Taking Time (Really) seriously.

Improving forecasts with time-series clustering

Thomas Chadefaux

Early draft — Please **do not circulate**

August 27, 2019

Abstract

Do conflict processes exhibit repeating patterns over time? And if so, can we exploit the recurring shapes of the time series to forecast the evolution of conflict? Here we study escalation patterns using recent machine-learning methods derived from information geometry, clustering, and pattern recognition in time series. Our goal is to supplement typical correlation-based approaches with clustering and prototyping methods to extract shapes and ideally better understand the patterns of escalation into war. We apply these methods to a particularly challenging task: forecasting the precise timing of Palestinian rocket and mortar attacks on Israel. Using four years of minute-level prices for 500 Israeli stocks, we find that financial asset prices react on average 30 minutes ahead of the launch of a rocket. We validate our result in a true out-of-sample manner. Using live market data and minute-level rocket attack data from Israel's home front command, we publicly broadcast our forecasts in real time for every minute of every day.

Do conflict processes exhibit repeating patterns over time? And if so, can we exploit the recurring shapes and structures of the time series to forecast the evolution of conflict? Theory has long focused on the sequence of events that precedes conflicts (e.g., escalation or brinkmanship). Yet, current empirical research is unable to represent these complex interactions unfolding over time. This is because it attempts to match cases on the value of covariates, and not on their structure or shape. In other words, it cannot easily represent real-world relations which may, for example follow a long alternation of escalation and détente, in various orders and at various speeds, before finally breaking down into conflict. Lags of a few order might be included in a regression to account for issues of autocorrelation, but they cannot represent the potentially complex trajectories that pre-conflict events take.

Here, we aim to address these issues using recent machine-learning methods derived from pattern recognition in time series to study escalation. The novelty is be to move away from the current reliance in the social sciences on covariance structures of the raw signals, and to supplement these typical approaches with clustering and prototyping methods to extract shapes and better understand the patterns of escalation into violence.

I apply these methods to a particularly challenging task: forecasting the precise timing of Palestinian rocket and mortar attacks on Israel. Since 2001, Palestinian militants from the Gaza Strip have launched thousands of rocket and mortar attacks on Israel. These attacks have killed dozens and injured thousands. They also create widespread psychological trauma, medical issues such as miscarriages, depression and post-traumatic stress. Can the timing of these rocket launches be predicted? In other words, do early warning signals exist

Using four years of minute-level prices for 500 Israeli stocks, I find that financial asset prices exhibit warning signs ahead of the launch of a rocket. These warning signs are not visible in changes in prices—or in fact in changes in any moment of the price series—but in more complex dynamics. To obtain this result, we relied on various distance measures for time series (shape-based such as Dynamic Time Warping, feature-based such as Wavelet decomposition, etc.). The distance from one time series to another is calculated for every two-hour period, every stock, and every method, and these distances are then aggregated in an Ensemble model, resulting in a forecasted probability of a rocket attack.

The early warning signals that we uncover are both more accurate (fewer false negatives), and provide earlier warnings than existing approaches. We also validate our result in a true out-of-sample manner. Using live market data and minute-level rocket attack data from Israel's home front command, we publicly broadcast our forecasts in real time for every minute of every day.

Patterns, Rockets and Markets

Conflict forecasting has received increasing attention in political science.[1] The availability of increasingly fine-grained spatiotemporal data, in particular, has allowed more refined predictions (Brandt, Freeman & Schrodt 2014, Weidmann & Ward 2010) using data such as stock market prices (Chadefaux, 2017b; Schneider, Hadar, & Bosler, 2017), news reports (Chadefaux, 2014), urban violence (Bhavnani, Donnay, Miodownik, Mor, & Helbing, 2014), climate data (Håvard Hegre et al., 2016; Witmer, Linke, O'Loughlin, Gettelman, & Laing, 2017), or night-light emissions (Weidmann & Schutte, 2016). Advances in computational methods have also made it possible to analyse larger and broader sources of information in real time, thereby moving from structural to short-term measures of tensions and other markers of conflict.[2]

Methodologically, however, existing approaches in the social sciences typically rely on analyzing the relationship between individual observations (e.g., country-year). They do so by matching a given situation with another that

[1] see Cederman & Weidmann (2017), Chadefaux (2017*a*), Hegre, Metternich, Nygård & Wucherpfennig (2017) and Schneider, Gleditsch & Carey (2011) for recent reviews). Of particular interest have been international conflicts (Beck, King & Zeng 2004, Gleditsch & Ward 2013, Chadefaux 2014), civil wars (Ward, Greenhill & Bakke 2010), coups (Goldstone, Bates, Epstein, Gurr, Lustik, Marshall, Ulfelder & Woodward 2010, Ward & Beger 2017), and mass killings (Rost 2013, Ulfelder 2012). Directly related to our focus on finance and rocket launches here is work on the reaction of financial markets to terror (Fleischer & Buccola 2002) and conflicts (Schneider & Troeger 2006), but also markets anticipation of conflict onsets (Chadefaux 2017*b*) and outcomes (Schneider, Hadar & Bosler 2017).

[2] For example, TABARI (Textual Analysis by Augmented Replacement Instructions), for example, uses the lead sentence of wire service reports (e.g. Reuters, Agence France press, etc.) to generate such event data (P. Schrodt, 2009). The World-Wide Integrated Crisis Early Warning System (ICEWS) is currently the most prominent of these event datasets. Sponsored by the Defense Advanced Research Projects Agency in the United States, it provides a detailed database of political events at the sub-daily and sub-national level (O'Brien, 2010). Phoenix, an open-source alternative to ICEWS, is also a becoming a central resource for real-time data collection. An advanced ERC project, ViEWS, also directly aims to build a real-time early warning system for war (Havard Hegre et al., 2018).

shares similar covariate values. Researchers may for example find that countries with high ethnic fractionalization, low GDP, or weak institutions are prone to war. Countries that share these attributes would in turn be deemed to have a high probability of civil war. Yet this 'matching' on the value of covariates does not make real use of time dependence in the sequence. While lags and first differences may be included to account for the role of the immediate past, these approaches are unable to incorporate the more complex dynamics of escalation that typically arise from geopolitical tensions.

Yet, most real-world interactions cannot simply be understood as a function of the current state of a variable or even of many variables. Modeling the type of complex back and forth, up and down, aggression–appeasement–further aggression that are common in international relations is difficult. Existing models (e.g., ARIMA) may include some additional lags, but are unable to account for this complexity. This may explain why, in existing approaches, the lagged dependent variable is almost always the best predictor—the immediate past best explains the current state of the system.

