The Triggers of War
Disentangling the Spark from the Powder Keg

Thomas Chadefaux*

August 10, 2016

*Special thanks to Michael Mäs and Sebastian Schutte. I also thank Bob Axelrod, Karsten Donnay, Hein Goemans, Rob Hyndman, Giamperro Marra, Amin Mazloumian, Jim Morrow, Jonson Porteux, Heiko Rauhut and Jason Reifler for useful comments.

Abstract

The onset of interstate conflict often hinges on seemingly random events ('sparks') such as the assassination of Franz-Ferdinand in 1914. However, the literature on the causes of interstate war has mostly focused on identifying fertile grounds ('powder kegs'), ignoring these intricacies of history that are typically treated as noise. Yet this approach cannot explain why certain fertile grounds remain peaceful, or why wars start precisely when and where they do. Here, we measure and demonstrate the importance of these idiosyncratic factors with monthly information about sparks from three different sources: a) one hundred years of newspaper articles; b) two hundred years of government bond yields; and c) fine-grained event-data. These measures of triggers significantly improved our ability to explain and predict conflict. In particular, we found that fertile grounds or the occurrence of a trigger are both prone to conflict, but it is their combination that is disproportionately dangerous.

Keywords: International Conflict; War; News; Government Bonds; Prediction; Triggers

Word Count ≈ 8,700

On June 28, 1914, a young man pressed a pistol's trigger and unleashed a series of consequences ultimately leading to one of the worst human tragedies in history. Whether World War I would have occurred regardless of Archduke Franz Ferdinand's death is indeterminable, but the assassination is widely accepted as the spark that lit the powder keg of European territorial rivalries.[1] Many conflicts follow a similar pattern, by which a 'trigger' or a collection of events leads to escalation and ultimately the outburst of violence between states. These events explain why certain fertile grounds degenerate into war, whereas seemingly similar ones do not, and why World War I started in 1914 rather than in 1915 or 1913.

Unfortunately, existing research on the causes of war has mostly focused on determining the fertility of the ground for conflict. Models typically rely on 'structural' variables—information about the state of the system, the attributes of the countries or of the dyad such as regime type, distance, or military capabilities—to estimate the stability or instability of the system. An unstable equilibrium is one in which a spark would suffice to set in motion a path to conflict. However, without information about that spark, the powder keg often remains just that—a dangerous confluence of risk factors indeed, but one that is harmless as long as it is not lit.

To be sure, the importance of short-term factors on the onset of war has long been recognized, as evidenced by extensive work on crisis negotiations, escalation, and the collection of large event data. However, researchers have so far not been able to show empirically the relative importance of these short-term changes as opposed to longer-term trends. In particular, what is the effect of short-term developments on the probability of war onset, and how much of the variance in that probability do they account for? Equally important, are these proximate changes sufficient to cause large-scale events? In other words, is a large 'trigger' sufficient to cause war? Or are short-term changes only consequential when paired with a dangerous structure?

[1] On the causes of the first World War, see for example Clark (2012).

Answering these questions is central to our understanding of the causes of war, and more generally of history. Is the entire historical process determined by long-term trends—a process in which short-term events would explain little of the variance—or rather one in which stochasticity plays a meaningful role? And if short-term factors matter, do they matter even more when the underlying structure is ripe? If small, seemingly random events play a crucial role in the onset of large-scale events such as wars, then our ability to anticipate them will be significantly hampered. If, on the other hand, the structure is the main determinant of the onset of war, then this indicates that the onset of conflicts, and more generally history, may be less stochastic than expected.

In this paper, we aim to show and quantify the importance of these short-term changes—the seemingly random intricacies of history that are typically treated as noise and relegated to the error term. What is the role of short-term events as opposed to structural factors in the onset of war, and how do the effect of triggers vary depending on the underlying risk of the structure? We show that this 'noise' significantly improves our ability to predict and explain which fertile grounds end up in conflict. In particular, we find that the combination of the spark and the powder keg is crucial: a spark in an unstable equilibrium is likely to lead to conflict, whereas a comparable one in a stable equilibrium will be of little consequence.

An obvious challenge of this research agenda, however—one that explains why it has received little attention—is the difficulty to measure these sparks. They can take many different forms, and it is almost impossible to predict when, where or how they will appear. In addition, their meaning and implications depend on the cultural context, or even on the order in which they occur, so that including them systematically in our models is particularly challenging.

Instead of attempting to measure these multi-faceted events directly, we therefore chose to observe a metric that reveals their occurrence. In particular, we argue that careful observers of international affairs would be able to detect and interpret the meaning and likely consequences of a trigger. What we need, therefore, are the opinions and analyses of these daily observers. We obtained such information from two main sources. First, by searching millions of newspapers for indications of growing tensions—a proxy for the occurrence of a trigger. Second by analyzing decades of financial data and in particular government bond yields as a reflection of market participants estimated risk associated with a given country. Out of concerns for robustness, we also validated our results using fine-grained event data obtained from King & Lowe (2003)'s 10 million international dyadic events

The paper proceeds in three steps. We first discuss the idea that an equilibrium can be dramatically affected by small disturbances. We then present our data on historical newspapers, government bond yields and dyadic events, as well as the structural variables and conflict data we use. Finally, we estimate models in which we add our measures of idiosyncrasy to existing 'structural' models. By measuring how much models are improved by the addition of these proxies for triggers, we are able to indirectly measure the impact of randomness on the onset of war. In particular, we show that the trigger level indicators are significant predictors of the onset of war. We confirm this finding by demonstrating how their inclusion significantly improves upon the out-of-sample predictions of existing models.

Stable vs. Unstable Equilibria

The idea that an equilibrium can be dramatically affected by small disturbances is well-known in many physical, biological and social contexts. 'Tipping points', 'phase transitions' or 'bifurcations' refer to situations in which a small—and typically unpredictable—change can lead a system to shift suddenly from one equilibrium to another. Water, for example, can remain in a liquid state all the way to temperatures well below 32 degrees Fahrenheit. It freezes only in the presence of a disturbance—a crystal seed. In this example, temperature—the 'structural' variable—is insufficient to explain freezing by itself since, without the unpredictable disturbance of the seed, water will simply remain in a state highly *prone to* freezing, but still liquid.

[Figure 1 about here.]

These brittle equilibria—whether it be water or dyadic peace—imply that information about the stability of the system is insufficient; it is also necessary to know about the shocks it receives to characterize and predict its evolution. In particular, the *combination* of the trigger—the perturbation—and the state of the system is of central importance. An unstable system without a trigger may remain in equilibrium, just as supercooled water remains in a liquid state, and two countries may be prone to war and yet remain peaceful. Conversely, a system should not be affected by a shock unless it is unstable to start with (Fig. 1). In other words, wars tend to befall those dyads with a fertile ground for conflict *and* in which a trigger has occurred. In a stable equilibrium, the assassination of the Archduke in 1914 may have been of no consequence; but in the context of intertwined alliances, rapid shifts in power, differing political regimes and rapidly arming countries with close geographic proximity, this spark led to war.

Unfortunately, the literature on the causes of war has focused mostly on determining what constitutes a fertile ground for interstate conflict. The stability of the system is characterized by 'structural' variables'—variables that tend to remain relatively stable over long periods of time. For example, military spending (Glaser 2000), long-standing territorial rivalries (Huth 1998), large and rapid shifts in power (Powell 2004, Chadefaux 2011) or particular alliance patterns (Signorino & Ritter 2002, Leeds 2003) are some of the factors that have been associated with an elevated risk of interstate conflict. In World War I, for example, the fertile ground was formed by the web of alliances, the growing power of Germany, the Anglo-German naval race and the distribution of political regimes in Europe. These factors tend to be relatively stable over time.

However, this approach cannot explain why certain fertile grounds remain peaceful despite their similarity to others that lead to war, or why wars start precisely when and where they do. While many models show the existence of a fertile ground for war in Europe in the early 20$^{\text{th}}$ century, few can explain why it occurred where and when it did, or even why it occurred at all. Indeed, the onset of war often depends on events or sequences of events that are idiosyncratic, unpredictable and, most importantly hardly quantifiable (Gartzke 1999). A protest, aggressive speeches, the mobilization of troops, an election, a prime minister visiting a shrine or a terrorist attack are all possible sparks.

Historians have long understood the impact of stochastic factors on the stability of the system. The role of chance in history is often referred to as Cleopatra's Nose theory (Pascal 1946)—the idea that Cleopatra's nose distracted Mark Anthony to the point where his performance at the Battle of Actium was affected and, with it, the fate of the Roman Empire. This view contrasts with the more teleological perspectives of Hegel or Marx and Engels who, while acknowledging the importance of accidents, see them as random fluctuations around a historical path. Hume and later Carr give even less credit to chance (Hume 2001, Carr 1964, Carr & Davies 1961). For them, it is merely the term that we apply to a phenomena when we are ignorant of its causes (Hume 2001, p.46).[2] Indeed, we could have predicted that Franz-Ferdinand would be assassinated had we known the road conditions that day, the plans of the conspirators, and the mindset of Gavrilo Princip, in addition to the fact that Franz Ferdinand's car engine stalled and the gears locked, giving Princip an opportunity to

[2] For an excellent review of the problem of historical contingency, see Cederman (1997, pp. 38–44).

fire his pistol. But this information is beyond anyone's reach, so that these events are practically unpredictable. It is in this sense that we understand "unpredictable" events, which we call here accidents, sparks or triggers: as "things and events whose inner interconnection is so remote or impossible of proof that we can regard it as non-existent, as negligible."[3]

Here, we want to address this debate of proximate causes in the onset of war—and in history more generally—by quantifying the role of each.[4] In particular, how well can we predict war given information about structural conditions, but without information about the spark, and vice-versa?

Measuring Triggers

A central difficulty is the detection and quantification of triggers: what constitutes a trigger, and what is its magnitude? One approach consists in collecting 'event data'—records of actions and responses within and between countries. There are however limitations to this approach. First, coding is typically labor-intensive and the data's time-coverages are limited as a result. Recent databases such as DARPA's ICEWS or King & Lowe (2003)'s 10 million international dyadic events circumvent this problem with automated collection and coding, and we do use these data as robustness checks. However, event-data also suffers from other limitations. First, triggers may not be coded as events in existing data—the accidental death of an influential politician, the election of an extremist, a famine, or the discovery of new gas or oil reserves can all be triggers, but may not be coded in existing datasets. Event data by necessity only include a finite set of possible events, corresponding to what history has taught us are possible or likely triggers. While continuously adjusted (ex post) to reflect such

[3]Letter from Engels to Bloch, September 21, 1890, quoted in Talbot (2009).

[4]For a related debate in biology, see for example Gould (2000) and Vermeij (2006).

new evidence, it may still miss potentially important triggers that are new in their manifestations. Continuous improvements in these data may alleviate these concerns in the future, but at the moment still limit our ability to study long trends and patterns.

