figure 3.1 is the significant number of edges between nodes of the same color, representing infighting within ethno-religious groups. In particular, we see hostilities between several Palestinian groups, the only 2 Shi'ites groups, and 2/3 Maronite groups. Sectarian infighting is a well-established feature of the Lebanese Civil War, thus it is reassuring that the network graph depicts it. This feature also differentiates the Lebanese Civil War from current conflicts in the region, where alliances and hostilities follow sectarian lines. For example, in Iraq and Syria there is no infighting among Kurdish or Shi'ite groups, while there is limited infighting among Sunni groups in Syria (e.g. IS vs al-Nusra) and non-violent competition between Palestinian groups in the Palestinian territories (Hamas vs. Fatah) (Christian and Druze militias are no longer active in the region).

Table 3.1: Degree Summary Statistics

Degree	Minimum	Mean	Median	Maximum	σ	ρ (Id, Od)
Indegree	0	1.68	0.5	8	2.46	0.84
Outdegree	0	1.68	1.5	5	1.61	

Moving on to a more thorough analysis of the network's structure, we turn to the distribution of hostilities across groups. Figure 3.2 shows that both indegree (number of groups attacked by) and outdegree (number of groups attacked) follow a right-skewed distribution, with some nodes (e.g. the 4 isolates) attaining the minimum of 0, and others attaining the maxima of 8 and 5, respectively. This skew is also reflected in Table 3.1, where the mean of both degree distributions is greater than the respective median and their standard deviations are relatively high. Taken together, this information suggests that there are few, central rebel groups in the network, a pattern that is also visible in the network graph. An interesting feature of the network that is not easily discernible in the graph is that there is more variation in the targets vs. initiators of hostilities (since $\sigma_{Id} > \sigma_{Od}$). Finally, it is

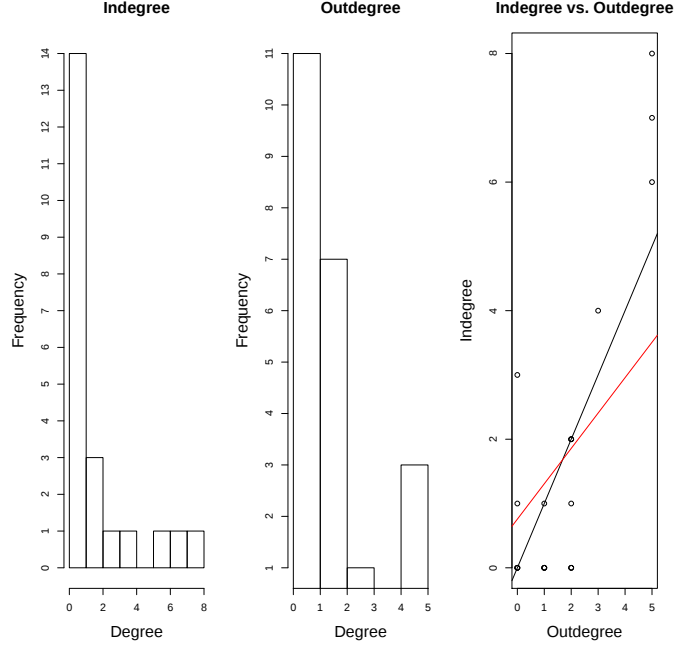


Figure 3.2: Degree Histograms & Scatterplot
Notes: Black line in scatterplot is 45 degree line.

worth noting that, owing to the aforementioned mutuality of hostilities, there is a strong positive correlation between indegree and outdegree (0.84), also mirrored in the steep slope of the scatterplot in Figure 3.2.

At this stage, we can focus on the node level. Table 3.2 displays four different centrality scores for each rebel group. Across most measures, three groups stand out: Amal, Fatah/PLO, and the South Lebanon Army (SLA). Unsurprisingly, these are three groups that are included in virtually every historical account of the conflict. Moreover, they are the biggest factions from the three ethno-religious groups most active in the conflict—Amal from the Shi’ites (though later surpassed by Hezbollah), Fatah from the Palestinians, and SLA from the Maronite Christians. Clearly, these groups’ centrality scores bode well with popular facts about the conflict, but what do the different measures reflect?

Betweenness centrality, a metric more complicated than the intuitive indegree and out-degree scores, measures a node’s propensity to act as a bridge between other nodes. In social networks, it is intended to measure the control that a node controls on communica-

tion between other nodes (Freeman, 1979). In the case of civil conflict, groups with high betweenness centrality stand in the way of smaller groups attacking other groups—for example, SLA attacks Amal and is attacked by 5 Palestinian groups, but none of *those* groups attacks Amal. Eigen centrality, in turn, measures a node’s ties to *central* nodes—a node’s ties with central nodes make that node eigen central (Bonacich, 1972). A good illustration of this concept is the Popular Nasserist Organization (PNO), which ranks fourth in eigen centrality, but only has 2 ties. This is because these ties are Amal and Fatah, the two most central groups by most measures. Returning to Amal, Fatah, and SLA, their high eigen centrality is owed to the fact that all they attack each other, and all are attacked by several other groups. A final centrality measure that is binary, and hence omitted from the table, is whether a node acts as a cutpoint. Cutpoints are nodes whose removal disconnects the network, dividing it into smaller networks. Again, 3 cutpoints are detected, corresponding to Amal, Fath, and SLA. In short, Table 3.2 shows how using multiple and complementary measures of actor centrality reenforces our understanding of the complex dynamics in the Lebanese conflict.

3.2.2 Predicting Inter-Rebel Hostilities

In this section, I predict hostilities in the conflict network by fitting an array of regression models, aiming to maximize fit and predictive power. In choosing terms to include in my models, I look to the emerging literature on inter-rebel violence. However, I also contribute to the literature by including novel node-level covariates from the MAROB data, as well as dyad-level covariates.

For this purpose, I employ the state-of-the-art model for predicting edges in networks, the ERGM. Crucially, this model treats the observed network – as a whole – as a draw from a (multivariate) distribution, and thus does not require nodes and edges to be independently distributed in order to estimate the effects of covariates on the network’s structure (Cranmer and Desmarais, 2011). In other words, because the unit of observation – from the model’s standpoint – is the whole network and not its nodes and edges, the ERGM does not require

Table 3.2: Centrality Scores by Group

Group	Indegree	Outdegree	Betweenness	Eigen
al-Ahbash	0	0	0	0
aJaI	0	1	0	0.13
al-Mourabitoun	0	2	0	0.23
al-Saíqah	0	2	0	0.20
Amal	7	5	60.67	0.48
Asbat al-Ansar	0	0	0	0
DFLP	0	1	0	0.12
FRC	1	2	0	0.25
Fatah Uprising	2	2	4.33	0.13
Fatah/PLO	8	5	56.50	0.39
Hezbollah	2	2	0	0.27
IUM	0	1	0	0.13
NLP	1	1	0	0.07
PLF	0	0	0	0
PPSF	0	1	0	0.12
Phalangists	4	3	27.67	0.22
PFLP	3	0	0	0
PFLP-GC	1	0	0	0
PNO	0	2	0	0.28
PSP	2	2	1.33	0.22
RPCP	0	0	0	0
SLA	6	5	42.50	0.37

Notes: Bold numbers denote the 3 highest values in each score. Groups are labeled as in the MAROB data and might be labeled differently in other sources.

the independence assumption for unbiased and consistent estimation. On the contrary, the classical regression model does require (conditionally) independent observations to deliver statistically sound estimates—whether from a frequentist or Bayesian perspective.