These challenges are likely to be even more pronounced in the context of conflict dynamics. The static value of covariates is unlikely to prove useful to predict the evolution of conflict. Here I address this by treating entire sequences of events as units of analysis. This makes it possible to extract potentially complex motifs and highly nonlinear patterns, as opposed to only basic trends or static values. In addition, conflict dynamics may take place over different time spans. A march on the capital that takes a week or a month is in essence the same phenomenon, but looks very different from the point of view of traditional time series analysis. Typical approaches are unable to deal with potentially distorted and time warped patterns.

To understand the importance of considering sequences as a whole rather than as a set of more or less independent observations, consider a simple illustrative example. Suppose that we observe a pre-conflict process, such as the evolution of the price of a financial asset (top plot in Fig. 1; vertical dotted lines denote the timing of the conflict event). We are then interested in finding similar patterns in the future, with the expectation that the similarities will help us anticipate the likely outcome of that sequence of events. Unfortunately, a logistic regression or a random forest applied to the data points is unable to detect any pattern here. This is regardless of the number of lags, first differences, or 'time since the last event' variables we include (Beck, Katz, & Tucker, 1998). The problem is not the logit itself, but rather that we do not treat sequences as the unit of observation. Instead, decom-

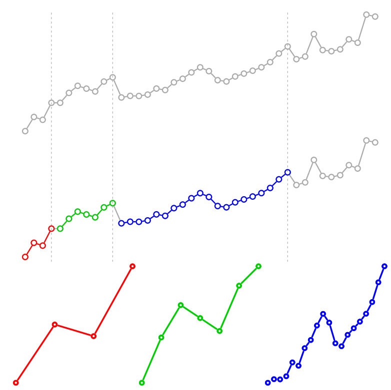


Figure 1: One-to-one measures of distance fail to discern patterns.

posing the series into sequences, we realize that the same pattern repeats three times, albeit over different durations (second row). Isolating each one, it then becomes clear that they are all noisy versions of a general motif of a form that could be summarized as ‘up 1, down 1/2, up 1’. Correlation would also fail here, because it requires the same number of observations. This is unfortunate, as escalation, power shifts, or terrorism may take place over different time horizon yet share a general underlying pattern. That pattern would be missed using standard approaches.

The importance of sequences has long been recognized in fields such as bioinformatics and molecular biology, where ‘motifs’ refer to repeated sequences in a DNA sequence or in amino acids (International Human Genome Sequencing Consortium, 2001). These sequences may also extend at different scales (Bernardi et al., 1985), just as we argue that pre-conflict patterns may emerge on different time scales.[3] Recent improvements in the availability

[3]To be sure, the idea that sequences might exhibit recurring patterns is not entirely new in social sciences either. For example, Marx’s historical materialism is a theory of repeating motifs throughout history (Marx, 2010), and Kondratieff waves describe long-term economic cycles (Kondratieff, 1979). In economics, the ‘J curve’ describes a country’s trade balance following a devaluation of its currency. These approaches, however, are ad hoc. The researcher hypothesizes a pattern, then proceeds to look for it in empirical data, but it is likely that many patterns will remain undetected. Worse, the risk is to find patterns where none actually exist (Pinker, 2011, p. 204). Power spectra approaches may help detect cycles in the data, but are unlikely to be useful to model the unfolding of geopolitical events, which are more chaotic and less regular than, say, waves in GDP growth (Korotayev & Tsirel, 2010).

of fine-grained time series related to conflict now allow us to use techniques which, so far, had been limited to data-rich fields such as finance or speech recognition, where identifying patterns is key. Now is therefore the ideal time to extend existing methodological approaches in political science to include not only observations treated more or less as independent, but also sequences of observations. This is in contrast with current approaches, which typically attempt to match observations on a one-to-one basis, but fail to detect patterns which may be stretched, distorted, or generally highly non-linear.

Markets and Rockets

Since 2001, Palestinian militants from the Gaza Strip have launched thousands of rocket and mortar attacks on Israel. Over the 2004–2014 decade, these attacks have killed 27 Israeli civilians, 5 foreign nationals, 5 IDF soldiers, and at least 11 Palestinians (rockets are often crude and sometimes land in Gaza) and injured thousands. They also create widespread psychological trauma, medical issues such as miscarriages, depression and post-traumatic stress.

The weapons, often referred to as Qassams[4], were initially crude and short-ranged, mainly affecting Sderot and other communities bordering the Gaza Strip. In 2006, more sophisticated rockets began to be deployed, reaching ever more distant cities until, in 2012, Jerusalem and Tel Aviv were targeted using locally and Iranian made rockets. In July 2014, the northern city of Haifa was targeted for the first time. Figure 2 illustrates the range of various rockets and the associated time given to local residents to seek protection. Residents in areas close to the Gaza strip only have 15 second warnings.

Qassam rockets are too inaccurate and prone to malfunction to be used against specific military targets in or near civilian areas, and are mainly launched for the purpose of harming civilians. The economic impact is mod-

Others have, on the contrary, sought to disprove the existence of these patterns. The efficient market hypothesis, for example, shows that any recurrent motif in markets would quickly be exploited by traders and therefore disappear. In conflict studies, Richardson also looked for patterns in the timing of the onset and the duration of wars, yet found that they are governed by a Poisson process—i.e., that they were random (Richardson, 1945). Yet this finding relies on coarse data on the year of onset of interstate conflict.

[4]'Qassam' rockets are named after the Izz ad-Din al-Qassam Brigades, the armed branch of Hamas.

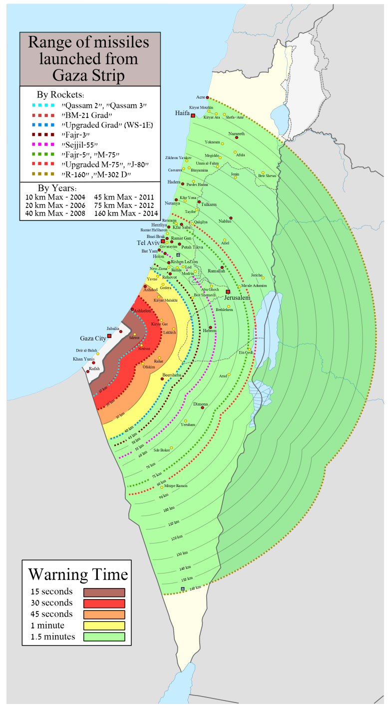


Figure 2: Range of missiles launched from the Gaza strip and associated warning time for residents

est, especially since a system called Iron Dome, designed to intercept the rockets before their landing, has been successful used to limit the number of rockets reaching their targets. However, rocket launches are a source of information for observers. They are a signal of the degradation of the political climate and of a resurgence of activism in Gaza. They might lead to large-scale retaliation by Israel–as was the case in 2014–and are such are valuable sources of information for market participants.