A related limitation of event data is that events and changes are often context-, history- and culture-dependent. How, for example, should a shoe thrown at President Bush be coded? Perhaps as an aggression, but much of the cultural undertones of the event would be missed. In other words, triggers can take countless forms and their implications are context-dependent, such that existing models cannot be amended by incorporating a variable for each of these possible variations. Again, improvements in text processing will keep refining the performance of these data, but we still think that at present they should be complemented by additional sources of information.

We therefore adopt a different approach. We let contemporaries decide what is a trigger and what is not, as well as its magnitude. A careful observer of international affairs should be able to make sense of ongoing events and would know, for example, that a shoe thrown in Chicago has a different meaning from one thrown in Baghdad. In other words, the consequences of specific events can often be better foreseen by those well acquainted with their context than by scholars attempting to classify these events after the fact.

What we need, then, is a measure of the perception that observers of international affairs have of events and international or domestic developments—or the absence thereof. We argue that such a measure can be derived from two different sources: news and financial markets. First a systematic analysis of *historical newspapers* for indications of tensions can reveal growing concerns about war, and hence act as an indicator of the occurrence of a trigger. Second, *government bond yields* reflect market participants' belief in the ability of the government to repay its debt—the probability of which is strongly affected by conflict. Hence the yield demanded by investors to hold these bonds should reflect at least partially the investors' belief about the probability of a coming war.

The reader may be concerned that our measures are not really capturing 'triggers', but rather some form of structural variables for which we simply do not have data or about which we may not even know. Or we may succeed in measuring proximate events, but ones that are really related to the structure and should therefore not be considered as triggers per se. A vote to increase the military budget, for example, would be a structural change which might lead to a cascade of news and increase in bond yields. In this case, the 'trigger' we might claim to measure would really be an alternate measure of the change in a given structural variable, but with finer-grained temporal resolution. In other words, our results might be driven by our access to data that is closer to the event. We therefore tested whether including variables measuring structural factors closer to the event would void our results. We show that this is not the case.

Second, our measure of triggers may simply be capturing the information from other missing variables. In that sense, they would not be a measure of triggers, but rather a proxy for missing structural variables. After all, we cannot include every conceivable variable, and there may even be others that have not yet been identified in the literature. This is, by its very nature, impossible to rule out entirely, but we propose a method to show that it is implausible. In particular, we show that the marginal gains from additional structural variables decreases rapidly, regardless of the order in which they are added, and that beyond a few, the predictive power of additional structural variables even decreases. This suggests that adding an increasing number of variables is unlikely to ever yield the same explanatory power as the addition of measures of triggers. These two issues are addressed in more detail below.

Finally, we are aware that the structural risk index and the trigger level index are likely to overlap. First, triggers tend to occur in fertile grounds. Conversely, some of the triggers become structural features of a country or a dyad—a coup for example is a trigger, but it also changes the regime and is therefore incorporated the next year into the structural variables (we address this point more directly below). In addition, the opinions of contemporaries are not only determined by current events, and hence are not a measure of triggers alone. Conciously or not, journalists and market participants form more or less formal models of the underlying probability of conflict based on structural variables, and interpret current events in light of this likelihood. For example, any journalist would seriously discount the probability of a conflict between Nicaragua and Burundi. This means that our proxy for the trigger does not only incorporate information about the trigger, but also about the fertility of the underlying ground. In short, then, we expect the structural risk index and the trigger index to be correlated. Yet what matters is that this correlation is imperfect, and it is that extra-information that we aim to separate and quantify using multivariate (logistic) regression. Finally, we believe that out-of-sample predictions are the best metric to judge whether proxies for triggers actually add any information to structural models. If they do, then the occurrence of a trigger is valuable information and cannot be easily dismissed as a simple consequence of a high underlying structural risk.

Data

Measures of Triggers

Newspapers. The press is an ideal source of information because it provides fast, accurate and in-depth coverage of events throughout the world. A database of news also avoids the problem of hindsight by using only information that was available at the time and, by consistently applying the same methodology to every war, prevents any temptation to cherry-pick the evidence. Finally, newspapers have an important advantage over event-base data: they can report tensions even when no actual event occurred (and hence nothing is recorded in, say, COPDAB). Conversely, an event may not be perceived as significant by its contemporaries. In other words, an analysis of news gives us information about the interpretation of events by their contemporaries, and not an event description from which meaning needs to be inferred *a posteriori*, with the benefit of hindsight.

To estimate changes in domestic and international tensions, we relied on data from Chadefaux (2014), in which the largest available database of newspapers, *Google News Archive*, was used.[5] This wide collection includes a large proportion of all English-speaking newspapers, ranging from major publications such as *The New York Times*, *The Washington Post* and *The Guardian*, to more obscure local ones such as the *California Oil Worker* or *The Cambridge City Tribune*.[6] In all, the database spans more than 200 years and consists of over 60 million pages. It also includes as subsets major providers of news archives such as Proquest Historical Newspapers, thereby making it the world's largest database in terms of the number of articles referenced. This comprehensiveness has the added advantage of smoothing out any particular newspaper's biases, such as those caused by their geographic location, their political orientation or their substantive focus (Chadefaux 2014).

Within these data, the entire text of every article was searched for every month from 1902 to 2010 for mentions of a given country together with a set of keywords typically associated with tensions.[7] Thus, a sample search would be "Pakistan AND tensions OR crisis OR conflict [...]" for newspapers published in March 1948. This search yielded 33 results, indicating that 33 newspaper articles mentioned at least one

[5] `http://news.google.com/newspapers?nid=t_XbbNNkFXoC`.

[6] see http://news.google.com/newspapers for a partial list.

[7] The keywords used were: *tension(s), crisis, conflict, antagonism, clash, contention, discord, dissent, disunion, disunity, feud, division, fight, hostility, rupture, strife, attack, combat, shell, struggle, fighting, confrontation, impasse*

of our keywords together with Pakistan in their text. This procedure was repeated for every month from January 1902 to December 2010, and for every country included in the Correlates of War dataset (Correlates of War Project 2008).[8]

The resulting data form a fine-grained measure of contemporaries' (journalists) perceptions of tensions and their evolution over time. We denote the number of conflict-related news for a given pair of countries in a given month by $News_{ij,t}$, corresponding to the average of the (logged) number of news for each country in that month. Note that we aggregated the number of conflict-related news at the dyadic level for consistency with the other variables we use below. However, disaggregating this variable into its country components has little effect on our results.

Government Bond Yields. News items are a strong proxy for the occurrence of a trigger, but they are not perfect. Journalists may influence each other, tailor their articles to their audience, or be subjected to government influence. In addition, data collection constraints limit us to news reported in English, which may introduce some bias. News from North Africa, for example, may be better reported in French than English. To strengthen our results, we therefore relied on a second proxy for the occurrence of a trigger: financial market data. In particular, we used government bond yields as indicators of the financial markets' expectations about the likelihood of war.[9]

Government bonds are the long-standing way by which governments borrow money to fund their expenses. The bond's interest rate—its yield—reflects, among other factors, the risk associated with it. The risks are twofold. First a credit-risk: the government might simply fail to repay part or all of the money borrowed. Second, inflation might reduce the real interest perceived by the investor. Conflicts, and war

[8] See Chadefaux (2014) for a more complete discussion of the data.

[9] On using market data to measure of investors' expectations of conflict, see also Ferguson (1999, pp. 131–5) and Schneider (2012).

in particular, are strongly associated with both risks, as they tend to be costly endeavors that imply large deficits, often financed through inflationary policies. As a result, an investor who expects a war to occur before the bond maturity will only buy the bond if its yield is sufficient to compensate the risk of holding it, and we therefore expect that bond yields will rise prior to war, and in particular in response to events that are likely to trigger or speed its onset.[10]

Here, we relied on government bond yields data provided by *Global Financial Data* (GFD).[11] Data on government bond yields from 45 countries were collected (10-year bonds when available, shorter maturity otherwise). The data date back to the early 19$^{\text{th}}$ century for the United States, the United Kingdom, Russia or the Netherlands but is more recent for other countries (see SI table A.1 for further details). The resulting variable measuring bond yields, $Bond\ Yield_{ij,t}$, is the average of the (logged) bond yield of each country. We also note that bond yields may be correlated with other macroeconomic variables such as inflation, central bank rates, debt levels or the flexibility of the exchange rate. We included these variables in alternate specifications of our model, with no substantial effect on our results (see SI and table A.6).

Minor Events. The news and bonds data are novel measures of the occurrence of a trigger—whether a tangible one such as a particular event, or a more impalpable one such as a change in mood. However, we also want to ensure the robustness of our results to regular event data, and therefore also tested our results on the "10 million international dyadic events" dataset (King & Lowe 2003), which automatically extracts information about international conflict and cooperation between 1990 and 2000 by 'reading' reports from news agency and automatically assigning them a classification. In particular, we gathered the Goldstein scale (positive and negative)

[10]The absolute value of the yield, however, may be less informative than its change over time, as the overall yield is influenced by many underlying economic variables (see below).

[11]http://www.globalfinancialdata.com.

assigned to each event, and average their values by dyad-month.[12] King & Lowe (2003)'s data have the advantage of being fine-grained and of directly recording the occurrence of events (i.e., potential triggers). Unfortunately, they are limited in time, which impairs our ability to draw inferences about the onset of rare large conflicts. Regardless, we find that our results are also supported by this additional datasets, further supporting the evidence from the two other, larger datasets.

In sum, we relied on three different datasets as independent variables: 'news', 'bonds' and King & Lowe (2003)'s '10 million events'. Our goal is to strengthen our findings by ensuring their robustness against different proxies for short-term shocks—triggers. While each proxy is imperfect, support for our hypothesis across several of these data would be strong evidence that structural variables alone are insufficient to understand the onset of interstate wars.

Measuring Structural Risk

To measure structural risk—how fertile the ground is to interstate war—we largely followed the existing literature, and in particular Beck, King & Zeng (2004) and De Marchi, Gelpi & Grynaviski (2004). The following variables were included: a dummy for geographic contiguity (*Contiguous*), and a measure (*Distance*) of geographic distance (Huth 1998); a dummy for the existence of a defense, neutrality or entente alliance (*Ally*) (Gibler 2009); an index from -1 to 1 (*Similarity*) which measures the similarity in alliance portfolios between dyad members (Signorino & Ritter 2002);[13] a dummy indicates whether the dyad includes at least one major power (*Major Dyad*). A measure of power imbalance within the dyad (*Asymmetry*), ranging from 0 (equality) to 1 (perfect inequality) is also added (Ray & Singer 1973, p.

[12] We further tested the robustness of our results using the GDELT dataset, and found the results to be entirely consistent with those presented here. Unfortunately, GDELT's legal status has recently become uncertain, and we therefore choose not to publish these results at this point.