I begin by attempting to exploit another advantage of the ERGM over the classical regression model: it can incorporate edge-, dyad-, and triad-level terms in the regression. However, adding terms for transitivity, cyclicity, or other triadic features causes the model to not converge. Similarly, the diagnostics for edgewise and dyadwise shared partners terms suggest they should not be included in the model. The only exception is the term for mutual ties, which improves model fit drastically. Recalling the network’s relatively large reciprocity score (0.54), based on the fact that hostilities are reciprocated in conflict, the significant effect

of the Mutual term is not surprising.

Moving on, I use the ERGM to estimate the effect of node-level covariates on hostilities, something which the literature attempts through ill-applied standard regressions. I search through the MAROB dataset for node-level covariates that appear in the literature as predictors of inter-rebel hostilities. I start by setting a baseline: a naive model that only includes the number of edges each node has as a predictor. After supplementing this model with a term for mutual ties – as noted above, the only higher-level covariate that improves model fit – I progressively add node-level covariates in an effort to minimize Residual Deviance. The latter is the most popular goodness-of-fit statistic for models estimated through Maximum Likelihood and nested within each other. Note that all of my nodal covariates terms are estimated effects for *out*-edges. This is because the literature on inter-rebel violence forms hypotheses about the effect of *group i having feature x_i* on its likelihood of attacking other groups. Table 3.3 shows the results of this search, through a series of Analysis of Variance (ANOVA) comparisons. Every model from the third one onwards adds a nodal covariate, and all models aside from the penultimate one make a statistically significant improvement in fit. Overall, at the cost of only 8 degrees of freedom (residual dof = 454), residual deviance drops by more than 75% between the baseline and final model (from 640 to 157).

Nevertheless, Residual Deviance is only one of many criteria for model selection. I supplement my search for the best-fitting model by using the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). Though the difference between the criteria is small, the former is better-suited to small samples and penalizes additional parameters more heavily (i.e. rewards parsimony). Table 3.4 shows output from the BIC- and AIC-minimizing ERGMs. Incidentally, the latter model is the same as the model minimizing Residual Deviance. However, the BIC-minimizing ERGM has four fewer nodal covariates. Substantively, this is an important difference, because the excluded covariates – Ethnic Cleansing, Control Resources, Control Territory, Agreement with State – are presented as determinants of inter-rebel violence in previous studies. I return to this point in Section 3.3. All of the terms in the models, including the nodal covariates, are statistically significant at the 5% or 1% level. Moreover, all of the diagnostics for both models indicate that the ERGMs

Table 3.3: ANOVA of ERGMs

Model	Deviance	Resid. DoF	Resid. Dev	Pr(> Chisq)
Edges	640.47	461	640.47	0***
+ Mutual	414.70	460	225.77	0***
+ Terrorist Tactics	24.53	459	201.24	0***
+ Popular (= BIC-min.)	30.48	458	170.77	0***
+ Ethnic Cleansing	3.49	457	167.28	0.06
+ Control Resources	5.03	456	162.24	0.02*
+ Control Territory	0.95	455	161.30	0.33
+ Agreement w/ State	4.36	454	156.93	0.04*

Notes: All variables are coded from the MAROB dataset (Asal, Pate and Wilkenfeld, 2008). A group employing Terrorist Tactics is one that attacks civilians and other non-security personnel (e.g. civil servants, elected representatives, other government representatives). A Popular group is one that has the support of the majority of the ethno-religious community it belongs to (Sunni, Shi'ite, etc). A group using Ethnic Cleansing is one that targets civilians of the same ethnicity/religion in the same geographic area. A group that Controls Resources is one that expropriates and markets scarce natural resources. A group that Controls Territory is one that controls movement into/out of/through an area and, potentially, establishes governing structures and maintains infrastructure in that area. A group that reaches Agreement with State is one that succeeds in negotiating with the state and extracting – but not necessarily implementing – concessions. BIC-min denotes the BIC-minimizing model. The Residual Deviance- and AIC-minimizing model is the last one. The p-value is with respect to the reduction in Residual Deviance being statistically significant. **p < .05; ***p < .01

converged.⁹ Interestingly, the coefficient for Popular is larger in the AIC-minimizing model, despite the model controlling for more covariates. In contrast, the other coefficients (aside from the Edges term) decrease in size, as would be expected if they were confounded with the additional covariates.

Table 3.4: Best-Fit Models

	BIC-Min	AIC-Min
Edges	−11.2*** (1.7)	−14.4*** (2.5)
Mutual	2.0*** (0.7)	1.6** (0.7)
Terrorist Tactics	1.7*** (0.3)	1.3** (0.6)
Popular	1.6*** (0.3)	2.3*** (0.5)
Ethnic Cleansing		−6.9** (3.4)
Control Territory		1.6** (0.7)
Control Resources		8.3*** (3.0)
Agreement with State		−3.8** (1.9)
BIC	195.3	206.2
AIC	178.8	173.2
Residual Deviance	170.8	156.9

Notes: See Table 3.3 for variable meaning.

p < .05; *p < .01

3.2.3 Clustering

The last aspect of network structure I explore is clustering. To do this, I employ the Latent Position Cluster Model (LPCM), which has the ability to identify clusters in the network and assign nodes to them. This is done by first placing nodes on a latent space, based on the Euclidean distance between nodes' and dyads' covariates, as well as higher-order terms

⁹I suppress diagnostics tables and plots to conserve space.

like transitivity (Hoff, Raftery and Handcock, 2002). Moreover, the LPCM can account for clustering based on “unobserved attributes or on endogenous attributes such as position in the network, ‘self-organization’ into groups or a preference for popular actors” (Handcock, Raftery and Tantrum, 2007, p. 302).

I fit an LPCM using all of the nodal covariates identified in Table 3.4 as significant predictors of hostilities. According to various diagnostics, the model converges.¹⁰ Figure 3.3 shows (posterior mean) estimates of the latent positions of rebel groups, projected onto 2-dimensional space. Two clusters are identified by the model: they are almost co-centric, with one subsuming the other, and are centered around 0 on both latent dimensions. The narrower cluster contains all but one of the Palestinian and Sunni groups, and it places them relatively close to each other. (It also contains 2/3 Maronite groups, also positioned adjacently, though far from the cluster’s center.) In turn, the broader cluster contains the only two Shi’ite groups in the network, though they are far from each other as well as the cluster’s center. Note that the broader cluster subsumes the narrower one—all groups in the latter cluster also belong to the former, but not the vice-versa. Finally, the only groups outside both clusters are the third Maronite group (SLA) and the only Druze group (PSP), but they are placed at maximal distance from each other. I interpret these findings in the next section.

