Forecasting Civil Wars: Theory and Structure in an Age of “Big Data” and Machine Learning

Online Appendix

Robert A. Blair*

Nicholas Sambanis†

A.1 CAMEO codes for predictors

We use the ICEWS dataset to construct four sets of predictors designed to capture the dynamics of conflict escalation. These predictors are operationalized as event counts at the country-month level, categorized as follows:

- ***Low-level violence***, comprising the following CAMEO events:

145:	Protest violently
1451:	Violent protest for leadership change
1452:	Violent protest for policy change

*Assistant Professor, Department of Political Science and Watson Institute for International and Public Affairs, Brown University, robert_blair@brown.edu

†Presidential Distinguished Professor, Department of Political Science, University of Pennsylvania, sambanis@upenn.edu

1453: Violent protest to demand rights

1454: Violent protest to demand change in institutions

170: Residual violent repression category

180: Use unconventional violence

183: Non-military bombing; outside of military engagements

171: Seize or damage property

175: Use repression

186: Assassinate

191: Impose blockade; restrict movement

1952: Remotely piloted aerial munitions

193: Fight with small arms

- ***Non-violent repression***, comprising:

123: Reject request or demand for political reform, not specified below

1232: Reject request to change policy

1233: Reject request for rights

1234: Reject request for change in institutions, regime

124: Refuse to yield, not specified below

1241: Refuse to ease administrative sanctions

1321: Threaten with restrictions on political freedoms

1322: Threaten to ban political parties or politicians

137: Threaten with repression

172: Impose administrative sanctions, not specified below

1721:	Impose restrictions on political freedoms

- ***Demands***, comprising:

104:	Demand political reform, not specified below
1042:	Demand policy change
1043:	Demand rights
1044:	Demand change in institutions, regime

- ***Accommodations***, comprising:

035:	Express intent to yield, not specified below
0342:	Express intent to change policy
0343:	Express intent to provide rights
0344:	Express intent to change institutions, regime
034:	Express intent to institute political reform, not specified below
080:	Yield, not specified below
081:	Ease administrative sanctions, not specified below
0811:	Ease restrictions on political freedoms
083:	Accede to requests or demands for political reform, not specified below
0832:	Accede to demands for change in policy
0833:	Accede to demands for rights
0834:	Accede to demands for change in institutions, regime
0812:	Ease ban on political parties

A.2 Classification of actors

A.2.1 Actors classified as "government"

We classify the following actors in the ICEWS dataset as "government:"

- Government
- Executive
- Executive Office
- Cabinet
- Foreign Ministry
- Interior/Home Ministry
- NGO Ministry
- Management/Budget/Planning/Organization Ministry
- Agriculture/Fishing/Forestry Ministry
- Finance/Economy/Commerce/Trade Ministry
- Industrial/Textiles/Mining Ministry
- Post/Telecoms Ministry
- Science/Tech Ministry
- State Owned Enterprises
- State-Owned Agricultural
- State-Owned Transportation

- State-Owned Utilities
- State-Owned Heavy Industrial/Chemical
- State-Owned Defense/Security
- State-Owned Durable Goods
- State-Owned Consumer Goods
- State-Owned Consumer Services
- State-Owned Consulting/Financial Services
- State-Owned Science/Tech/Knowledge/Innovation
- State-Owned Medical/Health/Pharmaceutical
- Women/Children/Social/Welfare/Development/Religion Ministry
- Education Ministry
- Energy Ministry
- Environment Ministry
- Transportation Ministry
- Food Ministry
- Disaster Ministry
- Health Ministry
- Science/Tech/Knowledge/Innovation Ministry
- Human Rights Ministry
- Elections Ministry

- Housing/Construction Ministry
- Justice/Law Ministry
- Tourism Ministry
- Drugs Ministry
- Labor Ministry
- Water Ministry
- Information/Communication/Transparency Ministry
- State Media
- Defense/Security Ministry
- Government Major Party (In Government)
- Government Minor Party (In Government)
- Government Provincial Party (In Government)
- Government Municipal Party (In Government)
- Government Religious
- Military
- Research and Design Wings
- Research and Design Wings Headquarters
- Research and Design Wings Education/Training
- Research and Design Wings Support
- Research and Design Wings Medical

- Army
- Army Headquarters
- Army Special Forces
- Army Infantry/Regular
- Army Mechanized (Ships, Tanks, Planes)
- Army Education/Training
- Army Support
- Army Medical
- Navy
- Navy Headquarters
- Navy Special Forces
- Navy Infantry/Regular
- Navy Mechanized (Ships, Tanks, Planes)
- Navy Education/Training
- Navy Support
- Navy Medical
- Air Force
- Air Force Headquarters
- Air Force Special Forces
- Air Force Infantry/Regular

- Air Force Mechanized (Ships, Tanks, Planes)
- Air Force Education/Training
- Air Force Support
- Air Force Medical
- Marines
- Marines Headquarters
- Marines Special Forces
- Marines Infantry/Regular
- Marines Mechanized (Ships, Tanks, Planes)
- Marines Education/Training
- Marines Support
- Marines Medical
- Coast Guard
- Coast Guard Headquarters
- Coast Guard Special Forces
- Coast Guard Infantry/Regular
- Coast Guard Mechanized (Ships, Tanks, Planes)
- Coast Guard Education/Training
- Coast Guard Support
- Coast Guard Medical

- Judicial
- National/Supreme Court
- Provincial Court
- Municipal/District Court
- Civil Court
- Religious Court
- Military/Tribunal
- Local
- Provincial
- Municipal
- Legal
- Parties (National)
- Major Party (National)
- Minor Party
- Provincial Party
- Municipal Party

A.2.2 Actors classified as "rebels"

We classify the following actors as "rebels:"

- Radicals/Extremists/Fundamentalists

- Organized Violent
- Rebel
- Insurgents
- Separatists

A.2.3 Actors classified as "opposition"

We classify the following actors as "opposition:"

- Dissident
- Protestors/Popular Opposition/Mobs
- Exiles
- Opposition Major Party (Out of Government)
- Opposition Minor Party (Out of Government)
- Opposition Provincial Party (Out of Government)
- Opposition Municipal Party (Out of Government)
- Banned Parties

A.3 Definition and coding of civil war

We use the Sambanis (2004) dataset to operationalize the onset of civil war. We choose the Sambanis dataset for two reasons. First, it offers the most careful coding of the month of civil war onset, with explicit coding rules and detailed documentation for each case, including a 256-page online supplement. Second, the Sambanis dataset conceptualizes civil war as a continuous process: most conflicts are coded as spanning the entire period from onset to termination without interruption, and so do not start and stop depending on whether they happen to reach a given threshold of battle-related deaths in a given year.

We believe this approach is of greater practical and theoretical interest than available alternatives. For example, in the UCDP-Prio dataset, Myanmar (Burma) is coded as experiencing seven different civil wars between 2000 and 2015, with twelve new onsets after 2000; India is coded as experiencing nine different civil wars over the same period, with eight new onsets after 2000; the DRC is coded as experiencing three different civil wars with three new onsets after 2000; etc. While we recognize the usefulness of this approach for some purposes, we believe predicting the first onset of a given conflict is more important, both theoretically and practically, than predicting whether it happens to reach a specific battle-related deaths threshold in a given year.

Datasets like UCDP-Prio strike us as particularly inappropriate for our purposes. We wish to assess whether the incidence of low-level violence between governments, opposition and rebel groups can predict the onset of civil war. In the UCDP-Prio dataset, however, the same civil war may start and stop repeatedly depending on the number of battle-related deaths it happens to generate each year, even when it "looks" like the same conflict in every other respect. As a result, months with a high but not-quite-high-enough incidence of low-level violence will be coded as peaceful, even if they occur after the month of onset. This will cause our model to erroneously associate low-level violence both with conflict (in the month preceding onset) *and* with peace (in the months between "new" onsets). The same problem will arise if a conflict does not cross the UCDP's battle-related deaths threshold for months or even years after it has obviously already begun. The Sambanis dataset avoids both of these problems.

Sambanis (2004, 829-31) defines civil war according to the following 11 criteria:

(a) The war takes place within the territory of a state that is a member of the international system with a population of 500,000 or greater.

(b) The parties are politically and militarily organized, and they have publicly stated political objectives.

(c) The government (through its military or militias) must be a principal combatant. If there is no functioning government, then the party representing the government internationally and/or claiming the state domestically must be involved as a combatant.

(d) The main insurgent organization(s) must be locally represented and must recruit locally. Additional external involvement and recruitment need not imply that the war is not intrastate. Insurgent groups may operate from neighboring countries, but they must also have some territorial control (bases) in the civil war country and/or the rebels must reside in the civil war country.

(e) The start year of the war is the first year that the conflict causes at least 500 to 1,000 deaths. If the conflict has not caused 500 deaths or more in the first year, the war is coded as having started in that year only if cumulative deaths in the next 3 years reach 1,000.

(f) Throughout its duration, the conflict must be characterized by sustained violence, at least at the minor or intermediate level. There should be no 3-year period during which the conflict causes fewer than 500 deaths.

(g) Throughout the war, the weaker party must be able to mount effective resistance. Effective resistance is measured by at least 100 deaths inflicted on the stronger party. A substantial number of these deaths must occur in the first year of the war. But if the violence becomes effectively one-sided, even if the aggregate effective-resistance threshold of 100 deaths has already been met, the civil war must be coded as having ended, and a politicide or other form of one-sided violence must be coded as having started.

(h) A peace treaty that produces at least 6 months of peace marks an end to the war.

(i) A decisive military victory by the rebels that produces a new regime should mark the end of the war. Because civil war is understood as an armed conflict against the government, continuing armed conflict against a new government implies a new civil war. If the government wins the war, a period of peace longer than 6 months must persist before we code a new war.

(j) A cease-fire, truce, or simply an end to fighting can also mark the end of a civil war if they result in at least 2 years of peace. The period of peace must be longer than what is required in the case of a peace agreement because we do not have clear signals of the parties' intent to negotiate an agreement in the case of a truce/cease-fire.

