Public and Private Information in International Crises: Diplomatic Correspondence and Conflict Anticipation

RESEARCH NOTE

Constantine Boussalis, Thomas Chadefaux, and Andrea Salvi
Trinity College Dublin, Ireland

AND

Silvia Decadri
Università degli Studi di Milano, Italy

Scholars of international conflicts have long emphasized the role of private information in the onset of interstate wars. Yet, the literature lacks direct and systematic evidence of its effect. This is largely due to challenges with accessing decision-makers' private and often confidential information and opinions. We compile a large corpus of declassified French diplomatic cables that span the period 1871–1914. Using these texts, we estimate a dynamic topic model to generate diplomatic thematic variables, which are then used to forecast the onset of French Militarized Interstate Disputes (MIDs). The inclusion of information from diplomatic correspondence greatly improves estimates of MID timing, compared to models that rely solely on public information such as structural determinants and revealed risk perceptions derived from financial markets or the press. These results emphasize the importance of private information in decisions to go to war and the limitations of empirical work that relies solely on publicly available data.

Los investigadores de conflictos internacionales llevan mucho tiempo destacando el papel de la información privada en el inicio de las guerras interestatales. Sin embargo, la bibliografía carece de pruebas directas y sistemáticas de su efecto. Esto se debe en gran medida a las dificultades para acceder a la información y las opiniones privadas, y a menudo confidenciales, de los responsables de la toma de decisiones. Recopilamos un amplio corpus de cables diplomáticos franceses desclasificados que abarcan el período 1871–1914. A partir de estos textos, estimamos un modelo de temas dinámicos para generar variables temáticas diplomáticas, que luego se utilizan para pronosticar el inicio de las disputas interestatales militarizadas (Militarized Interstate Disputes, MID) francesas. La inclusión de la información procedente de la correspondencia diplomática mejora en gran medida las estimaciones del período de las MID, en comparación con los modelos que se basan únicamente en la información pública, como los determinantes estructurales y las percepciones de riesgo reveladas que se obtuvieron de los mercados financieros o la prensa. Estos resultados ponen de relieve la importancia de la información privada en las decisiones de declarar una guerra y las limitaciones del trabajo empírico que se basa únicamente en datos públicos.

Les chercheurs en conflits internationaux ont longtemps insisté sur le rôle des informations privées dans le déclenchement des guerres interétatiques. Pourtant, la littérature manque de preuves directes et systématiques de leur effet. Ceci est en grande partie dû aux défis liés à l'accès aux options et informations privées et souvent confidentielles des décideurs. Nous avons compilé un vaste corpus de télégrammes diplomatiques français déclassés s'étendant sur la période 1871–1914. Nous nous sommes appuyés sur ces textes pour estimer un modèle thématique (topic model) dynamique afin de générer des variables thématiques diplomatiques ensuite utilisées pour prévoir le déclenchement des Conflits interétatiques militarisés français. L'inclusion des informations issues des correspondances diplomatiques améliore grandement les estimations des moments auxquels interviennent les conflits interétatiques militarisés par comparaison aux modèles qui ne reposent que sur des informations publiques, telles que les déterminants structurels et les perceptions de risquées révélées, qui sont issues des marchés financiers ou de la presse. Ces résultats mettent l'accent sur l'importance des informations privées dans les décisions d'entrer en guerre et sur les limitations des travaux empiriques qui ne reposent que sur des données publiquement accessibles.

Constantine Boussalis is an Assistant Professor of political science at Trinity College Dublin. His research lies at the intersection of computational social science and political communication.

Thomas Chadefaux is a Professor of political science at Trinity College Dublin. His research focuses on conflict forecasting, machine learning, and bargaining models of war.

Silvia Decadri is a Postdoctoral Researcher at the University of Milan. Her main research interests include legislative behaviour, political communication, natural language processing, and text analysis.

Andrea Salvi received his Ph.D. degree in political science from Trinity College Dublin (2022). He works on Civil Conflict, Reactive Violence, and Counterinsurgency. He recently joined the Joint Research Centre of the European Commission as a Scientific Project Officer.

Introduction

The onset of wars despite the existence of more efficient peaceful settlements is a long-standing puzzle in the international relations literature. If there are peaceful agreements that both parties would prefer to war, then why do states fail to locate them and instead resort to costly and risky conflicts?

A central explanation for this puzzle is the existence of private information. Leaders may know something that others do not and have an incentive to keep this information

private. Such asymmetric information may lead to incompatible demands and the ultimate breakdown of negotiations into conflict.

Unfortunately, the assumption that policymakers indeed hold privileged information is based on little empirical evidence. Beyond historical anecdotes and our intuition that leaders do actually have more valuable information than the public, we have, in fact, little to no systematic evidence that this private information exists. And if it does exist, can it also explain or predict the onset of conflict? In other words, would we anticipate wars better if we had access to leaders' private information? Or is private information largely irrelevant, and decisions are instead mostly dictated by other—publicly observable—factors?

This scarcity of empirical evidence is problematic given the central role that private information plays in theories of conflict, from ones about the decision to go to war (Fearon 1995; Gartzke 1999), to the duration of conflict (Slantchev 2003), or the ability to coerce others (Schultz 2001). Moreover, in the absence of an obvious way to measure private information, the empirical literature on the onset of war has had to rely on *public* information to explain conflict (Bennett and Stam 2004) or to forecast it (e.g., news, financial markets) (Chadefaux 2014, 2017a,b; Mueller, and Rauh 2017). While there has been some recent systematic empirical work on private information in crisis bargaining (Katagiri and Min 2019), we still know very little about the private information that is available to leaders and how it might affect their decisions.