3.3 Discussion

From the perspective of the literature on inter-rebel violence, this study produces several substantive insights. This is because the best-fit models in Table 3.4 include a number of covariates that other scholars have presented as causes of rebel hostilities. In this section, I discuss each of these variables’ marginal effects, and, where possible, connect my findings to the literature. I also interpret the clustering model’s output from the perspective of historical accounts of the conflict.

¹⁰I use the `latentnet` package in R to fit these models (Krivitsky and Handcock, 2008). Again, I suppress these diagnostics to conserve space.

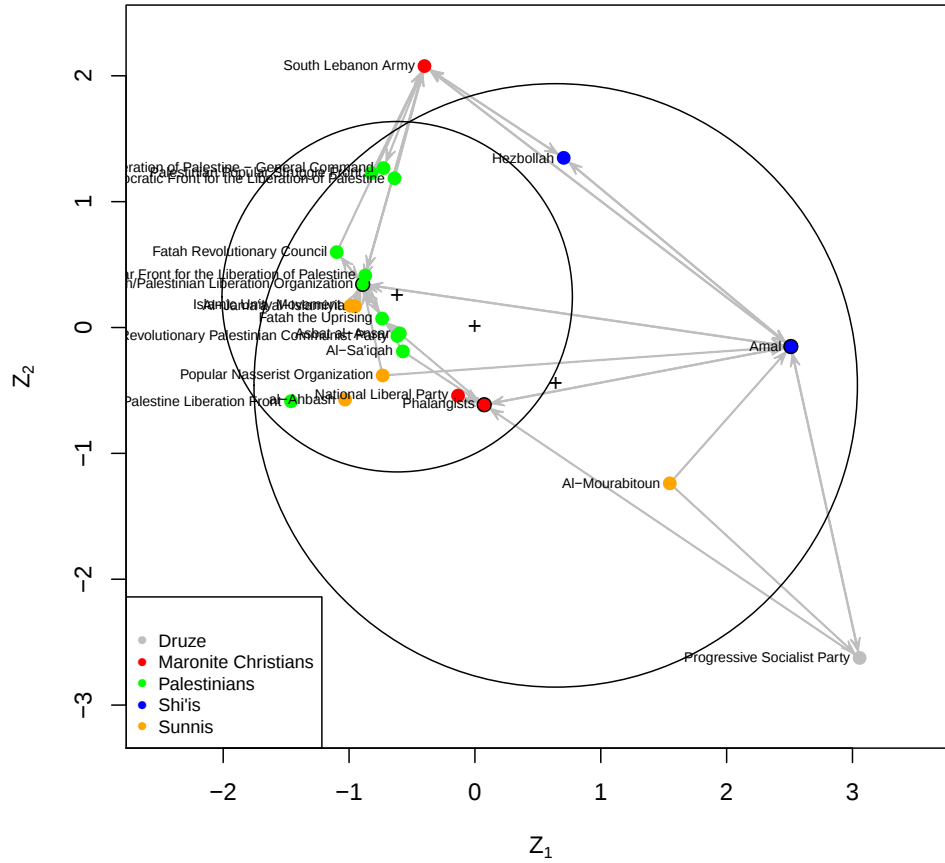


Figure 3.3: Latent Position Cluster Model

Notes: Latent dimensions are scale-free. 3 crosses mark the origin of the two latent dimensions (0,0) and the center of the 2 clusters.

A popular rebel group is, on average, roughly 9 times more likely to attack other (randomly chosen) rebel groups than a group that does not have majority support from its ethno-religious community.¹¹ This is in line with the argument of [Pischedda \(2015\)](#), that groups facing “windows of opportunity” – that is, which have an advantage over other co-ethnic rebels in reaping the support of their broader ethnic community – are more hostile towards co-ethnic groups.

¹¹All of the marginal effects I report are based on output from the second column of Table 3.4. Implied in each statement is that all other covariates (binary) are kept at their reference category. As such, the effects I report are what is sometimes referred to as “first differences”. Henceforth, to conserve space, I do not provide full interpretations of marginal effects.

Groups that control valuable natural resources are, on average, roughly 400 times more likely to attack another group than resource-poor groups. This supports the arguments of [Eck \(2010\)](#) and [Fjelde and Nilsson \(2012\)](#), that natural resources enlarge groups' capacity for violence, leading to more attacks on other rebels. [Eck \(2010\)](#) reports a positive coefficient for groups controlling drugs, while [Fjelde and Nilsson \(2012\)](#) report a positive coefficient for groups controlling oil reserves and a negative one for groups controlling gemstones.¹² Given that my own data does not distinguish between the type of resource, it is likely that the positive coefficient I estimate is due to the effect of the oil and drugs trades (there are few gemstone-endowed regions in Lebanon). Indeed, [Marshall \(2012\)](#) documents how Lebanese warlords financed their arms purchases by cultivating marijuana, processing cocaine and heroin, and, less so, by smuggling oil. The author also attributes some rebel hostilities to violent contests over international trafficking routes.

Controlling territory increases the probability that a group attacks another group by roughly 400%. Again, this finding is in line with [Fjelde and Nilsson \(2012\)](#), who argue that controlling territory adds to groups' strategic capacity, thereby enabling them to scale-up their hostilities towards competitors. This is because territorial control allows the group to harness the resources of civilians, including manpower and valuable information.

Reaching an agreement with the state means that a group is, on average, roughly 98% less likely to attack another group. This is in contrast with the argument of [Eck \(2010\)](#); that groups in negotiation with the government will try to eliminate other groups, in order to be the sole recipients of state concessions. That said, the author cautions us that the evidence in support of her argument is non-robust. As such, it is possible that the negative association I report is generalizable outside the Lebanese case.

Terrorist tactics increase a group's likelihood of initiating hostilities against another rebel group by roughly 270%, on average. On the other hand, a group that engages in ethnic cleansing is roughly 99% *less* likely to attack another group. Both of these findings are

¹²It is also highly likely that rebel groups that control valuable resources are also the target – not just the initiator – of hostilities, as other groups try to seize their wealth. However, my models cannot test this hypothesis, as the dependent variable is whether group i attacks group j and not the vice-versa.

novel, thus pointing to the necessity of incorporating rebel tactics into theories of inter-rebel violence. Puzzlingly, although there is a wide literature on rebels’ tactics in fighting against the state, it has not been integrated with the emerging literature on inter-rebel violence. Since theory-building is beyond this study’s scope, I leave it up to the literature to modify existing theories so as to account for the above correlations.