(k) If new parties enter the war over new issues, a new war onset should be coded, subject to the same operational criteria. If the same parties return to war over the same issues, we generally code the continuation of the old war, unless any of the above criteria for coding a war's end apply for the period before the resurgence of fighting.

Sambanis also codes the year and (importantly) month in which each conflict begins and ends. In some cases the month of onset follows immediately from the criteria above (e.g. in cases where fighting causes 1,000 deaths in the first month). In other cases, the month of onset indicates "the start of major armed conflict," regardless of whether the conflict crosses the 1,000 deaths threshold in the first month (Sambanis 2004, 830). For additional details, see the article and the 256-page online supplement.

A.4 Assessing ICEWS data quality: Iraq 2006 as a case study

Scholars have raised a number of concerns about the quality of ICEWS data (e.g. Schrodt and Van Brackle 2012), and of media-based event data more generally (e.g. Weidmann 2016). To explore the nature of measurement error in the ICEWS dataset—and, in particular, to assess the dataset's accuracy and comprehensiveness in identifying incidents of low-level violence—we look more closely at the events coded for a single country (Iraq) in a single year (2006). This is an exploratory exercise, and is intended only to help fix ideas on the limitations and idiosyncrasies of the ICEWS dataset. We focus on the coding of events surrounding the bombing of the al-Askari Mosque in the city of Samarra on February 22, 2006. This event is often referred to as the start of Iraq's sectarian civil war.

ICEWS data for the first three months of 2006 show a pattern of escalation. Daily events range from a low of 25 on January 10th to a high of 221 on February 22nd. The bombing is the first event coded on February 22nd: it is referred to indirectly in a BBC report of a demonstration called to condemn the bombing of the shrine. The next four rows in ICEWS also refer to the same incident, classified in two different ways: three are classified as "demonstrate or rally" (two of which are identical listings of the same headline) and one is classified as "make appeal or request." The bombing of the shrine is reported directly as "use conventional violence" in the fifth row, though the target is listed as an "Islamic Preacher" rather than the broader Shiite community; and the same text report is repeated as a different form of violence ("conduct suicide") with a different target ("Iraq"). About half of all events coded on that day are in some way related to the bombing, though many are duplicates of one another. Duplication takes several forms: the same event may be reported under different "event text" categories,[1] or the same (or similar) headlines may appear in two

[1]For example, the Samarra mosque bombing is reported as "Use unconventional violence" (at least five separate reports); "Conduct suicide, car, or other non-military bombing" (at

or more of the text sources from which the raw data are scraped. Casual inspection suggests that more than 75% of reports on this day are repeated 2 to 3 times and in some cases as many as 8 times.

Some events are also misclassified. For example, a headline about Muqtada al Sadr "cutting short a visit to Lebanon to return home after the bombing" is classified as an instance of "reducing or breaking diplomatic relations." There are 12 separate reports related to this event (e.g. reports of Muqtada crossing the border, or of the number of days his visit to Lebanon was supposed to last). In this case, the actions of a Shia leader are mischaracterized as reflecting the Iraqi government's diplomatic relations with its neighbors; further, that leader's response to a crisis at home is mischaracterized as a rupture in diplomatic relations abroad. Although duplication occurs across all "event text" categories, misclassification is more of a problem for some than for others. For example, "unconventional violence" (and other "low-level violence" categories that we use in our forecasts) are less likely than other categories to include content that is obviously unrelated to the escalation or de-escalation processes that they intend to capture.

Finally, one limitation of the ICEWS is that it cannot be used to capture dynamic action-reaction sequences between pairs of actors. Actions and reactions are not connected in a way that would allow for analysis of tit-for-tat escalation processes. For example, the Samarra mosque bombing in Iraq provoked a number of reprisal killings of Sunni clerics, but these reactions cannot be connected to the bombing except by reading the relevant headlines and hand-coding the connections—a prohibitively onerous task given the size of the dataset.

Overall, despite these limitations, a pattern of escalation is obvious in the Iraq data when one reads the headlines and event sentences scraped from the data sources. The bombing prompted diplomatic initiatives and statements, changes to planned meetings between po-

---

least four separate reports); "Make statement" (at least three separate reports). These counts do not include statements of blame attribution or other characterizations of the attack, nor do they include reports of actions that were a result of the bombing.

litical leaders, demonstrations and riots, violent retaliations, and increased security patrols to prevent further escalation. Although some events are duplicated and misclassified, the overall picture that emerges is of a country on the verge of escalating conflict.[2]

A.5 Descriptive statistics

Tables A.1, A.2 and A.3 report descriptive statistics for the ICEWS-based predictors in the *escalation*, *Quad* and *Goldstein* models, respectively. Because the *CAMEO* model includes over 1,000 predictors, we do not include descriptive statistics here. For each predictor we report the mean, median, minimum, maximum, standard deviation and skewness over all years in the dataset.

A.6 Time series plots for *escalation* model predictors

Figures A.1 through A.10 plot the frequency of each ICEWS-based predictor in the *escalation* model, focusing, for compactness, on countries that experienced at least one onset of civil war between January 1, 2008 and December 31, 2015 (the duration of the test set). Areas shaded in grey denote periods of civil war. The y-axis in each plot is scaled to accommodate the maximum frequency of the corresponding predictor.

A.7 Alternate coding rules for the dependent variable

In the manuscript we consider two alternate coding rules for the dependent variable. While the Sambanis dataset contains the most careful and explicit coding rules of any civil war dataset we are aware of, in some cases it is unclear when an ongoing civil war has ended or a

[2] Instead of being perceived as a type of measurement error, it might be possible to use duplication as a "weight" to indicate the importance of a given event. We leave this for future studies to explore.

new one has begun, and in a couple of cases it is unclear whether a civil war is ongoing at all. We therefore test the robustness of our results to four alternate coding rules corresponding to what we believe to be the most potentially problematic of these ambiguities:

1. the Colombian civil war is coded as ending in September 2015 (in the manuscript we code the Colombian civil war as ongoing through December 31, 2015);

2. Turkey is coded as experiencing a second onset of civil war in June 2014;

3. Russia is coded as not having a civil war at all (in the manuscript Russia is coded as having an ongoing civil war throughout the training and test sets); and

4. India is coded as not having a civil war at all (in the manuscript India is coded as having an ongoing civil war throughout the training and test sets as well).

We make these two latter changes because, while India and Russia both have ongoing conflicts with separatist groups, the conflicts are so localized and low-intensity that they are arguably qualitatively different from the others in the dataset. Results from models with this first alternate coding are labeled "Coding of DV 1."

Also, while the Sambanis dataset codes the month of onset for every civil war, these decisions inevitably involve some discretion. In general we were unable to identify any obvious, empirically-grounded alternatives to the months of onset that Sambanis codes. Nevertheless, as a robustness check we recode all civil wars as beginning in January of the year of onset, and as ending in December of the year of termination. This has the effect of lengthening all civil wars in the dataset. Results from models with this second alternate coding are labeled "Coding of DV 2."

A.8 Alternate performance metrics

As we discuss in the manuscript, there are a number of reasons to prefer the AUC over other potential metrics for gauging model performance, at least in our case. In particular, the

AUC is intuitive and easy to interpret even for non-specialists; it incorporates true and false positives (Type I errors) and true and false negatives (Type II errors) simultaneously; it does not require that we select an arbitrary discrimination threshold above which we predict civil war will occur; it provides a single summary metric to facilitate comparison across models; and it allows us to detect differences between models even when the dependent variable is extremely rare and the predicted probabilities of onset are, accordingly, extremely low.

AUC is not, however, the only option. In the manuscript we display separation plots as well, which are also intuitive, and which contain information that the AUC does not (for example, the difference in predicted probability between true positives and false negatives). We also provide precision-recall curves. Another alternative is the Brier score (Brier 1950), defined as the average squared deviation of each predicted probability from the true observed value of the dependent variable (0 or 1). The smaller the Brier score, the better the model. Brier scores have been used (albeit sparingly) in the conflict forecasting literature (e.g. Montgomery, Hollenbach and Ward 2012), and are easy to interpret. As we show in Table A.4, however, because our dependent variable is so rare and our predicted probabilities so small, the Brier score is indistinguishably close to zero for all models—a highly favorable but ultimately uninformative gauge of our models' relative performance. As others have noted (e.g. Chiba and Gleditsch 2017, 291), the Brier score simply is not useful for such highly imbalanced data.

Another alternative is the F1 score, which is the balanced harmonic mean of precision and recall.[3] The higher the F1 score, the better the model. As we show in Table A.4, the *escalation* model outperforms all the alternatives over 1-month windows, and outperforms two of the four alternatives over 6-month windows. However, the maximum F1 score is generally very similar across models—likely an artifact, again, of the rarity of the dependent

[3]Precision is the ratio of true positives to the sum of true and false positives at a particular discrimination threshold; recall is the ratio of true positives to the sum of true positives and false negatives. The F1 score is calculated $\frac{2\times(\text{Recall}\times\text{Precision})}{(\text{Recall}+\text{Precision})}$.

variable, and of the fact that the F1 score does not incorporate true negatives.

A.9 Alternate specifications for the *escalation* model

Our base specification of the *escalation* model consists of ten ICEWS-based predictors. We consider two alternative specifications here. First, given that several of the *escalation* predictors are very sparsely distributed (see Figures A.1 through A.10), we consider an alternate specification that relies on the four "low-level violence" predictors alone. Second, given that the process of escalation is likely to evolve over multiple months prior to the onset of civil war, we consider another alternate specification that adds three lags of our low-level violence predictors to the other ten (i.e. low-level violence two months, three months and four months prior to onset). Table A.5 reports AUCs for these two models over 1-month (top panel) and 6-month (bottom panel) windows. Neither model outperforms the base specification reported in the manuscript.