[13] We used the weighted S (binary alliance data) provided by Häge (2011).

422), together with a measure of the dyad members' absolute capability levels using the widely-used Composite Index of National Capabilities ($Cinc_i$) from the Correlates of War (Singer, Bremer & Stuckey 1972, Singer 1988, v4.0). We also include information about regime type (*Polity*) for each dyad member, ranging from -10 (autocracy) to +10 (democracy), as well as an interactive term of the two (*Joint Democracy*) (Marshall, Jaggers & Gurr 2002). In addition, we added a variable measuring the time in years (or fractions thereof) since the last conflict (*Peace Years*, or PY) as well as the square and cube of this variable (PY^2 and PY^3) as a measure of temporal dependence (Carter & Signorino 2010).[14] We also added a variable measuring the age of the dyad (*Dyad Age*), the idea being that for new dyads (such as those that emerged in the 1960s as a result of decolonization), the value of Peace Years will not be as informative as for older dyads, and hence the effect of Peace Years is likely to be interactive with the duration of the dyad.

Measures of Conflict

We use the dyadic Militarized Interstate Disputes dataset (Gochman & Maoz 1984, Jones, Bremer & Singer 1996). However, we depart from previous studies which relied on yearly data, and instead study it at the monthly level (dyad-month) to take advantage of our finer-grained independent variables. Using yearly data instead would imply relying on indicators of tensions from the preceding year, which would defeat the purpose of estimating the impact of the trigger. Following Beck, King & Zeng (2000), we also limit our analysis to politically relevant dyads. The set contains

[14]This variable controls for the fact that conflicts are not independent events, and that a conflict is more likely to occur shortly after another than after a long period of peace (Beck, King & Zeng 2000, Beck, Katz & Tucker 1998).

783,685 monthly dyads between 1816 and 2001.[15] The dependent variable is coded as one for dyad-months for which a war starts within the next twelve months, and zero otherwise. We limit our attention to wars (battle deaths $\geq$ 1,000), since our hypothesis relates to the onset of large conflicts and not the small skirmishes which can often be assimilated to triggers themselves.[16] In total, 1,972 dyad-months are coded as 1 for the onset of a conflict within the next year. Summary statistics of our variables are available in table A.2.

[Table 1 about here.]

Results

We now test our hypothesis that interstate wars tend to occur as the result of the co-occurrence of two factors: a fertile ground (high structural risk), as measured by indicators developed in the literature (e.g., geographic proximity, capabilities, regime type); and the occurrence or development of a 'spark' or 'trigger', which we measure using various proxies (news, government bond yields & event data).

Logistic Regression

To assess the respective role of triggers—short-term instabilities—and structure in the onset of conflict, we first estimated the probability of conflict in a given dyad-month using two different models: one based only on structural variables (the workhorse of conflict studies), and the other solely on proxies for triggers. If the latter model out-

[15] However, not all our measures of triggers extend over this entire period (see table 1), and hence our models will include fewer observations.

[16] A similar improvement in predictions also applies if we include small skirmishes as well, though as expected the improvement in forecasting is somewhat lower.

performs the first, then the discrepancy is evidence of the role of proximate causes—triggers—in the onset of conflict.

To derive the probability of conflict based on structural variables only, we estimated the following standard logistic regression model:

$$P(\text{Onset}_{ij,t} = 1) = \Lambda\left(\alpha_0 + \sum_{1}^{s} \alpha_s \text{Structural Var}_{s,t}\right), \tag{1}$$

where $\text{Onset}_{ij,t} = 1$ when a conflict occurs within one year of time t between countries i and j, $P(\text{Onset}_{ij,t} = 1)$ is the probability of that event, $\Lambda(z) = e^z/(1 + e^z)$ is the logistic cumulative distribution function, and 'Structural Var' are all the variables described in section .[17] All independent variables are lagged by one year. In this and all subsequent logits, the coefficients were corrected for the bias inherent to rare event logistic regression (King & Zeng 2001*b*, King & Zeng 2001*a*). In addition, standard errors are assumed to be clustered by country dyads.[18] The results of this model are reported in table 2 (model 'Structural').

To estimate the probability of conflict based on our proxies for triggers only, we relied on the following model:

$$P(\text{Onset}_{ij,t} = 1) = \Lambda\Big(\beta_0 + \beta_1 \text{Trigger}_t + \beta_2 \Delta\text{Trigger}_t\Big), \tag{2}$$

where 'Trigger' $\in$ $\{\text{News}_{ij}$, Bond yield_{ij}, $\text{Goldstein}_{ij}\}$ respectively, depending on which measure of triggers we use (see table 1), and ΔTrigger is the monthly change

[17] I.e., 'Structural Var' $\in$ {PY, PY^2, PY^3, CINC_i, CINC_j, Asymmetry, Asymmetry^2, Contiguous, Distance, Ally, Similarity, Major Dyad, Polity_i, Polity_j, Joint Democracy}.

[18] Failure to do so may underestimate standard errors, since observations for a given dyad are likely to be correlated across time. Because our data combines yearly and monthly data, we also estimated robust standard errors clustered by *both* dyad and year (Thompson 2011), with no notable difference in the results.

in that variable.[19] Each is lagged by one month.

We report in table 2 the results of this model ("Trigger") when 'trigger' is operationalized using both conflict-related news and government bond yields (our variables with the longest time-span). We find that the coefficients associated with our measures of trigger are significant and have the expected sign.[20] The same tables for each operationalisation of 'trigger' (news only, bonds only and Goldstein scale only) are reported in the SI (tables A.3–A.5). We find that each of the measures of trigger is significant, both in a model alone and when combined with the structural model ("Structural + Trigger").

[Table 2 about here.]

Yet our goal here is not to simply show the significance of our measure of triggers—although this alone reveals their importance—but rather that it is the combination of triggers and a fertile ground that is most conducive to war. We therefore calculated, for each joint value combination of the two models' estimated risk, the frequency at which conflict occurs. Figure 2 clearly validates our hypothesis: wars are most likely when both trigger and structural variables are high. The predicted values from the structural model alone explain some of the variance in the onset of conflict and confirms the expected result that an increase in structural risk leads to an increased probability of war onset in the dyad. However, a large portion of that variance is to be attributed not to the fertile ground, but rather to the size of the trigger.

[Figure 2 about here.]

[19] I.e., $\Delta\text{Trigger}_t = \text{Trigger}_t - \text{Trigger}_{t-1}$. In the case of King & Lowe (2003)'s data, 'Trigger' uses both the positive and negative side of the Goldstein Scale (see table A.5).

[20] The negative sign for Bond Yield confirms the difficulty of drawing general inferences from the absolute bond yield (see discussion above), and confirms that the change in yield is a more meaningful variable.

We confirm these findings more formally by using the fitted values from the structural model (1) and the trigger model (2) as composite measures of 'structural risk' and of the 'trigger level' respectively. Plugging these values back into a model of conflict onset, we can determine whether the predictions derived from the proxies for triggers are significant—in other words, if any of the variance in the occurrence of conflict is accounted for by short-term changes, taking into account the predictions of the structural models. An interaction term between structural risk and trigger level also takes into account the possibility that the effect of a trigger is larger when the underlying structural risk is high (see fig. 1 and associated discussion on p. 5).

$$\begin{aligned} P(\text{Onset}_{ij,t}) = \Lambda\Big(&\gamma_0 + \gamma_1 \text{Structural risk}_{ij,t} \\ &+ \gamma_2 \text{Trigger level}_{ij,t} \\ &+ \gamma_3 \text{Structural risk} \times \text{Trigger level}_{ij,t}\Big) \end{aligned} \tag{3}$$

If the estimates derived from news and market data do not differ from or improve upon those of the structural model, then the coefficient on the "trigger index" will be small or insignificant; if instead our hypothesis that short-term disturbances matter is correct, then predictions based only on an index of structural risk will underperform their combination with the predictions derived from indices of trigger levels.

The results of model (3) for different proxies for the size of the trigger (using news, bonds or the Goldstein scale) are reported in table 3. We find that both the trigger index and the structural risk index variables are significant and have the expected sign. The interaction term is not significant, but we know that a significant interaction term is neither necessary nor sufficient for variables to actually interact meaningfully in affecting our quantity of interest $Pr(Y)$ (see in particular Berry, DeMeritt & Esarey

(2010)).[21]

To test for an actual interaction, we therefore used second differences and estimated standard errors by bootstrapping:[22] We compute a second difference in Pr(war) involving changes in trigger and structure from their 75$^{\text{th}}$ to their 95$^{\text{th}}$ percentile, when the other independent variables are held at central values.[23] We obtain a second difference of 0.012, which is statistically significant at the .01 level (confidence interval of $[0.011, 0.00014]$). This means that an increase in the trigger from 'low' to 'high' when the level of structural risk is 'high' leads to an additional increase in $p(war)$ of 0.012 over the same increase when structural risk is low. This result is not specific to our choice of high and low percentiles, as shown in Figure 3, which displays the estimated second difference for all combinations of low and high values.

[Figure 3 about here.]

Figure 3 shows that the effect on $Pr(Y)$ of moving from a low to a high trigger is always higher in a high structural risk environment than in a low risk environment. In other words, the effect of a trigger varies as a function of the level of underlying risk: there is an interaction between trigger and structure. In a low-risk environment, a trigger has little effect on $Pr(Y)$—we are in a stable equilibrium. In a high-risk environment, however, a trigger has a large effect on $Pr(Y)$—the system is unstable.

[21] In particular, our argument is that triggers and the structure affect the probability of war ($Pr(Y)$) interactively, but not the underlying latent variable Y^*. While all authors agree that a non-significant interaction term says nothing about whether there is actually any meaningful interaction, Rainey (2015) suggest that an interaction should be included anyway, which we have done here (similar results apply without it).

[22] Calculating standard errors using CLARIFY instead instead yields almost exactly the same results (Tomz et al. 2003).

[23] We chose these values given our theoretical argument that triggers will have a larger impact for high values of the structural risks. The second difference is calculated as:

$$\begin{aligned}\Delta\Delta[Pr(war)] =&[Pr(War|trigger = High, structure = High)- \\ &Pr(War|trigger = Low, structure = High)]- \\ &[Pr(War|trigger = High, structure = Low)- \\ &Pr(War|trigger = Low, structure = Low)] \end{aligned} \tag{4}$$

[Table 3 about here.]

The interaction is also shown in figures 2 and 4: the onset of war is a function not only of an elevated structural risk, but also of a large short-term disturbance. Even a dyad with a high structural risk will experience little to no conflict in the absence of a trigger (Fig. 4B). Similarly, a large trigger in a low-risk environment is unlikely to disturb the peace (Fig. 4A). This confirms that the onset of wars depends not only on structural factors, but also on short-term, proximate causes, as well as on the interaction between the two. Conflicts are likely to occur not only when the ground is fertile for war, but also when a spark lights the powder keg.