I now return to the output from the LPCM (Figure 3.3). The lack of tight clustering around sectarian lines reflects the highly complex nature of the Lebanese Civil War: multiple rebel groups from each of several ethnic and religious sects, with cross-cutting political preferences within sects and substantial variation in resources and strategies. Naturally, this presents a tough setting for establishing clear patterns. Furthermore, the aggregated nature of the edges across 12 years of a dynamically evolving conflict might obscure clusters from distinct phases in the conflict. Nevertheless, there are some patterns in Figure 3.3 worth interpreting.¹³

The narrower cluster seems to pick-up the significant infighting among Palestinian groups, a distinctive feature of the Lebanese conflict. At the same time, the close positioning of the Palestinian groups emphasizes their similarities across many dimensions. Indeed, where it not for the extreme organizational fragmentation and “plague of initials” that civil war is associated with, we could imagine several of these Palestinian consolidating into one group (Bakke, Cunningham and Seymour, 2012). The presence of 4/5 Sunni groups in the first cluster is also interesting. In addition to the occasional alliances between Palestinian and Sunni groups (e.g. Popular Nasserist Organization joining PLO in 1976 Damour offensive), it is possible that shared religion creates more shared features between them, which the latent model detects. The only Sunni-labeled group outside the first cluster, Al-Mourabitoun, actually had mixed membership historically – containing, Sunnis, Shias, Maronites, and Druze – and also perpetually formed alliances with groups from all sects but the Maronites. Of further interest is the presence of 2/3 Maronite groups in the narrow cluster. Their

¹³As with Principal Component Analysis, where researchers often label and interpret the k most important principal components in a way that is informed by their theory, interpreting the clusters identified by LPCMs is at the researcher’s discretion. The historical information in this paragraph is from O’Ballance (1998).

adjacent presence might be explained by the fact that the National Liberal Party’s militia joined forces with the Phalangists in 1976, then were eliminated by them in 1980. As for these Maronite groups being placed in the same cluster as the Palestinians and Sunnis, this might be owed to their shared tactics, which are measured by two node-level covariates: terrorist tactics and ethnic cleansing. Indeed, like the more extreme Palestinian groups (e.g. Al-Sa’iqah), the Phalangists targeted civilians (terrorism) and eliminated whole areas of non-coethnics (ethnic cleansing) in the Karantina and Tel al-Zaatar massacres.

The wider cluster contains the only two Shi’ite groups, Amal and Hezbollah, though they are placed far apart. This could be owed to Hezbollah forming as a splinter of Amal due to disagreements over secularism, as well as their frequent infighting (esp. 1987 – 1989). Since all Palestinian (and Sunni) groups are included in the broader cluster with Amal and Hezbollah, the positioning of the two Shi’ite groups can also be interpreted through their attitude towards the Palestinian issue. Indeed, Amal exchanged attacks with Palestinian groups (see War of the Camps), with which Hezbollah was historically aligned. Amal’s hostile stance to Palestinian groups might also explain why it is placed much further from the narrow cluster than Hezbollah is. Finally, the only Druze group’s (PSP) placement outside the broader cluster and at maximal distance from other nodes can be interpreted through their shifting alliances and opportunistic tactics. Similarly, the only other group outside the broader cluster, the Maronite South Lebanon Armys, also differed drastically from other groups: though it was initially aligned with the other Christian groups, it broke away, and it acted mostly as a proxy actor for Israel.

3.4 Conclusion

Following the civil war literature’s recent shift to the study of inter-rebel conflict, this article reviewed and demonstrated the advantages of a network approach. This was done by applying several tools from network analysis to the case of Lebanon’s Civil War, specifically the period 1980 – 1991. In particular, a network graph and descriptive statistics at the node-, dyad-, triad-, and network-level were used to confirm several patterns detected in historical

accounts of the conflict: a dense pattern of hostilities, high reciprocity and low transitivity in hostilities, infighting within religious sects, and the existence of 3 central rebel groups. Furthermore, Exponential Random Graph Models were used to predict inter-rebel hostilities, and found that groups that command support from the ethno-religious sect they belong to, control valuable natural resources and territory, and use terrorist tactics are more likely to attack other rebels, while groups that are able to reach an agreement with the state are less likely to attack other rebels. Finally, a Latent Position Cluster Model was employed to uncover clusters in the network and detected 2 sub-conflicts: a narrow cluster that includes the infighting among Palestinian groups and their Sunni allies and a broader cluster that includes the hostilities between rival Shi'ite groups.

My approach has implications for (inter)national policy-makers seeking to predict or influence inter-rebel hostilities. For example, given foreign powers' diverging preferences over the ongoing Syrian conflict's outcome and the multitude of groups involved in the conflict, knowing how network structure affects hostilities among rebels is crucial to policy-makers on all sides. If Iran or Russia's objective is to maximize conflict among rebel groups, so as to divert hostilities away from the allied Syrian regime, they will want to know what covariates predict inter-rebel hostilities. Though foreign powers cannot influence structural covariates that predict hostilities, they might be able to influence the resources available to groups and their ability to negotiate with the state. Similarly, if the US's objective is to channel resources to groups that will use them against the regime and not other against rebels, policy-makers will want to know what covariates predict a reduction inter-rebel hostilities. By building on my approach with more fine-grained data, it is possible to tackle these policy questions.

My approach also suffers from two limitations, which lend themselves to an equal number of suggestions for future research. First, the aggregated nature of my edges across 12 years of conflict might cause me to pool distinct phases of the Lebanese Civil War into one phase. Though this is not improper from a statistical perspective and, in fact, provides more degrees of freedom and a denser network for estimating my ERGMs, it increases the risk of null results. For example, if nodal covariate x caused rebels to be more violent in the first half

of the period in question, but less so in the second half, we might find a null effect by estimating the effect of x using data from the whole period. Another issue that pooling network data creates relates to the elimination of rebel groups during the period in question. If elimination is driven by some of the nodal covariates in my regressions, predicting hostilities against groups that no longer exist might bias my estimates. A remedy to these issues is modeling the network dynamically and using longitudinal ERGMs to predict hostilities across time. Unfortunately, this task is highly challenging and largely an area of ongoing research (Krivitsky and Handcock, 2014).

The second limitation of my approach is that my estimates are correlational, not causal. This is owed to the observational nature of my network data and the lack of a causality-oriented research design. However, it might be possible to exploit sources of random variation at the rebel-group level to identify some causal effects. For example, shocks to rebels' resources due to poor weather, economic conditions, or unforeseen foreign intervention might allow us to estimate the causal effect of rebel groups' resources on inter-rebel hostilities. That said, one is hard-pressed to think of random shocks for other node-level covariates (e.g. agreement with state), let alone higher-level covariates for the network. For this reason, the most obvious way to advance the use of network analysis in civil conflict studies – with network graphs and descriptive statistics, ERGMs, and LPCMs – is to apply these tools to additional conflict networks. I leave this exciting task to future research.

