A Bayesian Forecasting Model of International Conflict

George W. Williford and Douglas B. Atkinson

October 31, 2018

Received: 31-Dec-2017

Revised: 06-May-2018

Accepted: 31-Oct-2018

Abstract:

Scholars and practitioners in international relations have a strong interest in forecasting international conflict. However, due to the complexity of forecasting rare events, existing attempts to predict the onset of international conflict in a cross-national setting have generally had low rates of success. In this paper, we apply Bayesian methods to develop a forecasting model designed to predict the onset of international conflict at the yearly level. We find that this model performs substantially better at producing accurate predictions both in and out of sample.

Keywords:

International conflict, forecasting, prediction, Bayesian statistics

Author Bios:

George W. Williford is a Ph.D. candidate in Political Science at the University of Georgia. Corresponding author.

Email: williford@uga.edu.

Address: Candler Hall 301
University of Georgia,
Athens, GA, 30602

Douglas B. Atkinson is a postdoctoral researcher at Cardiff University. He received his Ph.D. in Political Science from the University of Georgia.

Acknowledgments:

We would like to thank Andy Owsiak, Amanda Murdie, Chad Clay, Scott de Marchi, and Josh Jackson for helpful comments on previous versions of this paper.

1 INTRODUCTION

Scholars of international relations have long endeavored to create a means of accurately forecasting interstate conflict. Being able to accurately anticipate interstate conflict will not only provide the international community with the information of when and where conflicts are likely, but which of these conflicts has the highest probability of occurring. With this information the international community can better direct scarce resources in the hopes of mediating the potential conflict.

Although there have been a number of attempts to create forecasting models of international conflict (e.g., Beck, King, and Zeng 2000, 2004; Gleditsch and Ward 2013; de Marchi, Gelpi, and Grynaviski 2004), most previous work has focused on predicting whether states are engaged in conflict, not the onset of conflict. While these studies have undoubtedly improved our forecasting ability, as of yet, there has not been a model that can accurately predict the onset of conflict with a high enough level of precision or with enough warning for policy makers to act on its predictions. This is unsurprising given the inherent difficulties associated with forecasting rare events. Nonetheless, developing models that can more accurately forecast international conflict is an endeavor worth pursuing.

In this article we apply Bayesian methods and machine-learning techniques to build a better prediction model. We use Bayesian logistic regression to provide regularized estimates of our coefficients and combine it with a technique known as undersampling to enhance the predictive power of our model. By using weakly informative priors to constrain the size of the estimated effects, we can help reduce the extent to which variables introduce extraneous noise to the model. The use of undersampling allows us to reduce the computational burden associated with our Bayesian approach. In addition, the use of undersampling produces more realistic predicted probabilities in the face of rare events. This helps resolve a common problem associated with forecasting rare events, namely, that standard models fail to assign high predicted probabilities to any of the observations. This allows us to attempt to accurately forecast the onset of militarized interstate disputes (MIDs) at the dyad-year level.

In addition, we improve upon previous models by attempting to model relative power dynamics. Most forecasting models focus on incorporating structural factors that change slowly, such as the presence of a territorial dispute (e.g. Gleditsch and Ward 2013). These factors change slowly, if at all, and are not easily manipulated by policy makers. Although previous research has modeled the balance of capabilities between two disputants, many existing theories of conflict posit that *changes* in two disputant's relative capabilities are most likely to produce conflict by creating uncertainty about each other's capabilities and making it more difficult for states to commit to agreements (5,6). Our results suggest that modeling more dynamic factors of two state's relationship may provide better predictions about conflict behavior.

2 THEORETICAL JUSTIFICATIONS FOR MODEL INPUTS

Scholars have long argued that shifts in relative power are likely to lead to conflict (7–9). States in relative decline will be concerned about their security, being driven by the fear that in an anarchic world rising opponents will take advantage of the state's moment of relative weakness and capture some corresponding share of their resources and influence. In this scenario, it has been argued, that the declining power is better off fighting a preventive war from a position of power than to continue to decline and be forced to fight a war from a weaker position at some point in the future (7,8). It has been argued, that even if the states can come to some agreement short of war, they will be unable to credibly commit to the agreement, because the state with increasing power will be unwilling to abide by the agreement when they are in a more favorable position in the future. (5,6). This leads us to anticipate that large and rapid shifts in relative capabilities, will be moments when military conflict is most likely. It should be noted, that rapid shifts in relative power do necessarily have to be cases where one state threatens another state's position in the international system, a rapid shift merely implies that their position has changed and may have little impact on the system structure but play an important role in the relationship between the states of the dyad (10).

