

# Can levels of fear and anger expressed at the UN General Debate signal elevated military spending?

## Introduction

“[A] quick glance at the state of research on emotions in international relations shows that, while innovative studies have been undertaken on the topic, there is still need to develop, explore and especially test theories and methods.” (Ariffin, Coicaud, and Popovski 2016). This paper aims to address this gap by empirically testing whether government representatives’ emotions can be a significant predictor of military spending.

This paper will operationalise a non-rational, emotional variable – indicative of fear and anger expressed by governments – which can be used to explain military spending. A quantitative score of fear and anger relating to foreign policy (hereafter, a ‘sentiment score’) will be derived from analysis of United Nations General Debate speeches using topic modelling (Latent Dirichlet Allocation or ‘LDA’ (Blei, Ng, and Jordan 2003)), and sentiment analysis. The objective of the paper is twofold: firstly, to ascertain whether sentiment score is a statistically significant predictor of states’ military spending, and secondly, to develop predictive models for whether a state will spend highly on its military in a given year.

Three datasets were used: UN General Debate speeches between 1991 and 2000 (Mikhaylov, Baturo, and Dasandi 2017), data published in Böhmelt and Bove (2014a), and the Correlates of War militarised interstate disputes dataset (v4.1) (Palmer et al. 2015). To address the first aim of the paper, a linear model was fitted regressing military spending on sentiment score. For the second, a comparison of classification techniques – logistic regression, random forests and support vector machines – ascertained whether a predictive model of supernormal military spending is possible using sentiment scores and other predictors. I conclude that sentiment scores appear to have a small but notable effect on military spending but are of limited importance in predictive models.

## Literature review

As Ariffin, Coicaud, and Popovski (2016) note, a widespread assumption of rational expectations and the pursuit of “interests” among the dominant theoretical approaches to international relations has led to the neglect of emotional determinants of state behaviour. For example, regarding the neo-realist school, they note that a select few emotions like fear or envy have been “overrationalised” (Ariffin, Coicaud, and Popovski 2016, 1). Theorists like Waltz (1979) or Mearsheimer (2001), who espouse the rational pursuit of a balance of power or hegemony respectively, could also explain the empirical basis of their theories through emotional responses of fear, envy or anger among decision makers.

Crawford (2000) criticises realists’ neglect of human responses to fear and liberal institutionalist’s neglect of emotional motivations for co-operation beyond mutual interests. Crawford also notes that where psychology has been considered, the focus tends to be on cognition rather than emotion, emphasising rationality over irrationality. While historically, international relations theory has taken into account emotions, this style of theorising fell out of favour in the post-war era (Crawford 2000, 120–23). Where emotion has attracted the attention of international relations or policy theorists, it has mostly been examined using qualitative methods like discourse analysis (Clement and Sangar 2018). In contrast, this paper investigates whether a quantitative approach can contribute to a neglected part of the discipline.

This paper builds on studies by Böhmelt and Bove (2014b) and Nordhaus, Oneal, and Russett (2012). Methodologically, their predictors represent those implied by a rationalistic approach, explored further in the next section.

## Theoretical framework

We can broadly classify the determinants of military spending from a rational choice perspective into two categories. Firstly, capability based spending is pre-emptive and based on the resources available to a state, which allows for greater military investment. The second is imperative-based; if a country is at war or observes high spending by hostile powers, then it is incentivised to spend more. Both of these categories demonstrate a clear logic of incentives and self-interest.

Domestic politics may also play a role. For instance, we would expect autocracies to be more reliant on militaries to maintain law and order given that such regimes often lack a functional judiciary, or that democracies are less likely to fight. Therefore, we may also expect the extent to which a country can be considered democratic to be a potential predictor of military spending levels.

In addition to that driven by government characteristics or incentives in the international state system, I define ‘emotionally motivated’ military spending as that driven by fear and anger, as expressed by state representatives. Hence, the causal claim investigated in this paper is as follows: if high levels of these emotions are shown by state representatives then governments are likely to overestimate threats, leading to increased military expenditure.

We can now frame two hypotheses:

*H1: Higher levels of fear and anger, as expressed by state governments, are associated with increases in military spending.*

*H2: Models specified to include levels of fear and anger among other predictors can predict whether a country engages in supernormal levels of military spending in a given year.*