A.10 Variable importance

Figure A.11 ranks the 10 predictors in the *escalation* model by their importance scores. Figures A.12, A.13 and A.14 plot importance scores for the predictors in the *Quad*, *Goldstein* and *CAMEO* models, respectively. (For ease of exposition we report only the top 30 most important predictors in the *CAMEO* model.[4]) Importance scores are calculated by permuting the values of each predictor across observations in the training set, then estimating the impact of those permutations on "out-of-bag" accuracy over all trees in the forest. While importance scores have been widely used, including in several recent studies in political science (Hill and Jones 2014; Muchlinski et al. 2015), they are prone to bias and

[4]Definitions of specific CAMEO codes can be found in Philip A. Schrodt's "CAMEO Event and Actor Codebook," available at `http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf`, accessed February 23, 2017.

should be interpreted with some caution.[5] Following Strobl and Zeileis (2008), we report raw (unscaled) permutation-based scores, rather than scaled or Gini-based scores.[6] Importance scores do not lend themselves to easy substantive interpretation, but their ordinal ranking and differences in magnitude are nonetheless indicative of relative predictive power.

The patterns in Figure A.11 are generally consistent with theories of escalation described in the manuscript. Over both 1-month and 6-month windows, violent protests and other small-scale confrontations between government and opposition groups are the most reliable predictors of civil war onset. This is consistent with the logic of escalation, whereby increasing low-level violence marks the transition from mobilization to rebellion. Moreover, with only one exception (accommodations in the 6-month model), events between governments and opposition groups are consistently better predictors than events between governments and rebels. Over 6-month windows in particular, non-violent repression of opposition groups is a better predictor of onset than low-level violence committed by rebels against government, and a much better predictor than low-level violence committed by government against rebels. This, too, resonates with the logic of escalation: conflict between governments and opposition groups should be logically and empirically prior to conflict between governments and rebels, which should occur primarily during (rather than before) civil war.

[5]Bias occurs, for example, when predictors are highly correlated, or when predictors vary in their scale of measurement or number of categories (or, in the case of continuous predictors, their number of unique values) (Strobl et al. 2007).

[6]Another option is to compute a modified version of the importance score that adjusts for random forests' tendency to prefer correlated predictors. We found this modification empirically intractable in a dataset this large (and, in the case of the *CAMEO* model, with this many predictors), and opt for the raw permutation-based score as the best available alternative.

References

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78(1):1–3.

Chiba, Daina and Kristian Skrede Gleditsch. 2017. "The shape of things to come? Expanding the inequality and grievance model for civil war forecasts with event data." *Journal of Peace Research* 54(2):275–297.

Hill, Daniel W. Jr. and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(03):661–687.

Montgomery, Jacob M., Florian M. Hollenbach and Michael D. Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20(3):271–291.

Muchlinski, David, David Siroky, Jingrui He and Matthew Kocher. 2015. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24(1):87–103.

Sambanis, Nicholas. 2004. "What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution* 48(6):814–858.

Schrodt, Philip A. and David Van Brackle. 2012. Automated Coding of Political Event Data. In *Handbook of Computational Approaches to Counterterrorism*, ed. V. S. Subrahmanian. New York: Springer pp. 23–49.

Strobl, Carolin and Achim Zeileis. 2008. "Danger: High Power! – Exploring the Statistical Properties of a Test for Random Forest Variable Importance." University of Munich Department of Statistics Technical Report 017.
URL: *https://epub.ub.uni-muenchen.de/2111/*

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis and Torsten Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics* 8(1).

Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60(1):206–218.

Table A.1: Descriptive statistics for *escalation* model predictors

	Mean	Median	Min.	Max.	S.D.	Skewness
government -> opposition low-level violence	0.90	0.00	0.00	223.00	5.03	20.22
opposition -> government low-level violence	1.35	0.00	0.00	343.00	6.66	22.72
government -> rebel low-level violence	0.19	0.00	0.00	123.00	1.51	30.69
rebel -> government low-level violence	1.38	0.00	0.00	238.00	7.03	10.86
government -> opposition non-violent repression	0.04	0.00	0.00	17.00	0.35	16.67
government -> rebel non-violent repression	0.01	0.00	0.00	36.00	0.27	84.40
government -> opposition accommodations	0.04	0.00	0.00	33.00	0.37	34.64
government -> rebel accommodations	0.02	0.00	0.00	34.00	0.31	64.67
rebel -> government demands	0.00	0.00	0.00	2.00	0.02	65.11
opposition -> government demands	0.00	0.00	0.00	2.00	0.03	51.79

Notes: Descriptive statistics for CAMEO-based predictors in the *escalation* model.

Table A.2: Descriptive statistics for *quad* model predictors

	Mean	Median	Min.	Max.	S.D.	Skewness
government -> opposition verbal conflict	0.65	0.00	0.00	75.00	2.41	11.17
government -> rebel verbal conflict	0.57	0.00	0.00	152.00	3.38	17.02
government -> opposition material conflict	3.23	0.00	0.00	383.00	10.67	10.84
government -> rebel material conflict	2.12	0.00	0.00	326.00	11.28	10.82
opposition -> government verbal conflict	2.40	0.00	0.00	570.00	11.99	20.63
rebel -> government verbal conflict	0.44	0.00	0.00	110.00	2.34	14.13
opposition -> government material conflict	0.77	0.00	0.00	62.00	2.82	7.58
rebel -> government material conflict	2.10	0.00	0.00	464.00	11.17	13.06
government -> opposition verbal cooperation	2.72	0.00	0.00	305.00	8.05	10.08
government -> rebel verbal cooperation	1.94	0.00	0.00	347.00	9.34	12.38
government -> opposition material cooperation	0.47	0.00	0.00	58.00	1.61	9.79
government -> rebel material cooperation	0.14	0.00	0.00	52.00	0.98	20.98
opposition -> government verbal cooperation	1.41	0.00	0.00	193.00	5.26	13.09
rebel -> government verbal cooperation	1.06	0.00	0.00	234.00	5.74	15.37
opposition -> government material cooperation	0.06	0.00	0.00	40.00	0.47	35.88
rebel -> government material cooperation	0.13	0.00	0.00	65.00	1.04	25.55

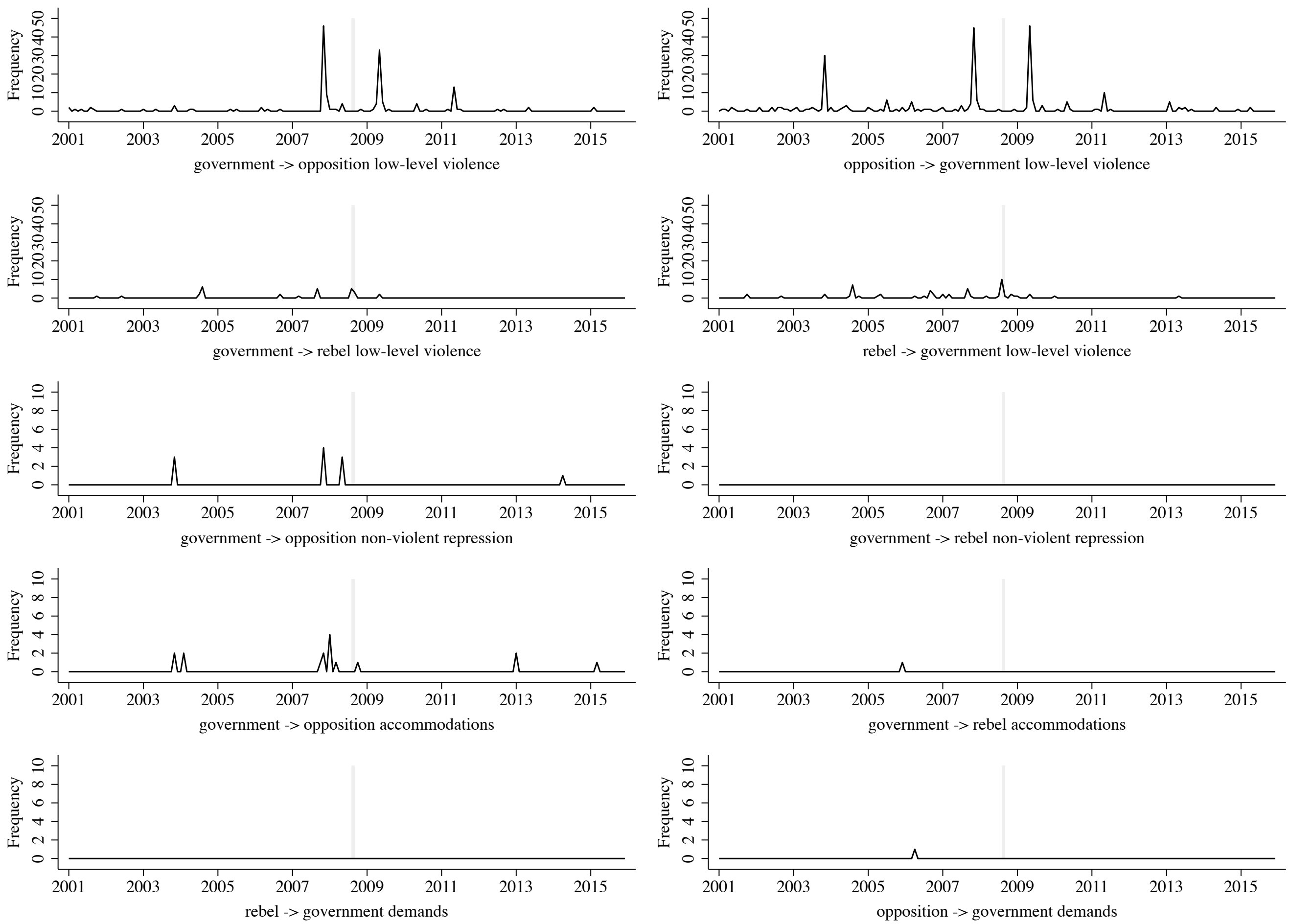
Notes: Descriptive statistics for CAMEO-based predictors in the *quad* model.