In this study, we address this gap by examining the perception that decision-makers themselves have of the underlying international tensions and, hence, of the risk of conflict. We offer the first direct empirical evidence of the predictive power of private information when forecasting war. Through a comparison of the relative predictive power of private information with respect to public information, we show that privileged information held by diplomats is indeed a relevant factor in determining international politics outcomes.

Using extensive data derived from French diplomatic cables from 1871 to 1914, we are able to observe decision-makers' own estimates and concerns through their private thoughts and reports. Data derived from this large set of diplomatic cables have two main advantages. First, they cover a long and important period in French history: from the defeat by Prussia in 1871 to the onset of World War I (WWI)—a period that redefined European Power politics. Second, these cables relay the genuine opinion of diplomats without any filter or bias. This gives us far more direct access to their thoughts than what might be relayed publicly in newspapers.

In this paper, we can therefore directly address the question of whether insiders indeed have access to privileged private information that enables them to anticipate conflict. In particular, by examining the diplomatic and military discourse prior to the onset of Militarized Interstate Disputes (MIDs), we can test whether this discourse differs in any significant way from their discourse during peaceful periods. We find that private information indeed improves our ability to forecast the onset of conflict. By comparing these estimates to the best public assessments of the risk of war based on financial markets and what is reported by the press, we can infer the added value provided by the private information that leaders have at their disposal. This reveals the importance of further studying private information in order to better understand why countries go to war and what leads them to take such costly decisions.

We first describe our diplomatic cables data and explain the unsupervised machine learning model used to explore and automatically classify the thematic content of the cables. We then use these estimated themes to provide a general overview of the content and evolution of the diplomatic discourse, from which we observe that some of these topics tend to precede the onset of conflicts. Finally, we use the estimated topics as predictors when forecasting MIDs. We find that the predictive power of our models improves relatively to a baseline model based only on public information.

Data

Our data are based on a large corpus of diplomatic documents—e.g., telegrams, letters, reports—sent between French diplomats and military actors from 1871 to 1914. The documents were extracted from the "Commission de publication des documents relatifs aux origines de la Guerre de 1914-1918" (henceforth *the Commission*)—a group of historians and diplomats appointed in 1928 to compile diplomatic documents related to international relations in the decades preceding the onset of WWI. The collection, named "Documents diplomatiques français relatifs aux origines de la guerre de 1914" (henceforth *DDF*) starts at the end of the Franco–Prussian War and ends with the onset of the First World War in 1914.[1]

The Commission published a total of 41 volumes of telegrams and various communications—about 21,000 in all, each with an average of 206.4 words (SD = 228.9). In these cables, French politicians, diplomats, and military actors convey facts, express opinions, and report encounters to each other or to the Ministry in Paris. The telegrams are often classified as confidential and typically reflect the candid opinion of the sender—e.g., the impression he had from a meeting or his thoughts about strategy.

Typical examples drawn from these cables are the two messages below. They illustrate that the type of matters routinely discussed ranged from discussions of power politics to reports of mundane events:

> "Count Munster came to find me to deny the rumor, circulated by the Parisian correspondent of the Times, about a conversation that would have taken place in Berlin between the Italian King and Prince Bismarck, according to which the Chancellor of the Empire would have pressed Victor-Emmanuel to arrange the return of Savoy and Nice to Italy. The German Ambassador explicitly told me that his communication was official. I rush to transmit it to your Excellency, with the character that Count Munster intended to give it and without adding anything to this guarantee."
>
> *Ambassador in London to Minister of Foreign Affairs, May 13, 1874.*

> "Should you deem it appropriate, and should other Powers do the same, you are authorized to bring or have someone bring to prince Bismarck our condolences for the heinous assassination attempt against him. It is good in any case that our feelings of indignation are known."
>
> *Minister of Foreign Affairs to Ambassador in Berlin, July 14, 1874*

The cables are mostly based not only on the archives of the French Ministry of Foreign Affairs but also on those

---

[1] Documents diplomatiques français relatifs aux origines de la guerre de 1914, Paris, Imprimerie nationale, 1929–1959.
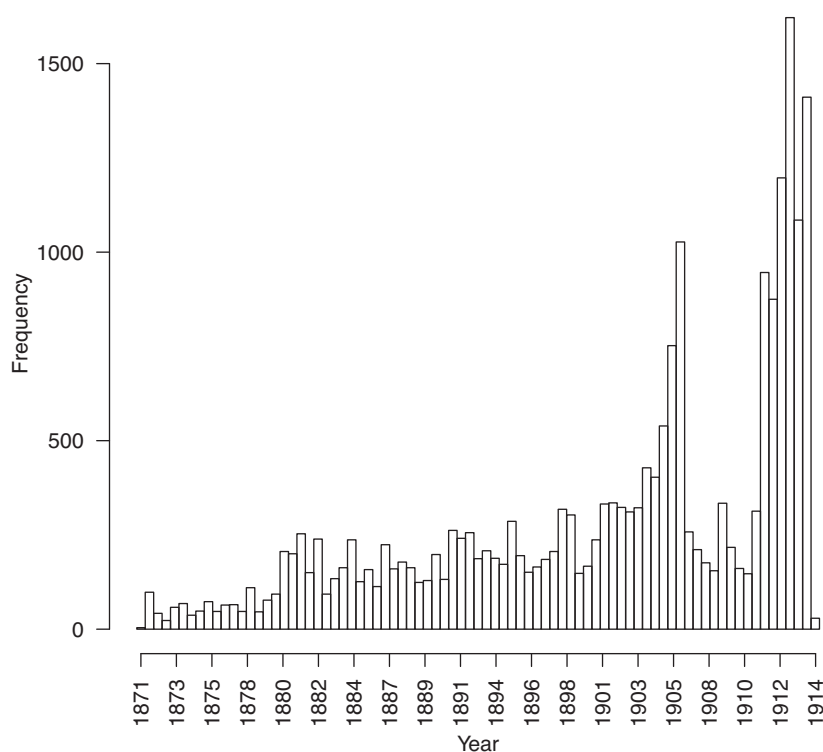
Figure 1. Frequency of French diplomatic cables in the corpus.