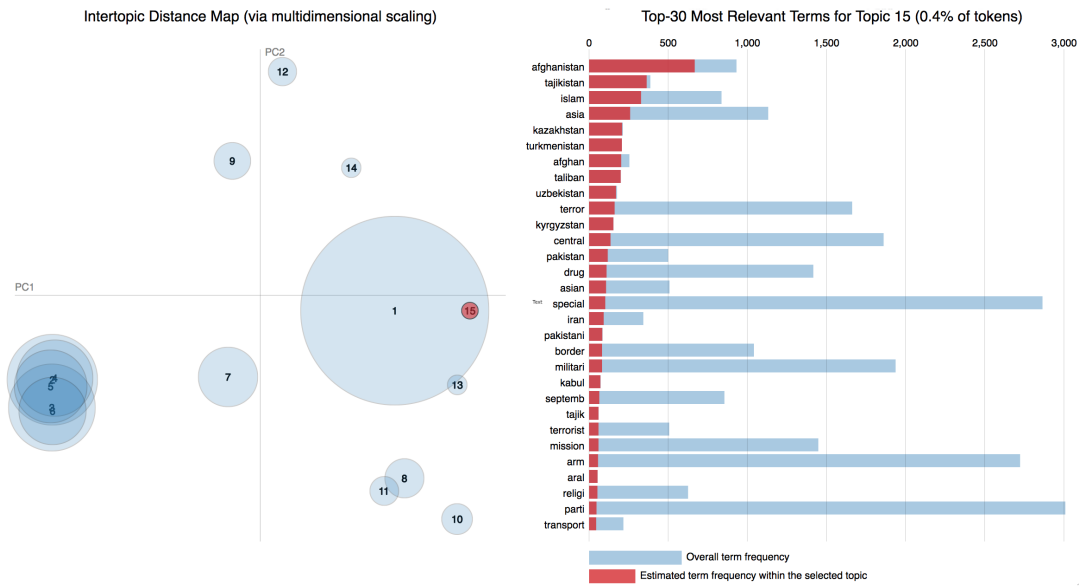
## Methodology and research design

This paper considers UN member states between 1991 and 2000 – the period between the end of the Cold War and the 9/11 attacks – given the period’s stability relative to the those that preceded and succeeded it. The data is panel-structured, with each observation corresponding to an individual state in a given year.

### Constructing a sentiment score

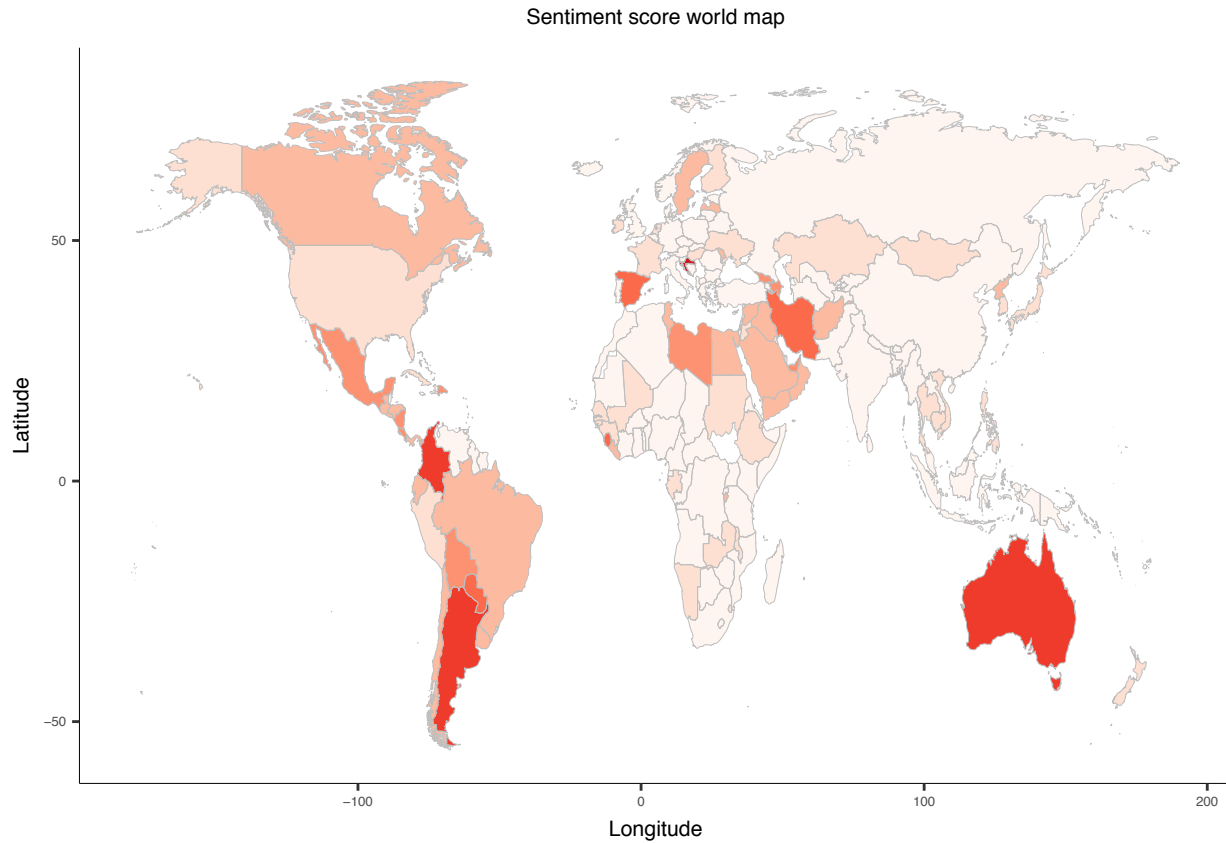
The NRC sentiment lexicon (Mohammad and Turney 2013) was used to identify words indicating fear or anger in the speech corpus. Word counts by speech were generated using these lexicons to give a raw sentiment score by country by year, which were then indexed against each country’s 1991 scores, set at 100. Scores were indexed because this enhances interpretability: we are interested in changes in spending in a given year so the change in sentiment in that year is more useful than a raw word count.