Markets therefore have an interest in anticipating the launch of rockets. Financial markets are the ideal time series for this project because they combine the forecasts of actors who have a financial stake in making accurate predictions (Arrow et al., 2008; Berg, Nelson, & Rietz, 2008; Wolfers & Zitzewitz, 2006). Securities are traded in a way that reflects the investors' beliefs about the probability of a certain event occurring. Large events such as wars are economically and financially costly, and market participants will therefore strive to anticipate them as early as possible and to react accordingly. For example, bonds are likely to be sold in anticipation of a war, as will be the stocks of industries most likely to be affected by it (Chadefaux, 2017b). As a result, financial assets often respond strongly to the expected occurrence of violence. We will use three main types of financial data:

However, the expected effect of a rocket launch on markets in not clear a priori. The effect may not be consistently positive or negative. It may change over time and across firms, as some may gain while others lose (Schneider). The effects may differ by sector, firm size, or geography. This makes traditional approaches all the more difficult because effects may cancel each other out.

Why would markets be able to anticipate these seemingly unpredictable events? There are three main reasons why this could be the case. First, rocket launches are planned. Militants do not act alone and often require the approval of their hierarchy. Rockets can also hardly be operated by a single individual and require coordination. Information may then travel to the markets through two mechanisms: (a) leaks: information may be inadvertently released or acquired by third parties, who then react by changing their trading on the stock market. This results in altered patterns on the market, which our algorithms would in turn exploit. (b) Second, and relatedly, militants or the organization itself may use their knowledge for for financial gains. Those with knowledge of the launch ('insider information') may trade—most likely through collaborators—on the Tel-Aviv stock exchange. Finally, (c) rocket launches and markets may in fact respond to external events. If these events affect both asset prices and the decision by militants to fire rockets, then markets may act as warning signals for rockets. In that case, the markets would be a proxy for events—possibly unobservable inside the government, etc.—that in turn lead to rockets. While a definitive answer as to which mechanism applies is not possible with our current design, our results strongly suggest that (a) or (b) are in fact the correct mechanisms. We elaborate on this argument below.

Data

Data on the timing of rocket launches was obtained from Israel's Home Front Command. The Red Colour alert programme is an early-warning radar system set up by the Israel Defense Forces in towns surrounding the Gaza Strip (Fig. 2). Its goal is to warn civilians of imminent attack by rockets (usually

Qassam rockets) using messages broadcasted on the internet, the radio, television and phone applications. Outside of areas serviced by the Red Color system, standard air raid sirens are used to warn of rocket attacks. In Sderot, it gives residents approximately 15 seconds' warning of an incoming missile.

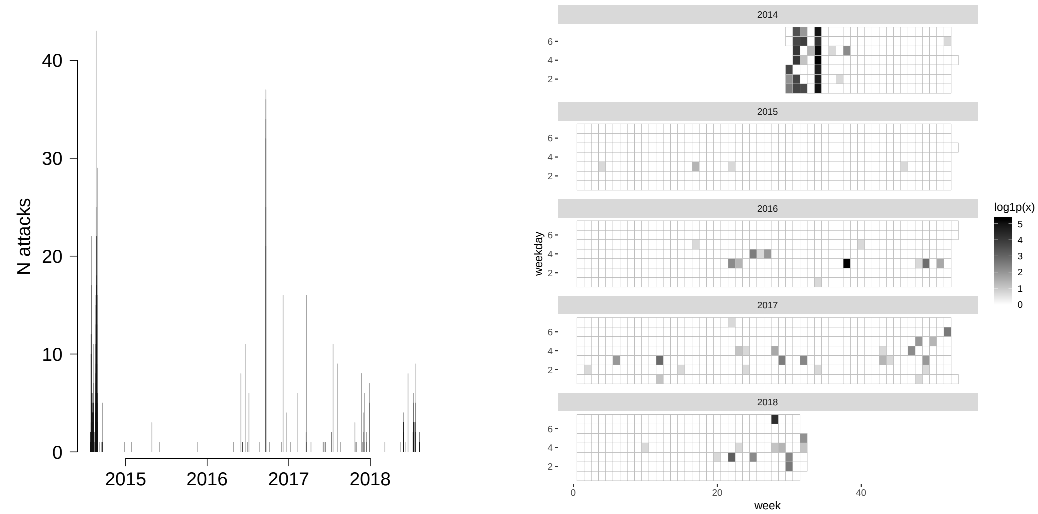
These real-time warnings and their historical records have a minute-level timestamp since 2014. Over the period of study here (2014–2018), 1,487 alerts were issued. The historical data is available at the minute level, which gives us great precision to match it with market data. A drawback of using financial market data is that it is limited to the market's opening hours (9:30 am–5:30 pm Monday to Thursday, 9:30 am–4:30 pm on Sundays). This excludes a large portion of attacks which happen in the evening, at night, on weekends or holidays, leaving us with only 651 attacks. We further exclude alarms which occur within less than 30 minutes of each other to avoid too much overlap between our sequences (see below), leaving us with about 150 cases.[5].

Figures 3 and 4b summarize simple information about the timing of attacks. Two empirical regularities emerge. First, attacks tend to cluster in time, for example in June 2014 (Fig. 3a and b). However, we found no pattern associated with days of the week or time of day (Fig. 3b). Second, attacks are very frequent. The probability of more than 8 consecutive days without a rocket being launched is only 50%, and 10% happen for a month. There has never been a 'peaceful' period longer than 3 months (we use peaceful as shorthand for the absnece of rocket), but make no assumptions about the state of the relations between Israel and Palestine otherwise..

Data on financial market prices covers 500 stocks listed on the Tel Aviv Stock market exchange.[6] However, we only rely on the 25 stocks with the highest liquidity, as other ones are simply not traded often enough for us to extract meaningful patterns. This has the advantage of significantly reducing what is already a computationally demanding task. Minute-level data on stock prices (closing price in each minute) from 2014-07-24 to 2018-08-09. were obtained from Tick Market Data.