of Ministries associated with foreign policy—e.g., the ministries of War, Maritime Affairs, and Colonies. They span the entire period 1871–1914, with at least dozens every month (figure 1). This allows us to draw inferences for every month of the 44 years covered. We do note a spike in activity prior to the onset of WWI, which is not surprising given the goals of the Commission.

Admittedly, there are concerns relative to the text selection process. First, parts of the correspondence may not have been available to the historians in charge of compiling the documents. This turns out not to be the case, however, as the Commission had access to all documents, including those classified as confidential. Second, the publication itself could not possibly include all the cables in that period—around 2 million documents. The Commission therefore selected documents based on scientific criteria. The selection was made by a team of academic researchers, including prominent historians, with complete access to all archives produced or received by the French Ministry of Foreign Affairs. It was made not with a particular goal or view in mind, but rather based on the relevance of the given cable to interstate relations over the period. Direct relevance to WWI was not a criteria, but the cables had to relate to meaningful events or facts that affected the states during this period.[2] Every text selected was included in its entirety; rare editions of the text (e.g., excerpts in the case of press clippings) were clearly noted by the Commission.

We retrieved the forty-one volumes in machine-readable format from the online collection of the National Library of France.[3] These unstructured text files were then passed through a series of preprocessing steps prior to data analysis. Online appendix A provides a more detailed description of the text preprocessing pipeline.

Unsupervised Learning of French Diplomatic Themes

The first empirical task of this study is to systematically explore and categorize the thematic content of the gathered French diplomatic cables. What topics were diplomats discussing and during which periods did certain themes rise or fall on the agenda? To answer this question, we content analyse the gathered cables. Clearly, the costs in terms of time and resources needed to manually annotate the themes in over 21,000 documents in our collection reach prohibitive levels. Further, we have no *a priori* knowledge of the span of latent themes found within the corpus.

Topic Modeling Using Dynamic Non-Negative Matrix Factorization

We address this classification problem via a computer-assisted *unsupervised* learning approach, which attempts to delineate substantive clusters of terms that appear across the corpus (see Grimmer and Stewart 2013). There are a number of useful computational methods at our disposal to accomplish this task, with Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) and the Structural Topic Model (Roberts et al. 2014) becoming increasingly popular methods in political science (Wilkerson and Casas 2017).

However, there are important considerations when analyzing a temporally long corpus such as ours. First, we wish to account for the changing nature of diplomatic discourse within a given theme. For instance, due in part to technological change, we expect the vocabulary for warfare in the 1870s to differ substantively from that used around the time

---

[2]In the words of the Commission: "We included all pieces that showed the overall orientation of French Diplomacy, and those which conveyed information about the general policies of other European Powers. We omitted texts that related to incidents which, regardless of their immediate importance, had no durable influence on the political life." (DDF Vol. 1, tome 1, p. XVII; own translation).

[3]The collection of diplomatic cables can be found at http://gallica.bnf.fr.

of the First World War. Also, it is possible that certain themes existed for a while but then disappeared naturally. Existing standard topic modeling approaches might not detect these ephemeral themes because they assume that the order of documents does not matter.

In this study, we employ the dynamic topic modeling method provided by Greene and Cross (2017) to help address these concerns. The authors' method explicitly models the temporal evolution of thematic content found within a given corpus using two levels of Non-Negative Matrix Factorization (NMF), which is a widely studied linear dimensionality reduction technique (Paatero and Tapper 1994; Lee and Seung 1999; Gillis 2014). More details on this model are provided in online appendix C.

We first split the corpus of 22,164 documents into time windows of 5-year segments from 1878 to 1914. Using our preprocessed text data as input (see online appendix A), we use the Dynamic NMF topic model to generate our main quantity of interest: per document normalized dynamic topic weights.[4] The number of topics to be estimated must be defined prior to running the model. Since we had no strict prior information about the likely number of topics that ran through the corpus, we relied on the following strategy: (1) we ran the model with varying topic numbers; (2) for each topic model run, we compared the calculated median coherence estimate of the topics; and (3) we performed a manual inspection of a set of topic model solutions with relatively high coherence values.[5] Based on this approach, we settled on $k = 26$ dynamic topics.

Topic Model Interpretation and Validation

Using the resulting model parameters of the topic model with twenty-six dynamic topics, we interpret the substantive meaning of these topics. First, we assign a descriptive label to each topic. To this aim, we review the twenty terms with the highest membership weighting for each dynamic topic to determine the unifying theme which binds these terms. In doing so, we also retrieve and review the top ten highest weighted documents for each topic to assist in the interpretation. A list of our topic labels along with the top ten highest weighted terms for each topic is in online appendix D.[6] As one might expect from a corpus of diplomatic correspondence, the majority of the themes are related to specific nations or regions. However, more abstract topics also emerge from the results, such as "Diplomacy/International Law" and "War/Alliances".