Having constructed an indexed sentiment score, an LDA topic model was used to identify fifteen topics within the corpus. The intuition behind topic modelling is that “documents exhibit multiple topics... [where] a topic [is] a distribution over a fixed vocabulary” (Blei 2012). LDA is a computational method that allows us to discover these hidden topic variables based on the terms observed in each document. Of the fifteen topics, those that reflected security or foreign policy issues in their top 30 terms were selected. An example is reproduced below, showing country names in a specific region and conflict-related terms, indicating foreign policy relevance.



The posterior document probabilities for the selected topics, derived from the estimated distribution of topic variables across documents conditional on observed terms, were extracted from the model. These were multiplied by the indexed sentiment scores to provide a measure of fearful and angry sentiment expressed in UN speeches, weighted by foreign policy topics.

The map below shows sentiment score by country. Darker red represents higher scores. The highest scoring countries are in southeast Europe, reflecting ethnic conflicts that took place there over the 1990s. Persistently conflictual regions like the Middle East, particularly Iran, score relatively highly. Surprisingly, Russia does not score highly despite the fall of the USSR, while Australia and Spain score highly with little historical justification. These anomalies may be due to the measure being arbitrarily based to 1991, which fails to capture elevated sentiment levels in the base year if subsequent increases do not take place, even if sentiment remains elevated.

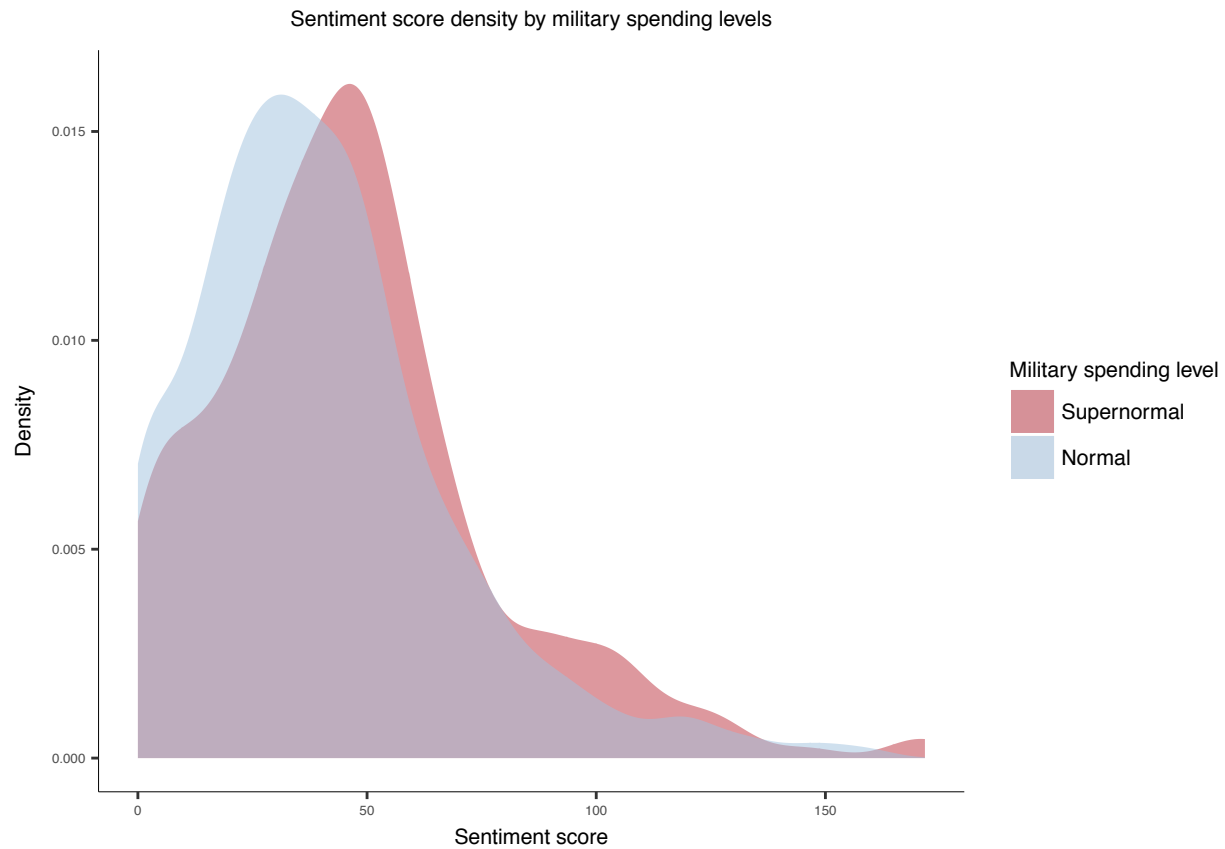


### Response variables: military spending and classes of expenditure

Military spending, the response variable for the first hypothesis, will be measured as log transformed dollar values, given the wide spread in the data (Böhmelt and Bove 2014a).