[Figure 4 about here.]

Out-of-Sample Predictions

Beyond significance tests, we agree with Beck, King & Zeng (2000) and Ward, Greenhill & Bakke (2010) that out-of-sample forecasting performance should always be one of the standards used to judge studies of international conflict. We therefore show in this section how the incorporation of proxies for triggers into a structural model dramatically improves our ability to *predict* the onset of conflict.[24]

[24] The prediction of conflict has recently received increasing attention, whether it be for interstate wars (Beck, King & Zeng 2000, Brandt, Freeman & Schrodt 2011, Gleditsch & Ward 2013, Ward, Siverson & Cao 2007), civil wars (Weidmann & Ward 2010), or other political disruptions, from state failure to political instability, genocides, human rights violations or ethnic conflicts (Bueno de Mesquita 2009, Goldstone et al. 2010, Schneider, Gleditsch & Carey 2010). Some work has focused on predicting the evolution of single conflicts (Pevehouse & Goldstein 1999, Schrodt & Gerner 2000), sometimes using fine-grained data as we have here (Schneider 2012). Prediction markets (Arrow et al. 2008, Berg, Nelson & Rietz 2008, Wolfers & Zitzewitz 2006) are relatively new and hence provide only a limited number of data points. Game-theoretic approaches focusing on prediction of individual conflicts have also yielded encouraging results (Bueno de Mesquita 2002, Feder 2002), but they typically rely on detailed information from issue or area experts.

We first follow Beck, King & Zeng (2004) in comparing both models' improvement over a baseline model. This baseline is a very simple model, only incorporating information about temporal dependency using a "Peace Years" variable, which records the number of months a dyad has been at peace at time t (together with its square and cube), as well as geographic information about the dyad (contiguity and distance). It is used as a benchmark against which both the structural model and the model incorporating triggers can be compared. Other baselines models will be considered below (section).

The dataset was split into two sets: a learning set, from [jan. 1902,t] and a test set from [$t+1$year, $t+2$years].[25] In other words we train our model on a learning set of past data, and then make predictions about the occurrence of conflict within the next twelve months, using the coefficients derived from the learning set. This procedure is repeated for every year from 1920 to 2001 (1902–1920 is used as the initial learning set).

We find that the model with structural variables *and* measures of triggers strongly outperforms the model with only structural variables, as measured by the Area Under the Receiver-Operating Characteristic Curve (ROC)—a classical way of evaluating binary forecasts. This is true whether we measure triggers using conflict-related news only, bond yields only, or a combination of the two (the combination does best).[26]

[Figure 5 about here.]

Indeed, when defining the dependent variable as the onset of conflict within one

[25] We exclude [t, $t+1$ years] from the sample to avoid contamination effects from yearly variables (although this is not essential for our results). Using a moving window (e.g., 20 years) for the learning set yields similar results.

[26] We do not, however, test it with King & Lowe (2003)'s data, since the time-span is much too short (and hence the number of conflicts far too limited) to be able to create learning and testing sets.

year, we find that the model with only structural variables reaches an area under the curve of 0.75, a gain of 0.11 over the baseline AUC of 0.64, whereas the model with measures for triggers (news + bonds) has an AUC of 0.83—a gain of 0.19, or more than a 70% additional increase in AUC, compared to the increase from the structural model alone (Fig. 5). The improvement is yet stronger when the onset of war is coded more narrowly as '1' only if the onset occurs within the next three or one month. The added value of the measures for triggers is even more obvious when measured using Precision and Recall—a metric often used in information retrieval as an alternative to ROC curves in the case of skewed data (Fig. 5B) (Manning & Schütze 1999, Bockhorst & Craven 2005). More precisely, we used the average F1 score, a measure that combines precision and recall.[27] Here we note that the addition of information about triggers more than doubles the added value of the structural variables over the baseline. Finally, separation plots (Greenhill, Ward & Sacks 2011) also convey the dramatic improvement gained by the inclusion of measures of triggers (Fig. 6).

[Figure 6 about here.]

A Straw Man?

While the results strongly support our hypotheses, two possible criticisms should be addressed at this point. First, we may be stacking the deck against structural variables since our measures of triggers include information much closer in time to the onset of war. Second, the reduction in the magnitude of the error term—which we interpret here as evidence of short-term instabilities or chance—might simply reflect the imperfection of existing structural models.

Coarse Structural Variables First, structural variables may face an unfair hand-

[27]The F1 score is defined as $2 \cdot (\text{precision} \cdot \text{recall})/(\text{precision} + \text{recall})$.

icap in our model since they are lagged by one year (given their yearly resolution), whereas our measures of triggers are lagged by one month only. Thus to predict the onset of conflict in, say, July 1914, we are using information about Cinc from the end of 1913, whereas our news and bonds data extends to June 1914. This implies that the improvement in out-of-sample prediction, which we attribute to short-term disturbances, may simply be due to the fact that journalists and market participants observe structural changes more often than structural variables are updated in available data. For example, they may observe a spike in military spending in May or June, and infer from it a rise in geopolitical risk.

To rule out this explanation we conducted the same analysis as above, but this time without lagging structural variables. This gives them a strong advantage, as it means that structural variables incorporate information which is *posterior* to the onset of conflict. The United Kingdom's Cinc, for example, is 0.117 in 1912, 0.113 in 1913 (i.e., stable), but jumps to 0.138 in 1914. This is because 1914 also includes the second part of the year following the onset of WWI, during which the U.K. dramatically increased its military spending.

We find that the added value of our measure of triggers is largely intact, even when structural variables incorporate future information. Thus the improvement in the out-of-sample predictions, as measured by the area under the ROC curve, is at 61%, as opposed to 73% with the lagged structural variables—a reduction indeed, but an improvement that remains substantial. In other words, even with an unfair advantage, structural variables cannot close the gap with measures of triggers. This strongly suggests that it is not the mere closer temporal distance to conflict that leads to the low performance of structural variables, but rather their inability to capture short-term disturbances.

Imperfect Structural Models Another objection is that structural models may retain a sizeable error term simply because they are imperfect, and not because of partly unpredictable historical shocks. It is particularly difficult to rule out this explanation, since we cannot prove that yet-to-be discovered structural variables will not account for all the variance in the onset of conflict. One way to approach this problem, however, is by asking whether previous structural improvements have progressively reduced the explanatory power of triggers. If so, then we may expect that the added value of triggers will disappear with the inclusion of the relevant structural variables.[28]

Yet we show here that the improvement in out-of-sample forecasts gained by adding structural variables is marginally decreasing. The performance of structural models quickly reaches a plateau and even starts to decrease beyond a certain number of variables (additional variables do not necessarily improve predictions, as they may lead to over fitting). Although this does not guarantee that so-far undiscovered variables would not dramatically improve structural models' predictive power, it does suggest that there may be inherent limits to the ability of structural variables to improve forecasts and hence to understand conflict.

To show this, we calculated the area under the ROC curve of models composed of every possible combination of the available variables. However, given the large number of variables (and hence the unwieldy number of combinations), we grouped variables by categories: temporal dependence variables (PY + PY^2 + PY^3 + Dyad Age + PY$\times$Dyad Age); measures of relative and absolute power (Cinc$_i$ + Cinc$_j$ + Major Dyad + Asymmetry + Asymmetry^2); measures of alliance patterns (Ally + Similarity); measures of geographic proximity (Contiguous + Distance); Regime type variables (Polity$_i$ + Polity$_j$ + Joint Democracy); and trigger indices (Bond Yield$_{ij}$+ΔBond Yield$_{ij}$+News$_{ij}$+ΔNews$_{ij}$).

We then combined the six groups of variables in every possible way, with any number of variables. For a model with one variable, for example, there are six possible

[28] See also Ward, Greenhill & Bakke (2010) for similar work in the context of civil wars.

models (e.g., a model with only power variables; only regime type variables; etc.). With two variables, there are 15 unique possibilities (Power and Alliances, Power and Geography, etc.). In total, 63 possible models were evaluated (a constant was added to each). We then calculated for each model's set of prediction the area under the ROC curve, according to the procedure detailed in section . This lets us find the best model with x variables. With 2 variables, for example a model that includes Polity and Power was the best-performing 'structural' model with an area under the curve (AUC) of 0.72. However, a model with two variables as well but replacing Power variables with information about triggers yielded an AUC of 0.82. The performance of the best structural models with x variables and the one of models incorporating measures for triggers (also with x variables) was plotted in figure 7.

[Figure 7 about here.]

Unsurprisingly, we find that raising the number of structural variables initially increases the AUC. Yet the increases are rapidly diminishing as the number of variables becomes larger. In fact, additional variables even end up diminishing the overall effectiveness of the model, as overfitting occurs. As a corollary, the improvement from the model including trigger over the structural model also reaches a plateau and never disappears. In other words, additional structural variables do not seem to gradually reduce the explanatory power of the variables measuring trigger, and we can hence conjecture that some of the variance in the onset of war will always remain out of the explanatory power of structural models.

Conclusion

Historians have long debated the relative importance of chance and necessity in history (Talbot 2009). Yet the debate has remained largely theoretical because the systematic measurement and quantification of accidents, chance, and imperceptible changes faced insuperable difficulties.

Instead of attempting to measure all idiosyncratic events directly we relied here on millions of contemporaries who observe current events and infer from them a likely course of events. Some, journalists, report on them—sometimes even on the most trivial ones. Others, market participants, rely on them to select their investments. Their decisions leave traces—articles and market data—that we use as indirect evidence of the occurrence of some form of trigger. These measures are far from perfect: our news analysis is crude, and bond yields often react to variables that are not always available. Yet, better measures would most likely only improve the predictive power, so that what we report here is likely to be a lower bound estimate of the improvement provided by information about triggers.

Overall, our results suggest that while conflicts may be predictable, there are limits to the ability of structural models to do so. Indeed, idiosyncratic factors—which we loosely refer to as 'triggers'—often explain the occurrence of specific crises, so that there might be an incompressible element of chance in conflict that limits our ability to systematically model their onset. In particular, we showed that it is the combination of a trigger and a powder keg that is particularly dangerous for peace, and hence that analyses of conflict without information about sparks will under-perform models that incorporate both. The consistency of our results across different proxies also serves to alleviate concerns over the flaws of a single measure.

Finally, our findings also have implications for the study of conflict, and in particular for the debate over large-N vs. qualitative methodologies (King, Keohane & Verba 2001). Structural models typically rely on large-N analyses, whereas qualitative work generally focus on reaching a more thorough understanding of the context, the actors and their interactions, and the accidents along the way. Our findings strongly suggest that both are complementary and, in fact, necessary to the understanding of conflict, and that one without the other will face major challenges in achieving strong results.