To fully capture the multifaceted nature of relative capabilities (exogenous and endogenous factors), we will be using the Composite Index of National Capabilities index (CINC) (11). The CINC index is a composite index that considers the nation's economic, demographic, and military attributes. To find the relative capabilities within the dyad, we take the capabilities of the weaker state and divide them by the combined capabilities of both states, giving us the proportion of shared capabilities held by the weaker state. We also include the squared value of this measure for each dyad. Rapid increases in either of these sends a signal to adversaries (or potential adversaries) of a likely military conflict (12–14). To capture this, we include a measure of the percent change from year t to year t-1 for each dyad year. In addition to the measures generated by the CINC scores, we include a dichotomous indicator of whether or not the states in the dyad are major powers and whether both states in the dyad possess nuclear weapons. Both of these inputs affect the ability and incentives for states to engage in militarized conflict. Major powers are able to more easily project force and employ military capabilities (15). Possession of nuclear weapons makes major conflict between two nuclear armed powers less likely but increases the likelihood that the same states will engage in lower level conflict (16,17).

Democracies have been shown to be less likely to engage in conflict with other democracies due to institutional and normative similarities (18,19). Similarly, autocracies have been shown to be less willing to engage in conflict with regimes that are similar to their own (20). Because of this, when the dyad is made up of states with a shared regime type, conflict is less likely. For this input, we employ the regime type measure developed by the Polity IV project (21). This measure ranges from -10, the most autocratic regime type, and 10, the most democratic. We include the Polity score for each state in the dyad. In addition, we include a measure of regime similarity. To operationalize this input, we multiply each state's Polity scores together. This produces a measure that ranges from -100 to 100, where 100 indicates two states have extremely similar regimes and -100 indicates that their regimes are extremely dissimilar. We use this measure rather than the standard dichotomous measures employed in the literature to avoid unnecessarily discarding variance based on arbitrary thresholds for what constitutes democracies and autocracies.[1]

All else equal, states that have similar security preferences should be less likely to fight each other. This includes states that have security alliances with each other. Although these states may have mutually incompatible preferences in some areas, they have a stronger incentive to cooperate with each other in order to maintain their alliance ties in the event that they should need them. We model this using a dummy variable to indicate whether two states have a defensive alliance using data from the Correlates of War Formal Alliance Dataset (22).

Scholars have also found that the previous use of force against the other state within the dyad, leads the two states within the dyad to be more likely to use violence as a means of resolving their disputes in the future (23). In cases where two states have engaged in militarized behavior in the past, shifts in relative power will be especially problematic as the side with increased

[1] De Marchi, Gelpi, and Grynaviski (2004) employ a similar measure. They also include the squared term of this measure. We find that including the squared term did not improve the predictive performance of the model and did not include it in the analysis presented here.

capabilities will take the opportunity presented by this scenario to resolve any outstanding disputes with the use of military force. To account for this, we include a count of the number of previous MIDs between two states. We expect that more previous MIDs are associated with a higher probability of future conflicts.

We also account for joint membership in international institutions. International institutions are a means that states can use in order to manage relative shifts in power. When the two states belonging to a dyad belong to the same international institution they will be less likely to engage in militarized conflict with each other (24). When shifts in relative power occur within a dyad where both states belong to the same institution, the joint membership should have restraining effect on the behavior of the states within a dyad that experiences a shift in relative power as resorting to violence would cause the state to sacrifice the values accrued to the states via the international institution. To employ this input, we use a count of the number of international institutions that both of the states in the dyad belong to (25). The more international institutions that both states belong to the stronger the restraining effect and the less likely a conflict between the states will occur.

3 MODELING APPROACH

Our dependent variable is a binary indicator of whether two states become involved in a new militarized interstate dispute (MID) in a given year. Data on MIDs comes from version 2 of (26) dataset. This dataset makes adjustments to remove extraneous MIDs and rectify coding errors in version 4.0 of the MID dataset (27). Because our dependent variable is binary, we use logistic regression to predict the probability of observing the onset of a MID between two states in a given year. Our sample is composed of all states in the Western Hemisphere from 1816-2001. In the long run, our goal is to develop a comprehensive forecasting model for international conflict across the globe. However, due to the computational requirements associated with Bayesian simulation, we limit our sample to the Western Hemisphere at this stage of our project. Using this region allows us to overcome some of the data limitations associated with other regions of the world and also facilitates comparisons with existing research, namely (3).