Table A.3: Descriptive statistics for *Goldstein* model predictors

	Mean	Median	Min.	Max.	S.D.	Skewness
government -> opposition Goldstein score	-22	0	-3,551	647	80	-12
government -> rebel Goldstein score	-17	0	-2,942	866	106	-11
opposition -> government Goldstein score	-18	0	-3,758	624	83	-17
rebel -> government Goldstein score	-17	0	-4,340	898	103	-13

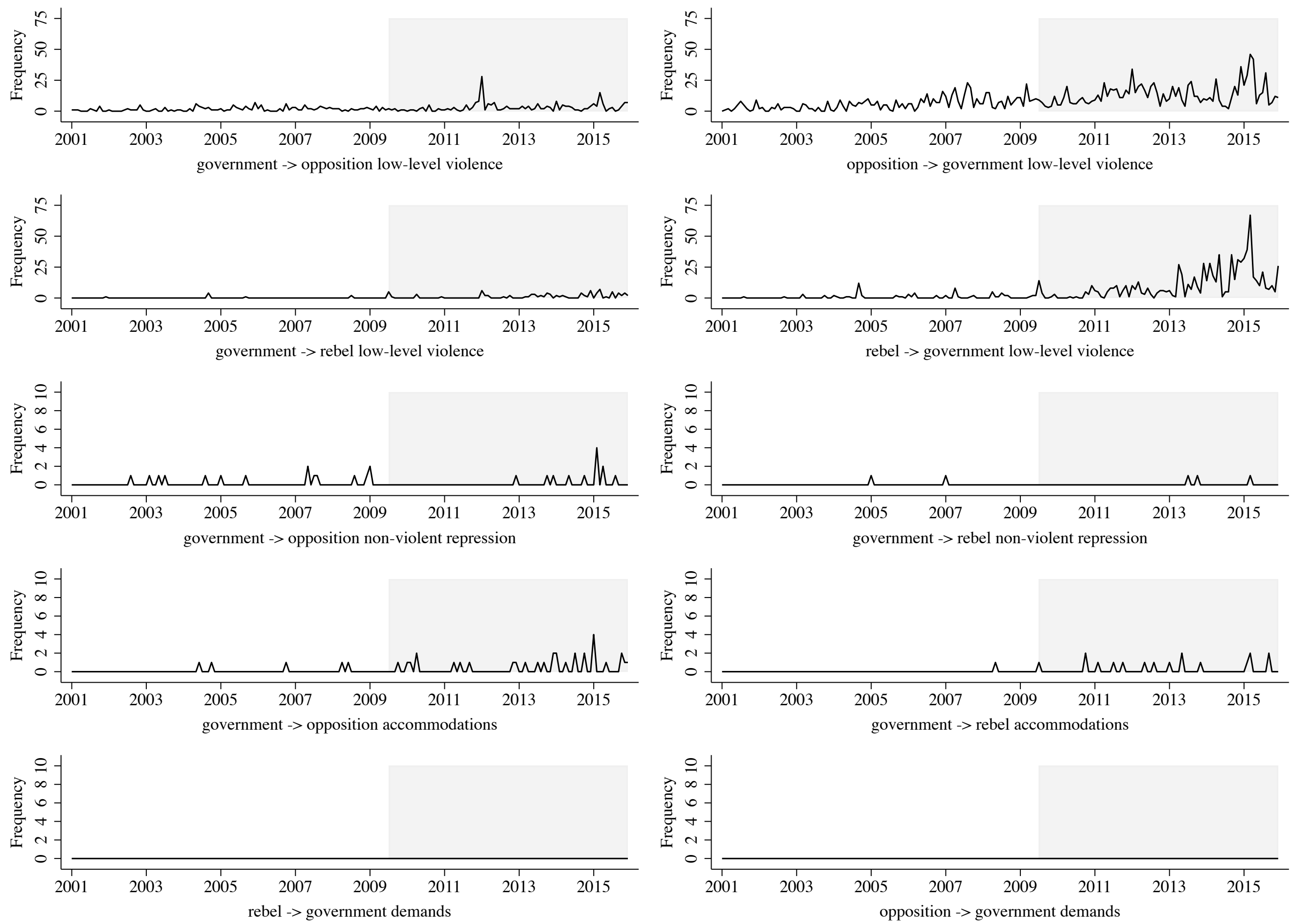
Notes: Descriptive statistics for CAMEO-based predictors in the *Goldstein* model.

Figure A.1: Time series plot of *escalation* model predictors for Georgia



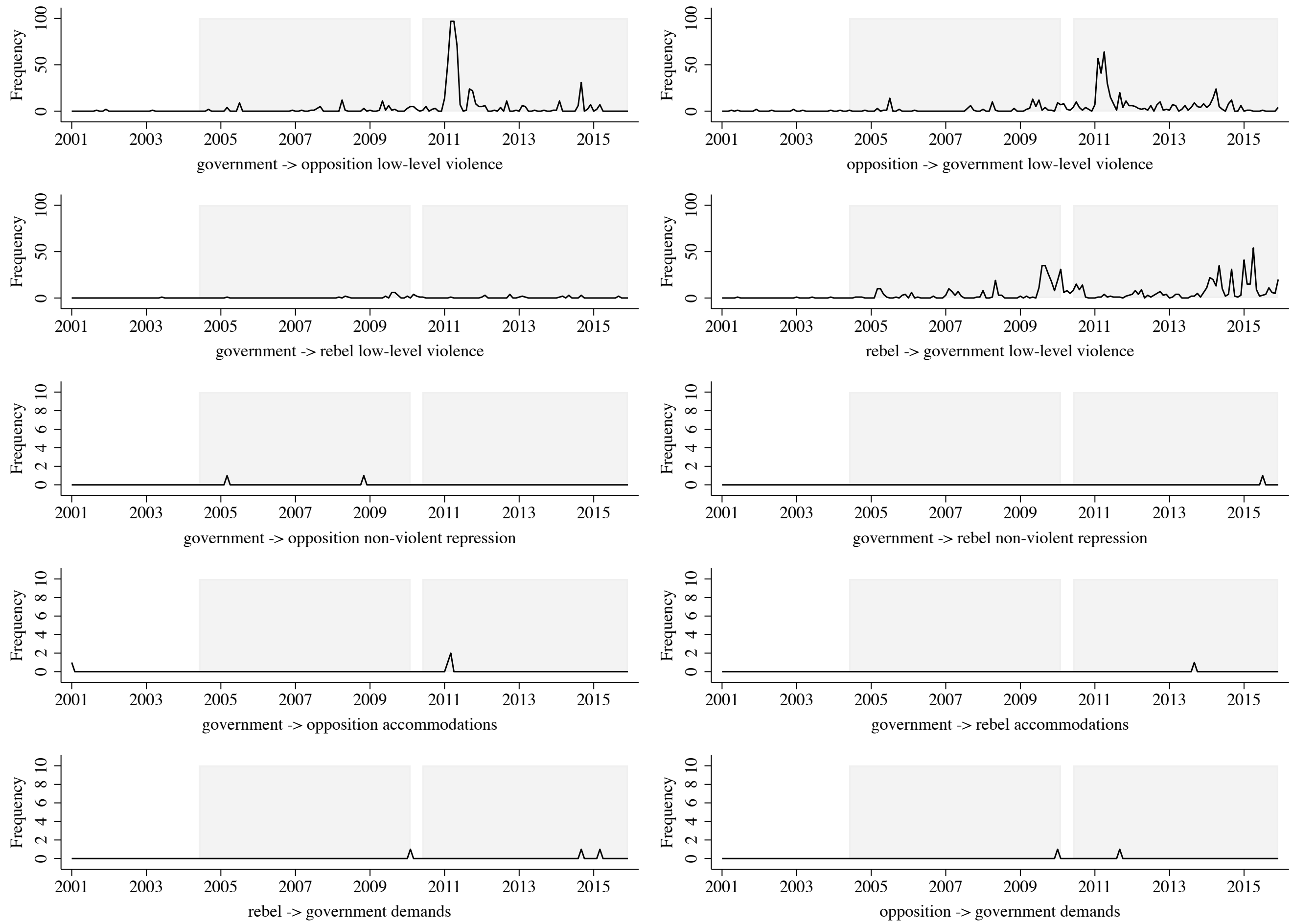
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Georgia. Months shaded grey denote periods of civil war.

Figure A.2: Time series plot of *escalation* model predictors for Nigeria



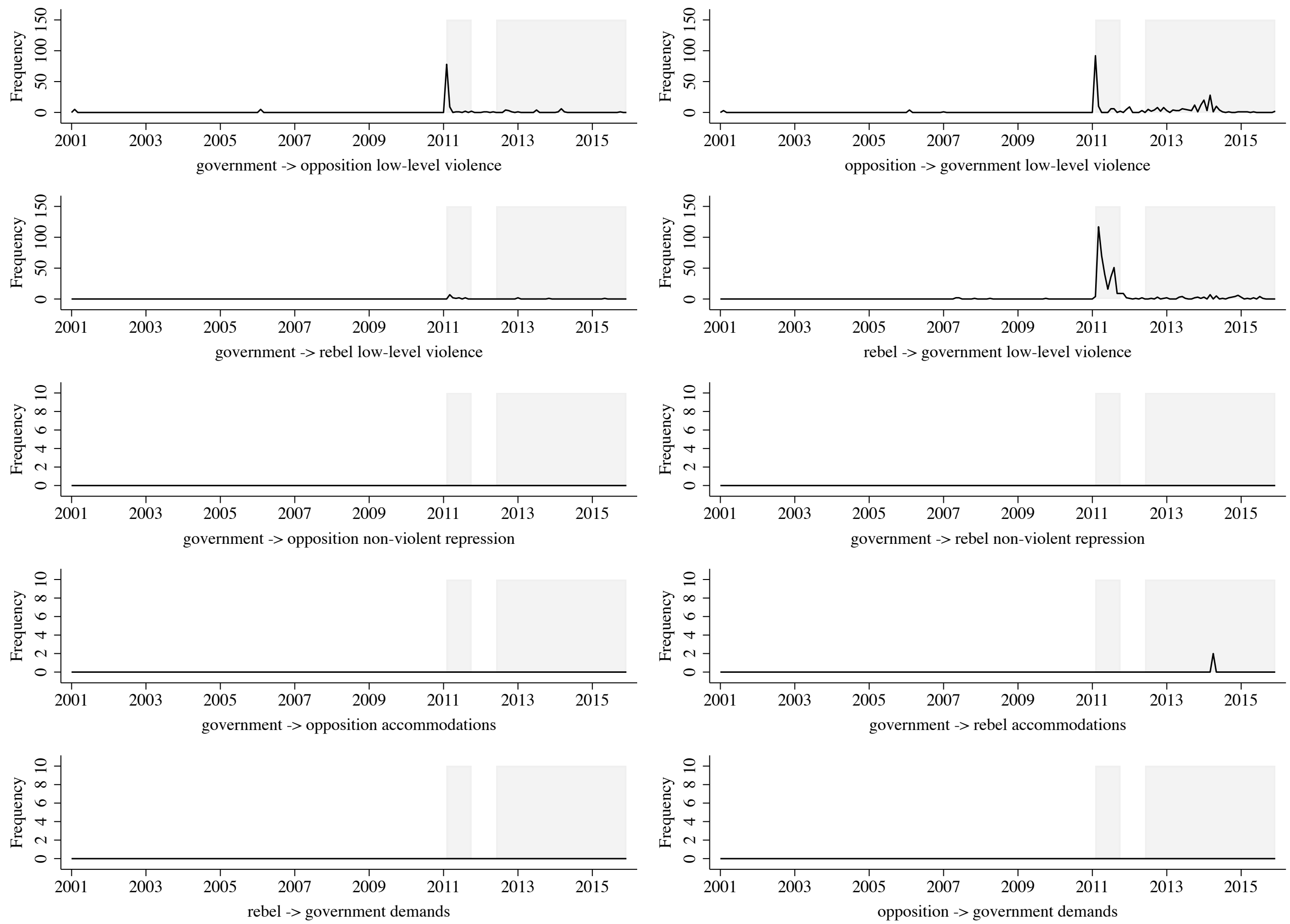
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Nigeria. Months shaded grey denote periods of civil war.