References

Arrow, Kenneth J., Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D. Nelson, George R. Neumann, Marco Ottaviani, Thomas C. Schelling, Robert J. Shiller, Vernon L. Smith, Erik Snowberg, Cass R. Sunstein, Paul C. Tetlock, Philip E. Tetlock, Hal R. Varian, Justin Wolfers & Eric Zitzewitz. 2008. "The Promise of Prediction Markets." *Science* 320(5878):877.

Beck, Nathaniel, Gary King & Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–36.

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.

Beck, Nathaniel, Jonathan N. Katz & Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Berg, Joyce E., Forrest D. Nelson & Thomas A. Rietz. 2008. "Prediction Market Accuracy in the Long Run." *International Journal of Forecasting* 24(2):285–300.

Berry, William D, Jacqueline HR DeMeritt & Justin Esarey. 2010. "Testing for interaction in binary logit and probit models: Is a product term essential?" *American Journal of Political Science* 54(1):248–266.

Bockhorst, Joseph & Mark Craven. 2005. Markov networks for detecting overlapping elements in sequence data. In *Neural Information Processing Systems*. Vol. 17.

Brandt, Patrick T., John R. Freeman & Philip A. Schrodt. 2011. "Real Time, Time Series Forecasting of Inter- and Intra-state Political Conflict." *Conflict Management and Peace Science* 28(1):41–64.

Bueno de Mesquita, Bruce. 2002. *Predicting Politics*. Ohio State University Press.

Bueno de Mesquita, Bruce. 2009. *The Predictioneers Game*. New York: Random House.

Carr, Edward H. 1964. *The Twenty Years' Crisis, 1919-1939*. New York: Harper Perennial.

Carr, Edward Hallett & Robert William Davies. 1961. *What is History?* Macmillan London.

Carter, David B & Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271–292.

Cederman, Lars-Erik. 1997. *Emergent actors in world politics: how states and nations develop and dissolve*. Princeton University Press.

Chadefaux, Thomas. 2011. "Bargaining over power: when do shifts in power lead to war?" *International Theory* 3(02):228–253.

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

Clark, Christopher. 2012. *The Sleepwalkers: How Europe Went to War in 1914*. Penguin UK.

Correlates of War Project. 2008. "State System Membership List, v2008.1." Downloaded from http://correlatesofwar.org.

De Marchi, Scott, Christopher Gelpi & Jeffrey D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(2):371–378.

Feder, Stanley A. 2002. "Forecasting for Policy Making in the Post-Cold War Period." *Annual Review of Political Science* 5(1):111–125.

Ferguson, Niall. 1999. *The Pity of War.* Basic Books.

Gartzke, Erik. 1999. "War is in the Error Term." *International Organization* 53(3):567–587.

Gibler, Douglas M. 2009. *International military alliances, 1648-2008.* CQ Press.

Glaser, Charles L. 2000. "The Causes and Consequences of Arms Races." *Annual Review of Political Science* 3(1):251–276.

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.

Gochman, Charles S. & Zeev Maoz. 1984. "Militarized Interstate Disputes, 1816–1976." *Journal of Conflict Resolution* 28(4):585–616.

Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted R. 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.

Gould, Stephen Jay. 2000. *Wonderful life: the Burgess Shale and the Nature of History.* New York: Random House.

Greenhill, Brian, Michael D Ward & Audrey Sacks. 2011. "The separation plot: a new visual method for evaluating the fit of binary models." *American Journal of Political Science* 55(4):991–1002.

Häge, Frank M. 2011. "Choice or circumstance? Adjusting measures of foreign policy similarity for chance agreement." *Political Analysis* 19(3):287–305.

Hume, David. 2001. *An Enquiry Concerning Human Understanding*. Vol. 3 New York: Oxford University Press.

Huth, Paul K. 1998. *Standing Your Ground: Territorial Disputes and International Conflict*. University of Michigan Press.

Jones, Daniel M, Stuart A Bremer & J David Singer. 1996. "Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15(2):163–213.

King, Gary & Langche Zeng. 2001*a*. "Explaining Rare Events in International Relations." *International Organization* 55(3):693–715.

King, Gary & Langche Zeng. 2001*b*. "Logistic Regression in Rare Events Data." *Political analysis* 9(2):137–163.

King, Gary, Robert O. Keohane & Sidney Verba. 2001. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

King, Gary & Will Lowe. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* 57(03):617–642.

Leeds, Brett A. 2003. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47(3):427–439.

Leetaru, Kalev & Philip Schrodt. 2013. "GDELT: Global Data on Events, Language, and Tone, 1979-2012." *International Studies Association Annual Conference, April 2013*. San Diego, CA.

Manning, Christopher D & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

Marshall, Monty G., Keith Jaggers & Ted. R. Gurr. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010. Version p4v2010." *Center for International Development and Conflict Management* .
URL: *http://www.cidcm.umd.edu/inscr/polity/index.htm*

Pascal, Blaise. 1946. *Pensées et Opuscules*. Hachette.

Pevehouse, John C. & Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions." *Journal of Conflict Resolution* pp. 538–546.

Powell, Robert. 2004. "The Inefficient Use of Power: Costly Conflict with Complete Information." *American Political Science Review* 98(02):231–241.

Rainey, Carlisle. 2015. "Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction." *Political Science Research and Methods* pp. 1–19.

Ray, James L. & David J. Singer. 1973. "Measuring the Concentration of Power in the International System." *Sociological Methods & Research* 1(4):403–437.

Reinhart, Carmen M & Kenneth Rogoff. 2009. *This time is different: Eight centuries of financial folly*. Princeton University Press.

Schneider, Gerald. 2012. Banking on the broker? Forecasting conflict in the Levant with financial data. In *Illuminating the Shadow of the Future: Scientific Prediction and the Human Condition*, ed. Paul Williamson Frank Wayman & Bruce Bueno de Mesquita. Ann Arbor, MI: University of Michigan Press.

Schneider, Gerald, Nils Peter Gleditsch & Sabine C. Carey. 2010. "Exploring the Past, Anticipating the Future: A Symposium." *International Studies Review* 12(1):1–7.

Schrodt, Philip A. & Deborah J. Gerner. 2000. "Using Cluster Analysis to Derive Early Warning Indicators for Political Change in the Middle East, 1979–1996." *American Political Science Review* 94(4):803–818.

Signorino, Curtis S. & Jeffrey M. Ritter. 2002. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43(1):115–144.

Singer, David J. 1988. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985." *International Interactions* 14(2):115–132.

Singer, David J., Stuart Bremer & John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills: Sage pp. 19–48.

Talbot, Ann. 2009. "Chance and Necessity in History: EH Carr and Leon Trotsky Compared." *Historical Social Research/Historische Sozialforschung* pp. 88–96.

Thompson, Samuel B. 2011. "Simple formulas for standard errors that cluster by both firm and time." *Journal of Financial Economics* 99(1):1–10.

Tomz, Michael, Jason Wittenberg, Gary King et al. 2003. "CLARIFY: Software for interpreting and presenting statistical results." *Journal of Statistical Software* 8(1):1–30.

Vermeij, Geerat J. 2006. "Historical Contingency and the Purported Uniqueness of Evolutionary Innovations." *Proceedings of the National Academy of Sciences* 103(6):1804–1809.

Ward, M.D., R.M. Siverson & X. Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.

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, Neils B. & Michael D. Ward. 2010. "Predicting Conflict in Space and Time." *Journal of Conflict Resolution* 54(6):883–901.

Wolfers, Justin & Eric Zitzewitz. 2006. Prediction Markets in Theory and Practice. Technical report National Bureau of Economic Research.

Figures

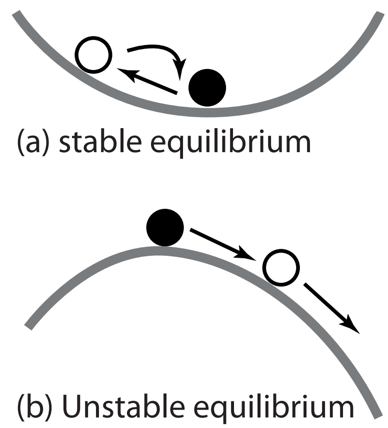

Figure 1: Stable and unstable equilibria. In a stable equilibrium, a small disturbance caused by frictions in a dyad has no long term consequences, as the system returns to equilibrium. In an unstable equilibrium, a small disturbance can degenerate into conflict.

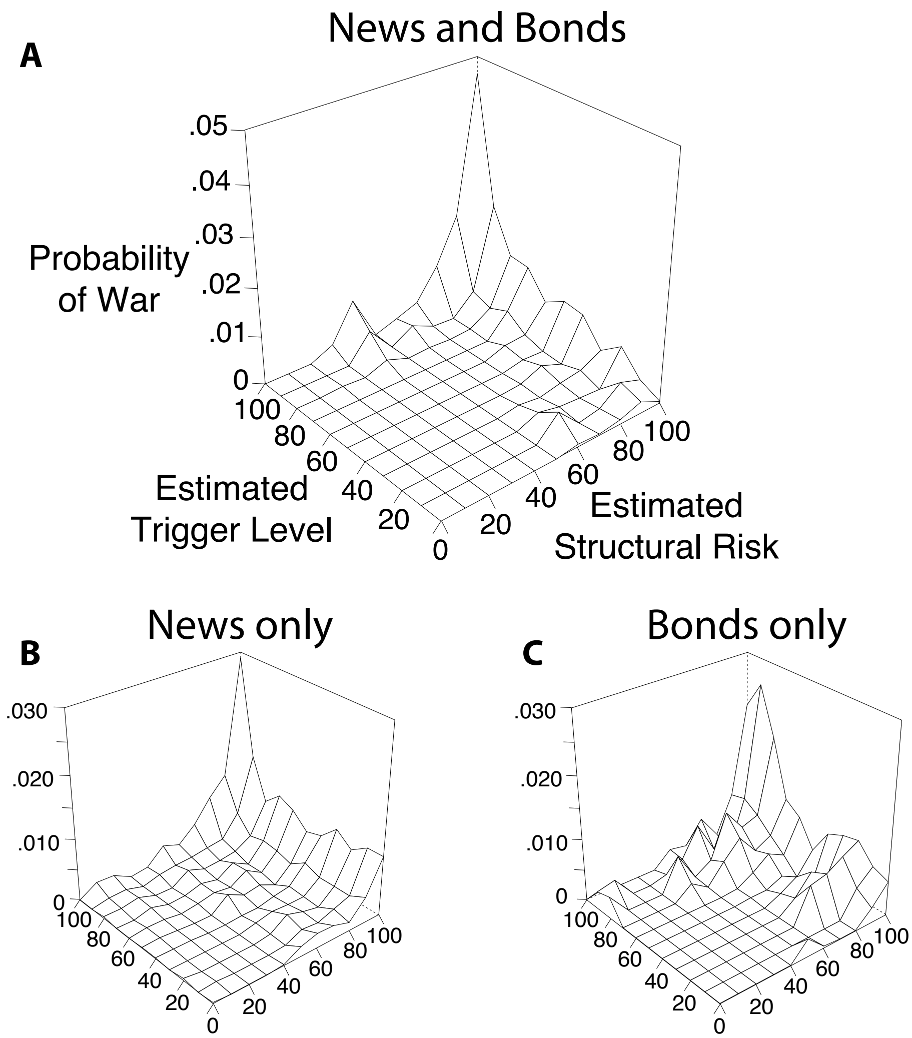

Figure 2: Observed values of the probability of conflict, as a function of the joint estimated value of the trigger (estimated from [1]) and of the structural risk (estimated from [2]). The figures show that war is most likely when a high structural risk (unstable environment) is associated with a large short-term disturbance (trigger). Panel A uses both conflict-related news and bonds yield variables as measures of triggers (table 2), whereas panel B and C use only news and only bonds, respectively (tables A.3 and A.4). Quantiles values of the independent variables are displayed on the x and y axes.