Our unit of analysis is the dyad-year, which provides one observation for each pair of states in our sample for each year that both states are members of the international system. Because our primary concern is with accurately forecasting conflict, we use split-sample cross-validation to evaluate the predictive ability of our model. Following Gleditsch and Ward, (2013), we divide our data into two sets: a training set composed of all observations up to and including 1989, and a test set of all subsequent observations. This allows us to evaluate our model's ability to predict the onset of conflict in the post-Cold War era. (e.g., 28), our results suggest that forecasting based on data in previous eras can still be effective despite this.

3.1 Rare Events Corrections

Because international conflicts are rare events, standard logistic regression is likely to underestimate the probability of a conflict occurring and produce biased estimates of the regression coefficients (29). To account for this, we employ two different corrections designed to remedy this problem. First, we use a sampling technique known as undersampling. This involves randomly selecting a subset of the observations that do not experience the event of interest to produce a balanced dataset that contains equal numbers of ones and zeros.[2]

Second, many of the problems associated with rare events can be corrected for by choosing an appropriate prior distribution. Several prior distributions have been studied in the context of

[2] In addition to undersampling, we estimated models using oversampling (i.e. increasing the size of the dataset by randomly sampling from the observations that experience the event) and two different algorithms designed to balance the data by creating synthetic cases based on the observed conflicts (SMOTE and ROSE). Of these, oversampling performed the best.

rare events. Among these, we selected the Cauchy prior suggested by (30).[3] The Cauchy distribution is a weakly informative prior distribution that imposes limited constraints on the size of logistic regression coefficients, which prevents the possibility of extremely large coefficients and mitigates the problems with separation that are often associated with rare events data. In addition, the use of a Cauchy prior helps to shrink the estimates of variables that do not have a strong influence on the dependent variable towards zero. The use of this prior naturally regularizes the estimates and helps prevent overfitting due to noise introduced by potentially extraneous covariates. Following the advice of (30) we use a Cauchy distribution with mean 0 and scale 10 as a prior on the constant term, and a Cauchy distribution with mean 0 and scale 2.5 for all other predictors. All predictors are centered, and continuous covariates are scaled to have a standard deviation of 0.5.

4 EMPIRICAL RESULTS

Table 1 presents the results of our analysis, including the mean of the posterior distribution, standard deviation, and the 95 percent credible intervals for each predictor. Our results are based on a sampling procedure that employs two Markov chains with 100,000 iterations each and a burn-in period of 50,000 iterations. We assessed the convergence of the two chains by ensuring that the Gelman-Rubin diagnostic ($\hat{R}$) values were above 1.01 for all estimated parameters.

[3] In addition, we tested models using several other prior distributions, including a Jeffreys prior, a normal prior, and a Laplace prior. Of these, the Cauchy prior produced the most accurate predictions.

Table 1: Posterior Distribution of Estimated Coefficients

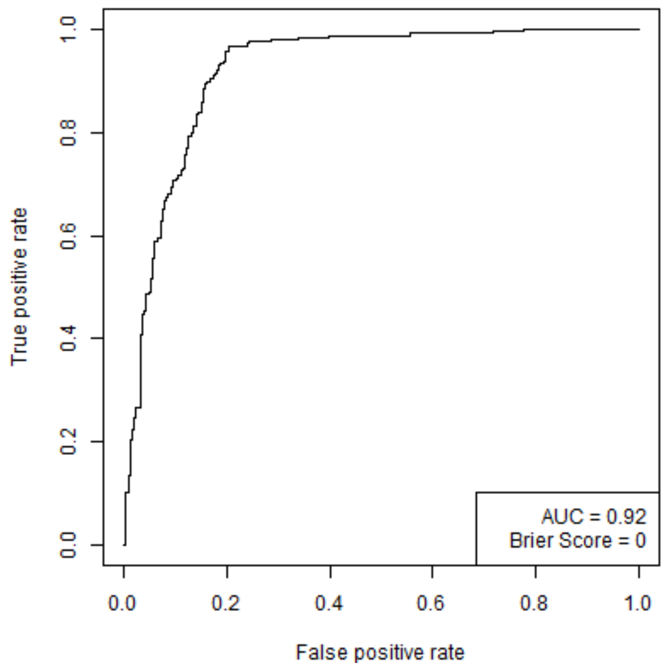
	Mean	Std. Dev.	2.5 Percent	97.5 Percent
Constant	-1.08	0.15	-1.39	-0.79
Capability Ratio	0.59	0.84	-1.04	2.26
Capability Ratio Squared	-0.43	0.78	-1.98	1.08
Percent Change Capabilities	0.12	0.25	-0.37	0.62
Major Power	0.63	0.52	-0.39	1.65
Polity A	-0.11	0.28	-0.66	0.44
Polity B	-0.25	0.28	-0.79	0.29
Polity A*B	-0.32	0.26	-0.82	0.18
Defensive Alliance	-0.78	0.34	-1.47	-0.12
Intergovernmental Organizations	0.98	0.35	0.32	1.67
Previous MIDs	2.52	0.28	2.00	3.09
Nuclear Power	0.68	0.62	-0.53	1.91

