UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Using Past Violence and Current News to Predict Changes in Violence

Hannes Mueller
Barcelon School of Economics

Christopher Rauh
University of Cambridge

Abstract

This article proposes a new method for predicting escalations and de-escalations of violence using a model which relies on conflict history and text features. The text features are generated from over 3.5 million newspaper articles using a so-called topic-model. We show that the combined model relies to a large extent on conflict dynamics, but that text is able to contribute meaningfully to the prediction of rare outbreaks of violence in previously peaceful countries. Given the very powerful dynamics of the conflict trap these cases are particularly important for prevention efforts.

Reference Details

2220 Cambridge Working Papers in Economics
2022/09 Janeway Institute Working Paper Series

Published 22 March 2022

Key Words Conflict, prediction, machine learning, LDA, topic model, battle deaths, ViEWS prediction competition, random forest

JEL Codes F21, C53, C55

Websites www.econ.cam.ac.uk/cwpe
www.janeway.econ.cam.ac.uk/working-papers

Using Past Violence and Current News to Predict Changes in Violence[1]

Hannes Mueller (IAE-CSIC, Barcelon School of Economics)

Christopher Rauh (University of Cambridge)

Abstract: This article proposes a new method for predicting escalations and de-escalations of violence using a model which relies on conflict history and text features. The text features are generated from over 3.5 million newspaper articles using a so-called topic-model. We show that the combined model relies to a large extent on conflict dynamics, but that text is able to contribute meaningfully to the prediction of rare outbreaks of violence in previously peaceful countries. Given the very powerful dynamics of the conflict trap these cases are particularly important for prevention efforts.

1. Introduction

The advance of machine learning into the Social Sciences promises new opportunities for both data generation (feature extraction) and data analysis. A real strength of standard implementations of machine learning methods is that they come with the ability for dimensionality reduction, regularization and cross validation and this has had a big impact on the field of conflict forecasting.[2]

The forecasting competition organized by the political violence early-warning system (ViEWS) offers a new step in this direction (Hegre et al 2022). Not only are models trained and tested through cross validation, but the competition offers a commitment device to use a so-called hold-out dataset by forcing the competitors to submit their predictions for the unknowable future. This strengthens the incentives to develop and submit models which are not overfitted to the existing data and thereby become useful for the actual policy task of preventing escalations.

In this article we present our forecasting model developed for the competition. Our methodology is based on previous work (Mueller and Rauh 2018, 2020) but changes crucial aspects to adapt to the task of forecasting continuous values of violence throughout the conflict cycle, i.e., even when conflict has already broken out. Our method combines features generated from the violence history with features extracted from newspaper text. According to the comparison article to this special issue our method achieved one of the lowest mean squared errors (MSE), the measure we trained for, at the country level (Vesco et al (2022)).

We argue that, depending on which part of the conflict cycle is to be predicted, the researcher faces different tasks. Within or after conflict, i.e., with ongoing conflict or in the years after a conflict has ended, the prediction can rely on a rich set of conflict history features. Models which are able to identify and combine these features effectively will provide a very good forecast. Here is where machine learning can help tremendously. But before conflict breaks out, the researcher actually faces a much harder task because conflict history features cannot be used for prediction. At the same time, the number of cases falls. The problem becomes a *small data* problem in which theoretical priors and

[1] We thank the organizers of the forecasting competition for their guidance when developing this project. This is an important public service. We also thank the editor and referees for their feedback to an earlier version of this article. All errors are ours. We thank Bruno Conte Leite and Luis Ignacio for excellent research assistance. Hannes Mueller acknowledges financial support from the Banco de España and the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D Barcelona School of Economics CEX2019-000915-S). The replication files to this article can be found in the Dataverse page https://doi.org/10.7910/DVN/BW7UV4.

[2] For overviews see, Schrodt et al (2013), Ward et al (2013), and Hegre et al (2017).

small models will be more useful and machine learning cannot provide a solution to the regularization problem but only discipline the approach.

We then present our methodology which is based on a corpus of millions of newspaper articles. Following our previous work, we perform dimensionality reduction using the Latent Dirichlet Allocation (LDA) (Blei et al 2003). We show that the topic model has the striking ability to incorporate new phenomena like the Covid-19 pandemic and endogenously interprets it correctly as a health emergency. This is a huge advantage compared to dictionary methods which might interpret increases in the mentions of words like "kill" or "death" as driven by conflict. Moreover, text is available in real-time and provides extensive within country variation over time, in contrast to, for instance, socio-economic variables.

Our results are in line with the idea that outbreaks of violence are particularly hard to predict. Our model shows relatively low forecast error when forecasting during peacetime and offers relatively minor fluctuations of escalation values in these periods. This means, however, that onsets of violence which occur during this time are not anticipated. Anticipating escalations that occur after long peaceful periods is extremely hard and leads to very large error values. Here is where text provides the most notable improvement – by making outbreaks of violence less of a surprise.

In what follows we first discuss the key role of conflict dynamics for the forecasting task and how these dynamics make forecasting escalations of violence after peacetime very difficult. Based on this discussion, we then present our methodology. Section 4 presents results.

2. Conflict Dynamics and the Small Data Problem

Figure 1 displays conflict dynamics graphically in the entire UCDP data between 1989 and 2020 by categorizing each observation by quintiles and displaying all transitions from one month to the next. Transitions to and from no violence are depicted separately as 0 as they represent the majority of data. Darker colours mean that there are more transitions in that cell. Most clearly visible is the fact that observations without any violence in T (first row) are extremely likely to transition to observations without violence in T+1 (dark top left square). This illustrates that peace is a very stable and transitions from no violence to any level of violence are relatively rare.