Bibliography

- Acemoglu, Daron, Simon Johnson and James A Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91(5):1369–1401.
- Acemoglu, Daron, Simon Johnson and James A Robinson. 2002. "An African Success Story: Botswana." .
- Albertus, Michael. 2015. *Autocracy and Redistribution*. New York: Cambridge University Press.
- Albertus, Michael and Victor Menaldo. 2012. "If You're Against Them You're With Us The Effect of Expropriation on Autocratic Survival." *Comparative Political Studies* 45(8):973–1003.
- Alstadsæter, Annette, Niels Johannesen and Gabriel Zucman. 2017*a*. "Tax Evasion and Inequality." *National Bureau of Economic Research* .
- Alstadsæter, Annette, Niels Johannesen and Gabriel Zucman. 2017*b*. "Who Owns the Wealth in Tax Havens? Macro Evidence and Implications for Global Inequality." *National Bureau of Economic Research* .
- Asal, Victor, Amy Pate and Jonathan Wilkenfeld. 2008. "Minorities at Risk Organizational Behavior: Data and Codebook." . Data retrieved from www.cidcm.umd.edu/mar/data.asp.
- Bagozzi, Benjamin E. 2015. "Forecasting Civil Conflict with Zero-Inflated Count Models." *Civil Wars* 17(1):1–24.
- Bagozzi, Benjamin E. 2016. "The Baseline-Inflated Multinomial Logit Model for International Relations Research." *Conflict Management and Peace Science* 33(2):174–197.
- Bagozzi, Benjamin E and Bumba Mukherjee. 2012. "A Mixture Model for Middle Category Inflation in Ordered Survey Responses." *Political Analysis* 20(3):369–386.

- Bagozzi, Benjamin E, Daniel W Hill, Will H Moore and Bumba Mukherjee. 2015. "Modeling Two Types of Peace: The Zero-inflated Ordered Probit (ZiOP) Model in Conflict Research." *Journal of Conflict Resolution* 59(4):728–752.
- Bagozzi, Benjamin E and Kathleen Marchetti. 2017. "Distinguishing Occasional Abstention from Routine Indifference in Models of Vote Choice." *Political Science Research and Methods* 5(2):277–294.
- Bakke, Kristin M, Kathleen Gallagher Cunningham and Lee JM Seymour. 2012. "A Plague of Initials: Fragmentation, Cohesion, and Infighting in Civil Wars." *Perspectives on Politics* 10(2):265–283.
- Barro, Robert J. 1989. "Economic Growth in a Cross Section of Countries." *National Bureau of Economic Research* .
- Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(01):21–35.
- Beck, Nathaniel, Jonathan N Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.
- Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer and Patrick Walsh. 2001. "New Tools in Comparative Political Economy: The Database of Political Institutions." *The World Bank Economic Review* 15(1):165–176.
- Besag, Julian. 1975. "Statistical Analysis of Non-Lattice Data." *The Statistician* 24(3):179–195.
- Blattman, Christopher and Edward Miguel. 2010. "Civil War." *Journal of Economic Literature* 48(1):3–57.
- Boix, Carles and Milan W Svolik. 2013. "The Foundations of Limited Authoritarian Government: Institutions, Commitment, and Power-Sharing in Dictatorships." *The Journal of Politics* 75(02):300–316.

- Bonacich, Phillip. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2(1):113–120.
- Braga, Igor and Maria Carolina Monard. 2015. "Improving the Kernel Regularized Least Squares Method for Small-Sample Regression." *Neurocomputing* 163:106–114.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1):5–32.
- Bueno De Mesquita, Bruce, Alastair Smith, Randolph M Siverson and James D Morrow. 2005. *The Logic of Political Survival*. Cambridge: The MIT press.
- Burgoon, Brian. 2006. "On Welfare and Terror Social Welfare Policies and Political-Economic Roots of Terrorism." *Journal of Conflict Resolution* 50(2):176–203.
- Byman, Daniel. 2015. "The Homecomings: What Happens When Arab Foreign Fighters in Iraq and Syria Return?" *Studies in Conflict & Terrorism* 38(8):581–602.
- Byman, Daniel L and Jeremy Shapiro. 2014. "Be Afraid. Be A Little Afraid: The Threat of Terrorism from Western Foreign Fighters in Syria and Iraq." *Foreign Policy at Brookings* 34.
- Cameron, A Colin and Douglas L Miller. 2015. "A Practitioners Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–372.
- Cárdenas, Julián. 2015. "Are Latin America's Corporate Elites Transnationally Interconnected? A Network Analysis of Interlocking Directorates." *Global Networks* 15(4):424–445.
- Cárdenas, Julián and Francisco Robles-Rivera. 2017. "Business Elite Networks for Corporate State Capture in Central American Countries."
- Cheibub, José Antonio, Jennifer Gandhi and James Raymond Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143(1-2):67–101.
- Chenoweth, Erica. 2013. "Terrorism and Democracy." *Annual Review of Political Science* 16:355–378.

- Collier, Paul and Anke Hoeffler. 1998. "On Economic Causes of Civil War." *Oxford Economic Papers* 50(4):563–573.
- Commission, European. 2015. "Syria Crisis." *ECHO Factsheets* .
URL: <https://ec.europa.eu/echo/node/441>
- Cook, Karen S and Russell Hardin. 2001. *Norms of Cooperativeness and Networks of Trust*. New York: Russell Sage Foundation.
- Cox, David R. 1983. "Some Remarks on Overdispersion." *Biometrika* 70(1):269–274.
- Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39(5):829–844.
- Cranmer, Skyler J and Bruce A Desmarais. 2011. "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis* 19(1):66–86.
- Credendo Group. 2017. "Country Risks Summary Table.". Data retrieved from <http://www.delcrederecroire.be/en/>.
- Crenshaw, Martha. 1981. "The Causes of Terrorism." *Comparative Politics* 13(4):379–399.
- Cunningham, Kathleen Gallagher, Kristin M Bakke and Lee JM Seymour. 2012. "Shirts Today, Skins Tomorrow: Dual Contests and the Effects of Fragmentation in Self-Determination Disputes." *Journal of Conflict Resolution* 56(1):67–93.
- Doner, Richard F and Ben Ross Schneider. 2000. "Business Associations and Economic Development: Why Some Associations Contribute More than Others." *Business and Politics* 2(3):261–288.
- Eck, Kristine. 2010. Fighting Our Friends Instead of Our Enemies: Explaining Armed Conflict Between Rebel Groups. In *Annual Meeting of the International Studies Association*.
- Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 17(1):1–26.