Following best practices in automated content analysis, we propose a few validation exercises (Grimmer and Stewart 2013). The aim of the validation exercise is to determine whether the topics generated by our model are indeed measuring the concepts we have ascribed to them. While the interpretation of the topics through descriptive labels, as discussed above, gives us a good sense of the semantic validity of such topics, it does not tell us much about their predictive and concurrent validity. We thus perform two additional validation procedures to increase our confidence

in the model's results: (1) we demonstrate the semantic and predictive validity of the model through illustrative charts (Quinn et al. 2010), and (2) we estimate the concurrent validity of the the model by calculating the agreement between human codings and machine predictions of topics (see Boussalis and Coan 2016). Each of these steps is described in detail in online appendix E, while here we summarize the main results.

First, we note that the terms clustered together form coherent themes (semantic validity). Obvious themes such as "Germany" and "German Emperor" are in close proximity, but also more nuanced ones such as "Greece", "Bulgaria" and "Russia" cluster together (see figure 1 in online appendix E). Moreover, we find that documents about a particular topic are more likely to appear during relevant periods, i.e., the prevalence of these themes varies over time in expected ways (predictive validity). Figure 2 shows the temporal variation of the prevalence of a selection of topics over time. In the graphs, we note that spikes in conversation appear during relevant episodes. For example, the salience of the "War/Alliances" peaks during known periods of international conflict like the Franco–Prussian War.

In addition, we find a high level of agreement between human judgement and machine prediction of a given cable's thematic content (concurrent validity). When we ask a human coder to pick the most relevant topic to assign to a random selection of cables from our corpus, the procedure results in an accuracy level of 0.7.

Forecasting MIDs with French Diplomatic Themes

Forecasting has recently become an important part of the conflict scholar's toolbox. Predictions of violent events have been increasingly accurate using various methods, ranging from expert knowledge to quantitative methods and formal modeling. Efforts to forecast conflicts have mostly typically on civil wars, but interstate conflicts have also received significant attention (Beck, King, and Zeng 2000) (for an extensive review, see Ward et al. 2013; Cederman and Weidmann 2017; Chadefaux 2017a). This new scholarship contrasts with much of the existing work in international relations, which has typically relied on explanatory approaches, focusing in particular on causal inference and often ignoring predictions (Puchala 1990; Shmueli 2010). Forecasting is particularly relevant here because of the scarcity of conflict onsets. With such a rare fraction of positive cases, there is a real danger of overfitting in a regression context. Significant *p*-values may be obtained by fitting just the right model, but this may not reflect any meaningful relationship but would rather be fitting noise. Instead, forecasting allows us to test our models out of the learning sample and hence to avoid overfitting (Ward, Greenhill, and Bakke 2010).

In this section, we test whether the topics extracted from the French diplomatic cable corpus are good predictors of the thirty-six MIDs involving France over the period.[7] For example, we might expect certain topics—e.g., ''War/Alliances''—to figure prominently in cables prior to disputes. For each month, we calculated the average weight of a given topic, resulting in 26 time series of 528 months each (44 years), for a total of 13,728 data points—one for each month-topic combination. Each of these time series is used as an independent variable in our analysis. Our dependent variable is the number of days until the next MID (*TimeToMID*).[8]

---

[4] Greene and Cross (2017) use Non-Negative Double Singular Decomposition (NNDSVD) to generate initial factors for their NMF algorithm (Boutsidis and Gallopoulos 2008). The code used to run the Dynamic NMF model was accessed at https://github.com/derekgreene/dynamic-nmf.

[5] We operationalize "topic coherence" as C_v, (Röder, Both, and Hinneburg 2015) and employ `gensim` to calculate coherence for topic solutions between 5 and 120, at steps of 5.

[6] Note that seven of the twenty-six topics were judged to be "junk" (AlSumait et al. 2009)—i.e., they represent themes of unclear theoretical substance. For transparency, we chose to keep them in the analysis nonetheless.

