The Hard Problem of Prediction for Conflict Prevention

Hannes Mueller[1] Christopher Rauh[2]

[1]IAE (CSIC), Barcelona GSE

[2]University of Cambridge

2022

Conflict facts

- Civil wars are a serious humanitarian and economic problem.
  - And we fail to prevent them.
- This is reflected in large expenditures on crisis response.
  - Humanitarian response: ca. 24.5 billion US dollars in 2014.
  - Peacekeeping: ca. 8 billion US dollars per year.
  - 44,400 people displaced daily in 2017.
- Review of the United Nations Peacebuilding Architecture (2015):
  - *If more global priority were consistently given to efforts at sustaining peace, might there not, over the course of time, be reduced need for crisis response?*

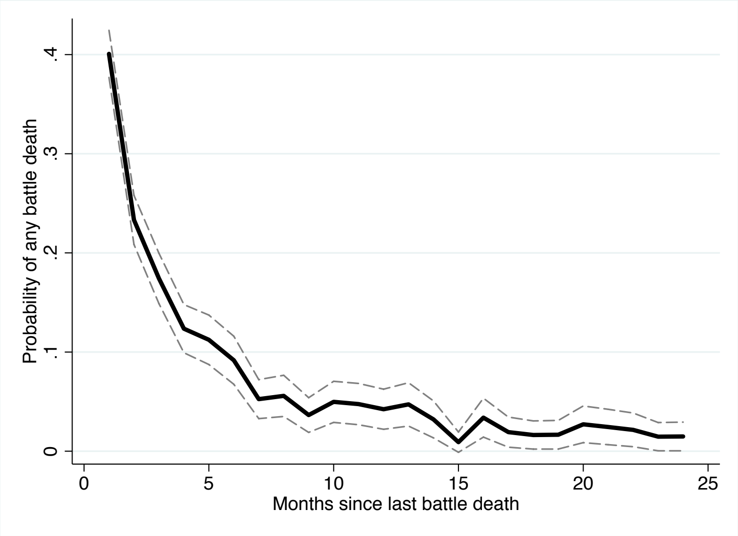
Our approaches

- We use vast amounts of newspaper text to predict the outbreak of violence
    - Reading Between the Lines: Prediction of Political Violence Using Newspaper Text (2018)
        - published in *American Political Science Review*
    - The Hard Problem of Prediction for Conflict Prevention (2022)
    - conflictforecast.org

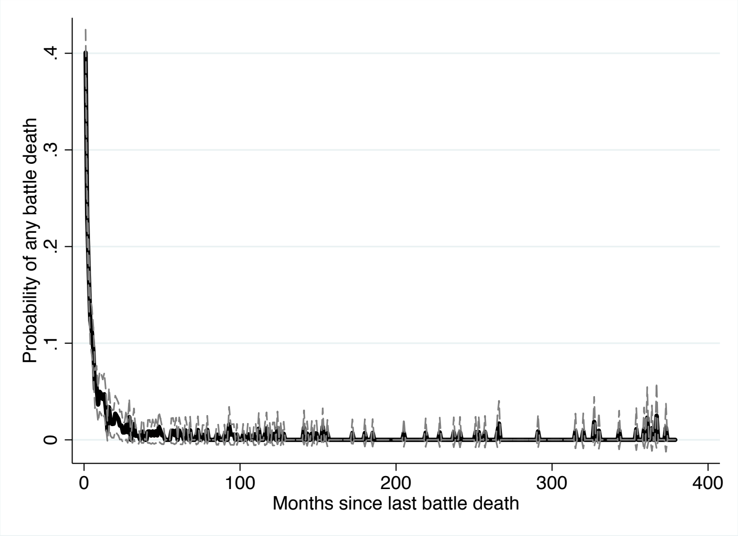
The hard problem – Imbalanced classes

- There were 2,271 onsets of any violence
- There were 97,807 peaceful months followed by peace.
- This means we have imbalanced classes.
  - Many more 0s (negatives) than 1s (positives).
- But it gets worse. Conflict risk:
  - in first 10 years after violence: 4.8%
  - after 10 peaceful years: 0.2% (hard problem)