For the predictive models, I constructed a binary categorical variable labelled 'Supernormal' for a country in a given year where military spending is one standard deviation above that country's long-run average (the average over the preceding decade), and 'Normal' otherwise (i.e. a level of military spending showing a modest increase, in line with the long run average or a decrease). These categories were chosen because the question is focused on military spending increases rather than cuts.

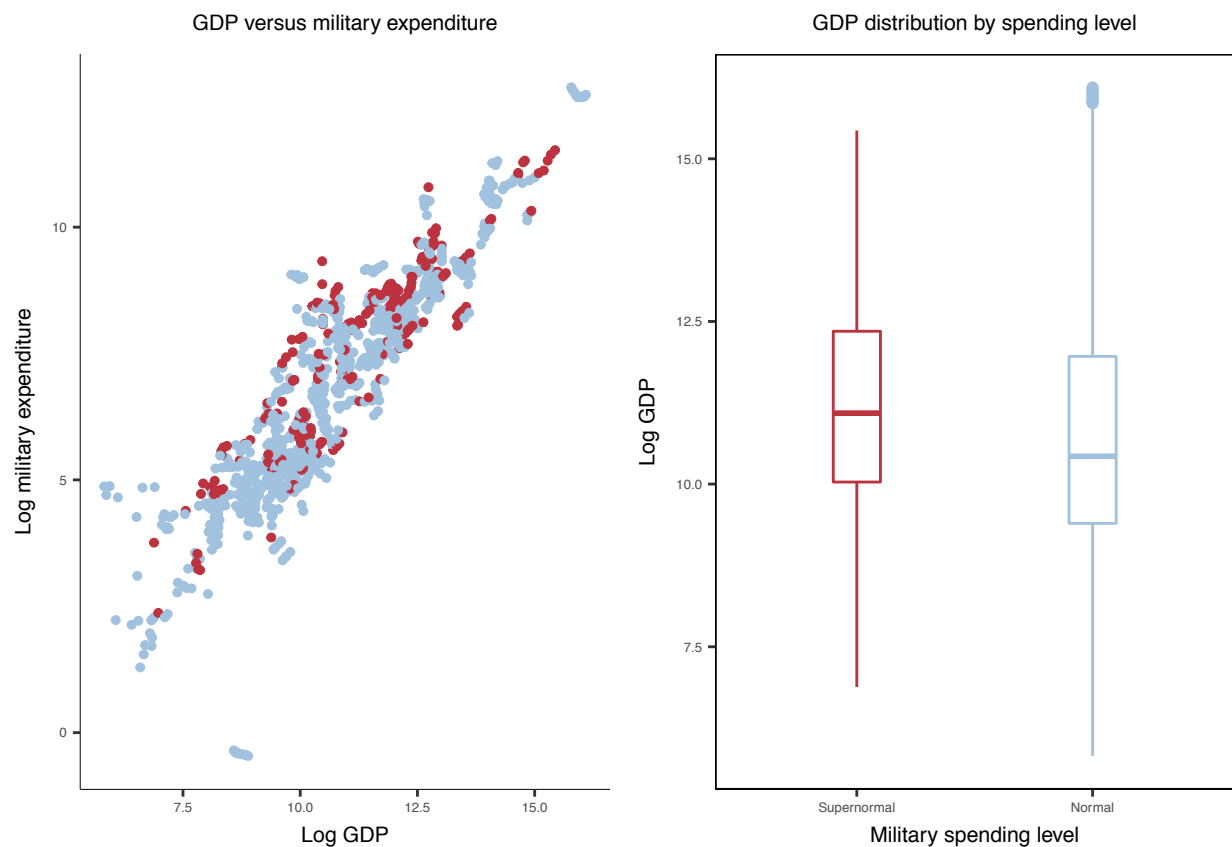
As an initial attempt to discern a relationship between these categories and sentiment score, density plots are shown below. Both distributions have a right skew but the supernormal category on average scores more highly on sentiment score, implying that there may be empirical evidence for the theoretically posited relationship.