---

[7] We include both the French mainland and its colonies. The data are extracted from the Correlates of War Project (Palmer et al. 2015).
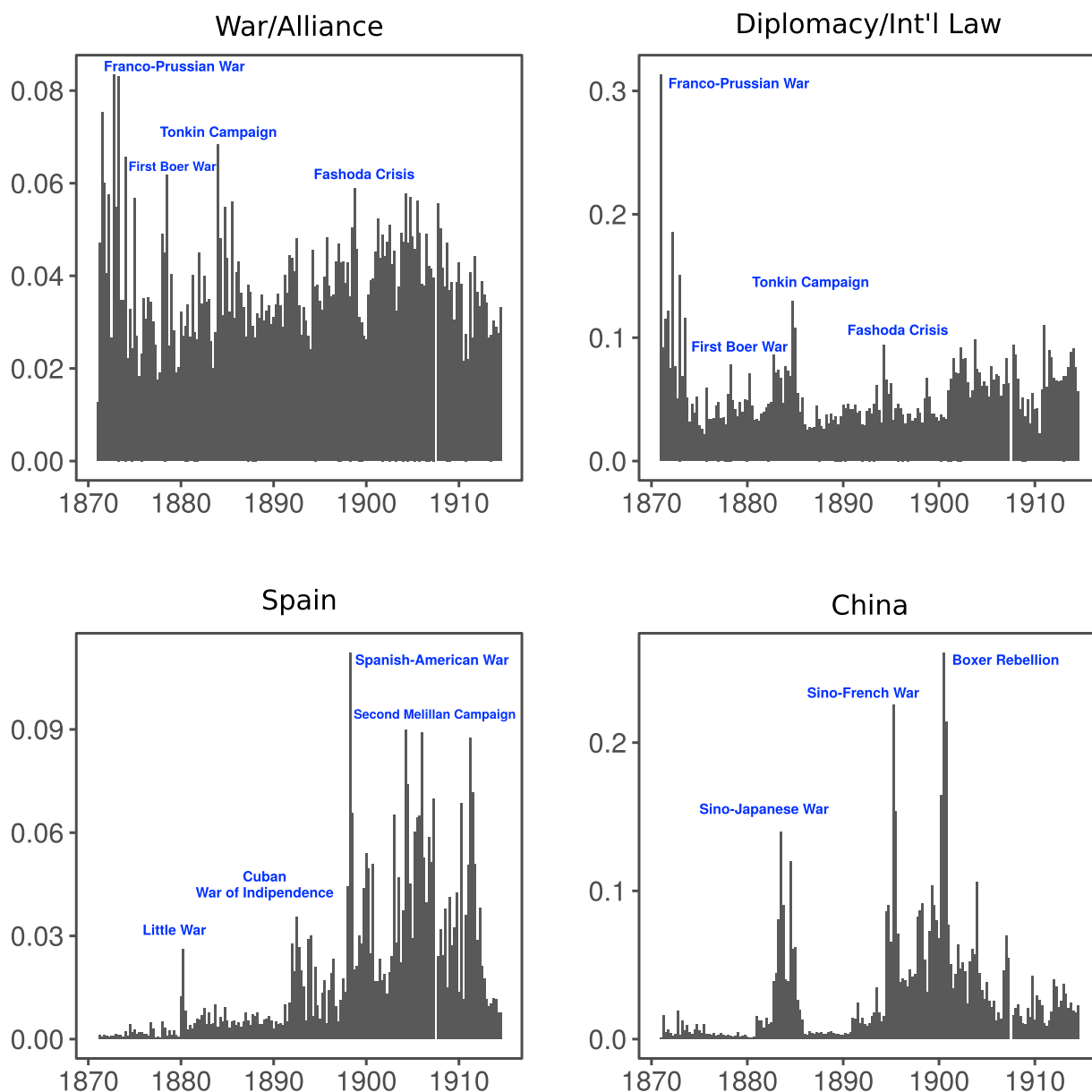
Figure 2. Predictive validity of topics. This figure plots the average normalized quarterly topic weights for "War/Alliance", "Diplomacy/Int'l Law", "Spain", "China" over the sample period. Note that these plots are on different scales.

We first combined our data on militarized disputes with as much publicly available information as possible. First,

[8]This approach is more suitable than a simpler binary approach (i.e., conflict/no conflict in a given month) for two main reasons. First, a binary approach implies that an observation preceding the onset by one month would be treated in the same way as another preceding it by a year or more. Both would be zeroes and a model would not differentiate between them. This is problematic in terms of learning, since covariates may be different close to an onset than far from it, yet will be treated the same way. It is also problematic when it comes to assessing the quality of the predictions, as a prediction off by one month will be treated as an error in the same way as another off by years. A second reason to rely on the continuous variable "time to MID" rather than the dichotomous MID/no MID is that our sample only contains 38 MIDs over the period of interest. This means that there is little opportunity for the model to learn if we only rely on binary measures of conflict onset. Worse, splitting the sample into a learning and testing set (where the learning set needs to temporally precede the testing set) means that we are only learning from a portion of that data (e.g., in expectation 19 if we split the data in two sets of equal duration). This would lead to large standard errors and poor predictions. On the other hand, drawing inferences about the time to the next MID allows us to use the variation in each observation and to avoid the problem of forecasting highly imbalanced classes.

we collected information on some of the well-established structural determinants of conflict, including *military expenditure*[9], *military personnel* (1,000s), *population* (1,000s), *iron and steel production* (1,000s of tons), *imports, and exports* in current millions USD.[10]

More importantly, we relied on financial market prices as predictors of conflict. Financial markets, and government bond yields in particular, are highly sensitive to geopolitical risk. Conflict and war are costly, tend to raise inflation and the risk of default, and therefore are typically associated with higher borrowing costs from the government. Because markets have an incentive to anticipate these changes, we can use their price fluctuations as predictors of conflict. This follows the approach of existing work on the subject

[9]In thousands current year GBP from 1816 to 1913 and in thousands current year USD for 1914.

[10]These data were obtained from the Correlates of War Project (COW) National Material Capabilities (v5.0) (Singer, Bremer, and Stuckey 1972) and COW Trade (v4.0) (Barbieri, Keshk, and Pollins 2008, 2009) data-sets.

(Schneider and Troeger 2006; Chadefaux 2017b). The main advantage of financial markets is that they are very good at aggregating large amounts of information and translating into a single variable—the price. These data incorporate all the public information and therefore address any issue of possible missing covariates. Here, following Chadefaux (2017b), we use government bond prices for France over the relevant period (''Bond Yield [lag]''), together with their daily return (''ΔBond Yield [log]'').

Finally, to offer an even more comprehensive view of the information available to the broader public and to distinguish it from the private information accessible to diplomats and policymakers, we also relied on discussions in the press. To capture the French public debate during the time frame analyzed, we have content-analyzed the issues discussed by the widely distributed French broadsheet *Le Figaro*. This involved the harvesting of all back issues of the newspaper that covered our sample period ($n = 15, 664$). In online appendix F, we offer a more detailed discussion of the topic model analysis of these newspaper textual data. Information reported by the press is added to the public assessments of the risk of war based on financial markets and gives us an even more accurate view of the information publicly available at the time.