[5] A possible way to address this would be to look at Israeli companies traded on the NY stock exchange, which is in a very different time zone. However, I fear that companies trading in NY will be less sensitive to geopolitical developments in Israel

[6] https://www.tase.co.il/en/market_data/index/142/components/index_weight



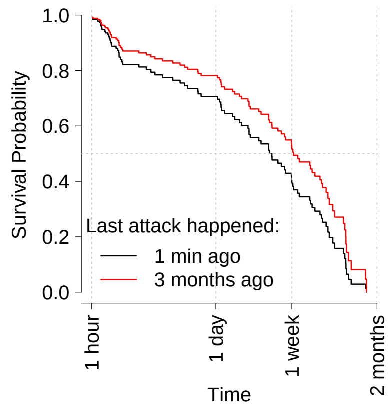
(a) (b)

Figure 3: Timing of attacks

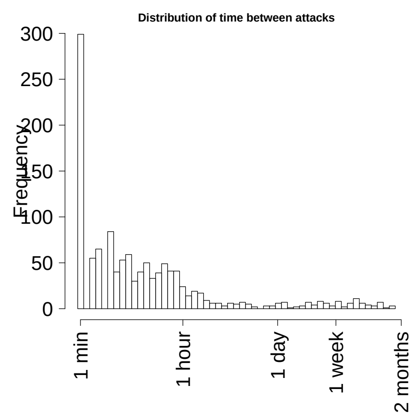
Methodology

We split each of the $N_s = 25$ liquid stocks' minute-level price time series (four years each) into small sequences of two hours (i.e., 120 minute-level observations—choosing a different size makes little difference), some of which precede a rocket launch, while others do not. The goal of this paper is to examine whether we can correctly classify stock prices sequences into pre-launch/not pre-launch. We use all pre-2015 sequences as a learning set and compare subsequent sequences to them. Each sequence is first Z-normalized —this is important to avoid matches solely due to similar raw values, and also because it is known empirically to lead to fewer classification errors (Keogh & Kasetty 2003).

A distance metric is calculated between each pair of sequences i and j occurring at time t_i and t_j respectively, where $t_i > t_j$ (one pre-2015, the other post-2015—we only compare sequences to past ones). For example, the sequence for stock prices of company TEVA on January 26th 2016 from 10:00 to 12:00 is compared to the stock prices of the same company on December 2014 from 2:00pm to 4:00pm; to the prices on September 2014 from 3pm to 5pm, etc. We compare all pre-rocket series to all pre-rocket series, but for computational reasons randomly sample a number of past and present



(a) Estimated probability of surviving 1 week is less than 40%. Longest stretch without an attack is 3 months. Predicted probabilities are the results of a Cox Proportional Hazard model of the launch of a rocket regressed on the time since the last attack and its square and cube.



(b) Distribution of the time between attacks

pre-peace series.

We denote by $d_{ij}^{\eta,s}$ the calculated distance between i and j using estimator η (see below for a list of estimators used) and data from stock s. Then the average proximity π of sequence i to sequences with outcomes $O \in \{\mathrm{R}, \mathrm{NR}\}$ (Rocket/no rocket) estimated using estimator η and data s, is calculated as

$$\pi_i^{\eta,s}(O) = \frac{1}{N_O} \sum_{j \neq i, t_j < t_i} 1/d_{ij}^{\eta,s}, \tag{1}$$

where N_O is the number of past observations with outcome O (here we have $N_R = 85$ and $N_{NR} = 200$). Note that this is simply the average (inverse) distance between i and every pre-rocket sequence preceding it. N is the number of distances calculated, that is, the number of observations j such that $t_j < t_i$ and $O_j = O$.

Then the probability of a rocket launch following a given two-hour sequence i, estimated using estimator η and data s, is calculated as the ratio

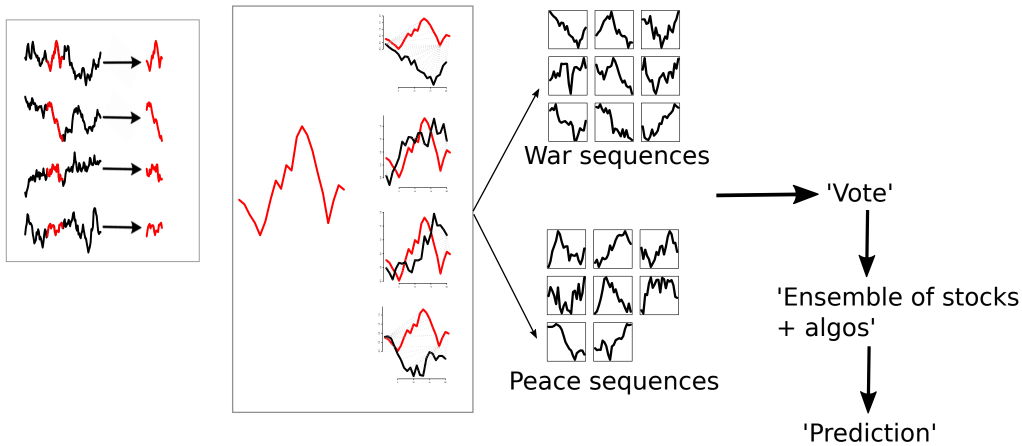


Figure 5: Overview of the work flow: (a) 4 year time series of stock prices are broken down into sequences of 2 hours each. (b) Each post-2015 sequence is compared to pre-2015 sequences; a distance metric is calculated for several algorithms and (c) the distances for each stock-sequence are averaged to obtain an 'average proximity to pre-attack sequences'. (d) These average proximities are combined into an ensemble model (using a random forest) to learn the weight to give to each stock-algorithm combination, using data from 2015–16. These weights are then used to make predictions over 2016–18. The results of these predictions are reported in section 'Results: Backward Forecasting' below.

of the proximity to pre-rocket sequences to the sum of the proximities:

$$p_i^{\eta,s} = \pi_i^{\eta,s}(R) / \sum_k \pi_i^{\eta,s}(O_k) \tag{2}$$

For each tuple $\{i, s, \eta\}$, respectively denoting a two-hour sequence i (e.g. June 6th, 10:32–12:32), a stock s (e.g., Teva Pharmaceutical Industries—TEVA), and method η (e.g., correlation), we thus obtain an estimate $p_i^{\eta,s}$ of the probability of a rocket launch. In all this gives us $68,250$ estimates.