## Selecting control variables

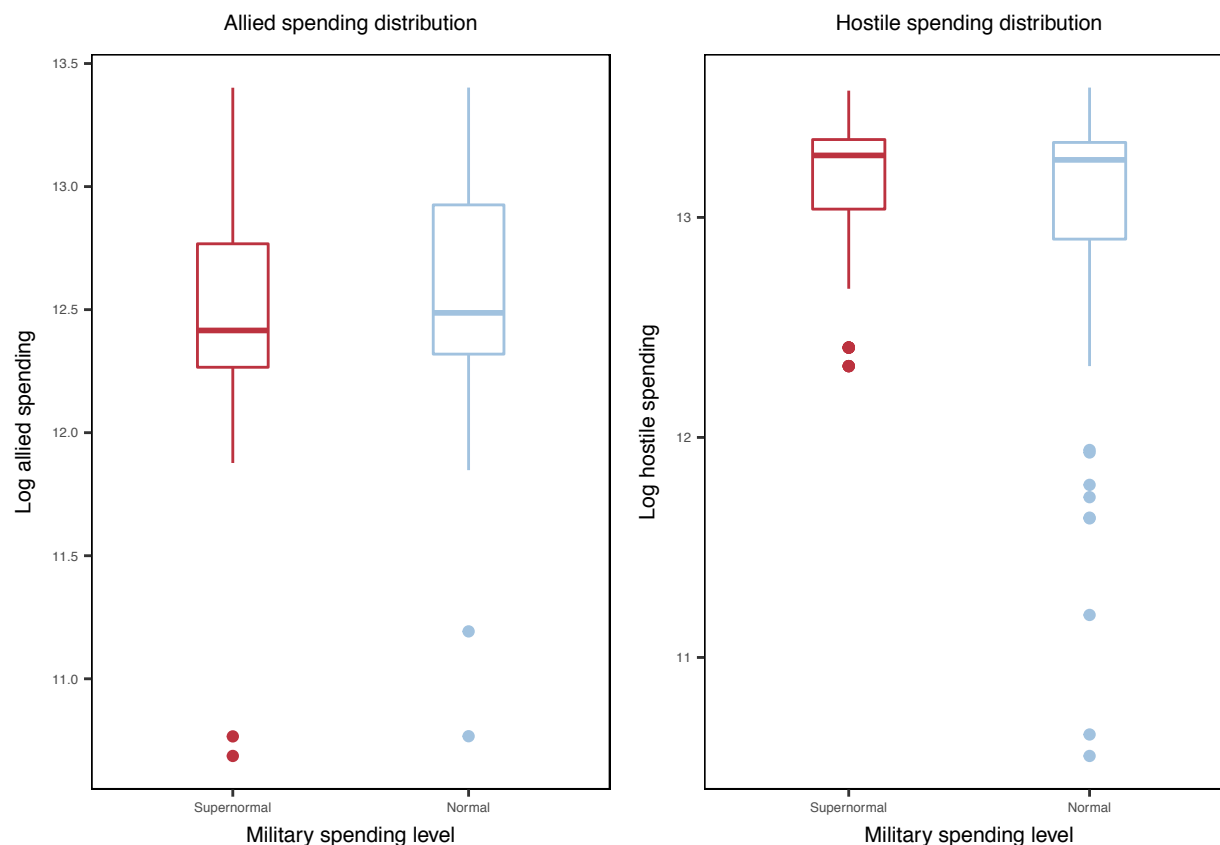
### GDP

The plots below show the relationship between GDP and military spending – I used log-transformed variables given the wide variation in the raw data. A clear relationship emerges from the scatter plot that suggests GDP is a likely predictor of military expenditure. We also find that the mean and distribution of GDP for countries in the supernormal spending class is higher than the normal class.



### Foreign military spending

The relationship between allied and hostile military spending and the response variables is consistent with a rational choice story. The boxplots show that, on average, high military spending by enemy states is likely to result in supernormal spending, while high allied spending is likely to result in normal spending. This is consistent with a ‘free-rider’ story, where countries benefit from allied expenditure, reducing the need to spend themselves (data sourced from Böhmelt and Bove (2014a)).



## Trade, regime type, alliances and conflict

I also controlled for trade as a proportion of GDP and level of democracy (measured by the polity score used by Böhmelt and Bove (2014a)). The data suggest that those with higher levels of trade tend to spend normally on their militaries, as do those with more democratic governments. This is in line with the expectation that more economically open and democratic countries are less likely to engage in conflict.

Finally, I controlled for NATO membership and current conflict. The latter has been derived from the Correlates of War militarised interstate dispute dataset (Palmer et al. 2015), containing all dyads of countries engaged in a militarised dispute by year. A categorical variable was constructed from this dataset, coded 1 where a country was engaged in a militarised dispute and 0 otherwise.

## Techniques and models

For the first hypothesis, a linear regression model was fitted to infer whether a significant relationship exists between sentiment score and military spending. Given the panel structure of the data, dummy variables for fixed and time effects were specified in the model. The response variable was log military expenditure and the predictor variable was sentiment score with controls as outlined above.