Together, structural variables, financial markets, and the press allow us to capture the information publicly available at the time. Our goal is then to test whether private information, as extracted from the cables, offers information beyond the publicly available information.

Our analysis proceeds in two steps. We first estimate the predictive value of our topics on a model using the entire sample of 44 years (1871–1914). We then further validate our model on out-of-sample data.

In-Sample Estimation

We estimated several models, each following the same general form but with different elements:

$$\text{Time To MID} = \overbrace{\underbrace{\mathbf{B}\boldsymbol{\beta_1}}_{\text{Base}} + \underbrace{\mathbf{S}\boldsymbol{\beta_2}}_{\text{Structure}} + \underbrace{\mathbf{F}\boldsymbol{\beta_3}}_{\text{Bonds}} + \underbrace{\mathbf{N}\boldsymbol{\beta_4}}_{\text{News}}}^{\text{Public Info.}}$$

$$+ \overbrace{\underbrace{\mathbf{C}\boldsymbol{\beta_5}}_{\text{Dipl. cables}}}^{\text{Private info.}} + \boldsymbol{\varepsilon}. \tag{1}$$

The Base variables only include a measure of time dependence, *"Time since the last MID"*, together with its quadratic and cubic terms. A second set of variables **S** (''Structure'') includes a matrix of structural covariates that are publicly observable, such as population or military expenditures, and $\boldsymbol{\beta_2}$, its associated vector of coefficients.[11] The next two sets of variables add assessments of the risk of war by contemporary observers using publicly observable data. First, we added financial data (**F**) in the form of French government bond yields over the period. Practically, **F** includes two variables: the value of the bond yield at the close of business and the change in that value from one period to the next. As bond yields reflect all available public information, this matrix adds a far more granular estimate of public information than structural variables alone. Second, we added information obtained from an analysis of French news (**N**) using

---

[11] While military expenditure or personnel may not be perfectly transparent, and hence not entirely public, the overall budget and general personnel are typically well known and therefore ''as if'' public.
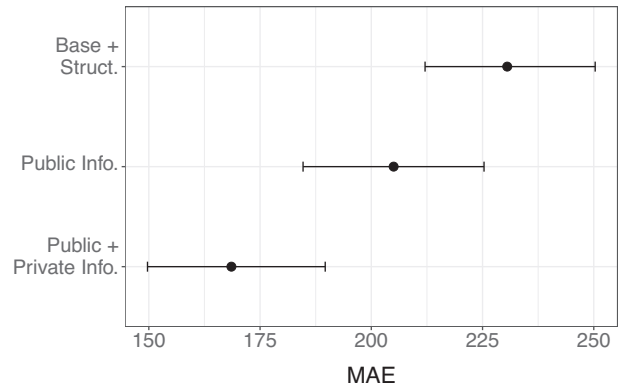

Figure 3. Out-of-sample Mean Absolute Errors and 95 percent confidence intervals (bootstrapped) for models with and without private information. For each observation in our sample, we compared our predicted value for the number of days to the next MID to the actual one. Larger errors reflect a poorer estimate of the timing of the next conflict. The models correspond to the ones in table 1. Additional models are reported in online appendix, figure 2.

broadsheet *Le Figaro*. These take the form of an $N \times 15$ matrix of fifteen topics. Finally, we added an $N \times 26$ matrix **C** of topics derived from diplomatic cables—each cell of which is a month-topic observation.

Table 1 reports the results of various regressions combining some or all of these variables. The coefficients associated with the cable and news topics are not reported here due to space limitations—see online appendix G, table 3 for the full results and additional specifications. We estimated all models with ordinary least squares. The in-sample results for each model (table 1) show that the model including topics derived from cables exhibits significantly larger R^2 values than models without them.[12] The addition of information derived from diplomatic cables improves upon not only the simple model (B+S), but even upon models that incorporate broad measures of public information, including information derived from financial markets and news reports. This suggests that diplomats have access to information that even those with a strong interest in forecasting wars do not.

Out-of-Sample Validation

Beyond in-sample results, we assessed the predictive ability of our models on out-of-sample observations using a moving window.[13] Practically, we start with a learning set using the first 100 months (1871–1879). We estimate coefficients on that set, and use these coefficients to forecast a value for month 101.[14] We then increment the learning set to the first 101 months and forecast month 102, and so on until all months are included. This results in 316 predicted values for the 316 months from 1888 to 1914.

For each observation from month 101 onward, we therefore obtain an estimated time to the next war. We then compare this estimate to the true time of war. Figure 3 displays the distribution of *Mean Absolute Errors* for the predictions of the time to the next MID. Note that, on average, a MID occurs about every 20 months (about 610 days) for France

---

[12] Not surprisingly, F-tests of joint significance of the topic variables also confirm these results ($p < 0.001$).

[13] We used a moving window as opposed to, say, random assignment to tests and learning sets because we did not want to include future observations to predict past ones.

[14] Varying the size of the initial learning set has little impact on the results.