(Figure 1)

But other levels of violence are also relatively persistent. This is immediately visible in the relatively dark colours along the diagonal. This is not only relevant for forecasting purposes but also because it confirms the existence of the conflict trap – violence begets violence and this makes conflict persistent.

Another interesting take-away from Figure 1 is that low values of conflict in quintile 1 are strongly mean-reverting. The second row shows that it is more likely that violence transitions back to 0 instead of staying in the first quintile. Beyond the first quintile violence becomes much more escalating – it is more likely that conflict escalates from quintile 2 to 3 than that it de-escalates back to quintile 1. This is also interesting because it suggests something about the nature of political violence. Figure 1 clearly shows that conflict can be a trap in the sense that countries have a much higher liklihood of ending up in the most intense violence, quintile 5, once they had an outbreak of violence – even if violence levels are minor.

But even in the first row of Figure 1, in transitions from peace to peace or conflict, the conflict trap plays a crucial role. This is shown in Figure 2. Here we show the likelihood of reaching a given level of violence on the y-axis after a certain number of months have passed without passing a threshold specified on the x-axis. In the left panel, the x-axis specifies the months since a month with 1 fatality, the middle panel since 50 fatalities, and the right panel since 500 fatalities. Month 0 means that the threshold was passed last month, while month indicates that there has been one month without passing th respective threshold. The three lines indicate different levels of monthly violence, where the black solid line indicates 1 fatality, the red dashed line 50 fatalities, and blue dotted line 500 fatalities.

(Figure 2)

The risk of continuation is extremely high with the likelihood being above 80 percent for 1 or 50 fatalities, and nearly 60 percent for 500 fatalities. However, in the first month after a given threshold of violence ended, we still see extremely high probabilities of relapse for 1 and 50 fatalities and elevated risk for 500 fatalities. We also see that after high levels of violence, lower levels are very likely to occur. The left panel of Figure 2 indicates that even within the relatively few transitions from 0 violence to violent quintiles in Figure 1, one can still use conflict history to forecast escalations. Moreover, the levels of past violence before peace predict different levels of relapse. Outside this range the likelihood of conflict eventually falls to low levels.

But, the conflict trap has two sides. Whereas conflict is very persistent and re-igniting, peace is also stable. Most countries are in peace most of the time. This means that even very low likelihoods of outbreaks in this state, less than one percent, imply that a large number of outbreaks occurr in a range outside the panels shown in Figure 2. These outbreaks are particularly important because they represent breaking points in which countries fall from stable peace into the conflict trap. In our discussion of results we will therefore also pay some attention to these hard-to-predict cases.

It is important to keep in mind this flipside of the conflict trap. If we manage to prevent countries from falling into the trap, we save human lifes and suffering and, eventually, also intervention costs and costs of reconstruction. UN/World Bank (2018) therefore stress the importance of moving prevention efforts ahead of the conflict cycle before conflict has broken out.

3. Methodology

Interpretation of Text in the LDA

Our prediction contribution is based on Mueller and Rauh (2020) who forecast the onset of violence using over 3.5 million newspaper articles written in the period January 1989 and August 2020. Our corpus is constructed from downloading articles from the BBC Monitor, the New York Times, the Washington Post, The Economists and the Latin News. We download text for over 195 countries by downloading a text if its headline mentions a country name or the name of the capital of a country. The reason we limit our corpus to these sources is their availability over the entire time period and downloadable format. A more comprehensive approach of covering local newspapers across the globe is constrained by availability. To some extent this shortcoming is overcome by the fact that the BBC monitor is a news aggregator which draws from many local sources and translates the respective articles. At the same time, the sources we use ensure some consistency in the coverage which makes learning from the experience of one country for another country possible. If each country was covered differently from the next, our supervised machine learning model would not be able to learn as much from the topics coming from the text. For the sake of dimensionality reduction, we pre-processed the articles with a Latent Dirichlet Allocation (LDA) which was introduced by Blei, Ng and Jordan (2003) model into 15 topics.

Before feeding the articles into the LDA, the raw text undergoes standard text analysis cleaning procedures. We filter out common expressions, also referred to as stop words, based on preexisting dictionaries, and remove tokens that appear in less than 100 articles and in more than 50% of the articles. Finally, we stem and lemmatize words, which transforms words into their roots, before also forming bigrams and trigrams, which are two- and three-word combinations. We are left with more than 100,000 unique tokens. The objectives of these cleaning procedures are to reduce computational complexity (the algorithm still has to deal with a 100,000 X 3.5 million matrix), to bunch common words together, and to capture some context by including word combinations.

The tradeoff once faces when choosing the number of topics is that a small number will provide general topics, which will not pick up subtle changes in risk, while a large number will provide specific topics, which might only be related to certain types of countries and violence. Given the strict forecasting deadline in 2020, we did not have time to experiment with the number of topics, but past research tells us that a low number of topics is most useful in forecasting outbreaks (Mueller and Rauh 2018). For the other two hyperparameters of the LDA, we use the default settings of the Gensim Python implementation of the dynamic topic model in Python by Rehurek and Sojka (2010). The weight variational hyperparameters for each document is inferred by the algorithm and the a-priori probability assigned to each topic is 1 divided by the number of topics.

The LDA is a method of unsupervised machine learning which provides probability distributions across all possible words and word combinations (tokens) by building topics around tokens that tend to appear together in documents. For example, when a journalist writes about business or economics, she will tend to use the words business, trade and markets together. The algorithm picks this latent semantic structure up and forms topics around it. As a disadvantage one could raise that LDA is a bag of words model, i.e., the order of words does not matter. However, compared to other text analysis techniques, LDA has several advantages. First, it allows to drastically reduce the dimensionality to very few dimensions compared to, for instance, term-frequency-inverse document frequency counts or Word2Vec. Of course, dimensionality reduction comes at the loss of information but the relatively small training sample for the conflict prediction exercise forces us to narrow down the number of predictors ex-ante. Second, we do not impose which part of the text is key as, for instance, keyword counts, or event coders do. We let the data speak. Third, topics are interpretable by humans and our discussion of the health and emergency topic, for example, exploits this important feature.