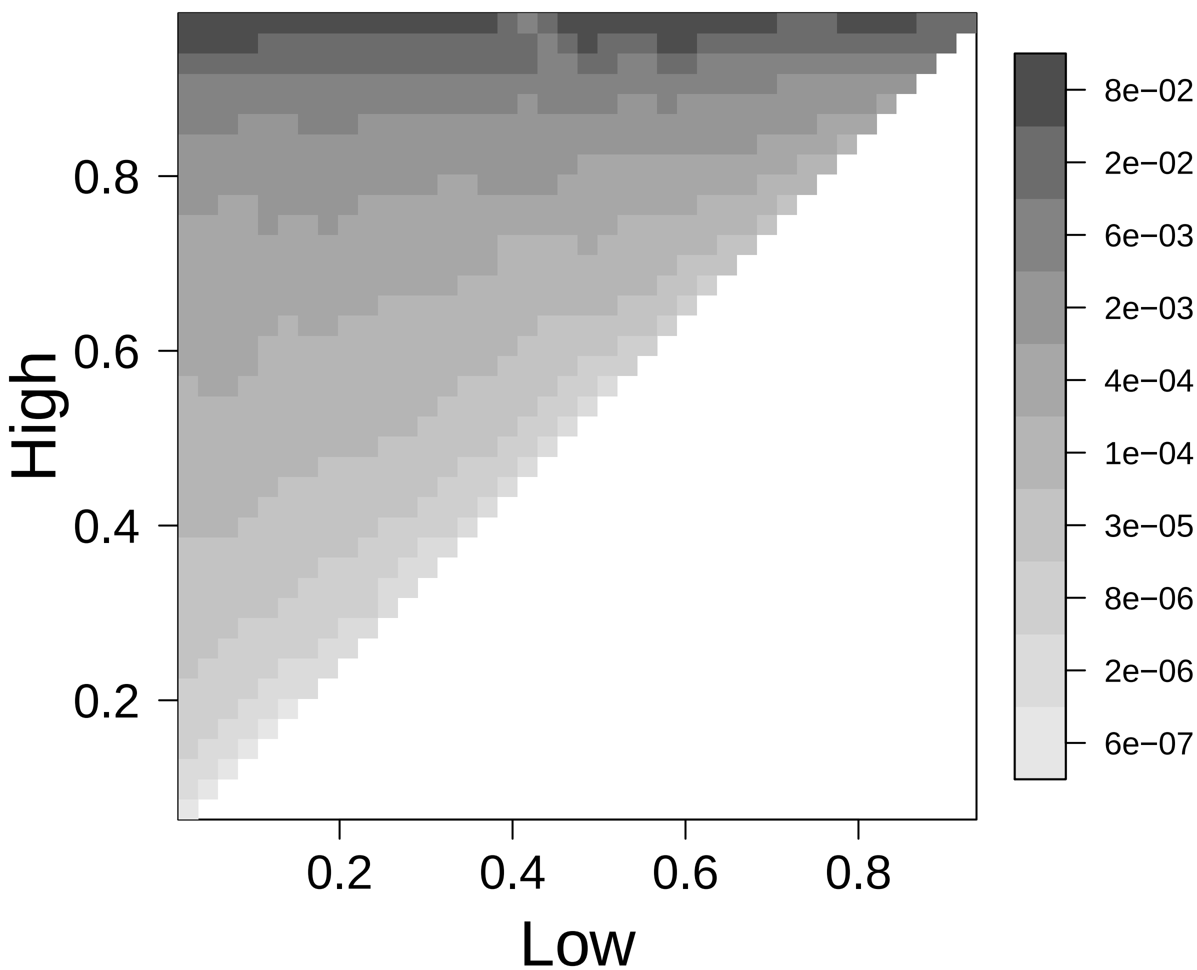

Figure 3: Heat map of second differences. Second differences are calculated as $\Delta\Delta[Pr(war)] = [Pr(War|HH) - Pr(War|LH)] - [Pr(War|HL) - Pr(War|LL)]$ The x axis has the value of 'low' and the y axis the value of 'high'. It can be read as: the effect of moving from a low trigger to a high trigger *given a high structural risk* is z percentage points larger than the same change in a low-structural risk environment, where z is represented by the color on the graph. More loosely, the darker the square, the more the underlying structural risk increases the effect of a trigger.

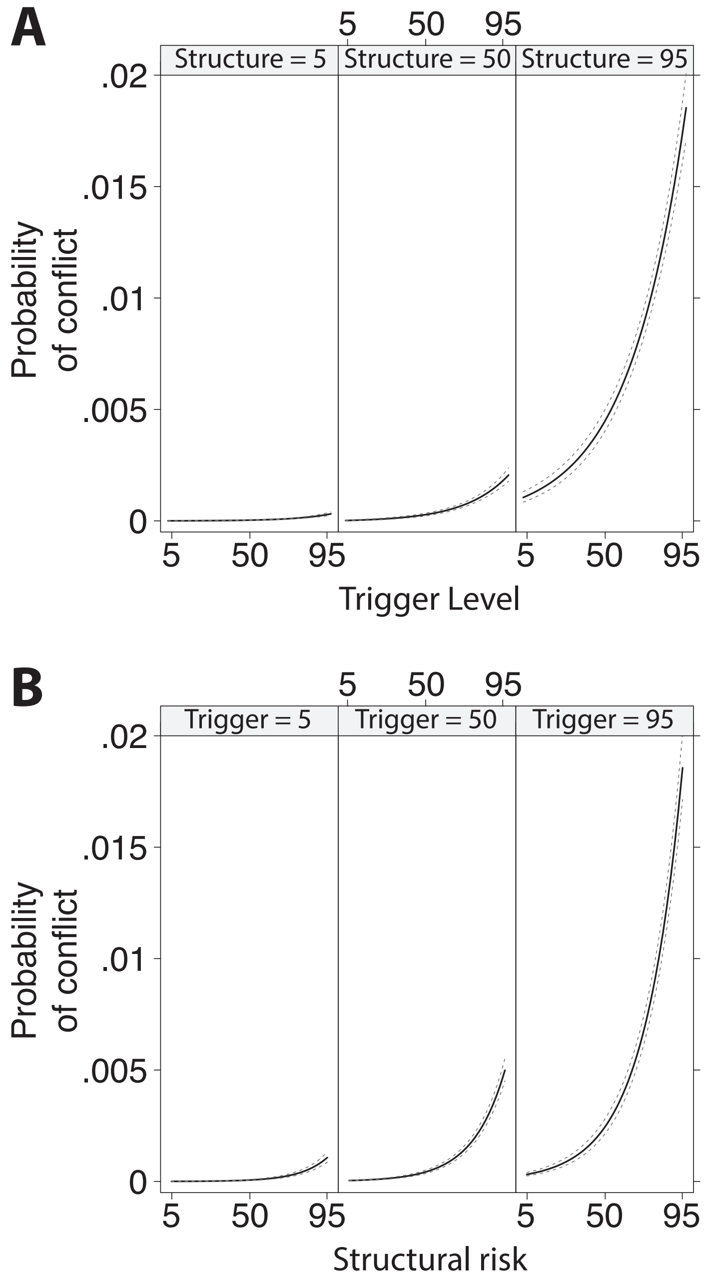

Figure 4: Fitted values of the probability of conflict, based on the joint value of the trigger and of the structural risk (see table 3), with 95% confidence interval. A combination of a high trigger in an unstable environment is most prone to war. Both News and Bonds were used to estimate the value of triggers, but similar results apply for each measure if used on its own.

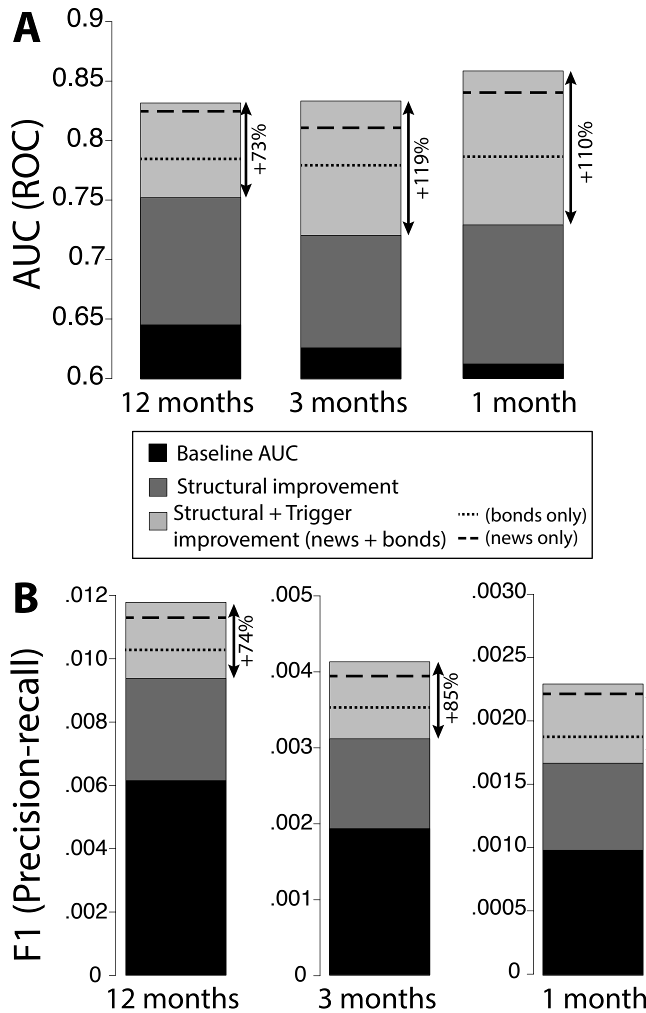

Figure 5: Measures of the out-of-sample predictive power of different models: the baseline model (a logistic regression with only a constant term, a correction for temporal dependence and measures of geographic proximity); the baseline augmented with structural variables ('Structural' in table 2), and the baseline with structural *and* measures of the magnitude of the trigger ('Structural + Trigger' in table 2). Different measures of the trigger were used: both News and Bonds (light gray); News only (dashed line); Bonds only (dotted line). **A**: Area under the Receiver-Operating Characteristic (ROC) curve; **B**: Average F1 score.