Table 1. Time to next MID as a function of public and private information. Coefficients on the 26 cable topics and on the 14 news topics not reported (see online appendix for full results and additional models). Starred coefficients are lagged

	Base + Struct. (B+S)	Public info. (B+S+Bonds+News)	Public + private info. (B+S+Bonds+News+Cables)
Intercept	−16248.87**	−27455.64**	−24370.46**
	(2042.92)	(4831.15)	(4716.69)
Military exp*	0.20**	0.33**	0.27**
	(0.03)	(0.06)	(0.06)
Military pers*	−54.37	203.48*	363.64**
	(56.98)	(92.41)	(92.47)
Imports*	143.71**	89.96**	65.21**
	(23.19)	(26.30)	(24.92)
Exports*	−126.93**	−124.71**	−88.58**
	(27.53)	(26.71)	(25.39)
Population*	44.83**	75.84**	71.95**
	(5.73)	(11.13)	(10.84)
Iron + steel prod.*	−37.55**	−35.04**	−38.24**
	(4.36)	(4.37)	(4.44)
Time since MID	0.44**	0.27	−0.03
	(0.17)	(0.17)	(0.15)
Time since MID^2	−0.00**	−0.00**	−0.00**
	(0.00)	(0.00)	(0.00)
Time since MID^3	0.00**	0.00**	0.00**
	(0.00)	(0.00)	(0.00)
Bond close*		349.34*	392.34**
		(143.87)	(139.41)
Bond return (log)*		289.38	−379.29
		(1292.57)	(1118.60)
Cables topics included	No	No	Yes
News topics included	No	Yes	Yes
R^2	0.26	0.40	0.60
Adj. R^2	0.24	0.36	0.54
Observations	411	411	411

$^*p < 0.05$; $^{**}p < 0.01$.

in that period. The estimates of the baseline model ("Base + Struct.") are, on average, 8 months off (about 245 days). As a reminder, the baseline model only includes temporal information about the time since the previous MID as well as various structural variables such as military expenditures or population. The best model that relies on public information only ("S + Bonds + News") does substantially better with a Mean Absolute Error of about 6 months and 12 days.

The model that combines all public and private information ("S + Bonds + News + Cables") substantially outperforms all other models, with an MAE of only about 5 months and a week. Pairwise *t*-tests were performed and confirmed that this model is better than the ones relying on public information not only substantially but also statistically (see online appendix H, table 4). The addition of private information, in other words, contributes to reducing the errors associated with models that rely solely on public information.

We also analyzed the contribution of each of the individual cable topics to the prediction. The results (see online appendix H, figure 3) suggest that the most useful cable topics are those related to Spain, England, Bulgaria, and the Franco–German conflict. The most useful press-related topics are the ones related to politics and defense. The contribution of each over time is also displayed in online appendix H, figure 4, while details of the overall model's performance are in online appendix H, figure 5.

These results suggest that information derived from cables in a given month allows us to forecast the timing of the next onset much more precisely than without it. The model

with the best out-of-sample performance is the one that combines public and private information. This confirms the importance of private information held by diplomats and "insiders". Thanks to this information, policymakers might be able to perceive signs of upcoming crises more clearly than actors with access to private information only.

Discussion

In this paper, we offer the first systematic attempt to produce empirical evidence of a longstanding assumption in the international conflict literature: that private information on the risks and costs of conflicts abounds among policymakers, and that it is crucial to understand why we go to war. To that end, we have gathered 40 years of French diplomatic cables to evaluate decision-makers' perceptions of the risks and costs of conflicts. If leaders and diplomats correctly identify risky situations for what they are, we would expect their beliefs to be reflected in the opinions, reports, and notes they send each other, often via confidential telegrams.

In our analysis, we proceeded in two main steps. First, we processed raw unstructured compilations of French diplomatic cables into machine-readable documents that were then automatically reduced to a set of relevant dynamic clusters of terms. We also included a similar series of time-series data for French newspaper discussions. Second, we used these clusters as predictors of the timing and onset of MIDs. Both in-sample and out-of-sample results show that the addition of private diplomatic information strongly contributes to improving our ability to predict the timing of the

onset of a MID. In particular, our best model significantly reduces the average timing error—from 6.4 months for the best model that relies on public information to 5.2 months for the one that includes private information encoded in cables.

Our analysis thus offers some initial empirical evidence that private information indeed matters when we study interstate conflict. This paves the way for future studies that could explore insiders' information to better understand why we go to war and how we can avoid costly and unnecessary conflicts. More recent secret information is obviously difficult to obtain. Documents that are of most interest will typically be classified or released so parsimoniously as to be of limited value. However, private information is at times made available—The Digital National Security Archive, for example, has over 100,000 declassified records of US policy decisions. At other times, the release of documents is involuntary. In 2010, for example, WikiLeaks released vast numbers of classified diplomatic cables of a similar nature to the French ones used here. These cables were sent to the US State Department from 1966 to February 2010 and include the views of foreign leaders as well as American diplomats' reports on their host countries. However, the legality and ethics of using such data remain contested (Michael 2015).

Future work may further improve upon our current results by implementing a number of changes. First, the messages may have value not only in the factual information they convey, but also in their tone. A sentiment analysis of the texts, for instance, could contribute to a more nuanced understanding of the documents. Second, our results rely on a bag-of-words approach, which does not consider word ordering and other sources of linguistic nuance; future work that employs an expanded set of natural language processing approaches, such as the use of word embeddings to better understand context (Rheault and Cochrane 2020; Rodman 2020; Rodriguez and Spirling 2021), could prove fruitful in further uncovering a link between private diplomatic discourse and conflict. Finally, our study is based only on French diplomatic cables, but future research could propose cross-country analyses. Data availability and accessibility is still an issue. To our knowledge, there are no publicly available diplomatic cable collections as complete and far-reaching as the one used here. Still, recent advances in data mining and language processing would greatly assist future replications. Such studies could further confirm the relevance of diplomatic correspondence, in particular, and private information more generally, in forecasting the onset of wars.