As an illustration, consider a simple estimator such as the correlation between two sequences. For stock TEVA, for example, we first calculate the distance using the correlation between sequence i at t_i 2015-01-01 from 10:00 to 12:00 and all preceding sequences j that ended up in a rocket launch. The average of these distances gives us $\pi_i^{\text{cor,TEVA}}(R)$. We repeat this for all j that did not end up in a rocket launch, giving us $\pi_i^{\text{cor,TEVA}}(NR)$. The estimated probability of a rocket launch at 12:00 on 2015-01-01, using estimator 'correlation' and stock TEVA, is then simply the fraction of these two values as defined in equation 2.

Each of these estimated probabilities $p_i^{\eta,s}$ is then fed into an 'ensemble' model. The basic idea is that estimates using data from various stocks and various methods can be combined to contribute to an improved forecast. We do this by splitting the data into a learning set, using the first half of the data (i.e., all observations prior to 20 September 2016 at 11:06AM), and a testing set using the second half.[7] In the learning set, we use a random forest to learn the best weights to give to each of the $p^{\eta,s}$. These weights are then used in the testing set (the second half of the data) to estimate the probability of a rocket. It is these probabilities that are reported in Fig. 2.

Distance Estimators

The core of our approach, as described above, is to compare sequences of two hours of stock prices to each other. Sequences that are 'similar' are expected to lead to the same outcome (rocket/no rocket), and dissimilar ones to different outcomes (see eqn. eqn:similarity). I now define precisely is meant by 'similarity'.

Time series clustering algorithms can loosely be categorized in two groups: shape-based and complexity-based approaches. I now review various approaches within each group.

Comparing Shapes

Shape-based approaches preserve the order of the sequences and compare their shapes either point by point, by some distortion and warping, or by decomposition into simpler geometric elements.

One-to-one measures of distance. A simple measure of distance uses the Pearson's correlation factor between two series X and Y. A *distance* measure can then be constructed as a function of the inverse of the correlation. Correlation between two series requires that they be of the same size—not a

[7]Using the first two thirds or three quarters of the data as learning sets instead makes little difference.

problem here—but only matches then one-to-one. This means that the red and the black series in Fig. 6 have a correlation of 0.

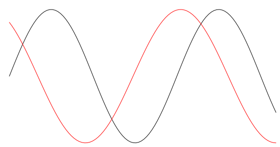
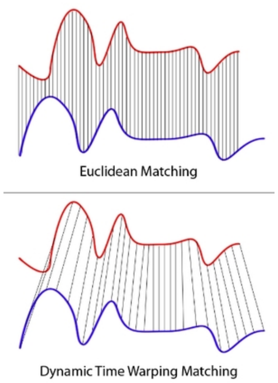


Figure 6: One-to-one measures fail to detect any similarities

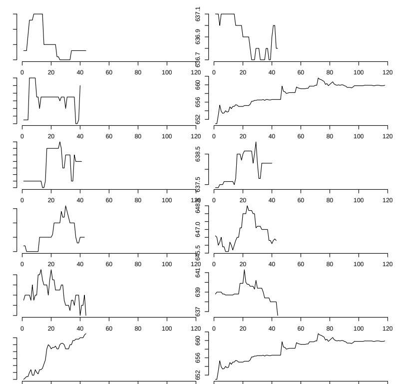
Measures of distance allowing for distortions. Another approach is to match the two series by allowing one observation in series i to be matched with several in series j.

Dynamic Time Warping (Berndt & Clifford, 1994) is an algorithm which measures the similarity between two temporal sequences that may vary in speed—which is particularly appropriate for the problem of escalation and the ups and downs of conflict, as these may take place over shorter or longer time periods but still present the same pattern. This type of problem cannot be easily addressed by standard approaches to time series. The basic idea underlying DTW is to shift, scale, stretch the series a way that minimizes their distance (Fig. 7).

Dimensionality Reduction. The idea here is that time series can be decomposed into their components—for example, a long-term trend plus seasonal variation plus a cyclical component and finally an irregular component (the residual). These transformations of the time series can help to search for patterns. For example, the discrete Fourier Transform (Agrawal, Faloutsos, & Swami, 1993), the discrete wavelet transform (Chan & Fu, 1999), Adaptive Piecewise Constant Approximation (APCA, Keogh, Chakrabarti, Pazzani, & Mehrotra, 2001), Singular Value Decomposition (SVD, Korn, Jagadish, & Faloutsos, 1997), Piecewise Aggregate Approximation (PAA, Keogh et al., 2001), or piecewise linear approximation (Morinaka, Yoshikawa, Amagasa, & Uemura, 2001). Figure **??** illustrates how these various methods convert time series into regular patterns that can then be matched and compared.



(a) Visual intuition.



(b) Example of pairs of series in our data with low distances according to DTW (left = pre-2015 series; right: post-2015).

Figure 7: Dynamic Time Warping

Here I use the discrete wavelet transform, which performs an unsupervised feature extraction using orthogonal wavelets. The distance is then calculated as the Euclidean distance between the wavelet approximations.

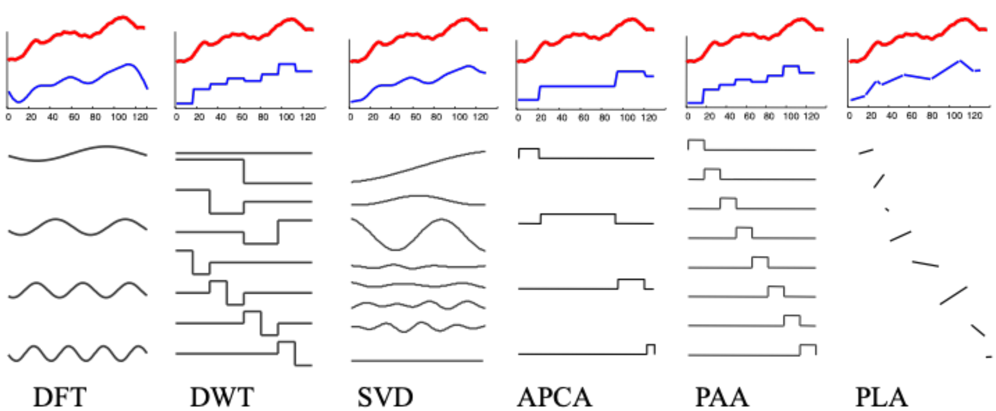


Figure 8: Dimensionality Reduction Algorithms

Complexity-based approaches

Complexity-based algorithms differ from shape-based ones in that they discard the order of the time series—either partially or completely. Instead, they compare underlying parameters of the series in the form of the data's periodogram, the number of peaks and troughs, etc.

Autocorrelation methods compute the dissimilarity between two time series as the distance between their estimated simple (ACF) or partial (PACF) autocorrelation coefficients.