This allows us to use the LDA to reduce the dimensionality from thousands of word counts to 15 topics. To solve the computational challenges implied by the need to re-estimate the LDA model from millions of articles for every month we rely on the dynamic LDA model (Blei and Lafferty 2006). In the dynamic LDA the results of a month are used as prior for the estimates the next month which saves computational time but still takes about two weeks in total.

Figure 3 displays the top tokens in four of the 15 topics.[3] The size of a token in these clouds captures the importance of the respective token inside the topic. The LDA does not understand the meaning of the tokens and so the headlines shown in Figure 3 are our interpretation of the word clouds. The final output of the LDA is the probability distribution across all tokens within each topic, and the 15 topic shares, which each are between zero and one, and together sum to one for each article.

[3] For an interactive representation of an LDA estimated on a similar corpus see the webpage https://conflictforecast.org/.

(Figure 3)

An important feature of the dynamic LDA model is that it guards us against changes in the keyword composition as topics are simply updated. As a result, topics are comparable across time. This is extremely useful also to visualize the workings of the LDA model and how it reacts to new developments. We were worried, for example, that the topic model might categorize articles covering the crisis emanating from the Coronavirus under *conflict* because the virus kills. But the topic model intuitively categorizes tokens related to the Corona crisis under a topic which seems to reflect health topics and emergencies. We show this in Figure 4 using the word clouds for this topic in January 2020 and February 2020.

(Figure 4)

We believe that this is a very useful feature. The unsupervised learning has interpreted tokens linked to Covid-19 through a topic that seems to capture health and natural disasters. The forecast based on these summaries is therefore robust to the appearance of a new phenomenon as it is interpreted in the right context. The reason the topic model is able to do this without supervision is that it exploits the underlying semantic patterns that news outlets use when writing articles. Given that journalists write articles that mention cases, hospitals and infections together with the coronavirus the topic model will categorize the token coronavirus in the topic that covers hospitals and infections.

Using Text and Conflict History as Predictors

Predictions were requested during periods of conflict which meant that months of intense violence needed to be covered and the question would be whether conflict de-escalates or escalates further. This changes the sample periods used for training in our previous work as we had excluded conflict periods, but these periods now stand at the center of attention. In some countries the training sample increases substantially due to this (Afghanistan, for example). It also provides a very rich conflict history to rely on when predicting conflict.

On the other hand, reporting in conflict countries is relatively one-dimensional and we could therefore not rely on the same rich variation when predicting de-escalations as when predicting conflict outbreaks. In addition, international news media attention tends to shift to countries when something new happens but moves away afterwards. Our prior was therefore that text would play a reduced role in the forecast of escalations and de-escalations in conflict periods. Predicting violence levels is relatively hard for text as reporting typically changes around the beginning of violence and towards the end but countries in conflict produce much less coverage. In addition, it is difficult to understand the dynamics of the conflict from topics which are designed to capture broad lines in reporting, i.e., what share of news coverage is on conflict vs. economics.

At the same time, we needed to add additional features from the conflict history as the dynamics of violence would play a key role in predicting violence in the future. It is true that recent conflict predicts a renewed outbreak during peaceful times, but it is even more true that current violence levels and recent dynamics are the best predictors of future levels.

Our experience also suggested that machine learning would not help solve most problems and that providing simple models with few variables with strong priors baked into the model would be better than providing long lists of variables at the country level as the data set is still very small for a machine learning application. We capture conflict history through three sets of variables:

1) Contemporaneous levels of violence and its recent lags (first to third lag)

2) Time since last conflict: months since last month with any violence (best>0), months since last month with over 50 fatalities (best>50), months since last month with over 500 fatalities (best>500)
3) Cumulative violence: cumulated best in the past 6 months, cumulated best in the past 12 months, cumulated best in the past 60 months and cumulated best in the past 60 months

To this *history* model we added our text variables to generate a *text+history* model. The text variables we use are our 15 topic shares at the country/month level plus the total token count. The 15 topic shares are the average topic shares across all articles about a country in each month weighted by the number of tokens while correcting for the prior. This ensures that topic shares at the country level again sum to one and all take a value close to 1/15 when little text is written. The idea is to capture how much news coverage is dedicated to each of the 15 topics in each country/month. We discuss the contribution of the text variables to our forecast in detail in section 4.

Building the Prediction Model

We use a random forest regression in order and predict levels of conflict intensity rather than changes. At a given node of a tree, the random forest searches a cutoff value of a predictor at each node in order to minimize the mean squared error of the prediction. For each observation that lies below (above) the cutoff, the algorithm computes the mean squared error relative to the mean number of fatalities for all observations that lie below (above) the cutoff. The algorithm then chooses the variable and the cutoff which minimizes the mean squared error.

A random forest is a collection of decision trees. The prediction of this forest of trees is the average across the final nodes of all trees. In order to avoid overfitting, the random forest bootstraps the sample from one tree to the next. Moreover, we also limit the number of variables that the tree can draw from at each node and which variables are available is random.

Given that the current level of violence was always observable, the escalation forecast task is equivalent to predicting violence levels in t+2, t+3, t+4, t+5, t+6, and t+7 and then subtracting the known level of violence in *t*. As levels in *t* are known and as predicting violence differences is much harder than predicting levels this significantly reduced errors.