Probability of relapse is high

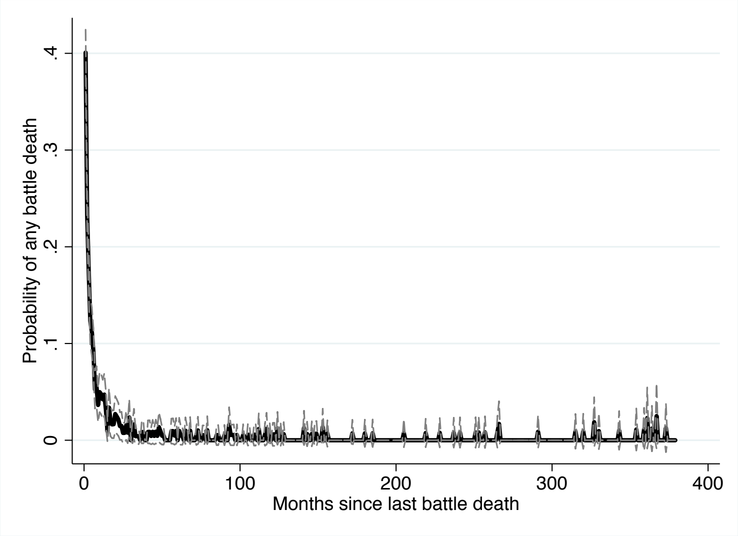


Probability of sudden outbreak close to 0



So how do we predict a sudden outbreak?

Probability of sudden outbreak close to 0



So how do we predict a sudden outbreak?

Conflict literature

Economics literature studying drivers of conflict:

- Ethnicity (Montalvo and Reynal-Querol 2005, Esteban et al 2014, Michalopoulos and Papaioannou 2016)
- Weather (Miguel et al 2004, Hsiang et al 2013, Ciccone 2011, Sarsons 2015)
- Commodities (Bazzi and Blattman 2014, Berman et al 2017)
- Political institutions (Besley and Persson 2011)

Conflict prediction literature using linear models:

- Goldstone et al (2010): Use slow-moving variables like political institutions or infant mortality.
- Chadefaux (2014) uses keyword counts of words clearly related to tensions and violence.
- Ward et al (2012, 2013) use a combination of event data and standard variables to make monthly predictions.

The hard problem – Why should we care?

- Pre-conflict stabilization efforts prevent conflict trap.
    - In the 5 years following an outbreak in a previously peaceful country, on average the country suffers:
        - 14 months of conflict.
        - Almost 7000 fatalities.
- Policymakers don't fully take this into account.
    - Conflict is much more salient than pre-conflict risk.
- But part of problem is forecasting.
- This makes prediction pre-conflict particularly important.

Our contribution

- Aggregate expert information to **forecast political violence**.
  - Focus on predicting conflict one quarter before it breaks out.
    - Task: 0, 0, **0**, 1, ·, ·, 0, 0...
- Use unsupervised machine learning to summarize text and supervised machine learning to predict outbreak of violence.
  - Unsupervised: Latent Dirichlet Allocation (topic model) to summarize 4million newspaper articles automatically.
  - Supervised: Use many different techniques. Previously not used much due to small sample problem with yearly data.
- Embed predictions in intervention framework for cost-benefit analysis.
- Great advantage of newspaper text: Available in real time.

Data – Predictors

Newssources

- 1 million articles from NYT, Washington Post, and The Economist.
- 3 million articles from BBC Monitor and LatinNews: monitors, and reports on, mass media worldwide and uses sources in multiple languages.
- Editors and journalists filter, translate and report news.
- We download an article if a country name or capital name is in the title.
- Total: 4 million articles on 190 countries dating from 1980m1 to 2020m8. By region

Topic models

- Latent Dirichlet allocation (LDA) introduced by Blei, Ng, and Jordan (2003).
- The LDA model in text analysis assumes that each document is a mixture of a small number of topics.
- Topics are nothing but probability distributions over words. Topic shares capture the share of words from a topic.
- Topic content is not predefined.
- Bag of words model: order of words does not matter.
- In this way we can transform the $100,000$ token counts of 4 million articles into topic shares and then aggregate these at the country/quarter level.