Figure A.3: Time series plot of *escalation* model predictors for Yemen



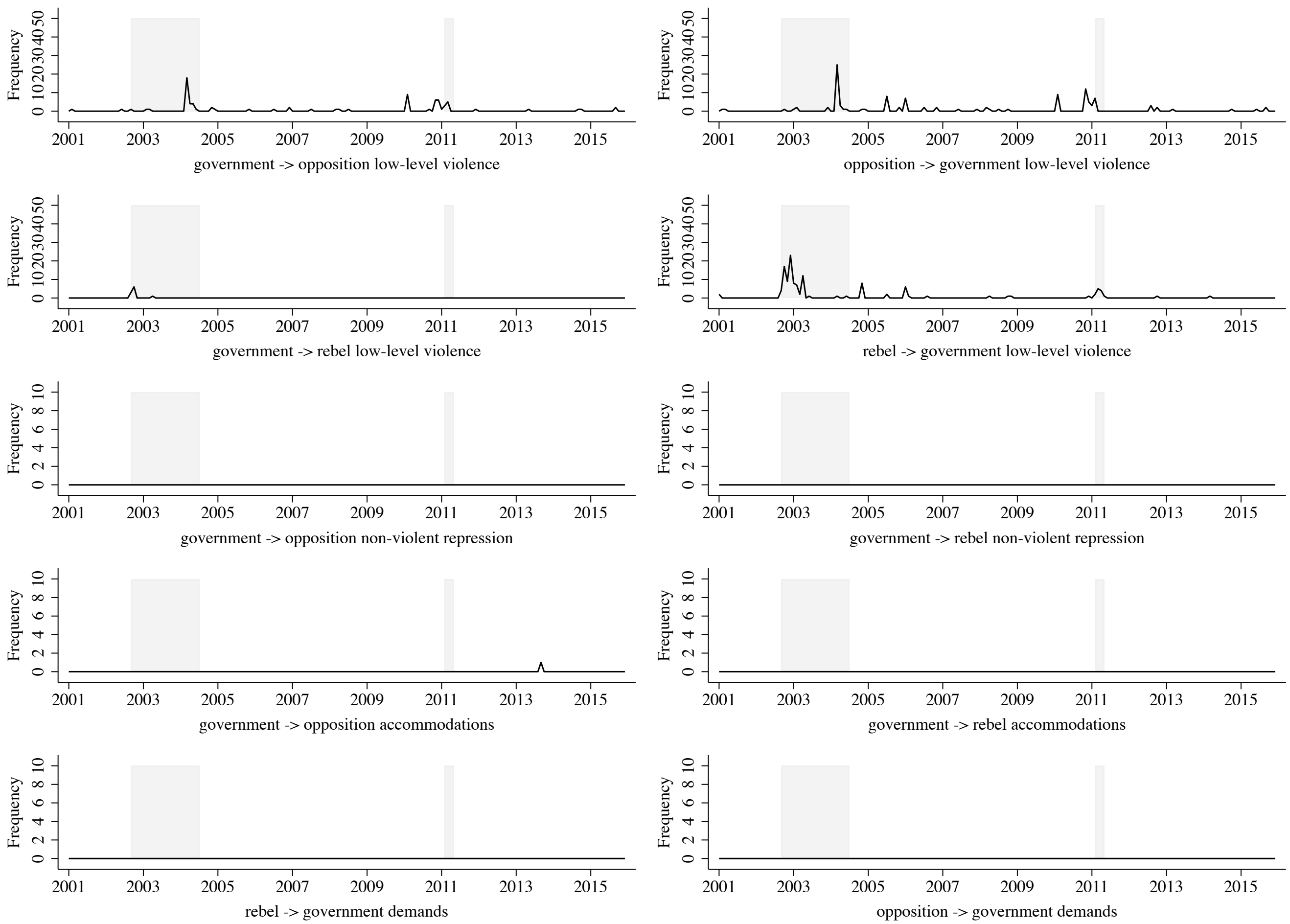
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Yemen. Months shaded grey denote periods of civil war.

Figure A.4: Time series plot of *escalation* model predictors for Libya



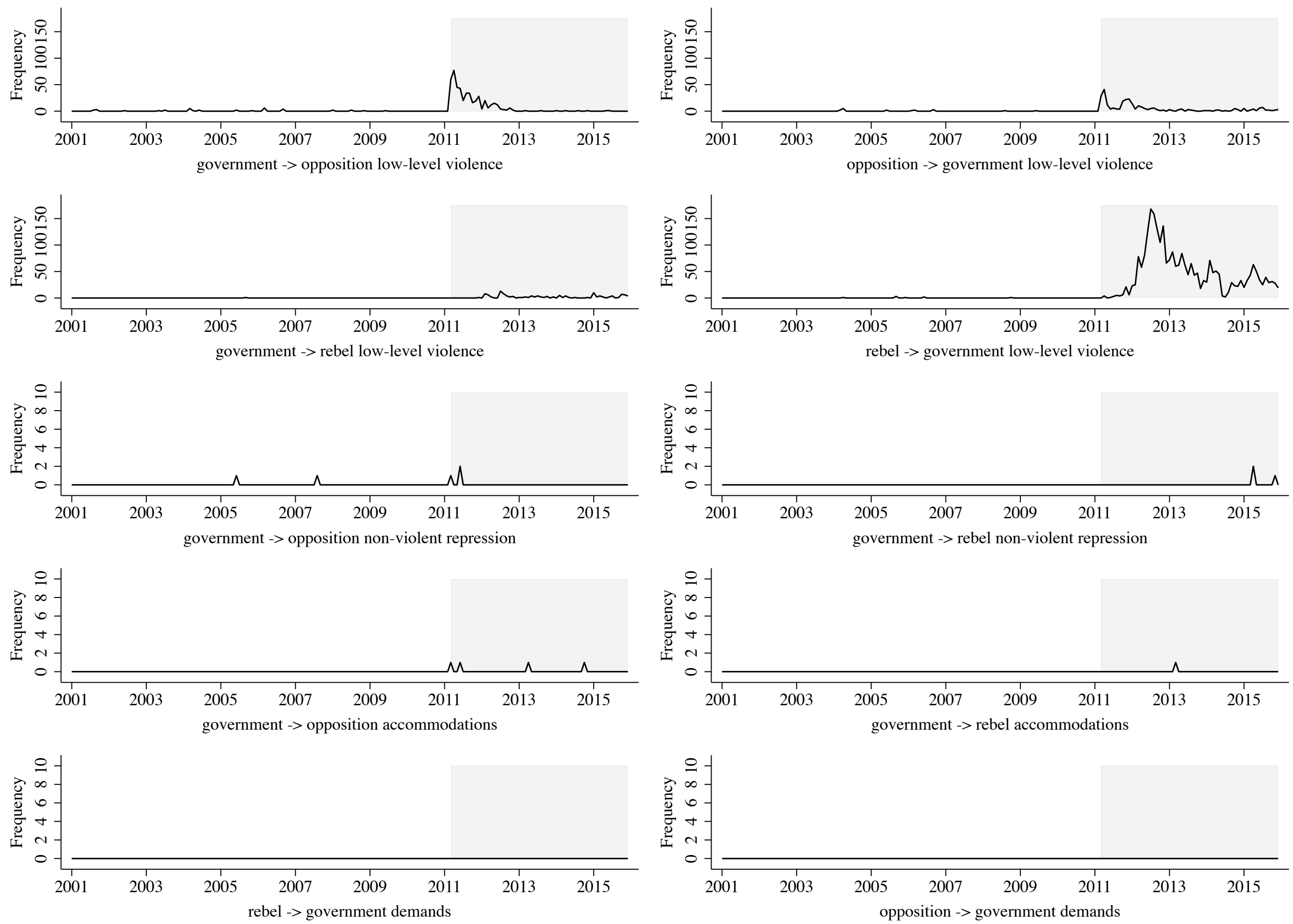
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Libya. Months shaded grey denote periods of civil war.