In summary, we estimate two models: a *history only* model and a *text+history* model. In each case we predict violence levels and run a grid search over the parameter space at three points in time: November 2013, November 2015 and August 2020 to capture the starting period of each forecast block. The grid search runs over the following parameters: Maximum depth of trees (6,7,8), number of trees (500,650). In addition, we fix that minimum sample split is 4, that the number of features from which we bootstrap at each node is the square root of the total number of features at each, and that the minimum sample at a terminal leaf is 4.[4]

[4] We limited the grid search to this set of hyperparameters by searching across a wider range in a preliminary analysis. Due to the computational complexity and processing times, we did not do an overly exhaustive grid search for every round.

4. Results

Main Results

Table 1 displays our mean square error for the two tasks with existing labels. As should be expected, the error increases when forecasting events that are further in the future. Note, however, that the increase is not linear, i.e., it first increases rapidly and then stabilizes.

(Table 1)

Interestingly, the error is lower in the later period (2017-2019). Given that we always train the model only with data available at the specific point in time this could indicate additional learning by the model from the earlier period to the later period with the larger training set available in the late period.

What do our results suggest for the main task of forecasting into the future? In Figure 5 we summarize our forecasts in August 2020 for the African continent. The Figure shows the forecast using the information available in August 2020. In Panel A we show the forecast two months ahead, October 2020, and in Panel B we show the seven months ahead forecast, March 2021. As can clearly be seen, both forecasts are very similar. They both suggest dramatic escalations for Libya, Mali and Niger and significant de-escalations for Mozambique, the Central African Republic and South Sudan.

(Figure 5)

In Figure 6 we show a detailed analysis of the forecasts over time for four examples from the African continent. Here we show the two months ahead forecast as a red dashed line. For example, in 2019m7 these figures show the predicted change of log violence between 2019m7 and 2019m9. The blue line shows the actual, realized violence change.[5] The model clearly has the ability to forecast broad tendencies of conflict in that the red and blue curves often coincide. This is remarkable given that we are showing out-of-sample forecasts in which only the information available at the time is used to forecast. For Mozambique our forecasts capture a whole series of escalations followed by de-escalations.

The figures reveal some very interesting features about the forecast methodology. First, fatalities are strongly mean reverting, and the model captures this. In Cameroon, for example, the model does not predict the first escalation but then predicts the following de-escalation quite well as it takes into account that after a sudden escalation violence will de-escalate again. Second, violence has escalating dynamics. This can be seen particularly well in Mozambique where sudden outbreaks and de-escalations are followed of peaceful periods which are accompanied by periods of slightly elevated risk which is slowly falling.

(Figure 6)

How Text Helps in the Forecast

Given the considerable effort that has gone into our text features it is particularly interesting to see how and when text increases forecast precision. In Figure 7 we show the mean square error (MSE) for each month between 2014m1 and 2020m8. The MSE for the model without text is shown as a solid line and the MSE for the model with text is shown as a dashed line. Clearly, there are very few months where the two deviate visibly from each other. This is an important finding as it suggests that errors

[5] Only for 2020 we do not report the actual data for better visibility.

due to the text features do not suddenly arise. However, Figure 7 also confirms our prior that in this forecast task, text adds less value overall.

(Figure 7)

However, this changes somewhat when we instead focus on the MSE at the country level in Figure 8. Here the forecast error is again extremely similar for most countries but for some countries the MSE with text is visibly below the MSE without text. The most clearly visible example is the country with the largest MSE (Libya) where adding text to the model makes some difference for the forecast error.

(Figure 8)

The example of Libya is an indicator of an important advantage of using text when forecasting new unforeseen outbreaks of violence. In Table 2 we illustrate this by providing summary statistics for our two months ahead forecasts in the *text+history* model and the *history* model. However, to illustrate the added value of text most clearly, we now focus on months which followed peaceful months with no fatalities.

In the first row of Table 2 we show summary statistics for all months following a peaceful month. The MSE in cases starting from peaceful months is now much lower than in the full sample (Table 1). This is because the huge persistence of peace (Figure 1) forces the forecast models to stick close to 0 in most forecasts and this is correct in the large majority of cases. This effect is so strong that the difference in the MSE is very small between the *history* and the *text+history* models.

In the second and third row we instead focus on outbreaks of violence, i.e., escalations of violence that occur after peaceful months. Here it becomes clear that the tendency of the model to predict violence close to 0 during peacetime produces large errors when an outbreak happens. Forecast errors here are very large (between 2.5 and 4.1 log points). To the forecasting model, outbreaks of violence are surprises. However, now the model with text is performing visibly better.

In order to highlight the role of hard-to-predict onsets we show the MSE of outbreaks that follow four consecutive months of peace. In these cases, conflict dynamics are even less predictive of whether a situation will escalate or de-escalate. Again, the number of cases falls and the MSE increases. Whereas these cases are increasingly important to identify for prevention purposes they are also increasingly hard to predict.

(Table 2)

We see the role in our newspaper topics in these latter cases where the text can pick up some signals of approaching outbreaks even when the history does not indicate high levels of risk. This might not affect the MSE very much as fitted values will only vary slightly but text here offers at least the chance to provide early warning.[6]

To highlight this, we display two histories of risk for two countries which did not experience any onset of violence but where, arguably, risk of violence has varied considerably over the last few years. Our first case is Venezuela (Figure 9) which, according to the definition used in the competition, has not suffered from armed violence. The fitted values from our model however produce several swings in

[6] Mueller and Rauh (2018, 2020) analyse the role of text in predicting onsets in detail finding that topics associated with violence (like the military, terror or even peace agreements) directly increase before conflict breaks out. Topics that capture judicial procedures, diplomacy, economic activity or trade tend to decrease before violence breaks out.

the predicted violence. The protests of 2019 lead to a very clear uptick in expected violence, for example.

(Figure 9)

The second example is Belarus (Figure 10). Here the risk dramatically increases with the outbreaks of violent protests in August 2020. Our model predicts that this would lead to a further escalation in violence. This is in itself an important insight from the prediction model – as conflict is persistent, changes in violence always mean the model will predict higher levels of violence later. Violence will not prevent future violence but trigger more.