From Articles to Topics

Example: NYT - March 29, 1991. Libya

The exiled Prince Idris of Libya has said he will take control of a dissident Libyan paramilitary force that was originally trained by American intelligence advisers, and he has promised to order it into combat against Col. Muammar el-Qaddafi, the Libyan leader. The United States' two-year effort to destabilize Colonel Qaddafi ended in failure in December, when a Libyan-supplied guerrilla force came to power in Chad, where the original 600 commandos were based. The new Chad Government asked the United States to fly the Libyan dissidents out of the country, beginning a journey that has taken them to Nigeria, Zaire and finally Kenya. So far, no country has agreed to take them permanently. The 400 remaining commandos, who have been disarmed, were originally members of the Libyan Army captured by Chad in border fighting in 1988. They volunteered for the force as a way of escaping P.O.W. camps. "Having received pledges of allegiance from leaders of the force, Prince Idris has stepped in to assume responsibility for the troops' welfare," said a statement released in Rome by the royalist Libyan government in exile. It was overthrown in 1969.

Example: NYT - March 29, 1991. Libya (Stopwords)

the exiled prince idris of libya has said he will take control of a dissident libyan paramilitary force that was originally trained by american intelligence advisers and he has promised to order it into combat against col muammar el qaddafi the libyan leader the united states two year effort to destabilize colonel qaddafi ended in failure in december when a libyan supplied guerrilla force came to power in chad where the original 600 commandos were based the new chad government asked the united states to fly the libyan dissidents out of the country beginning a journey that has taken them to nigeria zaire and finally kenya so far no country has agreed to take them permanently the 400 remaining commandos who have been disarmed were originally members of the libyan army captured by chad in border fighting in 1988 they volunteered for the force as a way of escaping camps having received pledges of allegiance from leaders of the force prince idris has stepped in to assume responsibility for the troops welfare said a statement released in rome by the royalist libyan government in exile it was overthrown in 1969

Example: NYT - March 29, 1991. Libya

exiled prince idris libya control
dissident libyan paramilitary force originally trained
american intelligence adviser promised order
combat col muammar qaddafi libyan leader
united state year effort destabilize colonel qaddafi ended
failure libyan supplied guerrilla force came
power chad original commando based
new chad government asked united state fly libyan
dissident country beginning journey taken
nigeria zaire finally kenya far country agreed
permanently remaining commando
disarmed originally member libyan army captured
chad border fighting volunteered force
way escaping camp having received pledge allegiance
leader force prince idris stepped assume
responsibility troop welfare statement released
rome royalist libyan government exile overthrown

Example: NYT - March 29, 1991. Libya (Lemmatizing)

exiled prince idris libya control
dissident libyan paramilitary force originally trained
american intelligence adviser promised order
combat col muammar qaddafi libyan leader
united state year effort destabilize colonel qaddafi ended
failure libyan supplied guerrilla forced came
power chad original commando based
new chad government asked united state fly libyan
dissidnet country begining journey taken
nigeria zaire finally kenya far country agreed
permanentry remaining commando
disarmed originally member libyan army captured
chad border fighting volunteered forc
way escaping camp having received pledge allegiacne
leader force prince idris stepped assume
responsibility troop welfare statement released
rome royalist libyan government exile overthrown