Figure A.5: Time series plot of *escalation* model predictors for Cote d'Ivoire



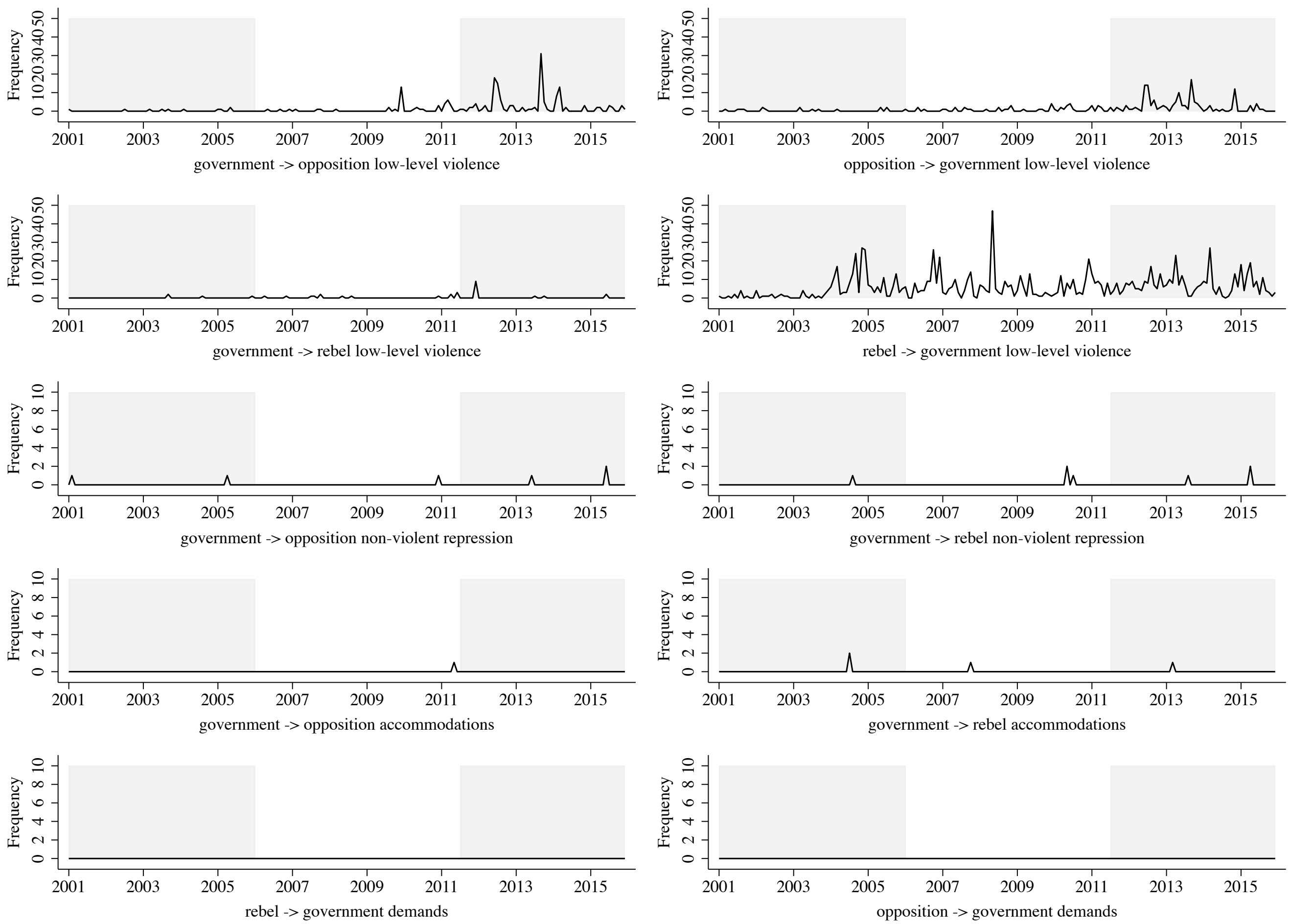
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Cote d'Ivoire. Months shaded grey denote periods of civil war.

Figure A.6: Time series plot of *escalation* model predictors for Syria



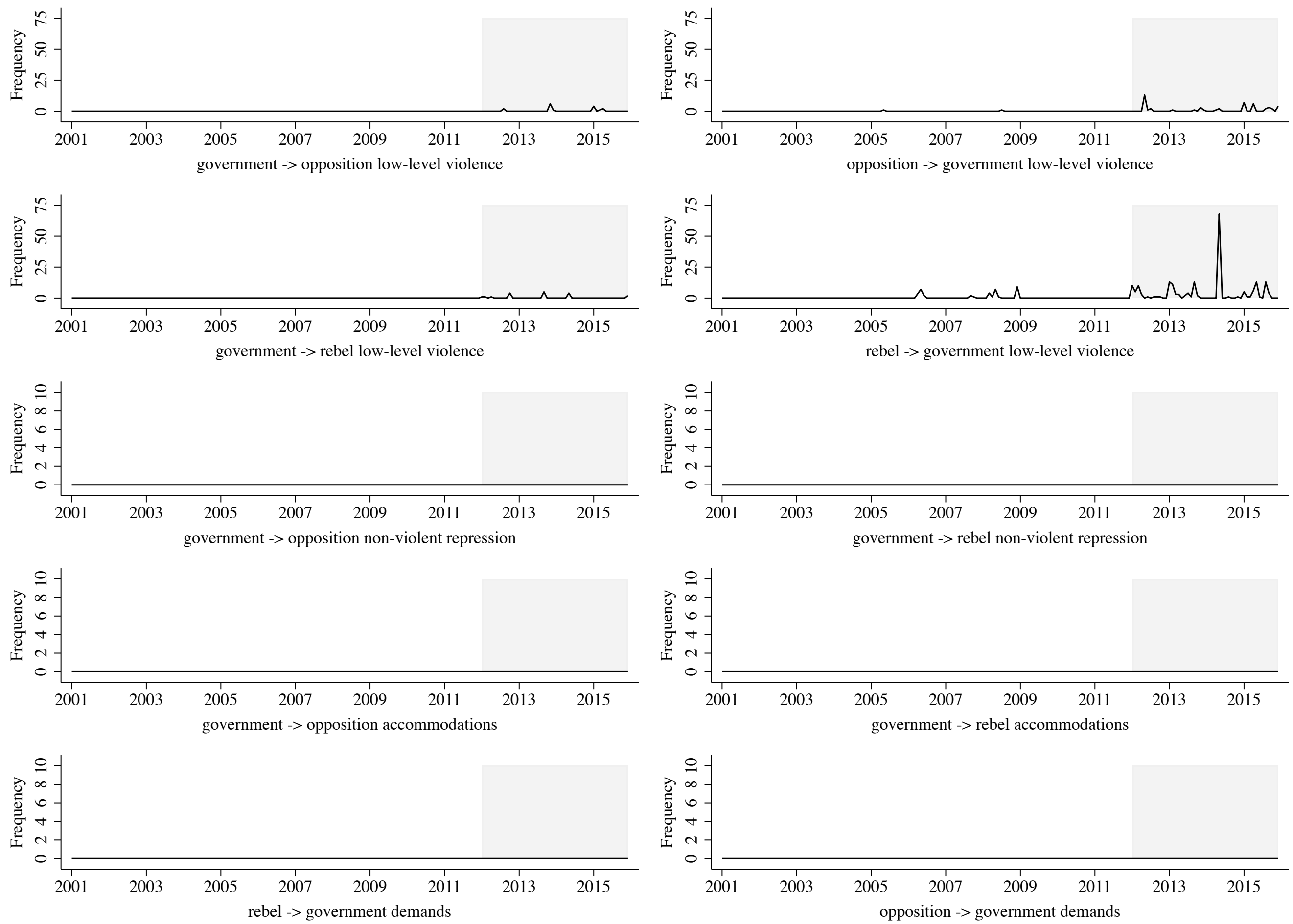
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Syria. Months shaded grey denote periods of civil war.

Figure A.7: Time series plot of *escalation* model predictors for Sudan



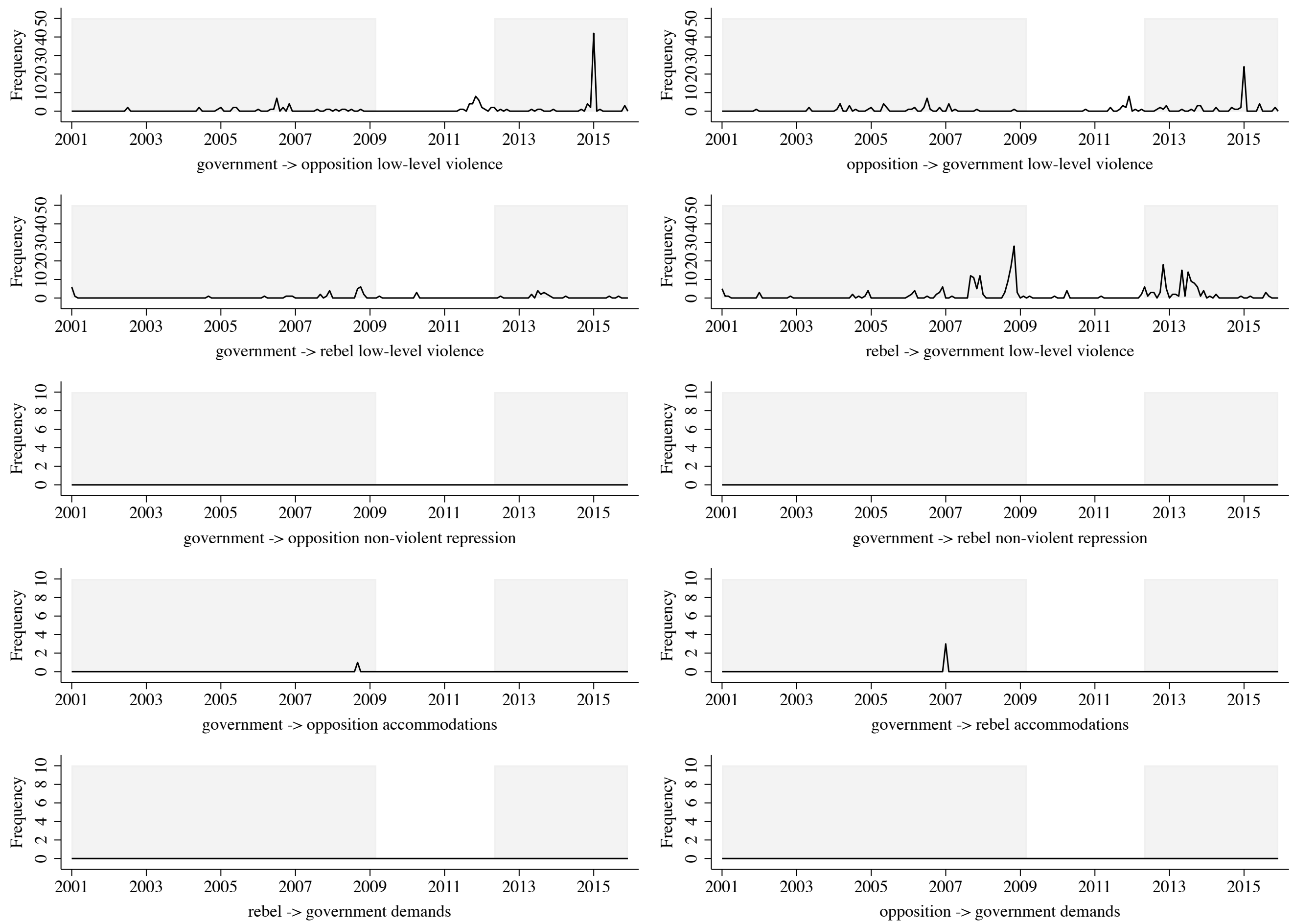
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Sudan. Months shaded grey denote periods of civil war.