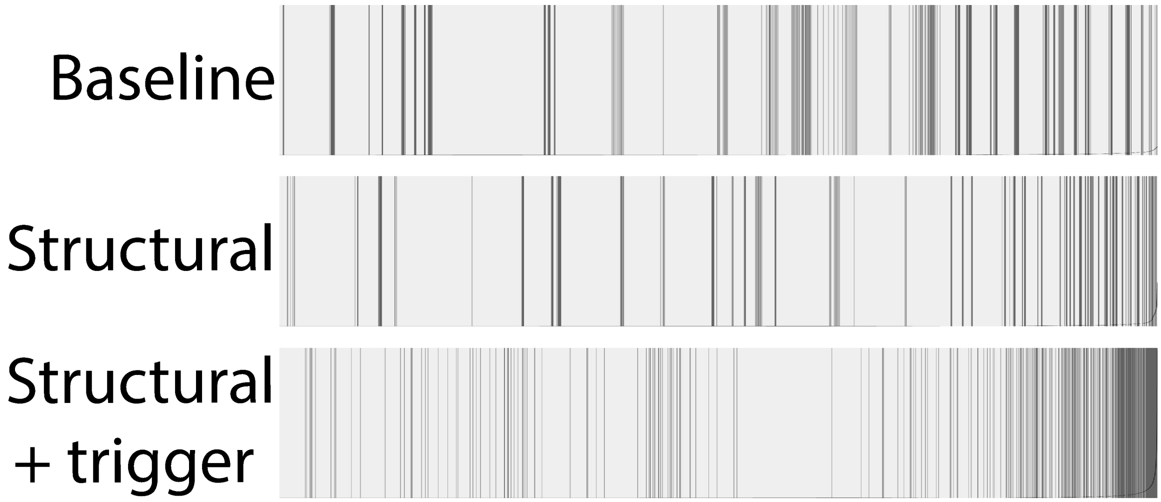

Figure 6: Separation Plots. The 'structural + trigger' model uses both government bond yields and conflict-related news to estimate the magnitude of the trigger.

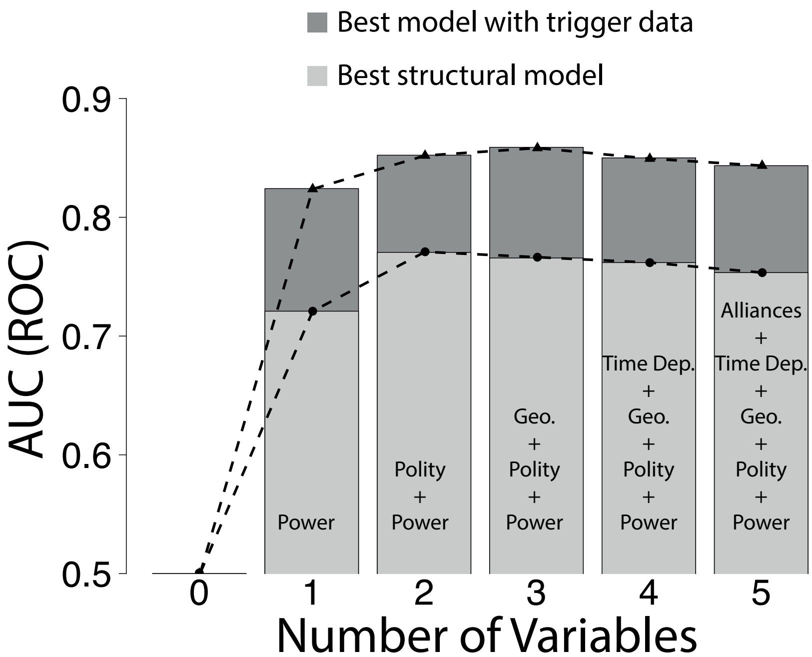

Figure 7: Area under the ROC curve of models with n variables. For each number of variables, the out-of-sample area under the curve (AUC) of each possible model including any combination of n variables was computed. The plot displays the AUC of the best model with n variables excluding measures of triggers (light gray) and including them (dark gray).

Tables

Name	Temp. range	N	Type	Source
Government bond yields	1816–2001	526,371	Monadic	Global Financial Data
Conflict-related news	1902–2001	591,018	Monadic	Chadefaux 2014
10 mil. dyadic events	1990–2000	112,927	Dyadic	King & Lowe 2003

Table 1: Data sources used to estimate the occurrence and size of a trigger.

	'Triggers'	'Structural'	'Structural + Triggers'
(Intercept)	-2.865 $(0.583)^{***}$	-7.920 $(1.216)^{***}$	-3.338 $(1.452)^{*}$
News_{ij}	0.149 $(0.029)^{***}$		0.179 $(0.032)^{***}$
ΔNews_{ij}	0.908 $(0.187)^{***}$		0.574 $(0.170)^{***}$
Bond Yields_{ij}	-2.506 $(0.371)^{***}$		-2.896 $(0.585)^{***}$
$\Delta\text{Bond Yields}_{ij}$	1.641 $(0.377)^{***}$		1.713 $(0.500)^{***}$
Peace Years		0.003 (0.022)	0.025 (0.020)
Peace Years^2		0.000 (0.000)	0.000 (0.000)
Peace Years^3		0.000 (0.000)	0.000 (0.000)
Dyad Age		-0.002 (0.006)	-0.004 (0.006)
Peace Years$\times$Dyad Age		0.000 $(0.000)^{*}$	0.000 $(0.000)^{*}$
Cinc_i		6.926 $(1.465)^{***}$	1.909 (1.829)
Cinc_j		9.349 $(3.616)^{**}$	5.944 (3.440)
Asymmetry		8.530 $(2.367)^{***}$	7.173 $(2.230)^{**}$
Asymmetry^2		-8.891 $(1.746)^{***}$	-7.811 $(1.673)^{***}$
Major Dyad		0.904 (0.475)	0.463 (0.480)
Contiguous		1.150 $(0.451)^{*}$	1.287 $(0.464)^{**}$
Distance		-0.179 (0.099)	-0.179 (0.105)
Ally		-0.087 (0.422)	-0.247 (0.407)
Similarity		-0.995 $(0.389)^{*}$	-0.842 $(0.360)^{*}$
Polity_i		-0.098 $(0.023)^{***}$	-0.090 $(0.021)^{***}$
Polity_j		-0.075 $(0.023)^{***}$	-0.072 $(0.022)^{***}$
Joint Democracy		-0.006 $(0.003)^{*}$	-0.006 $(0.003)^{*}$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Logistic regression of the occurrence of war within one year, with correction for rare events and robust standard errors clustered on dyads in parenthesis. Model 'Triggers' corresponds to eqn 2; model 'Structural' to eqn. 1. $N = 391,745$. Measures of triggers used: Conflict-related news and government bond yields. Coefficients and standard errors are rounded to three decimal places.

	Bonds and News	News	Bonds	10 mil. events
(Intercept)	-6.586	-6.464	-7.03	-7.727
	$(0.138)^{***}$	$(0.117)^{***}$	$(0.178)^{***}$	$(0.382)^{***}$
Structural Risk	0.376	0.482	0.344	-0.005
	$(0.028)^{***}$	$(0.039)^{***}$	$(0.071)^{***}$	(0.005)
Trigger Level	0.348	0.204	2.213	0.298
	$(0.058)^{***}$	$(0.052)^{***}$	$(0.416)^{***}$	$(0.047)^{***}$
Struct. Risk × Trigger	0.060	-0.042	0.719	0.063
	(0.033)	(0.045)	$(0.065)^{**}$	$(0.002)^{***}$
N	$391,745$	$591,018$	$526,371$	$112,927$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Logistic regression of the occurrence of war within one year (model [3]), with correction for rare events and robust standard errors clustered by dyad in parenthesis. Each column uses a different operationalisation of the 'Trigger' variable. 'Structural risk' and 'Trigger Level' are the in-sample fitted values from models 1 and 2, used as composite measures of the variables of interest.

A Supporting Information (Online SI)

A.1 Supplementary Tables

Country	First record
Argentina	1859
Belgium	1834
Brazil	1861
Bulgaria	1993
Canada	1920
Chile	1839
Colombia	1899
Denmark	1880
Egypt	1862
Finland	1987
France	1880
Germany	1880
Greece	1863
Hungary	1997
Iceland	1993
India	1947
Indonesia	1997
Ireland	1928
Italy	1862
Kenya	1987
Malaysia	1961
Mexico	1872
Morocco	1996

Country	First record
Netherlands	1816
New Zealand	1925
Norway	1963
Panama	1997
Peru	1997
Philippines	1997
Portugal	1851
Russia	1820
Singapore	1998
South Africa	1920
Spain	1850
Sri Lanka	1951
Switzerland	1899
Thailand	1979
Tunisia	1991
Turkey	1997
United Kingdom	1816
United States	1816
Uruguay	1882
Venezuela	1914
Zambia	1995
Zimbabwe	1965

Table A.1: Countries for which government bond yield data is available, and earliest year of record.