Example: NYT - March 29, 1991. Libya

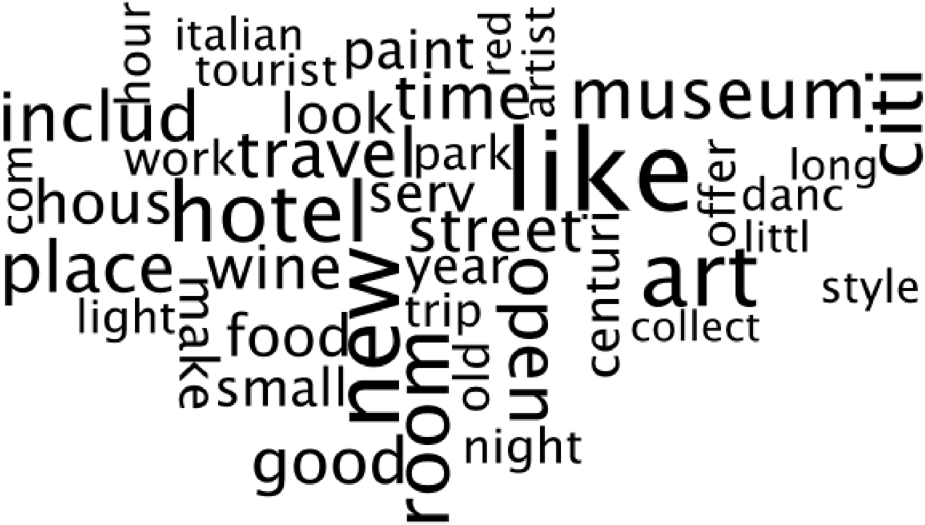
exil princ idri libya control dissid libyan paramilitari forc origin train american intellig advis promis order combat col muammar qaddafi libyan leader unit state year effort destabil colonel qaddafi end failur libyan suppli guerrilla forc came power chad origin commando base new chad govern ask unit state fli libyan dissid countri begin journey taken nigeria zair final kenya far countri agre perman remain commando disarm origin member libyan armi captur chad border fight volunt forc way escap camp have receiv pledg allegi leader forc princ idri step assum respons troop welfar statement releas rome royalist libyan govern exil overthrown

Word frequency matrix

	Article nr								
	1	2	3	4	5	6	7	...	4,000,000
control	1	0	...						
origin	3	...		...					
failur	1	...			...				
forc	2	...				...			
way	1	...							
commando	1	...							
exil	1	...							
....									
last_word	0	...							

Economics topic (φ_1)

Tourism topic (φ_2)



Conflict topic (φ_3)

Example: NYT - March 29, 1991. Libya

exil princ idri libya control dissid
libyan paramilitari forc origin train american intellig
advis promis order combat col
muammar qaddafi libyan leader unit state year effort
destabil colonel qaddafi end failur libyan
suppli guerrilla forc came power chad origin
commando base new chad govern ask unit state fli
libyan dissid countri begin journey taken
nigeria zair final kenya far countri agre
perman remain commando
disarm origin member libyan armi captur chad
border fight volunt forc way escap camp
have receiv pledg allegi leader forc princ idri step
assum respons troop welfar statement releas
rome royalist libyan govern exil overthrown

NYT - March 29, 1991. Libya (Tourism topic $\eta_1 = 4\%$)

exil princ idri libya control dissid
libyan paramilitari forc origin train american intellig
advis promis order combat col
muammar qaddafi libyan leader unit state year effort
destabil colonel qaddafi end failur libyan
suppli guerrilla forc came power chad origin
commando base new chad govern ask unit state fli
libyan dissid countri begin journey taken
nigeria zair final kenya far countri agre
perman remain commando
disarm origin member libyan armi captur chad
border fight volunt forc way escap camp
receiv pledg allegi leader forc princ idri step
assum respons troop welfar statement releas
rome royalist libyan govern exil overthrown

NYT - March 29, 1991. Libya (Economics topic $\eta_2 = 2\%$)

exil princ idri libya control dissid
libyan paramilitari forc origin train american intellig
advis promis order combat col
muammar qaddafi libyan leader unit state year effort
destabil colonel qaddafi end failur libyan
suppli guerrilla forc came power chad origin
commando base new chad govern ask unit state fli
libyan dissid countri begin journey taken
nigeria zair final kenya far countri agre
perman remain commando
disarm origin member libyan armi captur chad
border fight volunt forc way escap camp
receiv pledg allegi leader forc princ idri step
assum respons troop welfar statement releas
rome royalist libyan govern exil overthrown

NYT - March 29, 1991. Libya (Conflict topic $\eta_3 = 27\%$)

exil princ idri libya control dissid
libyan paramilitari forc origin train american intellig
advis promis order combat col
muammar qaddafi libyan leader unit state year effort
destabil colonel qaddafi end failur libyan
suppli guerrilla forc came power chad origin
commando base new chad govern ask unit state fli
libyan dissid countri begin journey taken
nigeria zair final kenya far countri agre
perman remain commando
disarm origin member libyan armi captur chad
border fight volunt forc way escap camp
receiv pledg allegi leader forc princ idri step
assum respons troop welfar statement releas
rome royalist libyan govern exil overthrown

From raw topics to country/quarter panel (example)