Figure A.8: Time series plot of *escalation* model predictors for Mali



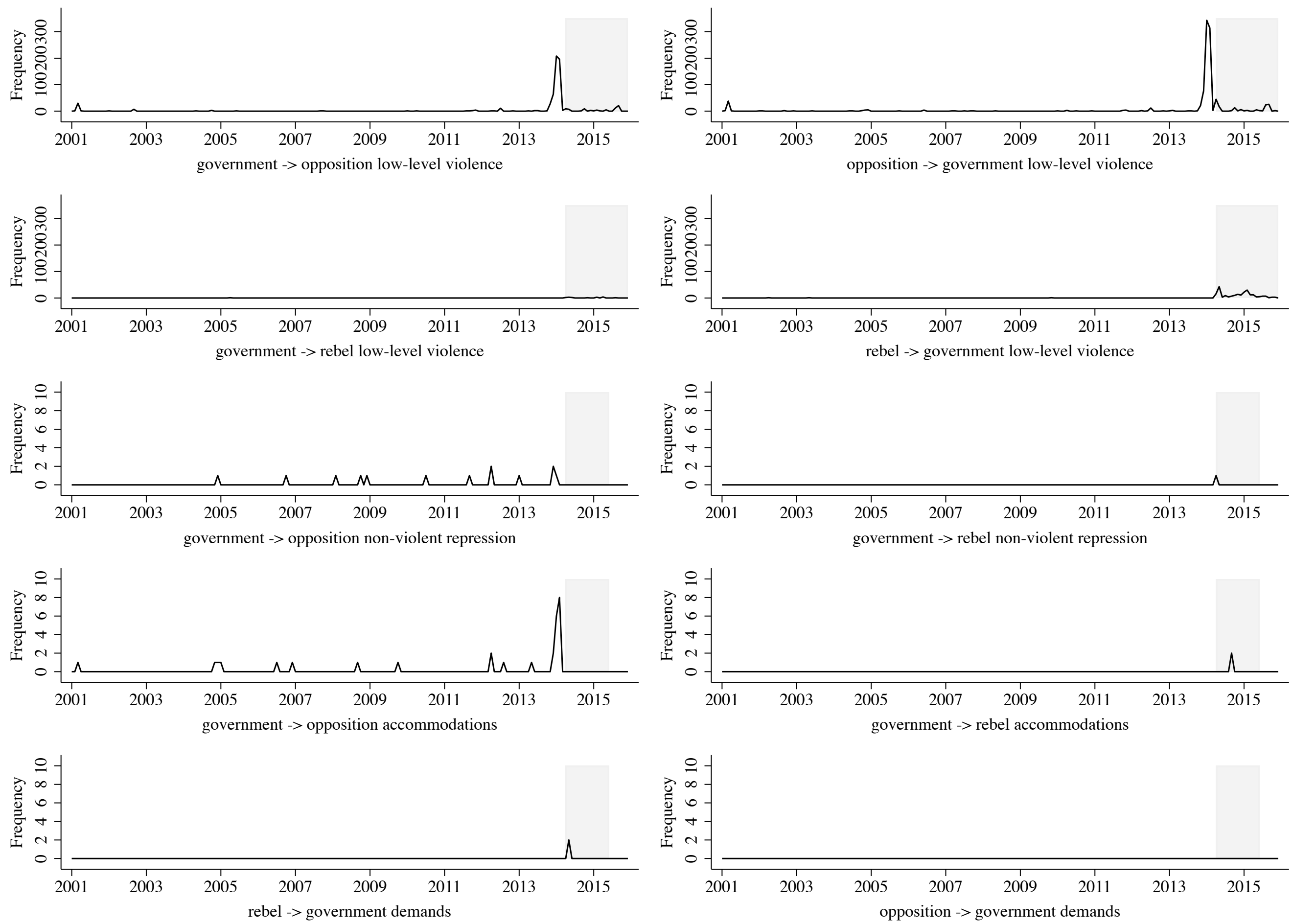
Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Mali. Months shaded grey denote periods of civil war.

Figure A.9: Time series plot of *escalation* model predictors for DRC



Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for the Democratic Republic of Congo. Months shaded grey denote periods of civil war.

Figure A.10: Time series plot of *escalation* model predictors for Ukraine



Notes: Time series frequency plot of CAMEO-based predictors in the *escalation* model for Ukraine. Months shaded grey denote periods of civil war.

Table A.4: Out-of-sample Brier and F1 scores

1-month forecasts

	Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
Brier score	0.0032	0.0032	0.0032	0.0032	0.0032
Maximum F1 score	0.0615	0.0513	0.0426	0.0335	0.0377

6-month forecasts

	Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
Brier score	0.0070	0.0070	0.0070	0.0070	0.0070
Maximum F1 score	0.1875	0.1622	0.2222	0.1714	0.2000

Notes: Brier and maximum F1 scores for our five random forests models. The top and bottom panels report results from 1-month and 6-month forecasts, respectively.

Table A.5: Out-of-sample AUCs for alternate *escalation* model specifications

1-month forecasts

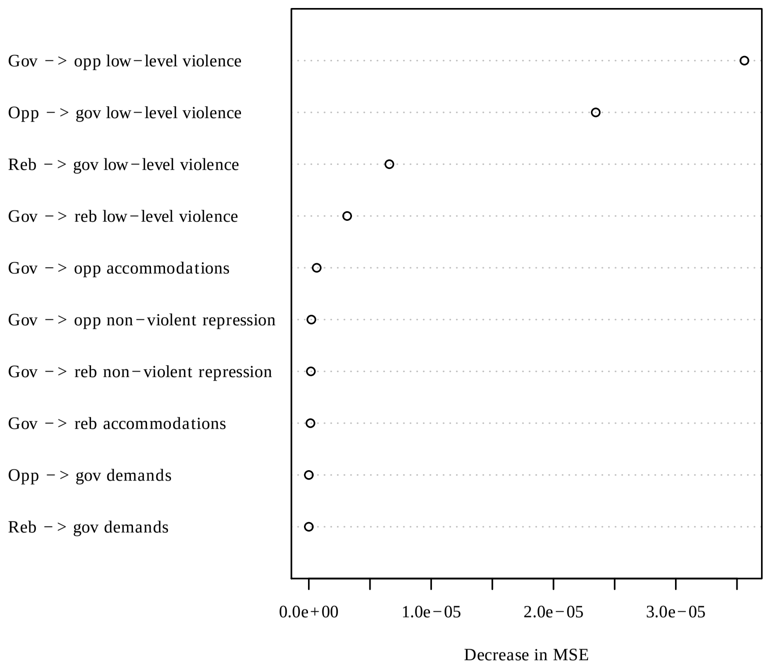
Low-level violence only	*Escalation with lags*
0.84	0.81

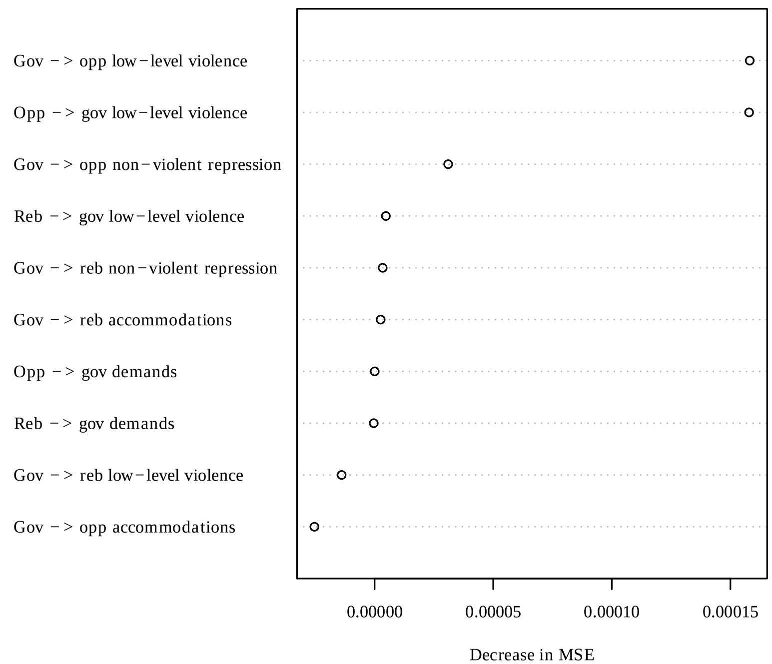
6-month forecasts

Low-level violence only	*Escalation with lags*
0.78	0.74

Notes: AUCs for alternate *escalation* models. The top and bottom panels report results from 1-month and 6-month forecasts, respectively.