The *power spectrum* of a time series describes the distribution of power into frequency components composing that signal. According to Fourier analysis, any physical signal can be decomposed into a number of discrete frequencies, or a spectrum of frequencies over a continuous range. The idea is to decompose a time series into a combination of sinusoids and their associated coefficients. This is referred to as 'spectral analysis' or analysis in the 'frequency domain,' in contrast to the time domain approach we have considered so far. Here I used power spectrum based approaches to compute

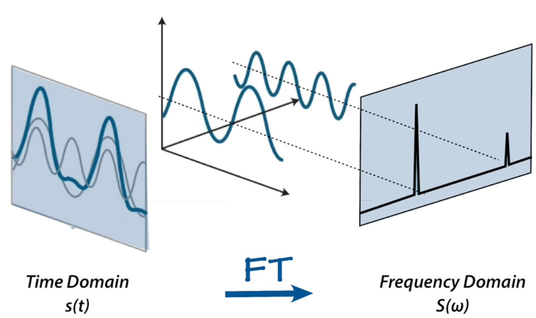


Figure 9: A time series (in blue on the left) can be decomposed into discrete frequencies (grey) using a Fourier transform (FT). The distribution of these frequencies defines the spectrum of that series (right)

a dissimilarity measure based on the ratio of local linear spectral estimators.

Compression-based dissimilarity approaches. Compression-based dissimilarity method (CDM) computes the dissimilarity based on the sizes of the compressed time series. The series are transformed to a text representation prior to compression. As a result, small numeric differences may in fact produce significantly different text representations, since 0.99 will have a different text representation than 1.

Permutation distribution clustering is an algorithm that attempts to cluster or match series based on the distribution of their sub-components. Loosely, it splits each series into short sequences (e.g., 'up-down-up', or 'down-down-up', etc.) and matches series based on the relative entropy of these sequences (Fig. 10). This approach is particularly suitable for our questions here because it is model-free and can discover patterns in highly nonlinear environments.

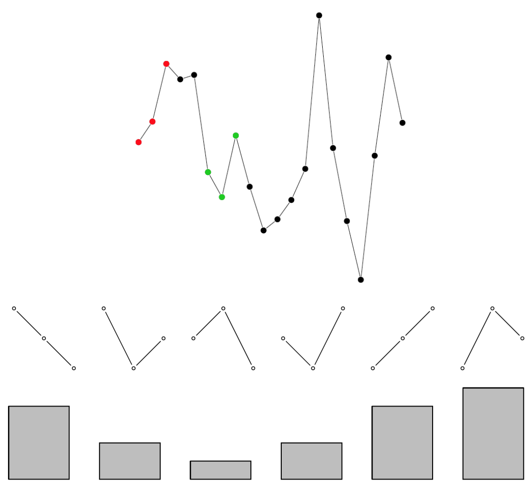


Figure 10: Complexity based clustering: Permutation distribution clustering

Complexity-invariant dissimilarity measures (Batista, Wang & Keogh 2011). Complexity invariance uses information about complexity differences between two time series as a correction factor for existing distance measures. It is based on the physical intuition that if we could "stretch" a time series until it becomes a straight line, a complex time series would result in a longer line than a simple time series (fig. 11). Intuitively, this means that a series with more peaks and valleys will be more complex than a 'straighter' one.[8]

[8]Formally, The Euclidean distance, ED(Q,C), between two time series Q and C, can be made complexity-invariant by introducing a correction factor: $CID(Q,C) = ED(Q,C) \times CF(Q,C)$ where CF is a complexity correction factor defined as: $CF(Q,C) = \frac{max(CE(Q),CE(C))}{min(CE(Q),CE(C))}$ where $CE(T)$ is a complexity estimate of a time series T. $CE(T)$ is in

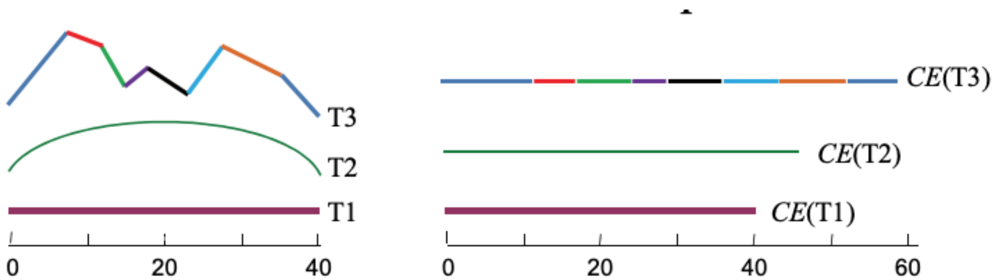


Figure 11: A measure of complexity for Complexity Invariant Distance. Complexity is measured by stretching time series (left) and measuring the length of the resulting lines (right). Source: Batista, Wang & Keogh 2011.

Results: Backward Forecasting

'Backward' forecasting was performed on the timing of rocket launches from 2014 to 2018 onset of inter- and intra-state conflict events for the past. We first define a 'learning' window (2014–2015) of 2-hour sequences. All subsequent series will be compared to these sequences and, depending on their average similarity to pre-rocket or pre-no-rocket, will be assigned an estimated probability of being a pre-rocket sequence (see section on distance estimators above).

The result is a distance metric for each possible pair of sequences and each algorithm between 2015 and 2018. Different data sources and algorithms will yield different estimates. Each of these estimates contains new information, and our second step will be to combine each distance measure into an ensemble forecast of all distances. We train this ensemble on the first half of the 2015-2018 test sample, and predict on the remaining half (i.e., mid-2016–mid-2018) For example, we might have learned from the learning set that the combination of the distance between a certain stock price using algorithm a and another stock's algorithm b yields particularly good predictions. We will therefore rely heavily on that estimate, and less on other. The ensemble allows us to aggregate these various measures and make the best use of each of their contributions.

---

turn computed as $CE(Q) = \sqrt{\sum_{i=1}^{n-1}(q_i - q_{i+1})^2}$

Overall Results

I now present the results of the final prediction, using the ensemble, on the 2016-2018 sample. A simple way of looking at results is to compare the number of correct predictions to the number of incorrect ones. A confusion matrix shows, for a given threshold, the number of observations correctly and incorrectly predicted as 0s and 1s. In table 1, for example, we choose a threshold of 0.2. This means that estimated probabilities of more than 0.2 are labeled as "1" and those less than 0.2 as 0. The table then reports how many '0' predictions were also '0' in the actual data; how many 0s were actually 1s, and so on. This simple confusion matrix shows that our ensemble produces more than twice as many correct predictions than incorrect ones ($\chi^2 \approx 7, p < 0.01$).