ID	Year	Month	Day	Tokens	η_1	η_2	η_3	...	η_{30}
					Topic shares				
11	1992	8	22	41	.035	.035	.047	...	.047
11	1992	8	29	274	.05	.307	.025	...	.041
11	1992	8	29	228	.018	.059	.018	...	.527
11	1992	8	29	320	.107	.038	.014	...	.044
11	1992	9	15	804	.071	.21	.009	...	.021
11	1992	9	15	480	.046	.019	.015	...	.048
11	1992	12	26	427	.388	.03	.059	...	.021

From raw topics to country/quarter panel (example)

ID	Year	Month	Day	Tokens	η_1	η_2	η_3	...	η_{30}
					$\overbrace{\hspace{1em}}^{\text{Topic shares}}$				
11	1992	8	22	41	.035	.035	.047	...	.047
11	1992	8	29	274	.05	.307	.025	...	.041
11	1992	8	29	228	.018	.059	.018	...	.527
11	1992	8	29	320	.107	.038	.014	...	.044
11	1992	9	15	804	.071	.21	.009	...	.021
11	1992	9	15	480	.046	.019	.015	...	.048
11	1992	12	26	427	.388	.03	.059	...	.021

Table: Collapsed country/quarter topic distribution

ID	Year	Month	Tokens	θ_1	θ_2	θ_3	...	θ_{30}
11	1992	3	2162	.092	.089	.029	...	.094
11	1992	4	...					

Forecasting

Data – Dependent variable

- Violence data from Uppsala Conflict Data Program (UCDP).
- Georeferenced Event Dataset (GED), Sundberg and Melander (2013).
- Internal conflicts: state-based conflict, non-state conflict, one-sided violence.
- Use monthly data for all countries 1980-2020.
- Focus on onset of at least one fatality, i.e. code a one following a one as missing.

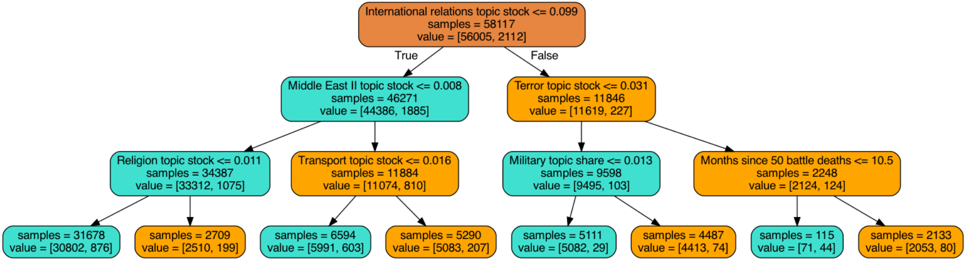
Model selection

- Predictors:
  - Text
    - 30 topics
    - 30 topic "stocks" (discounted past topic shares)
    - Number of tokens
  - Conflict history. Time since last conflict
    1. with at least one fatality
    2. with at least 50 fatalities
    3. with at least 500 fatalities
- Prediction model:
  - Random forest (because it outperforms everything).
- Train model with information until t, then predict $t + 1$.
  - Do this for all quarters $t = 2005m1 - 2020m8$.

A decision tree in a random forest

- Algorithm
  - Recursive binary splitting.
  - Numerical procedure where all values are lined up and different split points are tried and tested using a cost function.
  - For classification entropy measure is used – provides indication of how “pure” the leaf nodes are (how mixed the training data assigned to each node is).
    - Worst split: 50-50.
    - Best split: 100-0.
- Decision given by final leaf node.
- Tradeoff
  - More depth $\rightarrow$ overfitting.
  - Less depth $\rightarrow$ too general.