- Fan, Jianqing and Runze Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties." *Journal of the American Statistical Association* 96(456):1348–1360.
- Fariss, Christopher J and Zachary M Jones. 2018. "Enhancing Validity in Observational Settings When Replication is Not Possible." *Political Science Research and Methods* 6(2):365–380.
- Fearon, James D and David D Laitin. 2003. "Ethnicity, Insurgency, and Civil war." *American Political Science Review* 97(1):75–90.
- Felter, Joseph and Brian Fishman. 2007. "Al-Qa'ida's Foreign Fighters in Iraq: A First Look at the Sinjar Records."
- Fernández, Andrés, Michael W Klein, Alessandro Rebutti, Martin Schindler and Martin Uribe. 2016. "Capital Control Measures: A New Dataset." *IMF Economic Review* 64(3):548–574.
- Findley, Michael G, Daniel L Nielson and Jason Campbell Sharman. 2014. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. New York: Cambridge University Press.
- Finke, Roger. 2010. "International Religious Freedom Data, 2008." Data retrieved from <http://www.thearda.com/Archive/Files/Descriptions/IRF2008.asp>.
- Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *The American Economic Review* 91(4):1095–1102.
- Fjelde, Hanne and Desirée Nilsson. 2012. "Rebels Against Rebels: Explaining Violence between Rebel Groups." *Journal of Conflict Resolution* 56(4):604–628.
- Fowler, James H. 2006. "Legislative Cosponsorship Networks in the US House and Senate." *Social Networks* 28(4):454–465.

- Freedom House. 2016. “Freedom in the World.”. Data retrieved from <https://freedomhouse.org/report/freedom-world/freedom-world-2016>.
- Freeman, Linton C. 1979. “Centrality in Social Networks: Conceptual Clarification.” *Social Networks* 1(3):215–239.
- Friedman, Jerome, Trevor Hastie and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. New York: Springer.
- Gandhi, Jennifer. 2008. *Political Institutions under Dictatorship*. New York: Cambridge University Press.
- Garay, Aldo M, Elizabeth M Hashimoto, Edwin MM Ortega and Víctor H Lachos. 2011. “On Estimation and Influence Diagnostics for Zero-Inflated Negative Binomial Regression Models.” *Computational Statistics & Data Analysis* 55(3):1304–1318.
- GCCS. 2014. “Addressing the Foreign Terrorist Fighters Phenomenon from a European Union Perspective.” *Global Center on Cooperative Security, Human Security Collective, and International Centre for Counter-Terrorism* .
- Geddes, Barbara. 1999. “Authoritarian Breakdown: Empirical Test of a Game-Theoretic Argument.”.
- Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. “Autocratic Breakdown and Regime Transitions: A New Data Set.” *Perspectives on Politics* 12(02):313–331.
- Geddes, Barbara, Joseph Wright and Erica Frantz. 2018. *How Dictatorships Work*. New York: Cambridge University Press.
- Gehlbach, Scott and Philip Keefer. 2012. “Private Investment and the Institutionalization of Collective Action in Autocracies: Ruling Parties and Legislatures.” *The Journal of Politics* 74(02):621–635.
- Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. New York: Cambridge University Press.

- Gerges, Fawaz A. 2009. *The Far Enemy: Why Jihad Went Global*. New York: Cambridge University Press.
- Gibbs, Jack P. 1989. "Conceptualization of Terrorism." *American Sociological Review* 54(3):329–340.
- Graham, Benjamin AT, Noel P Johnston and Allison Kingsley. 2017. "Even Constrained Governments Steal: The Domestic Politics of Transfer and Expropriation Risks." *Journal of Conflict Resolution* pp. 1–30.
- Graham, David A. 2016. "What Is Mossack Fonseca, the Law Firm in the Panama Papers?" *The Atlantic* .
- Green, Donald P and Holger L Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly* 76(3):491–511.
- Greif, Avner, Paul Milgrom and Barry R Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102(4):745–776.
- Grim, Brian J and Conrad Hackett. 2006. "The Global Religious Landscape: A Report on the Size and Distribution of the World's Major Religious Groups as of 2010." *The Pew Forum on Religion and Public Life* .
- Grim, Brian J and Roger Finke. 2006. "International Religion Indexes: Government Regulation, Government Favoritism, and Social Regulation of Religion." *Interdisciplinary Journal of Research on Religion* .
- Guriev, Sergei and Konstantin Sonin. 2009. "Dictators and Oligarchs: A Dynamic Theory of Contested Property Rights." *Journal of Public Economics* 93(1):1–13.
- Gwartney, James, Robert Lawson, Joshua Hall, James Gwartney, Robert Lawson and Joshua Hall. 2016. "Economic Freedom of the World: 2016 Annual Report." . Data retrieved from <https://www.fraserinstitute.org/economic-freedom>.

- Haber, Stephen, Noel Maurer and Armando Razo. 2003. *The Politics of Property Rights: Political Instability, Credible Commitments, and Economic Growth in Mexico, 1876-1929*. New York: Cambridge University Press.
- Haddad, Bassam. 2011. *Business Networks in Syria: The Political Economy of Authoritarian Resilience*. Palo Alto: Stanford University Press.
- Hadenius, Axel and Jan Teorell. 2007. "Pathways from Authoritarianism." *Journal of Democracy* 18(1):143–157.
- Hainmueller, Jens and Chad Hazlett. 2014. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach." *Political Analysis* 22(2):143–168.
- Hajzler, Christopher. 2012. "Expropriation of Foreign Direct Investments: Sectoral Patterns from 1993 to 2006." *Review of World Economics* 148(1):119–149.
- Handcock, Mark S, Adrian E Raftery and Jeremy M Tantrum. 2007. "Model-Based Clustering for Social Networks." *Journal of the Royal Statistical Society: Series A* 170(2):301–354.
- Harding, Luke. 2016a. "Mossack Fonseca: Inside the Firm that Helps the Super-Rich Hide their Money." *The Guardian* .
- Harding, Luke. 2016b. "What are the Panama Papers? A Guide to History's Biggest Data Leak." *The Guardian* .
- Hegghammer, Thomas. 2011. "The Rise of Muslim Foreign Fighters: Islam and the Globalization of Jihad." *International Security* 35(03):53–94.
- Hegghammer, Thomas. 2013a. "Number of Foreign Fighters from Europe in Syria is Historically Unprecedented. Who Should Be Worried?" *Washington Post* .
- Hegghammer, Thomas. 2013b. "Should I Stay or Should I Go? Explaining Variation in Western Jihadists' Choice between Domestic and Foreign Fighting." *American Political Science Review* 107(01):1–15.

- Hewitt, Christopher and Jessica Kelley-Moore. 2009. "Foreign Fighters in Iraq: A Cross-National Analysis of Jihadism." *Terrorism and Political Violence* 21(2):211–220.
- Hill, Daniel W and Zachary M Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(03):661–687.
- Ho, Daniel E, Kosuke Imai, Gary King and Elizabeth A Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3):199–236.
- Hoerl, Arthur E and Robert W Kennard. 1970. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12(1):55–67.
- Hoff, Peter D, Adrian E Raftery and Mark S Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97(460):1090–1098.
- Honaker, James, Gary King, Matthew Blackwell et al. 2011. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7):1–47.
- Howell, Llewellyn D. 2012. "International Country Risk Guide Methodology." . Data retrieved from <https://epub.prsgroup.com/products/political-risk-services>.
- Ilzetzi, Ethan, Carmen M Reinhart and Kenneth S Rogoff. 2017. "Exchange Arrangements Entering the 21st Century: Which Anchor Will Hold?" *National Bureau of Economic Research* .
- International Consortium of Investigative Journalists. 2016. "The Panama Papers." . Data retrieved from <https://panamapapers.icij.org/>.
- International Monetary Fund. 2014. "World Economic Outlook Dataset." . Data retrieved from <http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/weoselgr.aspx>.
- Jackson, John E and Ken Kollman. 2012. "Modeling, Measuring, and Distinguishing Path Dependence, Outcome Dependence, and Outcome Independence." *Political Analysis* 20(2):157–174.

- Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton: Princeton University Press.
- Jenne, Erin K and Milos Popovic. 2016. Managing Internationalized Civil Wars. In *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press.
- Jensen, Nathan. 2008. “Political Risk, Democratic Institutions, and Foreign Direct Investment.” *The Journal of Politics* 70(4):1040–1052.
- Jensen, Nathan M, Edmund Malesky and Stephen Weymouth. 2014. “Unbundling the Relationship Between Authoritarian Legislatures and Political Risk.” *British Journal of Political Science* 44(03):655–684.
- Johannesen, Niels and Gabriel Zucman. 2014. “The End of Bank Secrecy? An Evaluation of the G20 Tax Haven Crackdown.” *American Economic Journal: Economic Policy* 6(1):65–91.
- Johns, Leslie and Rachel L Wellhausen. 2016. “Under One Roof: Supply Chains and the Protection of Foreign Investment.” *American Political Science Review* 110(01):31–51.
- Jones, Zachary and Linder Fridolin. 2015. Exploratory Data Analysis Using Random Forests.
- Keller, Franziska Barbara. 2014. Networks of Power: Using Social Network Analysis to Understand who Will Rule and who Is Really in Charge in the Chinese Communist Party. PhD thesis New York University.
- Kenkel, Brenton and Curtis S Signorino. 2013. “Bootstrapped Basis Regression with Variable Selection: A New Method for Flexible Functional Form Estimation.”
- Kimeldorf, George and Grace Wahba. 1971. “Some Results on Tchebycheffian Spline Functions.” *Journal of Mathematical Analysis and Applications* 33(1):82–95.
- Kimeldorf, George S and Grace Wahba. 1970. “A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines.” *The Annals of Mathematical Statistics* 41(2):495–502.

- King, Gary. 1989a. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33:123–147.
- King, Gary. 1989b. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press.
- Klausen, Jytte. 2015. "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq." *Studies in Conflict & Terrorism* 38(1):1–22.
- Kolaczyk, Eric D. 2009. *Statistical Analysis of Network Data: Methods and Models*. New York: Springer Science & Business Media.
- Krivitsky, Pavel N and Mark S Handcock. 2008. "Fitting Position Latent Cluster Models for Social Networks with Latentnet." *Journal of Statistical Software* 24(5):1–23.
- Krivitsky, Pavel N and Mark S Handcock. 2014. "A Separable Model for Dynamic Networks." *Journal of the Royal Statistical Society: Series B* 76(1):29–46.
- Kroll, Luisa, M Miller and T Serafin. 2010. "The World's Billionaires: 2010." *Forbes* .
- La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer and Robert Vishny. 1999. "The Quality of Government." *The Journal of Law, Economics, and Organization* 15(1):222–279.
- Lawless, Jerald F. 1987. "Negative Binomial and Mixed Poisson Regression." *Canadian Journal of Statistics* 15(3):209–225.
- Lewis, Bernard. 2004. *The Crisis of Islam: Holy War and Unholy Terror*. New York: Random House.
- Li, Quan. 2005. "Does democracy promote or reduce transnational terrorist incidents?" *Journal of Conflict Resolution* 49(2):278–297.
- Mahdavi, Paasha. 2014. "Why Do Leaders Nationalize the Oil Industry? The Politics of Resource Expropriation." *Energy Policy* 75:228–243.

- Mahdavi, Paasha. 2017. "Scraping Public Co-Occurrences for Statistical Network Analysis of Political Elites." *Political Science Research and Methods* .
- Mahdavi, Paasha and John Ishiyama. 2018. "Dynamics of the Inner Elite in Dictatorships: Evidence from North Korea." .
- Malet, David. 2010. "Why Foreign Fighters?: Historical Perspectives and Solutions." *Orbis* 54(1):97–114.
- Malet, David. 2015. "Foreign Fighter Mobilization and Persistence in a Global Context." *Terrorism and Political Violence* 27(3):454–473.
- Marschall, Melissa J, Anirudh VS Ruhil and Paru R Shah. 2010. "The New Racial Calculus: Electoral Institutions and Black Representation in Local Legislatures." *American Journal of Political Science* 54(1):107–124.
- Marshall, Jonathan. 2012. *The Lebanese Connection: Corruption, Civil War, and the International Drug Traffic*. Palo Alto: Stanford University Press.
- Marshall, Monty G, Keith Jagers and Ted Robert Gurr. 2016. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2012." . Data retrieved from <http://www.systemicpeace.org/polity/polity4.htm>.
- Mayer, Thierry and Soledad Zignago. 2011. "The GeoDist Database." . Data retrieved from http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6.
- Miller, Terry, Anthony B Kim and KR Holmes. 2015. "2015 Index of Economic Freedom." . Data retrieved from <https://www.heritage.org/index/download>.
- Montgomery, Jacob M and Santiago Olivella. Forthcoming. "Tree-Based Models for Political Science Data." *American Journal of Political Science* .
- Montgomery, Jacob M, Santiago Olivella, Joshua D Potter and Brian F Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* 23(4):488–505.

- Mroz, Thomas A and Timothy H Savage. 1999. "Overfitting and Biases in Nonparametric Kernel Regressions Using Cross-Validated Bandwidths: A Cautionary Note."
- Muchlinski, David, David Siroky, Jingrui He and Matthew Kocher. 2016. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24(1):87–103.
- Mullahy, John. 1986. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics* 33(3):341–365.
- Neumann, Peter R. 2015. "Foreign Fighter Total in Syria/Iraq Now Exceeds 20,000; Surpasses Afghanistan Conflict in the 1980s." The International Centre for the Study of Radicalisation and Political Violence.
URL: <http://icsr.info/2015/01/foreign-fighter-total-syriairaq-now-exceeds-20000-surpasses-afghanistan-conflict-1980s/>
- Nilsson, Marco. 2015. "Foreign Fighters and the Radicalization of Local Jihad: Interview Evidence from Swedish Jihadists." *Studies in Conflict & Terrorism* 38(5):343–358.
- Noor, Farish A and James Dorsey. 2014. "Responding to the Islamic State's Foreign Fighters: Retribution or Rehabilitation?" *Washington Post* .
- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
- North, Douglass C and Barry R Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *The Journal of Economic History* 49(4):803–832.
- O'Ballance, Edgar. 1998. *Civil War in Lebanon, 1975-92*. London: Macmillan.
- O'Donovan, James, Hannes F Wagner and Stefan Zeume. 2016. "The Value of Offshore Secrets: Evidence from the Panama Papers."

- Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87(03):567–576.
- Pearlman, Wendy and Kathleen Gallagher Cunningham. 2012. "Nonstate Actors, Fragmentation, and Conflict Processes." *Journal of Conflict Resolution* 56(1):3–15.
- Pepinsky, Thomas. 2014. "The Institutional Turn in Comparative Authoritarianism." *British Journal of Political Science* 44(03):631–653.
- Piazza, James A. 2011. "Poverty, Minority Economic Discrimination, and Domestic Terrorism." *Journal of Peace Research* 48(3):339–353.
- Pischedda, Costantino. 2015. Wars Within Wars: Understanding Inter-Rebel Fighting. PhD thesis Columbia University.
- Porter, Michael D, Gentry White et al. 2012. "Self-Exciting Hurdle Models for Terrorist Activity." *The Annals of Applied Statistics* 6(1):106–124.
- RAN. 2014. Conference Paper. In *Radicalisation Awareness Network Cities Conference Foreign Fighters*.
- Razo, Armando. 2009. *Social Foundations of Limited Dictatorship*. Palo Alto: Stanford University Press.
- Riedl, Maximilian and Ingo Geishecker. 2014. "Keep it Simple: Estimation Strategies for Ordered Response Models with Fixed Effects." *Journal of Applied Statistics* 41(11):2358–2374.
- Roodman, David et al. 2009. "How to Do xtabond2: An Introduction to Difference and System GMM in Stata." *Stata Journal* 9(1):86–136.
- Ross, Michael. 2013. "Oil and Gas Data, 1932-2011.". Data retrieved from <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZTPW0Y>.
- Sageman, Marc. 2008. "A Strategy for Fighting International Islamist Terrorists." *The Annals of the American Academy of Political and Social Science* 618(1):223–231.

- Schmid, Alex P. 1992. "Terrorism and Democracy." *Terrorism and Political Violence* 4(4):14–25.
- Schölkopf, Bernhard and Alexander J Smola. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press.
- Schwab, K and X Sala-i Martin. 2016. The Global Competitiveness Report 2015–2016. In *The World Economic Forum*.
- Shim, Jooyong. 2012. "Kernel Poisson Regression for Mixed Input Variables." *Journal of the Korean Data and Information Science Society* 23(6):1231–1239.
- Shim, Jooyong and Changha Hwang. 2011. "Kernel Poisson Regression Machine for Stochastic Claims Reserving." *Journal of the Korean Statistical Society* 40(1):1–9.
- Solomon, Matthew and Joseph Spanjers. 2017. "Illicit Financial Flows from Developing Countries: 2005-2014.". Data retrieved from <http://www.gfintegrity.org/report/illicit-financial-flows-to-and-from-developing-countries-2005-2014/>.
- Standaert, Samuel. 2015. "Divining the Level of Corruption: A Bayesian State-Space Approach." *Journal of Comparative Economics* 43(3):782–803.
- Stenersen, Anne. 2011. "Al Qaeda's Foot Soldiers: A Study of the Biographies of Foreign Fighters Killed in Afghanistan and Pakistan Between 2002 and 2006." *Studies in Conflict & Terrorism* 34(3):171–198.
- Stone, Mervyn. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B* 36(2):111–147.
- Sundberg, Ralph, Kristine Eck and Joakim Kreutz. 2012. "Introducing the UCDP Non-State Conflict Dataset." *Journal of Peace Research* 49(2):351–362.
- Svolik, Milan W. 2009. "Power Sharing and Leadership Dynamics in Authoritarian Regimes." *American Journal of Political Science* 53(2):477–494.

- Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. New York: Cambridge University Press.
- Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko and Richard Svensson. 2013. “The Quality of Government Standard Dataset, Version Dec13.”. Data retrieved from <http://qog.pol.gu.se/data/datadownloads/data-archive>.
- The Economist. 2012. “Shells and Shelves.” *The Economist*.
- The Economist. 2016a. “Anti-Globalists: Why they’re Wrong.” *The Economist*.
- The Economist. 2016b. “The Panama Papers: A Torrential Leak.” *The Economist*.
- Tibshirani, Robert. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society. Series B* 58(1):267–288.
- Treisman, Daniel. 2007. “What Have we Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?” *Annual Review of Political Science* 10:211–244.
- UNOCHA. 2015. “Updated Overview: 2015 Syria Response Plan and 2015-2016 Regional Refugee and Resilience Plan.” *United Nations Office for the Coordination of Human Rights*.
- Volden, Craig, Alan E Wiseman and Dana E Wittmer. 2013. “When Are Women More Effective Lawmakers Than Men?” *American Journal of Political Science* 57(2):326–341.
- Wang, Zhu, Shuangge Ma and Ching-Yun Wang. 2014. “Variable Selection for Zero-Inflated and Overdispersed Data with Application to Health Care Demand in Germany.”.
- Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.
- Weggemans, Daan, Edwin Bakker and Peter Grol. 2014. “Who Are They and Why Do They Go? The Radicalization and Preparatory Processes of Dutch Jihadist Foreign Fighters.” *Perspectives on Terrorism* 8(4).

- Weymouth, Stephen. 2011. “Political Institutions and Property Rights: Veto Players and Foreign Exchange Commitments in 127 Countries.” *Comparative Political Studies* 44(2):211–240.
- Wilson, Matthew C and James A Piazza. 2013. “Autocracies and Terrorism: Conditioning Effects of Authoritarian Regime Type on Terrorist Attacks.” *American Journal of Political Science* .
- Wilson, Matthew Charles and Joseph Wright. 2017. “Autocratic Legislatures and Expropriation Risk.” *British Journal of Political Science* 47(1):1–17.
- World Bank. 2014. “Worldwide Governance Indicators.”. Data retrieved from <http://info.worldbank.org/governance/wgi/index.aspx/countryReports>.
- World Bank. 2016. “World Development Indicators.”. Data retrieved from <https://data.worldbank.org/data-catalog/world-development-indicators>.
- Wright, Joseph. 2008. “Do Authoritarian Institutions Constrain? How Legislatures Affect Economic Growth and Investment.” *American Journal of Political Science* 52(2):322–343.
- Yeyati, Eduardo Levy. 2006. “Financial Dollarization: Evaluating the Consequences.” *Economic Policy* 21(45):62–118.
- Zeileis, Achim, Christian Kleiber and Simon Jackman. 2008. “Regression Models for Count Data in R.” *Journal of Statistical Software* 27(8):1–25.
- Zhu, Ji and Trevor Hastie. 2005. “Kernel Logistic Regression and the Import Vector Machine.” *Journal of Computational and Graphical Statistics* 14(1):185–205.
- Zorn, Christopher JW. 1998. “An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications.” *Sociological Methods & Research* 26(3):368–400.