For the second hypothesis, a logistic regression model, a random forests model and a support vector machine model ('SVM') were fitted to determine whether the above variables are capable of predicting normal or supernormal military spending for a given country in a given year. For the random forests model, the square root of the number of predictors was used as the number of predictors to be considered at each decision tree split. For the SVM model, a radial kernel was set to ensure adequate flexibility and the optimal cost parameter was determined through cross-validation.

Given that thresholds for the spending categories are unique to each country and year, and represent changes in military spending rather than absolute values, the risk of autocorrelation is lower than in the model fitted for the first hypothesis. Therefore, simplified models assuming cross-sectional structured data have been fitted, using the normal/supernormal classes as the response variable, with sentiment score and control variables as predictors. Predictive performance was evaluated using a validation set approach, randomly dividing the data such that 70% of observations were used to train the models and 30% held out as a test dataset.

## Results

### Hypothesis 1

A Breusch-Godfrey test on the initial linear model, which is shown on the left side of the table below, showed evidence of autocorrelation. No evidence of cross-sectional dependency was found. An adjusted model, with heteroscedasticity and autocorrelation consistent standard errors, is shown on the right side of the table. Sentiment score shows a small but positive association with military spending, increasing log military spending by 0.0009 for each unit increase in sentiment score, all else equal, significant to a 90% confidence level. The p-value for sentiment score suggests that at a 90% confidence level we can reject the null hypothesis, providing limited support for H1.

As expected, GDP shows the highest level of statistical significance, with higher GDP positively correlated with military spending to a 99% confidence level. Other controls are not significant in the model with adjusted standard errors. However, the 5% R squared suggests that the model does not explain much variance in the data and so we must approach its results with caution.

Table 1: Military expenditure linear regression

	<i>Dependent variable:</i>	
	Military spending	
	<i>panel linear</i>	<i>coefficient test</i>
	(1)	(2)
At_war	−0.0380* (0.0229)	−0.0380 (0.0294)
Sentiment_score	0.0009** (0.0004)	0.0009* (0.0005)
NATO_membership	−0.0498 (0.1493)	−0.0498 (0.0490)
Democracy_score	−0.0040 (0.0043)	−0.0040 (0.0069)
Trade_ratio	0.0194** (0.0084)	0.0194 (0.0186)
GDP	0.4814*** (0.0773)	0.4814*** (0.1506)
Hostile_spending	−0.0242 (0.1103)	−0.0242 (0.0833)
Allied_spending	−0.1001 (0.0746)	−0.1001 (0.0661)
Observations	1,083	
R <sup>2</sup>	0.0511	
Adjusted R <sup>2</sup>	−0.0934	
F Statistic	6.3230*** (df = 8; 939)	
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		



## Hypothesis 2

The best performing classifier was the random forests model, followed by the SVM model. The confusion matrices for each classifier are shown below. The focus here is on predictive performance using the test dataset – see appendix for logistic regression results.

A confusion matrix and model performance measures are shown below for the logistic regression. The error rate (76%) is high. This is driven by an extremely large number of false positives (incorrectly predicted supernormal country-years). The performance of this classifier is poor, unsurprising given the model’s rigid functional structure that is unlikely to reflect the structure of a complex prediction problem.

Predicted/Observed	Supernormal	Normal
Supernormal	73	247
Normal	1	4

Error rate:	76.3%
Correctly predicted:	23.7%
False positive rate:	98.4%
False negative rate:	1.4%

The same measures for the SVM model are shown below. The performance of this model is better, with an error rate of 21%. However, errors are mainly due to false negatives, at 85%. The classifier is consistently misclassifying supernormal cases as normal. Supernormal cases only make up 23% of the test set, so an error rate of 21% with 85% false negatives in the context of the underlying distribution suggest that this is not a successful prediction model.

Predicted/Observed	Supernormal	Normal
Supernormal	11	6
Normal	63	245

Error rate:	21.2%
Correctly predicted:	78.7%
False positive rate:	2.4%
False negative rate:	85.1%

Finally, the random forests model provides an incremental improvement on the SVM model, with an error rate of 20%, albeit with a comparable false negative rate of 81%. The random forests model is therefore the best performing classifier, but again the high false negatives given the test set distribution suggest that this cannot be considered a successful predictor either (see appendix for ROC plots of SVM and random forests models, which corroborate the latter’s superior predictive performance).