4.1 In-Sample Performance

We begin by examining how our model performs at predicting conflicts within the training set. We start by plotting the Receiver-Operating Characteristic (ROC) curve for the training predictions, as first suggested by (4). ROC Curves plot the proportion of correctly predicted 0s (False Positive Rate) on the *x*-axis and the proportion of correctly predicted 1s (True Positive Rate) on the *y*-axis. This allows for a comparison of the proportion of correctly classified 0s and 1s at different thresholds. The greater the area under the ROC curve, the better a classifier performs regardless of the prediction threshold specified. This can be summarized using the Area Under the Curve (AUC) statistic, which provides a succinct summary of how well the model performs. The closer this value is to one, the better a model performs, where values of 1 indicate that the model perfectly predicts the value of all observations. We also present the Brier score, which is the mean squared difference between an observation's predicted probability and its observed binary outcome. The closer this value is to 0, the less incorrect predictions the model makes.

Figure 1 presents the ROC curve for the training sample. The summary statistics illustrate that the model performs very well overall, with an AUC value of 0.92 and a Brier score indistinguishable from 0. The AUC can be interpreted as the probability that a randomly selected conflict observation has a 92 percent probability of being assigned a higher predicted probability than a randomly selected peace observation. From the ROC curve itself, it is apparent that whatever threshold chosen must produce a relatively high false positive rate to correctly classify most of the conflict observations. For example, to correctly classify 90 percent of the observed ones, it is necessary to accept a roughly 20 percent false positive rate.

Figure 1: ROC Curve, In-Sample



In terms of predictions, a model that performs well will assign higher predicted probabilities to observations that experience conflict than those that do not. We can evaluate this graphically by examining a separation plot of the data (31). Separation plots consist of a series of panels representing each observation in the data arranged from left to right in order of increasing predicted probability, with different colors used to indicate whether an event occurred. Dark panels represent observations where an event actually occurred, while light panels represent observations where no event occurred. The solid black line running from left to right represents the value of the predicted probability of each observation. Figure 2 displays the separation plot for our in-sample results. As illustrated, the model performs reasonably well, with dark panels clustered on the right-hand side and light panels clustered on the left.

Figure 2: Separation Plot, In-Sample

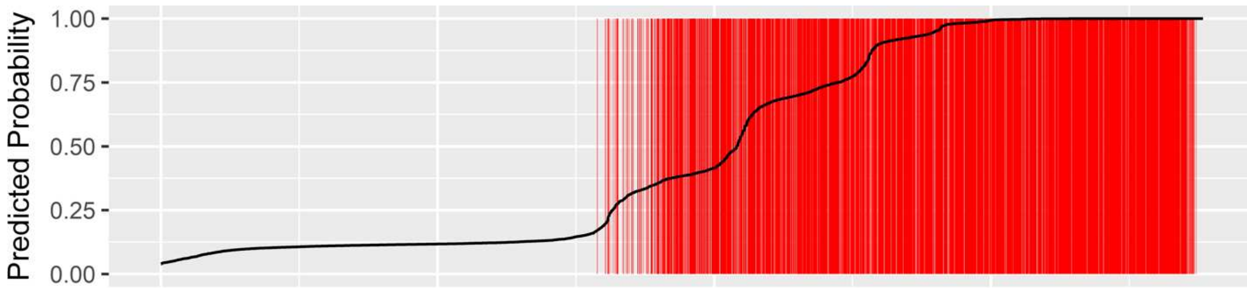


Table 2 summarizes the performance of our model by comparing the predicted values for each observation (0 or 1) in our test data with their observed values. Although any choice of threshold is possible, the use of undersampling to balance the training set makes it so that a threshold of 0.5 can be used to classify the observations into one category or the other. Without this technique, it is unlikely that many (if any) observations would be assigned predicted probabilities above 0.5. By contrast, our model correctly predicts the onset of roughly 73 percent of conflicts at this threshold. Similarly, the model predicts roughly 88 percent of peace years accurately. Although evaluating the performance of these models in the test sample is necessary to get a true sense of the model's forecasting performance, this provides a benchmark for what to expect in the test set. Analysts interested in shifting the balance of true positives and negatives may be interested in selecting an alternative threshold.