(Figure 10)

5. Conclusion

We have presented a prediction model using text and conflict history in a random forest to predict escalations and de-escalations of violence at the country level. We have argued that our text features are most useful in circumstances which are hard to detect in the overall MSE levels like in Venezuela but could be used to provide early warning in those cases as the changes in escalation risk seem useful. We have therefore integrated our model into the forecasting webpage conflictforecast.org.

The other aspect that is worth stressing is that the random forest model is able to exploit conflict dynamics in its forecast which can help understand the dynamics shown in our summary Figure 1. Conflict is escalating but mean reverting. A sudden increase of violence as experienced in Belarus is accompanied by upward shift in the expected level of violence but the model almost always expects some de-escalation after escalation.

References

Blei, David M, and John D Lafferty. 2006. "Dynamic topic models." 113–120, ACM.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet allocation." *The Journal of Machine Learning Research*, 3: 993–1022.

Hegre, Håvard and Paola Vesco, and Michael Colaresi. 2022. "Lessons from an Escalation Prediction Competition." *International Interactions* 48. Introduction to the Special Issue.

Hegre, Håvard, Nils W Metternich, Håvard Mokleiv Nygård, and Julian Wucherpfennig. 2017. "Introduction: Forecasting in peace research." *Journal of Peace Research.* 54(2): 113–124.

Mueller, Hannes and Christopher Rauh. 2022. "The Hard Problem of Prediction for Conflict Prevention," *Journal of the European Economic Association*. Forthcoming.

Mueller, Hannes, and Christopher Rauh. 2018. "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text." *American Political Science Review*, 112(2): 358–375.

Rehurek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." 45–50. Valletta, Malta:ELRA. http://is.muni.cz/publication/884893/en.

Schrodt, Philip A, James Yonamine and Benjamin E Bagozzi. 2013. "Data-based computational approaches to forecasting political violence." In Handbook of computational approaches to counterterrorism. Springer, pp. 129-162.

United Nations and World Bank. 2017. "Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict-Main Messages and Emerging Policy Directions." World Bank, Washington.

Vesco, Paola , Håvard Hegre, Michael Colaresi, Remco Bastiaan, Jansen, Adeline Lo, Gregor Reisch, and Nils Weidmann. 2022. "United They Stand: Findings from an Escalation Prediction Competition." *International Interactions* 48: 1-37. DOI: 10.1080/03050629.2022.2029856.

Ward, Michael, Nils Metternich, Cassy Dor , Max Gallop, Florian M Hollenbach, Anna Schultz, and Simon Weschle. 2013. "Learning from the past and stepping into the future: Toward a new generation of conflict prediction." *International Studies Review*, 15(4):473{490.