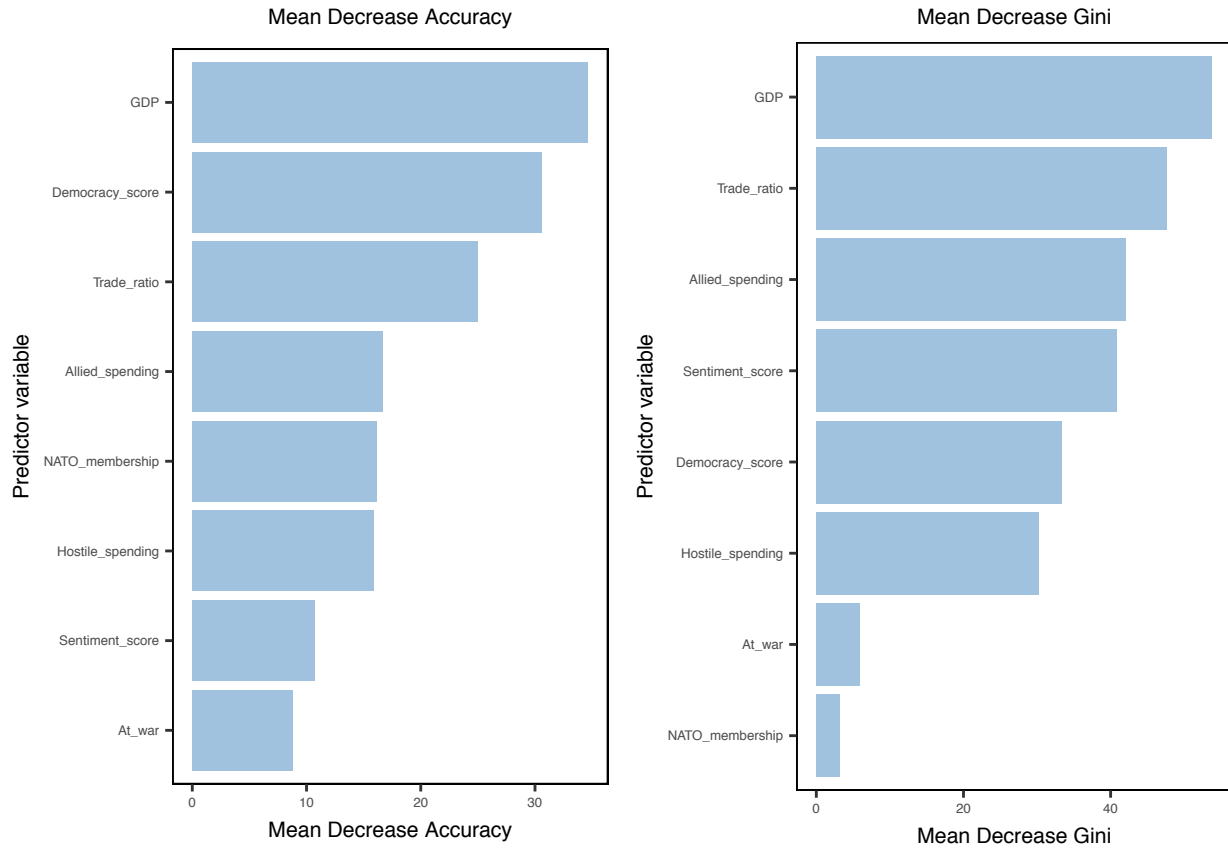
Predicted/Observed	Supernormal	Normal
Supernormal	14	6
Normal	60	245

---

Error rate:	20.3%
Correctly predicted:	79.7%
False positive rate:	2.4%
False negative rate:	81.1%

---

As the best performing classifier, the plot below analyses predictor importance in the random forests model. The highest mean decrease in accuracy results from the exclusion of GDP, the most important predictor on this measure. Sentiment score, in contrast, is second least important. However, in terms of node purity as measured by the Gini measure, sentiment score is ranked fourth after GDP, trade and allied spending. Overall, it is a predictor of limited importance to our model, being definitely less important than GDP, allied spending and trade, and likely less important than democracy score. Given that sentiment score is of doubtful importance in a relatively unsuccessful prediction model, we cannot reject the null hypothesis that predictions from the model are insignificant.



## Conclusion

Sentiment score showed a higher level of significance than other variables apart from GDP and was significant to a 90% confidence level in a regression on military spending. However, predictive models performed poorly. While we cannot consider any of these findings definitive, the findings regarding H1 are a useful precursor to further possible research, particularly given that the empirical study of emotional, non-rational motivations in international relations is in its infancy.

However, there are some limitations to address. The sentiment score itself focuses too much on changes

from year to year, particularly given its arbitrary indexing to a base year. This fails to capture prolonged periods of emotional linguistic usage and hence some important information may be lost in its construction. Furthermore, UN General Debate speeches are arguably too crafted to contain genuinely emotive language – electoral campaign speeches, for example, may be better for this purpose. In addition, clearer trends might emerge over a longer time series, and further consideration of time effects should be worked into future research. For instance, the impact of foreign military spending or increases in trade may act on military expenditure on a lag. Finally, the inclusion of arms imports and exports in the trade metric may result in an endogeneity risk, so may be worth adjusting. Nevertheless, this paper makes progress in demonstrating that linguistic usage and turns of phrase, in addition to substantive content of speeches, may generate useful data for both academic study and foreign policy prediction.