Supplementary Information

Supplementary information is available at the *International Studies Quarterly* data archive.

References

AlSumait, Loulwah, Daniel Barbará, James Gentle, and Carlotta Domeniconi. 2009. "Topic Significance Ranking of LDA Generative Models." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, edited by Ulf Brefeld, Elisa Fromont and Andreas Hotho, 67–82. Berlin: Springer.

Barbieri, Katherine, Omar Keshk, and Brian Pollins. 2008. "Correlates of War Project Trade Data Set Codebook." Codebook Version 2.

Barbieri, Katherine, Omar M.G. Keshk, and Brian M. Pollins. 2009. "Trading Data: Evaluating Our Assumptions and Coding Rules." *Conflict Management and Peace Science* 26(5): 471–91.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1): 21–35.

Bennett, D. Scott, and Allan C. Stam III, 2004. *The Behavioral Origins of War.* Ann Arbor, MI: University of Michigan Press.

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

Boussalis, Constantine, and Travis G. Coan. 2016. "Text-Mining the Signals of Climate Change Doubt." *Global Environmental Change* 36: 89–100.

Boutsidis, Christos, and Efstratios Gallopoulos. 2008. "SVD Based Initialization: A Head Start for Nonnegative Matrix Factorization." *Pattern Recognition* 41(4): 1350–62.

Cederman, Lars-Erik, and Nils B. Weidmann. 2017. "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science* 355 (6324): 474–6.

Chadefaux, Thomas. 2014. "Early Warning Signals for War in the News." *Journal of Peace Research* 51(1): 5–18.

Chadefaux, Thomas. 2017a. "Conflict Forecasting and Its Limits." *Data Science* 1(1–2): 7–17.

Chadefaux, Thomas. 2017b. "Market Anticipations of Conflict Onsets." *Journal of Peace Research* 54(2): 313–27.

Fearon, James D. 1995. "Rationalist Explanations for War." *International organization* 49(3): 379–414.

Gartzke, Erik. 1999. "War is in the Error Term." *International Organization* 53(3): 567–87.

Gillis, Nicolas. 2014. "The Why and How of Nonnegative Matrix Factorization." *Regularization, Optimization, Kernels, and Support Vector Machines* 12(257): 257–91.

Greene, Derek, and James P. Cross. 2017. "Exploring the Political Agenda of the European Parliament Using A Dynamic Topic Modeling Approach." *Political Analysis* 25(1): 77–94.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3): 267–97.

Katagiri, Azusa, and Eric Min. 2019. "The Credibility of Public and Private Signals: A Document-Based Approach." *American Political Science Review* 113(1): 156–72.

Lee, Daniel D., and H. Sebastian Seung. 1999. "Learning the Parts of Objects by Non-Negative Matrix Factorization." *Nature* 401(6755): 788–91.

Michael, Gabriel J. 2015. "Who's Afraid of WikiLeaks? Missed Opportunities in Political Science Research." *Review of Policy Research* 32(2): 175–199.

Mueller, Hannes, and Christopher Rauh. 2017. "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text." *American Political Science Review* 112(2): 1–18.

Paatero, Pentti, and Unto Tapper. 1994. "Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estimates of Data Values." *Environmetrics* 5(2): 111–26.

Palmer, Glenn, Vito d'Orazio, Michael Kenwick, and Matthew Lane. 2015. "The MID4 Dataset, 2002–2010: Procedures, Coding Rules and Description." *Conflict Management and Peace Science* 32(2): 222–42.

Puchala, Donald J. 1990. "Woe to the Orphans of the Scientific Revolution." *Journal of International Affairs* 44(1): 59–80.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1): 209–28.

Rheault, Ludovic, and Christopher Cochrane. 2020. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* 28(1): 112–33.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4): 1064–82.

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, edited by Xueqi Cheng, 399–408. New York, NY: ACM.

Rodman, Emma. 2020. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis* 28(1): 87–111.

Rodriguez, Pedro, and Arthur Spirling. 2021. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *Journal of Politics* 84(1): 101–15.

SCHNEIDER, GERALD, AND VERA E. TROEGER. 2006. "War and the World Economy: Stock Market Reactions to International Conflicts." *Journal of Conflict Resolution* 50(5): 623–45.

SCHULTZ, KENNETH A. 2001. *Democracy and Coercive Diplomacy*, vol. 76. Cambridge: Cambridge University Press.

SHMUELI, GALIT. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289–310.

SINGER, J. DAVID, STUART BREMER, AND JOHN STUCKEY. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820–1965." In *Peace, War, and Numbers*, edited by Bruce Russett, 19–48. Beverly Hills: Sage.

SLANTCHEV, BRANISLAV L. 2003. "The Principle of Convergence in Wartime Negotiations." *American Political Science Review* 97(4): 621–32.

WARD, MICHAEL D., BRIAN D. GREENHILL, AND KRISTIN M. BAKKE. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4): 363–75.

WARD, MICHAEL D., NILS W. METTERNICH, CASSY L. DORFF, MAX GALLOP, FLORIAN M. HOLLENBACH, ANNA SCHULTZ, AND SIMON WESCHLE. 2013. "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction." *International Studies Review* 15(4): 473–90.

WILKERSON, JOHN, AND ANDREU CASAS. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20(1): 529–44.