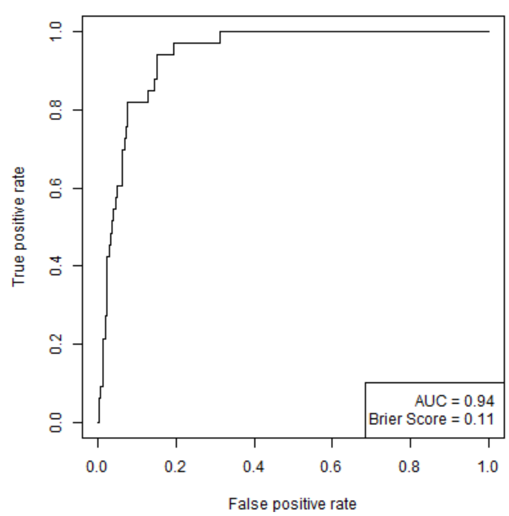
Table 2: Actual vs. Predicted Disputes, In-Sample

	No Dispute Predicted	Dispute Predicted
No Dispute Observed	225 (88.24%)	30 (11.76%)
Dispute Observed	69 (27.06%)	186 (72.94%)

4.2 Out-of-Sample Performance

We now turn to discuss the predictive power of our model out-of-sample. Our validation set consists of a total of 3,900 dyad-years in the period from 1990 to 2001. During this time, a total of 33 militarized disputes were observed within the sample. Figure 3 presents the ROC curve for our test sample, as well as the AUC and Brier scores. The AUC of 0.94 indicates that the model performs extremely well overall. This indicates that a randomly selected observation where conflict is observed has a 94 percent chance of being assigned a predicted probability higher than a randomly selected peace observation. As before, correctly identifying 90 percent of the 1s requires accepting a false positive rate of about 20 percent. As such, analysts are necessarily required to make tradeoffs when interpreting such models depending on whether they care more about identifying most potential conflicts or weeding out false positives. For example, an 80 percent true positive rate yields a slightly lower false positive rate of roughly 10 percent. These results are roughly comparable to those in the training sample.

Figure 3: ROC Curve, Out-of-Sample



Although the Brier score of 0.11 is substantially larger than the in-sample Brier score, it still indicates that the model has strong predictive performance. In addition, Figure 4 presents the separation plot for the test sample. Although there is considerably more gray space than before due to the much higher number of 0s in the test sample, the red lines still cluster on the right-hand side of the graph, indicating that the model systematically assigns high predicted probabilities to the observations that experience conflict.

Figure 4: Separation Plot, Test Sample

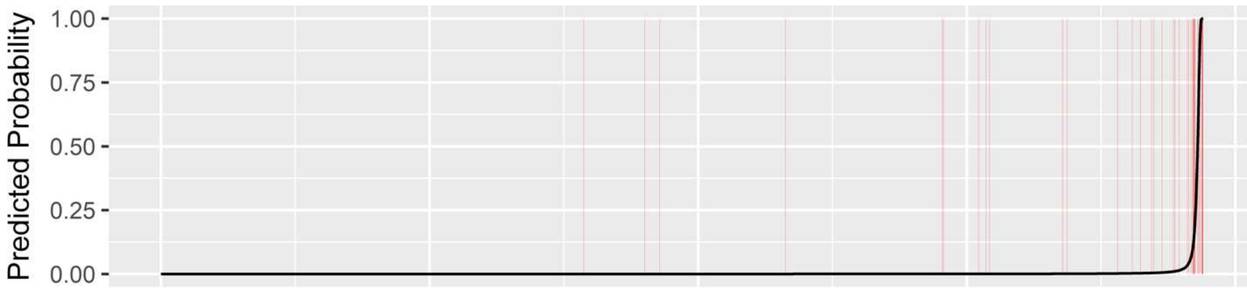
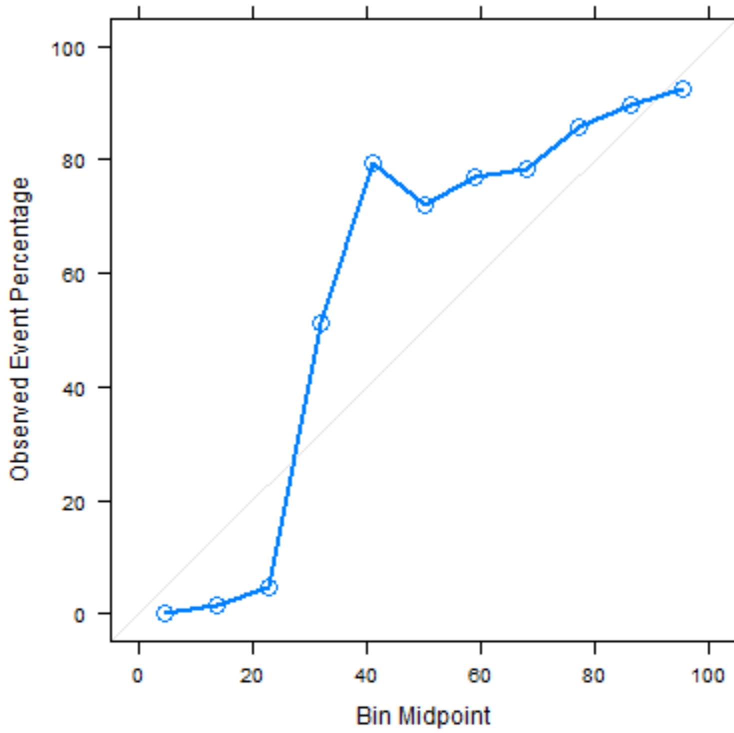