One (simplified) random tree from a random forest



Results: Trade-offs in forecasting

- Choose cut-off c to evaluate model: $\hat{y}_{i,t+1} > c \rightarrow$ forecast conflict.
- Trade-off:
  - high cutoff c implies more false negatives.
  - low cutoff c implies more false positives.
- ROC curves as a way to illustrate the trade-off.
- On the y-axis report the true positive rate (TPR):

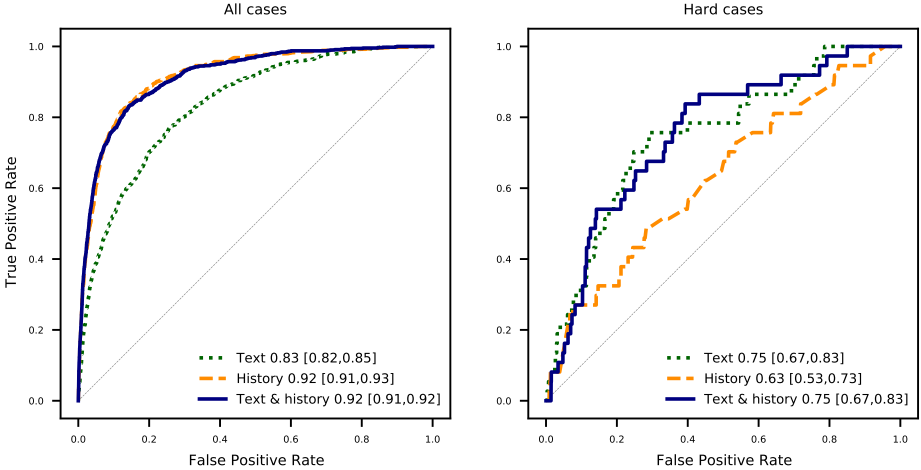
$$TPR_c = \frac{TP_c}{FN_c + TP_c}$$

- On the x-axis report the false positive rate (FPR):

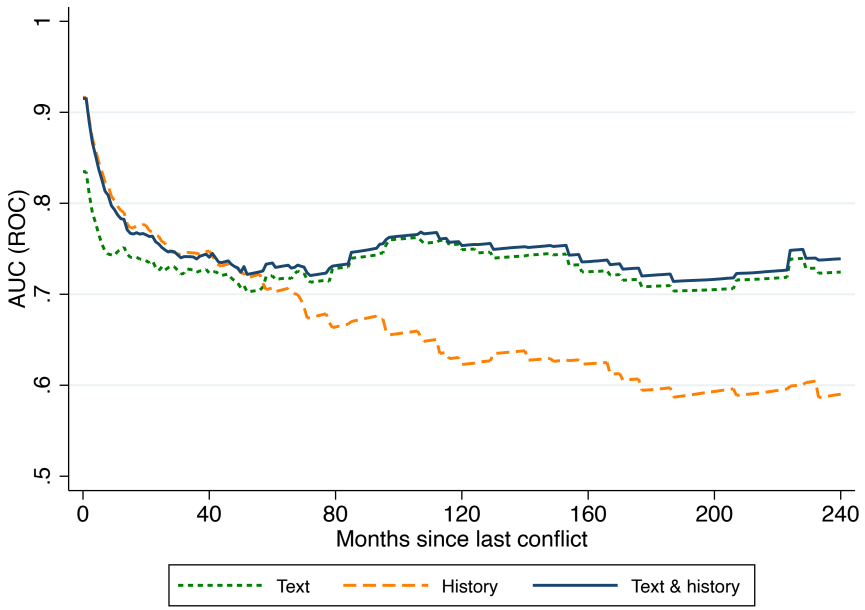
$$FPR_c = \frac{FP_c}{FP_c + TN_c}$$

Results

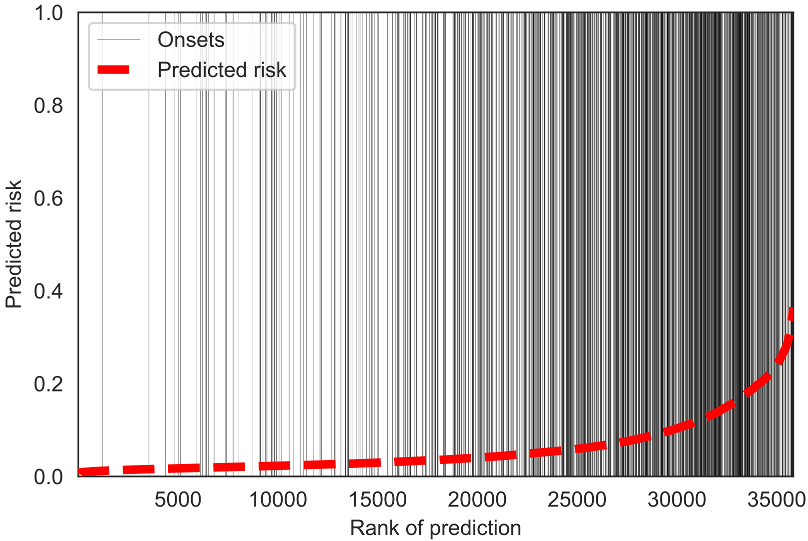
ROC curves for any violence (> 0) onsets 2000-2020