Table A.2: Summary Statistics of main variables used

Statistic	N	Mean	St. Dev.	Min	Median	Max
Onset_{ij}	783,685	0.003	0.050	0	0	1
News_{ij}	598,228	7.192	3.483	0.000	6.782	30.095
ΔNews_{ij}	591,018	0.031	0.371	−4.432	0.024	5.940
Bond Yields_{ij}	530,733	1.939	0.420	0.0003	1.859	5.958
$\Delta\text{Bond Yields}_{ij}$	526,371	−0.001	0.135	−3.716	−0.001	2.654
Peace Years	783,685	38.734	34.988	0.083	28.583	185.333
Dyad Age	783,685	40.470	36.144	1	30	184
Cinc_i	783,685	0.051	0.077	0.00001	0.015	0.384
Asymmetry	783,685	0.782	0.265	0.000	0.910	1.000
Major Dyad	783,685	0.766	0.423	0	1	1
Contiguous	783,685	0.317	0.465	0	0	1
Distance	783,685	3.177	2.955	0.000	2.931	11.989
Ally	783,685	0.169	0.375	0	0	1
Symmetry	783,685	0.471	0.373	−0.697	0.487	1.000
Polity_i	769,995	2.676	7.378	−10	5	10
Joint Democracy	783,685	7.624	58.902	−100	9	100

	'Triggers'	'Structural'	'Structural + Triggers'
(Intercept)	$-6.376\ (0.235)^{***}$	$-5.308\ (0.936)^{***}$	$-5.526\ (0.928)^{***}$
News_{ij}	$0.018\ (0.027)$		$0.038\ (0.029)$
ΔNews_{ij}	$0.932\ (0.149)^{***}$		$0.685\ (0.155)^{***}$
Peace Years		$-0.039\ (0.019)^{*}$	$-0.040\ (0.019)^{*}$
Peace Years^2		$0.000\ (0.000)$	$0.000\ (0.000)$
Peace Years^3		$0.000\ (0.000)$	$0.000\ (0.000)$
Dyad Age		$-0.003\ (0.006)$	$-0.004\ (0.006)$
Peace Years$\times$Dyad Age		$0.000\ (0.000)^{**}$	$0.000\ (0.000)^{**}$
Cinc_i		$6.508\ (1.510)^{***}$	$6.775\ (1.525)^{***}$
Cinc_j		$7.398\ (3.542)^{*}$	$7.719\ (3.489)^{*}$
Asymmetry		$3.902\ (2.091)$	$3.750\ (2.054)$
Asymmetry^2		$-5.111\ (1.702)^{**}$	$-4.962\ (1.677)^{**}$
Major Dyad		$0.525\ (0.415)$	$0.430\ (0.423)$
Contiguous		$0.462\ (0.418)$	$0.395\ (0.434)$
Distance		$-0.286\ (0.097)^{**}$	$-0.295\ (0.099)^{**}$
Ally		$-0.271\ (0.360)$	$-0.284\ (0.356)$
Similarity		$-1.217\ (0.359)^{***}$	$-1.172\ (0.363)^{**}$
Polity_i		$-0.085\ (0.020)^{***}$	$-0.083\ (0.020)^{***}$
Polity_j		$-0.045\ (0.020)^{*}$	$-0.046\ (0.019)^{*}$
Joint Democracy		$-0.008\ (0.002)^{***}$	$-0.008\ (0.002)^{***}$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.3: News: Logistic regression of the occurrence of war within one year, with correction for rare events and robust standard errors clustered by dyad in parenthesis. Model 'Trigger' corresponds to eqn 2; model 'Structural' to eqn. 1. $N = 591{,}018$. Measure of triggers used: Conflict-related news.

	'Triggers'	'Structural'	'Structural + Triggers'
(Intercept)	$-2.730\ (0.592)^{***}$	$-5.698\ (0.714)^{***}$	$-1.368\ (1.292)$
Bond Yields$_{ij}$	$-1.823\ (0.341)^{***}$		$-2.123\ (0.559)^{***}$
ΔBond Yields$_{ij}$	$1.202\ (0.404)^{**}$		$1.568\ (0.537)^{**}$
Peace Years		$-0.047\ (0.018)^{**}$	$-0.039\ (0.017)^{*}$
Peace Years^2		$0.000\ (0.000)$	$0.000\ (0.000)$
Peace Years^3		$0.000\ (0.000)$	$0.000\ (0.000)$
Dyad Age		$0.006\ (0.004)$	$0.003\ (0.005)$
Peace Years$\times$Dyad Age		$0.000\ (0.000)^{*}$	$0.000\ (0.000)^{**}$
Cinc$_i$		$3.082\ (1.085)^{**}$	$0.118\ (1.254)$
Cinc$_j$		$2.521\ (2.323)$	$-0.235\ (2.390)$
Asymmetry		$2.994\ (1.457)^{*}$	$3.213\ (1.454)^{*}$
Asymmetry^2		$-4.385\ (1.199)^{***}$	$-4.798\ (1.207)^{***}$
Major Dyad		$0.335\ (0.408)$	$0.217\ (0.429)$
Contiguous		$0.902\ (0.336)^{**}$	$0.943\ (0.344)^{**}$
Distance		$-0.017\ (0.068)$	$0.003\ (0.066)$
Ally		$-0.081\ (0.392)$	$-0.025\ (0.392)$
Similarity		$-0.736\ (0.266)^{**}$	$-0.857\ (0.262)^{**}$
Polity$_i$		$-0.076\ (0.020)^{***}$	$-0.073\ (0.021)^{***}$
Polity$_j$		$-0.044\ (0.018)^{*}$	$-0.048\ (0.018)^{**}$
Joint Democracy		$-0.008\ (0.002)^{***}$	$-0.009\ (0.002)^{***}$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.4: Logistic regression of the occurrence of war within one year, with correction for rare events and robust standard errors clustered by dyad in parenthesis. Model 'Trigger' corresponds to eqn 2; model 'Structural' to eqn. 1. $N = 526,371$. Measure of triggers used: Government bond yields.

	'Triggers'	'Structural'	'Structural + Triggers'
(Intercept)	$-8.005\ (0.369)^{***}$	$-10.449\ (2.608)^{***}$	$-10.855\ (2.769)^{***}$
Goldstein Scale(-)	$1.486\ (0.512)^{**}$		$1.277\ (0.543)^{*}$
ΔGoldstein Scale(-)	$-0.464\ (0.385)$		$-0.433\ (0.311)$
Goldstein Scale(+)	$-0.447\ (0.664)$		$-1.045\ (0.735)$
ΔGoldstein Scale(+)	$0.183\ (0.396)$		$0.490\ (0.399)$
Peace Years		$-0.127\ (0.171)$	$-0.077\ (0.212)$
Peace Years^2		$0.004\ (0.005)$	$0.004\ (0.006)$
Peace Years^3		$0.000\ (0.000)$	$0.000\ (0.000)$
Dyad Age		$0.006\ (0.039)$	$-0.021\ (0.036)$
Peace Years$\times$Dyad Age		$0.000\ (0.001)$	$0.000\ (0.001)$
Cinc$_i$		$45.557\ (41.811)$	$44.410\ (39.217)$
Cinc$_j$		$-36.036\ (33.518)$	$-34.912\ (39.248)$
Asymmetry		$15.732\ (4.787)^{**}$	$15.068\ (4.669)^{**}$
Asymmetry^2		$-17.937\ (5.178)^{***}$	$-17.663\ (5.074)^{***}$
Major Dyad		$-2.926\ (6.819)$	$-2.211\ (6.309)$
Distance		$-0.060\ (0.249)$	$-0.033\ (0.319)$
Ally		$1.322\ (0.809)$	$1.281\ (0.714)$
Similarity		$-0.627\ (1.670)$	$-0.345\ (1.814)$
Polity$_i$		$0.086\ (0.092)$	$0.090\ (0.101)$
Polity$_j$		$-0.242\ (0.177)$	$-0.242\ (0.168)$
Joint Democracy		$0.014\ (0.013)$	$0.016\ (0.011)$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.5: Logistic regression of the occurrence of war within one year, with correction for rare events and robust standard errors in parenthesis. Model 'Trigger' corresponds to eqn 2; model 'Structural' to eqn. 1. $N = 112,927$. Measure of triggers used: King & Lowe (2003)'s 10 million dyadic events. Note that the variable Contiguous was removed from this particular regression, as 35,365 failures are completely determined by a combination of Contiguous and Major Dyad (i.e., war never happened between 1990 and 2000 in dyads that are neither contiguous nor include at least one major power).

	Model 1	Model 2	Model 3
(Intercept)	-1.314 (0.776)	-6.140 $(0.897)^{***}$	-0.759 (1.442)
Bond Yields$_{ij}$	-1.897 $(0.416)^{***}$		-2.502 $(0.625)^{***}$
ΔBond Yields$_{ij}$	1.604 $(0.355)^{***}$		2.207 $(0.482)^{***}$
Inflation$_{ij}$	-0.003 $(0.000)^{***}$		0.043 $(0.001)^{***}$
GDP per capita$_{ij}$	-0.159 $(0.035)^{***}$		-0.071 (0.042)
On Gold$_{ij}$	-1.209 $(0.273)^{***}$		-1.207 $(0.264)^{***}$
Central Bank Rate$_{ij}$	-0.001 (0.012)		0.003 (0.016)
Peace Years		-0.052 $(0.021)^{*}$	-0.030 (0.022)
Peace Years^2		0.001 (0.000)	0.000 (0.000)
Peace Years^3		0.000 $(0.000)^{*}$	0.000 (0.000)
Dyad Age		0.004 (0.004)	-0.001 (0.005)
Peace Years$\times$Dyad Age		0.000 $(0.000)^{**}$	0.000 $(0.000)^{**}$
Cinc$_i$		2.710 $(1.176)^{*}$	-0.009 (1.393)
Cinc$_j$		3.399 (2.451)	0.979 (2.651)
Asymmetry		3.098 (1.826)	2.974 (1.788)
Asymmetry^2		-4.390 $(1.438)^{**}$	-4.578 $(1.452)^{**}$
Major Dyad		1.058 $(0.495)^{*}$	1.159 $(0.517)^{*}$
Contiguous		0.898 $(0.373)^{*}$	0.851 $(0.387)^{*}$
Distance		-0.041 (0.075)	-0.024 (0.075)
Ally		0.088 (0.390)	0.111 (0.388)
Similarity		-0.947 $(0.324)^{**}$	-0.975 $(0.284)^{***}$
Polity$_i$		-0.086 $(0.023)^{***}$	-0.064 $(0.026)^{*}$
Polity$_j$		-0.047 $(0.020)^{*}$	-0.045 $(0.020)^{*}$
Joint Democracy		-0.008 $(0.003)^{**}$	-0.008 $(0.003)^{**}$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.6: Logistic regression of the occurrence of war within one year, with correction for rare events and robust standard errors in parenthesis. Model 'Trigger' corresponds to eqn 2; model 'Structural' to eqn. 1. $N = 454,376$. Measure of triggers used: Government bond yields with control variables.

A.2 Control variables for Government bond yields

To test the robustness of our findings, we included a number of control variables in our model to separate the prediction of markets per se, and the extent to which they might simply follow other indicators. High inflation (*Inflation*), for example, reduces the net yield of any investment, and hence investors will demand higher yields as a compensation. The central bank's lending rate (*Central Bank Rate*) is also likely to affect the bond market.[29] Additional variables include whether the country is on the gold standard (*On Gold*) and its Gross Domestic Product per capita (*GDP per capita*), since government bond yields tend to be lower for wealthier countries (Reinhart & Rogoff 2009). For clarity and to show that the results are not dependent on our particular choice of variables, the results we report in the main text exclude these additional variables (inflation, etc.). However, the results with the addition of these control variables can be found in the SI and are consistent with the other findings (table A.6).[30]

[29] The central bank has various tools to influence markets, such as the marginal lending rate, the main refinancing rate, or the deposit rate. Discount rates are the interest rates charged on loans from the central bank to eligible institutions such as commercial banks. In reality, these instruments are highly correlated and all provide a good estimate of what we intend to measure: the bank's inflationary policy. We relied whenever possible on the daily discount rate defined by central banks (or other monetary policy authority), but also used other indicators such as the target rate when not available.

[30] Data on inflation and exchange rate regimes are derived from Reinhart & Rogoff (2009); data on central bank rates come from Global Financial Data.