	0	1
0	145	56
1	45	89

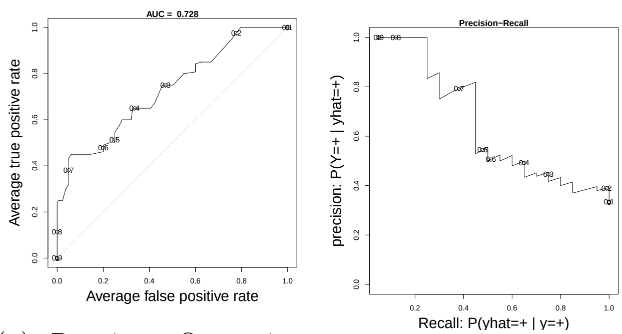
Table 1: Confusion matrix. Threshold = 0.2

While easy to understand and interpret, these results may depend on the choice of threshold (0.2 here). Receiver-Operating Characteristic (ROC) curves address this problem by plotting, for all possible thresholds $\gamma \in [0, 1]$, the number of false positives (i.e., an event is predicted but none occurs) against the number of true positives (an event is predicted and does occur). There is a tradeoff in the choice of threshold. With a low threshold, an alarm is frequently raised and all positive events are trivially correctly detected, but many false alarms are issued. With a high threshold, many events will incorrectly be predicted as 0s but there will be few false alarms. The ROC curve plots this trade-off in one convenient plot (fig. 12a). The ideal case is of course where there are no false positives and 100% true positives—i.e., where the curve traces out the top left corner of the plot. For comparison purposes, the ROC curve can be summarized in one statistic, the area under the ROC curve. Perfect predictive power results in an area under the curve of 1, whereas random predictions would lead to a curve following the diagonal line, with an area under the curve of 0.5.

Our results here are of an area under the ROC curve of around 0.73. Similar results are obtained using the precision-recall curve (fig. 12b). Whereas the ROC curve plots $P(Y = j|\hat{Y} = j)$, the PR curve plots $P(Y = j|\hat{Y} = j)$

against $P(\hat{Y} = j|Y = j)$. In other words the x axis is the probability that I issue a warning given that an event will happen; and the y axis is the probability that an event happens given that a warning was issued.

This is a far cry from some exercises in conflict incidence prediction, which can routinely reach an area under the ROC curve of 0.9. However, these are hardly comparable exercises. Forecasting the yearly incidence (not onset) of conflict is a far simpler task than predicting the timing of rockets. While 0.7 is not large, it is in fact striking that markets have any predictive power at all. After all, militants aim to take their targets by surprise. Moreover, each rocket launch is a relatively small event with limited impact on financial markets, so the signal to be extracted from markets is expected to be drowned in all the noise of other events affecting the market.



(a) Receiver Operating Characteristic (ROC) curve

(b) Precision-Recall curve

Figure 12: Forecasting results

Disaggregated Performance

The results presented above aggregate the predictions of a large number of combinations of algorithms and stocks. But how does the predictive power vary by algorithm and by stock? I examined the performance of each algorithm and stock in isolation (Fig. 13a). For example, calculating the distance between two time series using their Euclidian distance alone (using all stocks), we find little evidence of forecasting power (AUC under the ROC curve ≈ 0.45). The ensemble on the other hand has the best performance, with an AUC of 0.7. Periodogram-based methods also perform remarkably well.

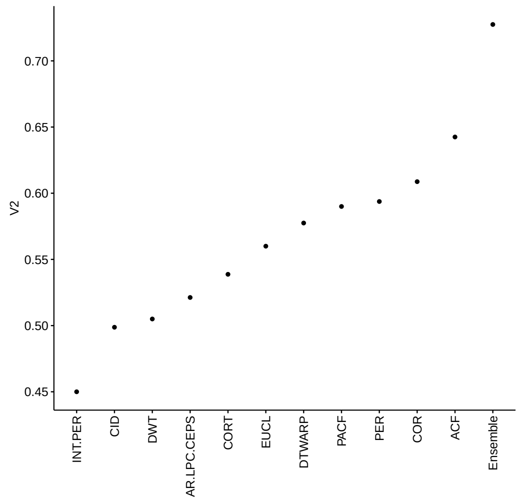
Importantly, we can compare the performance of the algorithms presented here to more classical approaches to time series, such as logits. The logit was estimated on data from 2014-2015 for each stock (just as the other algorithms used this period as an initial learning set) and predictions were made on the 2015–2018 period. Just as for the other algorithms-stock combinations were then combined into an ensemble, we also combined all predictions from the logit models (i.e., predictions for company TEVA, company AFIL, etc.) using 2015–2016 data and evaluated the combined forecast using data from 2016–18. The results show an area under the curve of 0.57—better than random but barely, and well off the 0.7 reached by the overall Ensemble. Similar results were obtained using random forests or neural network.

Results also vary by stock (fig. 13b). We classified these stocks by sector (as classified by Google Finance) and found that two sectors appear to be particularly sensitive to the expectation of rocket launches: Aerospace and defense; and oil and gas producers (fig. 13c) Real estate is also affected, whereas chemicals and technology very little. This is in line with what we'd expect. (more needed here).

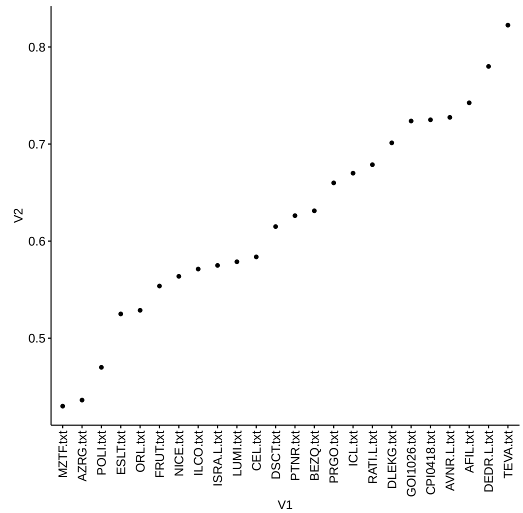
Lastly, we tested as a robustness check the possibility that the results depend on the stocks' market capitalization. Large companies might be more sensitive to political and geopolitical vicissitudes given their broad exposure to international trade and large number of subsidiaries who might be affected. Conversely, larger companies may be more resilient to shocks and hence less affected by rocket launches and their associated political effects. We find little evidence of either effect (fig. 13d)