ROC curves for any violence with fading history



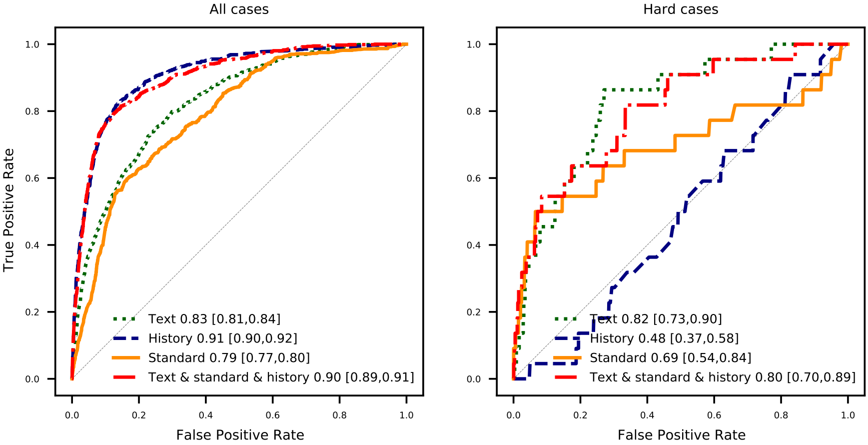
Separation plot for any violence with text



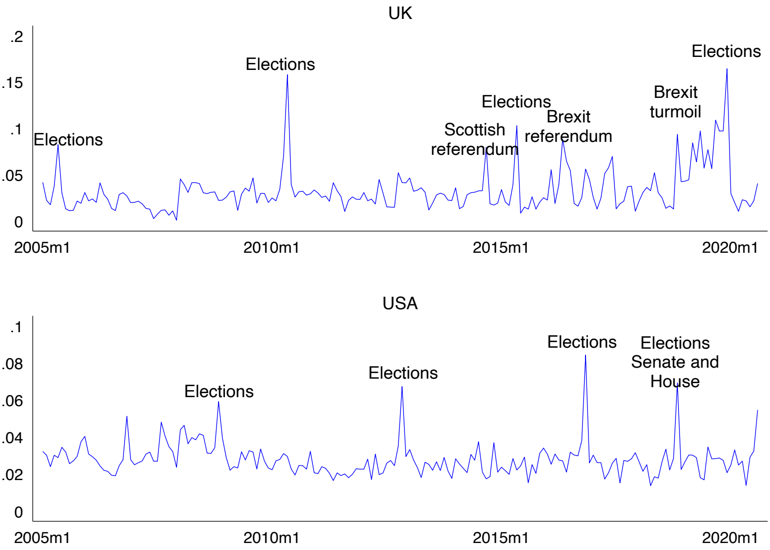
Comparing to other predictors

- Standard variables from Goldstone et al (2010):
    - Five different democracy dummies.
    - Number of neighboring countries in conflict.
    - Child mortality.
    - Share of discriminated population.
- Commodity prices:
    - 50+ commodities and their export weights.
    - Create sub-indexes capturing shocks in minerals, agricultural products, and energy revenue.

ROC curves compared to standard variables



Other events (politics topic in UK and US)



Summarizing the conflict prediction exercises

- Conflict history is an excellent forecaster.
  - Without conflict history, forecast becomes a hard problem.
- Use topics to summarize massive amount of BBC Monitor, NYT, WP, LatinNews, and Economist across 190 countries.
- Text can predict conflict with short prediction horizon.
- Text provide useful forecast without conflict history.
  - Cases without history are particularly interesting from a cost perspective (Mueller 2018).
- Text is available in real time.
- Machine learning helpful at prediction stage thanks to increased trainings sample at monthly level.

conflictforecast.org