Table 3 presents a cross-tabulation of the predicted and observed values for each dyad-year in our dataset. In total, our model predicts 352 dispute onsets. In doing so, we correctly predict the onset of 27 out of 33 disputes, thereby correctly predicting 82 percent of the conflicts that actually occurred during this period. As an additional measure of predictive performance, Figure 5 presents the calibration plot of the test sample predictions (1,4). This figure is produced by binning the observations according to predicted probability, rescaled between 0 and 100, by intervals of 10. For each bin, the number of observed events is calculated. The median predicted probability for each bin is plotted on the *x*-axis with the proportion of observations that experienced the event plotted on the *y*-axis. Points that fall on the 45-degree line indicate that the proportion of events that occurred within that bin is equal to the expected number of events (e.g. roughly 85 percent of the observations should fail within the 80-90 percent probability bin). From this plot, we can see that the model tends to overpredict low probability conflicts below the 30 percent cut point (i.e., less conflicts occur than should occur within these bins) and underpredict conflicts above this point. This illustrates the fact that the model tends to assign too low probabilities to conflicts, as is to be expected with rare events.

Figure 5: Calibration Plot, Test Sample



To get a sense of the disputes that our models fails to predict, we examine each of the false negatives in detail. Table 4 lists each of the seven observed disputes that our model failed to predict. Because we use data on the same countries and time period as (3), we can directly compare the results of our model to theirs. This comparison allows us to build on their already excellent work to produce an even more accurate forecasting model. Two of our false negatives, Canada vs. Haiti and Haiti vs. Argentina, stem from the same conflict, MID 4016, which involved an attempt by several countries to restore Haitian President Jean-Bertrand Aristide to power following a military coup. Notably, our model did accurately predict the occurrence of MIDs between Haiti and the Dominican Republic and the USA in this year, both of whom were participants in MID 4016.

Table 3: Actual vs. Predicted Disputes, Test Sample

	No Dispute Predicted	Dispute Predicted
No Dispute Observed	3542 (91.6%)	325 (8.4%)
Dispute Observed	6 (18.18%)	27 (81.82%)

In addition, our model did not accurately predict the onset of a conflict between Venezuela and Guyana. This incident occurred due to the escalation of a border dispute between the two countries, and although it led to the mobilization of troops in a display of force, no actual use of force was observed. Each of the remaining false negatives involved minor disputes over maritime boundaries or fishing rights, none of which presented a serious risk of escalation or casualties. In addition, our model predicts several disputes that (3) does not. In addition to the two mentioned above, our model accurately predicts a series of disputes between Belize and Guatemala, several other disputes between Trinidad and Tobago and Venezuela, a dispute between the U.S. and Peru, and a dispute between the U.S. and Venezuela.