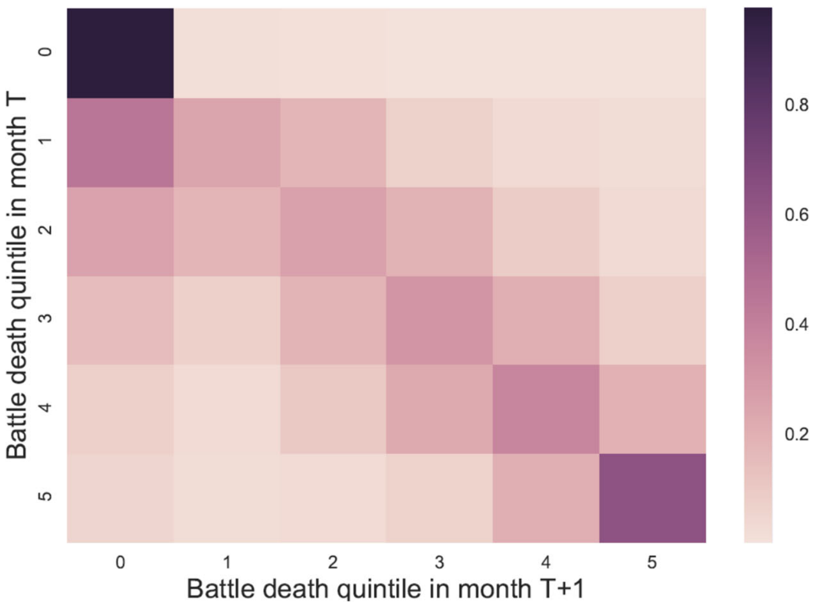

Figure 1: Escalation and De-escalation Risks Across the Conflict Cycle

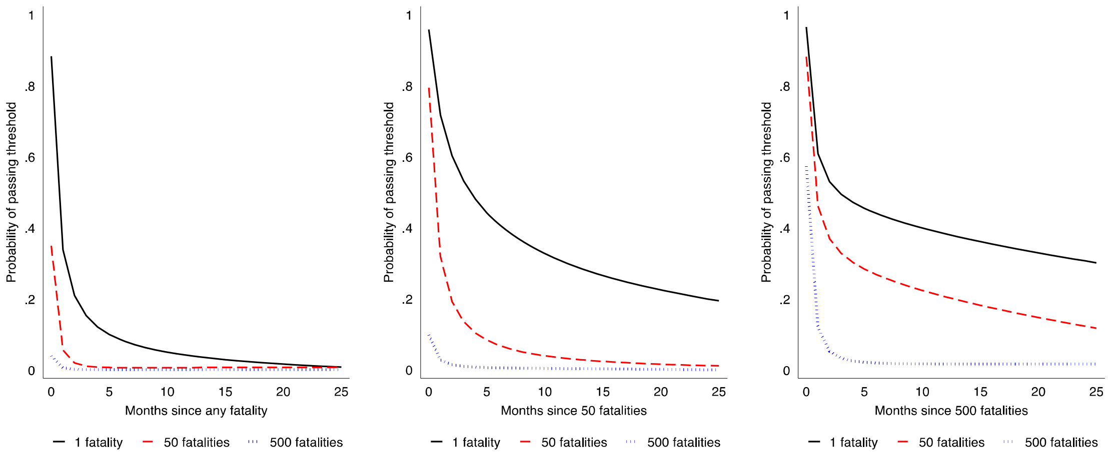

Figure 2: Continuation and post-conflict risk

Figure 3: Topic World Clouds from the LDA Model

Figure 4: Change in Topic from January 2020 to February 2020

Panel A: Forecast in August 2020 for October 2020

Panel B: Forecast in August 2020 for March 2021

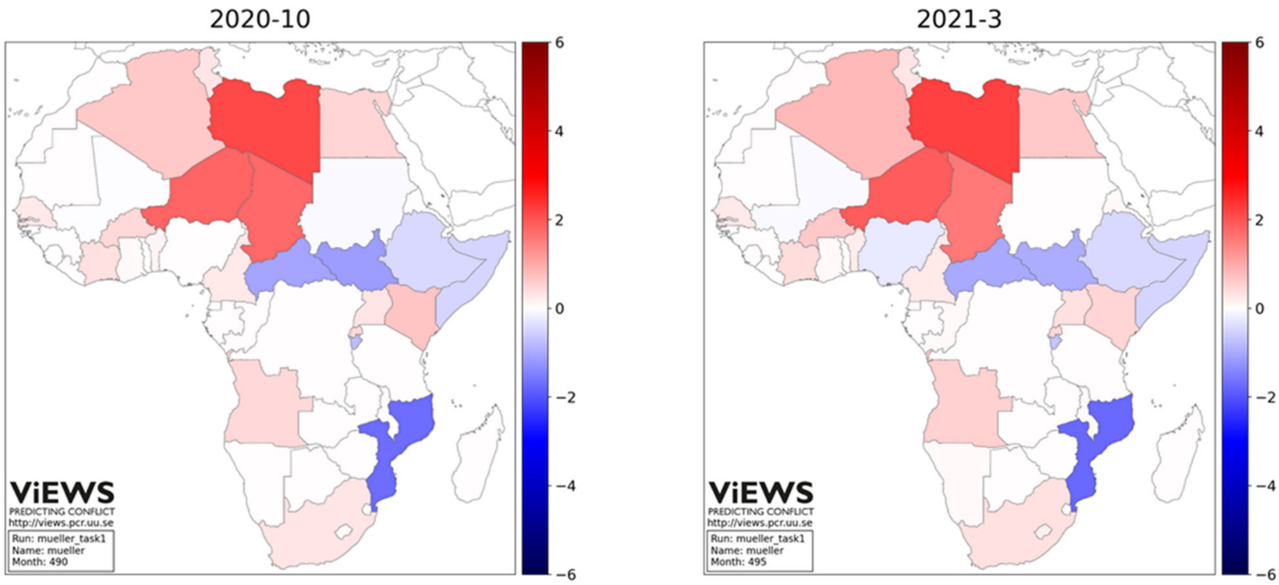

Figure 5: True out-of-sample forecasts

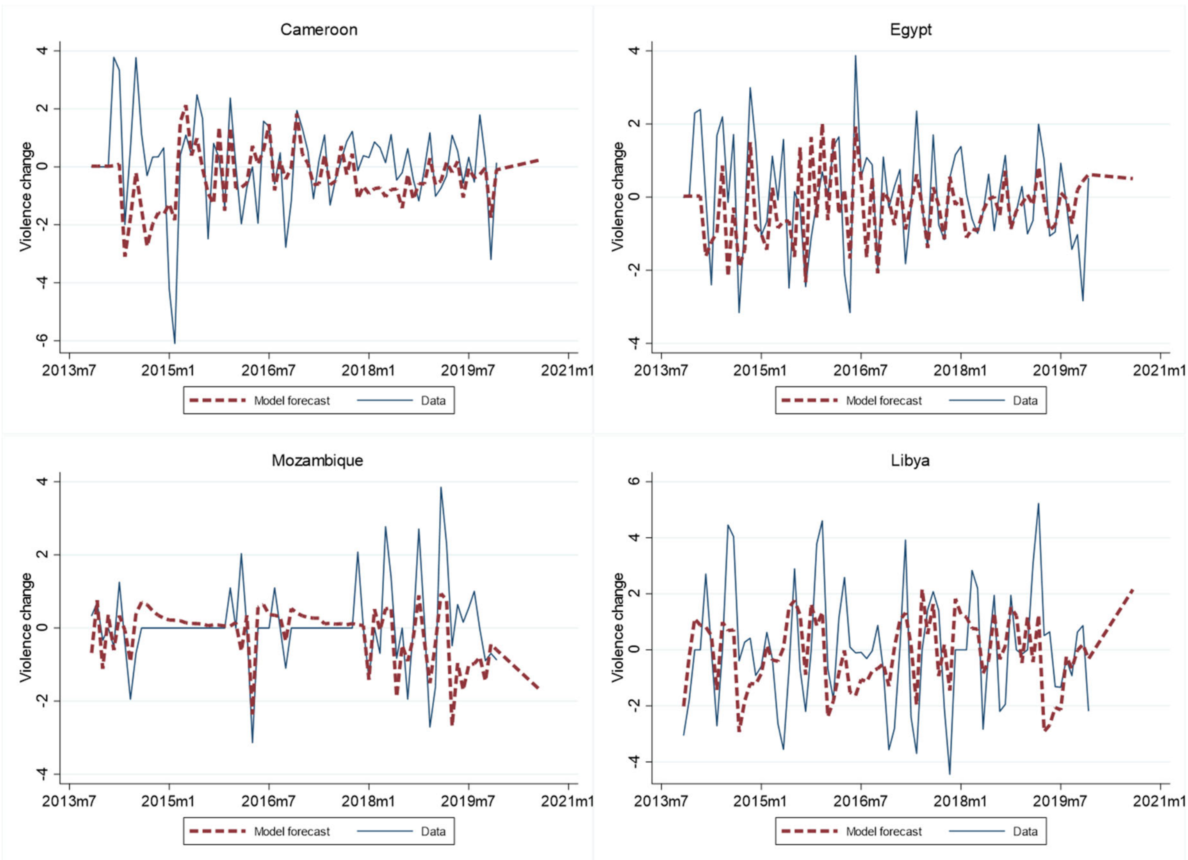

Figure 6: Country examples over time (two months ahead forecast)