Conclusion

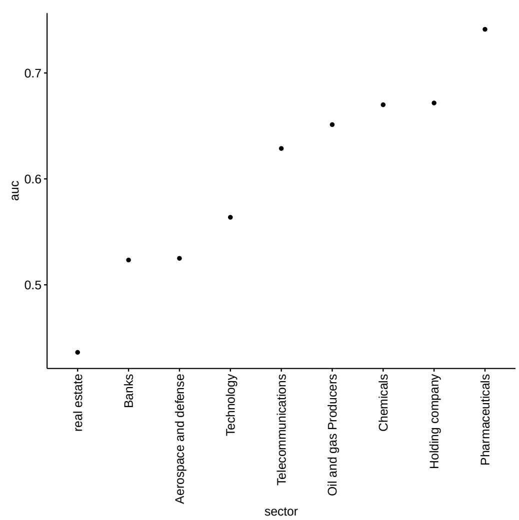
Our field suffers from a discrepancy between advanced theoretical models, which can well account for nonlinear patterns and in fact aim to identify motifs, and empirical data, which treats observations independently without modeling complex sequences as such. The reliance on one-to-one analysis of each observation (e.g, country-year) is unable to incorporate the more complex dynamics of escalation that typically arise from geopolitical tensions and may take place over different time spans. What we proposed here is an approach that extracts potentially complex dynamics from data



(a) By method



(b) By stock



(c) By sector

Figure 13: Area under the ROC curves

Beyond their methodological contribution, these results also have practical uses, as Israel's Home Front Command only gives its citizen less than a minute to seek cover. Our forecast can significantly extend this early warning

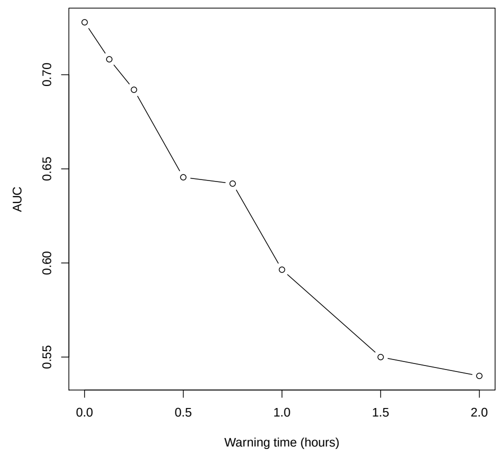


Figure 14: Area under the ROC curve as a function of warning time. Stock prices are no longer informative more than 2 hours prior to an attack.

time.

Finally, this project could also have important repercussions for the field of political science more generally. Matching time series, in particular, can be an important tool for causal inference. Existing work on matching, such as synthetic matching, typically relies on one-to-one methods such as logistic regressions (Abadie & Gardeazabal, 2003). This excludes the possibility that a common pattern might underlie time series of different length or more practically geopolitical changes that may take place slower or faster, but display the same shapes and attributes. Dynamic Time Warping, Permutation Distribution Clustering and other clustering methods based on dimensionality reduction will contribute to solving these problems and providing better matching techniques for time series.

References

Batista, Gustavo EAPA, Xiaoyue Wang & Eamonn J Keogh. 2011. A complexity-invariant distance measure for time series. In *Proceedings of the 2011 SIAM international conference on data mining*. SIAM pp. 699–710.

Beck, Nathaniel, Gary King & Langche Zeng. 2004. "Theory and evidence in international conflict: A response to de Marchi, Gelpi, and Grynaviski." *American Political Science Review* 98(2):379–389.

Brandt, Patrick T, John R Freeman & Philip A Schrodt. 2014. "Evaluating forecasts of political conflict dynamics." *International Journal of Forecasting* 30(4):944–962.

Cederman, Lars-Erik & Nils B. Weidmann. 2017. "Predicting armed conflict: Time to adjust our expectations?" *Science* 355(6324):474–476.

Chadefaux, Thomas. 2014. "Early warning signals for war in the news." *Journal of Peace Research* 51(1):5–18.

Chadefaux, Thomas. 2017*a*. "Conflict forecasting and its limits." *Data Science* 1(1-2):7–17.

Chadefaux, Thomas. 2017*b*. "Market Anticipations of Conflict Onsets." *Journal of Peace Research* 54:313—327.

Fleischer, Aliza & Steven Buccola. 2002. "War, terror, and the tourism market in Israel." *Applied Economics* 34(11):1335–1343.

Gleditsch, Kristian Skrede & Michael D Ward. 2013. "Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes." *Journal of Peace Research* 50(1):17–31.

Goldstone, Jack A, Robert H Bates, David L Epstein, Ted Robert Gurr, Michael B Lustik, Monty G Marshall, Jay Ulfelder & Mark Woodward. 2010. "A global model for forecasting political instability." *American Journal of Political Science* 54(1):190–208.

Hegre, Håvard, Nils W Metternich, Håvard Mokleiv Nygård & Julian Wucherpfennig. 2017. "Introduction: Forecasting in peace research." *Journal of Peace Research* pp. 113–124.

Keogh, Eamonn & Shruti Kasetty. 2003. "On the need for time series data mining benchmarks: a survey and empirical demonstration." *Data Mining and knowledge discovery* 7(4):349–371.

Rost, Nicolas. 2013. "Will it happen again? On the possibility of forecasting the risk of genocide." *Journal of Genocide Research* 15(1):41–67.

Schneider, Gerald, Maya Hadar & Naomi Bosler. 2017. "The oracle or the crowd? Experts versus the stock market in forecasting ceasefire success in the Levant." *Journal of Peace Research* 54(2):231–242.

Schneider, Gerald, Nils Petter Gleditsch & Sabine Carey. 2011. "Forecasting in international relations: One quest, three approaches." *Conflict Management and Peace Science* 28(1):5–14.

Schneider, Gerald & Vera E Troeger. 2006. "War and the world economy: Stock market reactions to international conflicts." *Journal of Conflict Resolution* 50(5):623–645.

Ulfelder, Jay. 2012. *Forecasting onsets of mass killing.* Working paper (https://ssrn.com/abstract=2409005).

Ward, Michael D & Andreas Beger. 2017. "Lessons from near real-time forecasting of irregular leadership changes." *Journal of Peace Research* 54(2):141–156.

Ward, Michael D., Brian D. Greenhill & Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4):363–375.

Weidmann, Nils B & Michael D Ward. 2010. "Predicting conflict in space and time." *Journal of Conflict Resolution* 54(6):883–901.