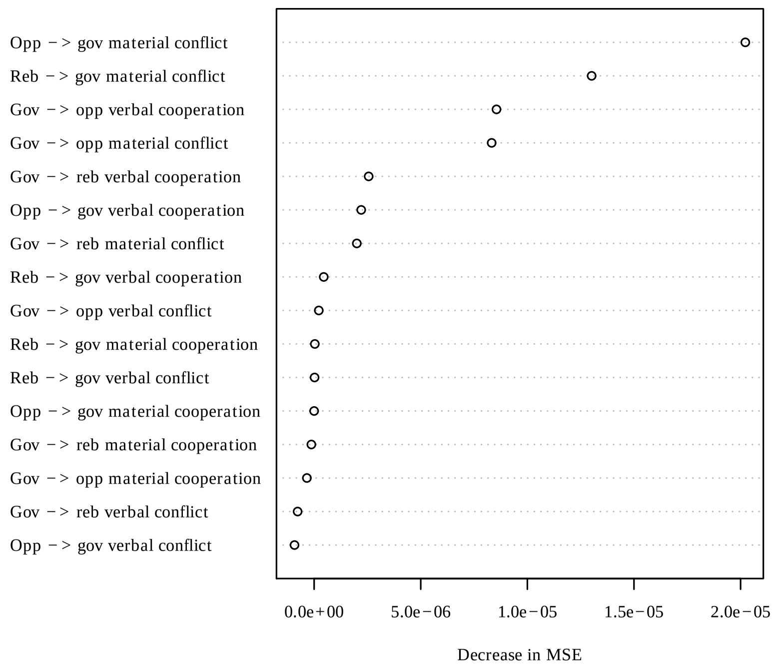
Figure A.11: Random forest importance score rankings for *escalation* model

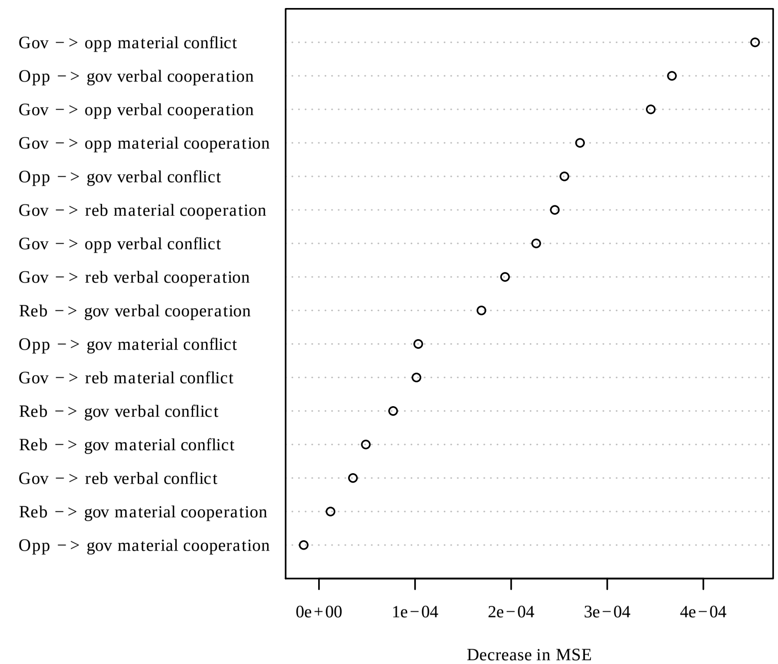




Notes: Ranking of predictors from the *escalation* model by their importance scores. The left and right plots show results from 1-month and 6-month forecasts, respectively.

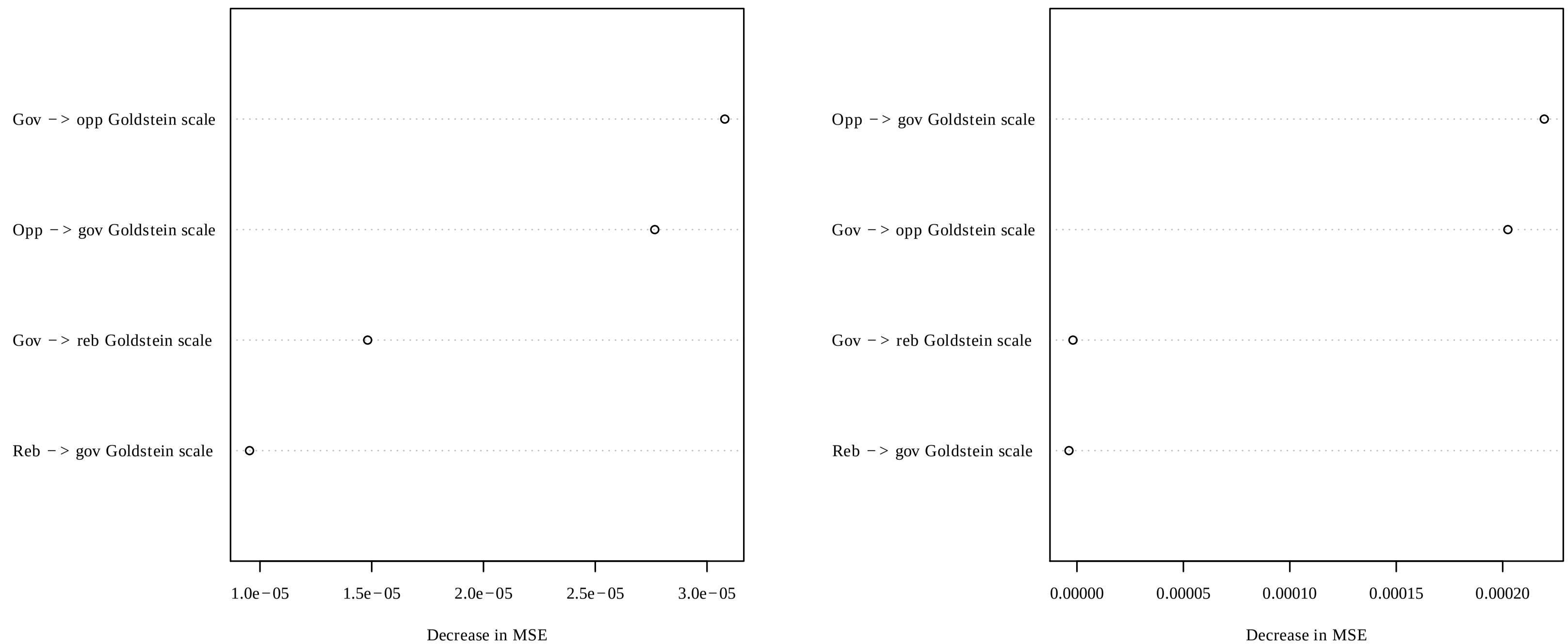
Figure A.12: Random forests importance score rankings for *Quad* model





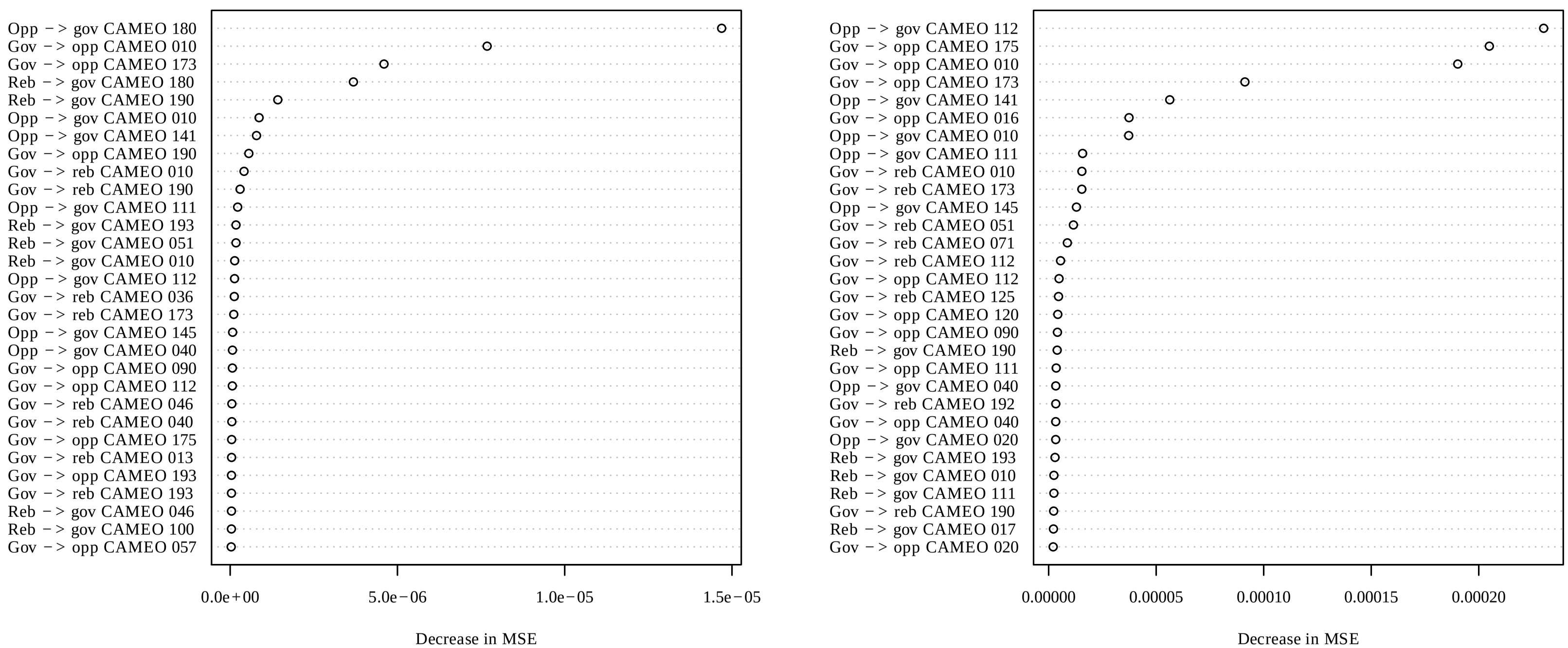
Notes: Ranking of predictors from the *Quad* model by their importance scores. The left and right plots show results from 1-month and 6-month forecasts, respectively.

Figure A.13: Random forest importance score rankings for *Goldstein* model



Notes: Ranking of predictors from the *Goldstein* model by their importance scores. The left and right plots show results from 1-month and 6-month forecasts, respectively.

Figure A.14: Random forest importance score rankings for *CAMEO* model (top 30 only)



Notes: Ranking of predictors from the *CAMEO* model by their importance scores. The left and right plots show results from 1-month and 6-month forecasts, respectively.