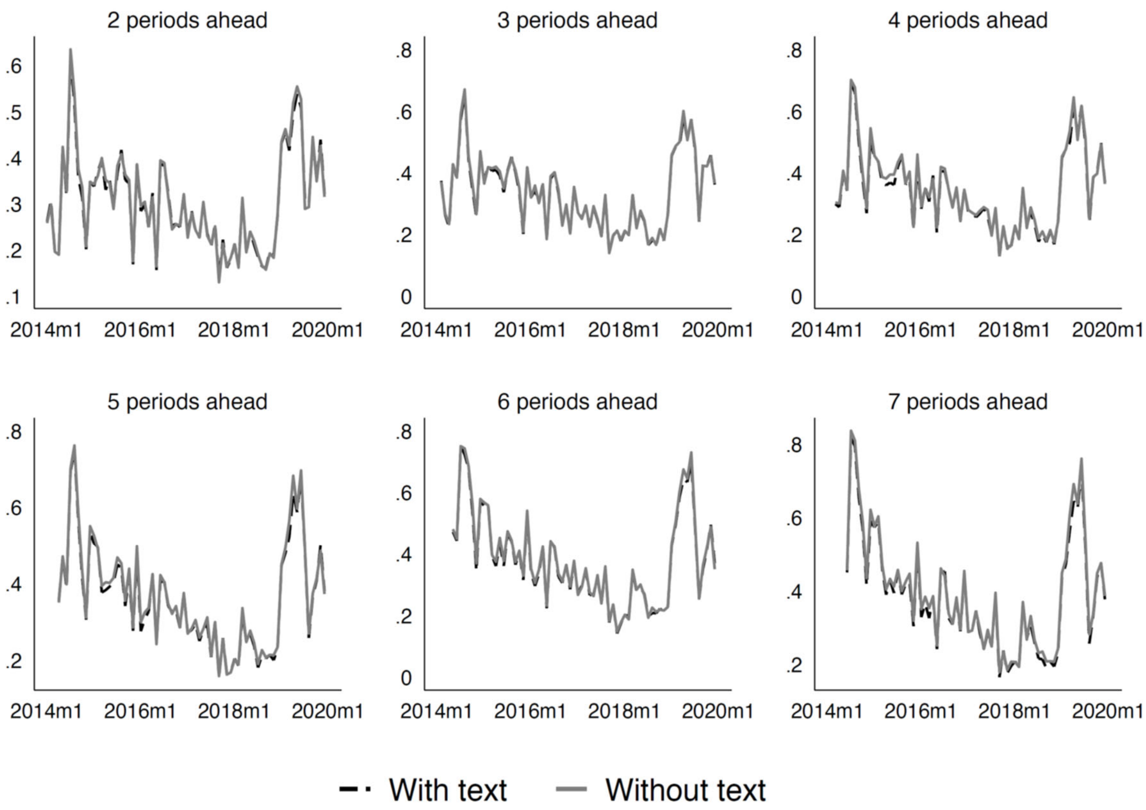

Figure 7: Mean square error (MSE) over time

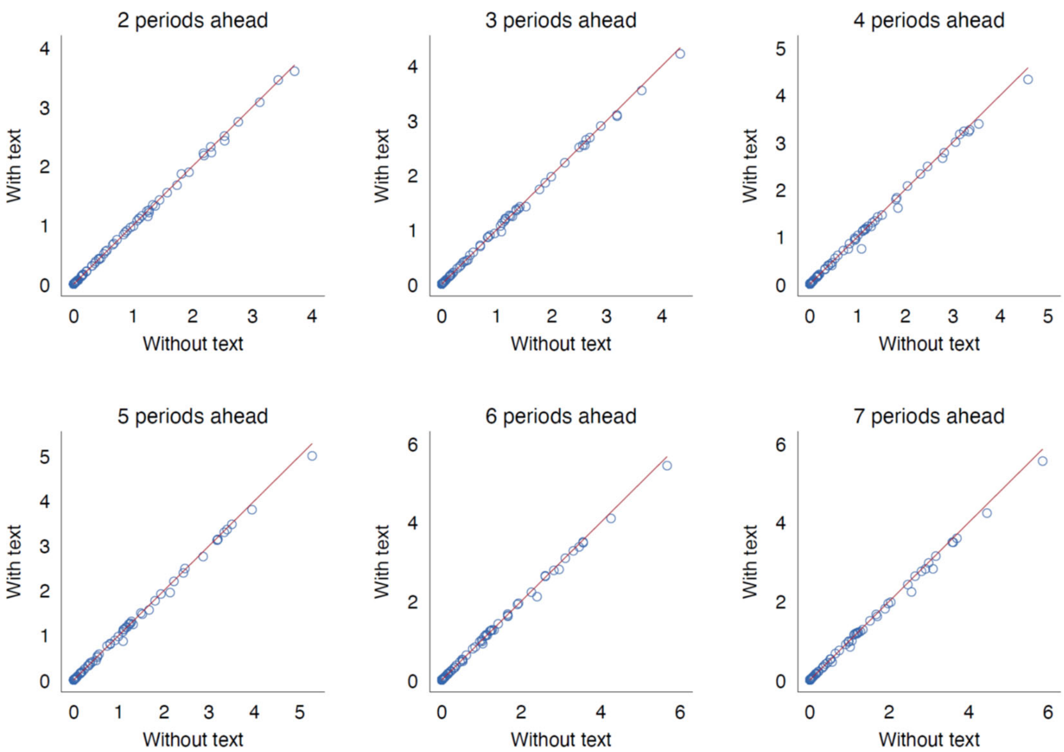

Figure 8: Mean square error (MSE) across countries

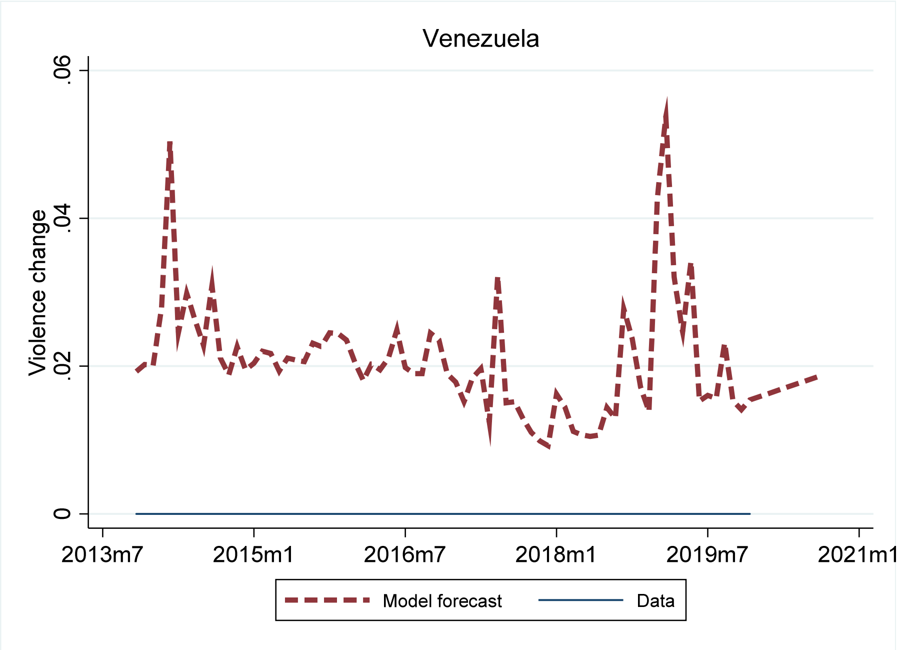

Figure 9: Example of past escalation forecast (6 months ahead)

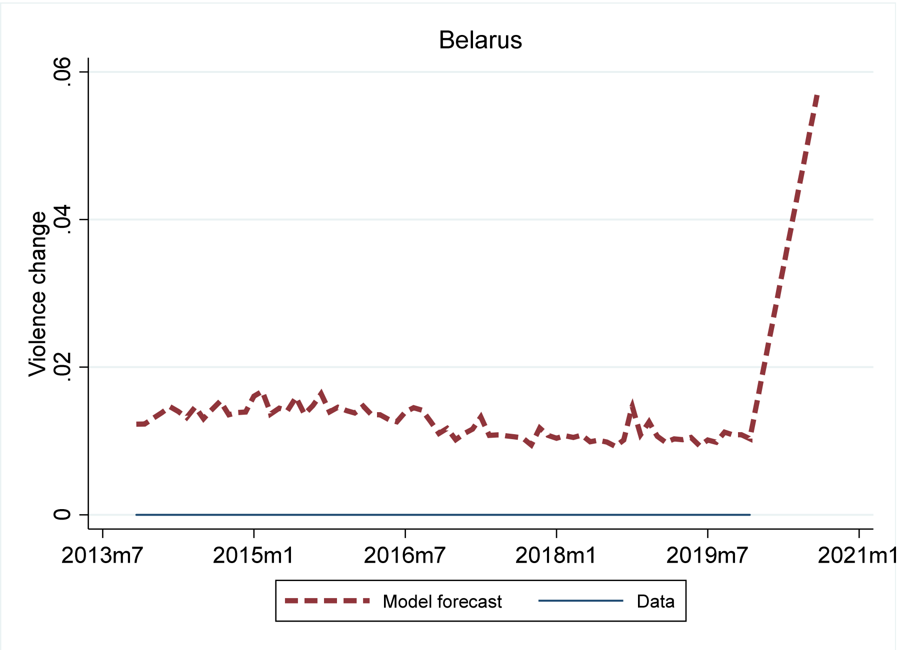

Figure 10: Example of past escalation forecast (6 months ahead)

Table 1: Overall mean square error in the forecasting tasks

	Obs	MSE: 2014-2016	MSE 2017-2019
2 months ahead	7002	0.322	0.293
3 months ahead	7002	0.357	0.317
4 months ahead	7002	0.370	0.326
5 months ahead	7002	0.390	0.339
6 months ahead	7002	0.408	0.348
7 months ahead	7002	0.407	0.354

Table 2: Mean Square Error in Months Without Violence

	Obs: 2014-2016	MSE: 2014-2016 (model with text)	MSE: 2014-2016 (model without text)	Obs: 2017-2019	MSE: 2017-2019 (model with text)	MSE: 2014-2016 (model without text)
2 months ahead (entire sample)	6096	0.124	0.125	6057	0.109	0.109
2 months ahead (outbreaks)	163	3.989	4.066	201	2.499	2.533
2 months ahead (outbreaks without recent conflict)	57	4.776	4.839	81	3.278	3.341