## Appendix

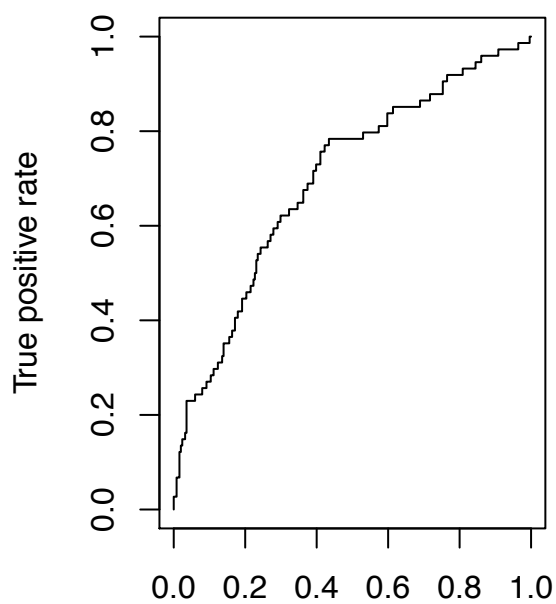
Table 7: Logistic regression model

	<i>Dependent variable:</i>
	Supernormal military spending
At_war	0.0167 (0.1986)
Sentiment_score	−0.0078*** (0.0029)
NATO_membership	0.7790** (0.3913)
Democracy_score	0.0437*** (0.0148)
Trade_ratio	0.0339 (0.1070)
GDP	−0.2492*** (0.0551)
Hostile_spending	0.1657 (0.5052)
Allied_spending	0.5949 (0.4041)
Constant	−5.5462 (11.0462)
Observations	758
Log Likelihood	−381.6918
Akaike Inf. Crit.	781.3837

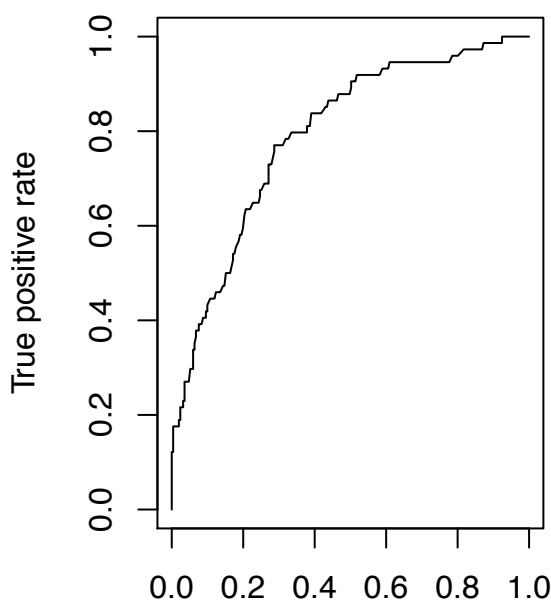
Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**SVM ROC**



**Random forest ROC**



**AUC**

Random forest	79.1%
SVM	69.7%

## References

- Ariffin, Yohan, Jean-Marc Coicaud, and Vesselin Popovski. 2016. *Emotions in International Relations: Beyond Mainstream International Relations*. Cambridge: Cambridge University press.
- Blei, David. 2012. “Probabilistic Topic Models.” *Communications of the ACM* 55 (4): 77–84. doi:10.1145/2133806.2133826.
- Blei, David, Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.
- Böhmelt, Tobias, and Vincenzo Bove. 2014a. “Forecasting Military Expenditure.” Harvard Dataverse. doi:10.7910/DVN/24941.
- . 2014b. “Forecasting Military Expenditure.” *Research and Politics* 1 (1): 1–8.
- Clement, Maeva, and Eric Sangar. 2018. *Methodological Perspectives on the Emotional Turn*. London: Palgrave Macmillan.
- Crawford, Neta C. 2000. “The Passion of World Politics: Propositions on Emotion and Emotional Relationships.” *International Security* 24 (4): 116–56.
- Mearsheimer, John J. 2001. *The Tragedy of Great Power Politics*. New York, New York: WW Norton & Company.
- Mikhaylov, Slava, Alexander Baturo, and Niheer Dasandi. 2017. “United Nations General Debate Corpus.” Harvard Dataverse. doi:10.7910/DVN/0TJX8Y.
- Mohammad, Saif M., and Peter D. Turney. 2013. “Crowdsourcing a Word-Emotion Association Lexicon.” *Computational Intelligence* 29 (3): 436–65.
- Nordhaus, W, JR Oneal, and B Russett. 2012. “The Effects of the International Security Environment on National Military Expenditures: A Multicountry Study.” *International Organization* 66 (3): 491–513.
- Palmer, Glenn, Vito D’Orazio, Michael Kenwick, and Matthew Lane. 2015. “The Mid4 Data Set: Procedures, Coding Rules, and Description.” Conflict Management; Peace Science. <http://cow.dss.ucdavis.edu/data-sets/MIDs>.
- Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, Massachusetts: Addison Wesley.