Table 4: False Negatives in Test Sample

Country A	Country B	Year	Mid Number	Fatality	Max Duration
Canada	Haiti	1993	4016	0	335
Haiti	Argentina	1993	4016	0	335
Honduras	El Salvador	1993	4010	0	25
Trinidad and Tobago	Venezuela	1996	4149	0	0
Trinidad and Tobago	Venezuela	1999	4155	0	3
Venezuela	Guyana	1999	4260	0	6

Although our model does correctly predict most of the disputes that occur in our validation sample, our model does produce a high rate of false positives. Of the 352 disputes predicted by our model, 325 constitute false positives. Admittedly, this is substantially higher than similar models (e.g., 3), and in our view, constitutes the most serious problem with our current model. To a certain extent, this is offset by the superior performance of our model at accurately predicting the onset of observed conflicts. From a pragmatic perspective, we would rather accurately predict as many conflicts as possible, even if doing so produces a higher number of false positives. Since the ultimate goal of forecasting is to take steps to prevent conflicts from happening or mitigate their effects, we would rather identify countries that are potentially at risk of conflict. We contend that this is valuable, even if many of these conflicts do not occur. By contrast, models that produce lower numbers of false positives and higher numbers of false negatives are of less use in planning for potential conflicts. Moreover the model classifies 92 percent of the 0s correctly. Given the overwhelming number of 0s in the dataset, this performance is substantial. Nonetheless, we plan to focus on reducing the number of false positives generated by the model in future work.

Finally, since shifts in the distribution of power between disputants are likely to influence the onset of international conflict, we included this measure in the model. Although this variable is itself insignificant, it does enhance the predictive performance of our model. Compared to a model that does not include this measure, our model predicts an equal number of conflicts in the test sample. However, it also predicts less false positives. Although we expected this measure to have a more measurable substantive effect, we suspect that the rough nature of this measure precludes finding more substantial results. However, our results do suggest that considering power dynamics may be an important avenue for future research. For example, using disaggregated measures of capabilities may be a fruitful endeavor.

5 CONCLUSION

In this paper, we develop a forecasting model that is capable of predicting the onset of international conflict at the yearly level. To our knowledge, this is one of the first papers to do so. Using Bayesian regularization techniques and undersampling, our model can accurately predict the onset of 80-90 percent of conflicts using the conventional 0.50 threshold and may thus be useful in providing policymakers with early warning regarding potential international disputes. In future work, we hope to expand this model to develop a global forecasting model capable of facilitating cross-national risk mapping of countries that are at risk of experiencing conflict in the future.

In addition, we provide some preliminary evidence that incorporating better measures that capture the dynamic relationship between two disputants may be a fruitful avenue for future research. By incorporating the change in the balance of capabilities between disputants in our model, we were able to reduce the false positive rate associated without our model. Although this variable does not have a significant effect in-and-of-itself, the fact that it improves the predictive performance of the model suggests that future models should make greater efforts to model dynamics between disputants. We suspect that more nuanced measures of changes in the balance of military capabilities would provide even greater predictive leverage and facilitate better predictions.

In future work, we hope to extend our approach to allow for more natural methods of making predictions about future conflict. Unlike frequentist forecasting models, our Bayesian forecasting model incorporates uncertainty into the predictions made in the forecasting stage, which can provide policymakers with more realistic information regarding the probability of international conflict and the risks they should take based upon the forecasts generated from the model. This is a step that has been called for in previous critiques of forecasting models of international conflict (32).

REFERENCES

1. Beck N, King G, Zeng L. Improving quantitative studies of international conflict: A conjecture. American Political Science Review. 2000;94(01):21–35.

2. Beck N, King G, Zeng L. Theory and evidence in international conflict: a response to de Marchi, Gelpi, and Grynaviski. American Political Science Review. 2004;98(02):379–389.

3. Gleditsch KS, Ward MD. Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes. Journal of Peace Research. 2013 Jan;50(1):17–31.

4. de Marchi S, Gelpi C, Grynaviski JD. Untangling Neural Nets. The American Political Science Review. 2004;98(2):371–8.

5. Fearon J. Rationalist Explanations for War. International Organization. 1995;49(3):379–414.

6. Powell R. War is a Commitment Problem. International Organization. 2006 Winter;60:169–203.

7. Copeland DC. The origins of major war [Internet]. Cornell University Press; 2000 [cited 2016 Dec 9]. Available from: https://books.google.com/books?hl=en&lr=&id=D1Se3ofJzM4C&oi=fnd&pg=PR7&dq=copeland+origins+of+major+war&ots=ofy-0w1fmr&sig=rfS6IRkmW81KyHunQ-tmGUZ5v2k

8. Gilpin R. War and change in world politics [Internet]. Cambridge University Press; 1983 [cited 2017 Jul 17]. Available from: https://books.google.com/books?hl=en&lr=&id=2iKL7zr3kl0C&oi=fnd&pg=PR9&dq=gilipin+1983&ots=W2PYEf7UOn&sig=4KM6ie7LIaoFy6fo_e9V48Kmbtg

9. Kennedy P. The rise and fall of the great powers [Internet]. Vintage; 2010 [cited 2017 Jul 17]. Available from: https://books.google.com/books?hl=en&lr=&id=_oF66b7HOyMC&oi=fnd&pg=PT15&dq=kennedy+1987+decline+of+empire&ots=ou5386e0_a&sig=hrysE-vrxmI98A39xMc0EFhVPbI

10. Debs A, Monteiro NP. Known unknowns: Power shifts, uncertainty, and war. International Organization. 2014;68(1):1–31.

11. Singer JD, Bremer S, Stuckey J. Capability distribution, uncertainty, and major power war, 1820-1965. Peace, war, and numbers [Internet]. 1972 [cited 2016 May 22];19. Available from: http://www.owlnet.rice.edu/~stoll/P570/sbs.pdf

12. Gibler DM, Rider TJ, Hutchison ML. Taking Arms Against a Sea of Troubles: Conventional Arms Races During Periods of Rivalry. Journal of Peace Research. 2005 Mar;42(2):131–47.

13. Jervis R. Perception and Misperception in International Politics. Princeton University Press; 1976. 468 p.

14. Rider TJ, Findley MG, Diehl PF. Just part of the game? Arms races, rivalry, and war. Journal of Peace Research. 2011 Jan;48(1):85–100.

15. Bremer SA. Dangerous dyads conditions affecting the likelihood of interstate war, 1816-1965. Journal of Conflict Resolution. 1992;36(2):309–341.

16. Gartzke E, Kroenig M. Nukes with Numbers: Empirical Research on the Consequences of Nuclear Weapons for International Conflict. Annual Review of Political Science. 2016;19(1):397–412.

17. Waltz KN. Nuclear Myths and Political Realities. The American Political Science Review. 1990;84(3):731–45.

18. Bueno de Mesquita B, Morrow JD, Siverson RM, Smith A. The Logic of Political Survival. Cambridge, MA: MIT Press; 2003.

19. Oneal JR, Russett BM. Clear and clean: The fixed effects of the liberal peace. International Organization. 2001;55(02):469–485.

20. Peceny M, Beer CC, Sanchez-Terry S. Dictatorial peace? American Political Science Review. 2002;96(1):15–26.

21. Marshall MG, Jaggers K. Political Regime Characteristics and Transitions, 1800-2012. Polity IV Project. 2013;

22. Gibler DM, Sarkees MR. Measuring Alliances: the Correlates of War Formal Interstate Alliance Dataset, 1816–2000. Journal of Peace Research. 2004 Mar;41(2):211–22.

23. Maoz Z. Pacifism and fightaholism in international politics: A structural history of national and dyadic conflict, 1816–1992. International Studies Review. 2004;6(4):107–133.

24. Mitchell SM, Hensel PR. International Institutions and Compliance with Agreements. American Journal of Political Science. 2007;51(4):721–37.

25. Pevehouse J, Nordstrom T, Warnke K. The Correlates of War 2 international governmental organizations data version 2.0. Conflict Management and Peace Science. 2004;21(2):101–119.

26. Gibler DM, Miller SV, Little EK. An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001. ISQ. 2016 Dec 1;60(4):719–30.

27. Palmer G, D'Orazio V, Kenwick M, Lane M. The MID4 dataset, 2002–2010: Procedures, coding rules and description. Conflict Management and Peace Science. 2015 Apr;32(2):222–42.

28. Jenke L, Gelpi C. Theme and Variations: Historical Contingencies in the Causal Model of Interstate Conflict. Journal of Conflict Resolution. 2016 Jan 7;0022002715615190.

29. King G, Zeng L. Explaining Rare Events in International Relations. International Organization. 2001 Summer;55(3):693–715.

30. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. 2008 Dec;2(4):1360–83.

31. Greenhill B, Ward MD, Sacks A. The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models. American Journal of Political Science. 2011 Oct 1;55(4):991–1002.

32. Brandt PT, Freeman JR, Schrodt PA. Real Time, Time Series Forecasting of Inter- and Intra-State Political Conflict. Conflict Management and Peace Science. 2011 Feb 1;28(1):41–64.