1 Appendix

1.1 Alliances and War Costs

The DiCE results shown in this paper have been predicted battle deaths for bilateral conflicts. One immediate counterpoint would be that wars are not typically fought between two states. That is, if North Korea and South Korea were to fight a war, we would not expect that conflict to be confined to those two states. We would reasonably expect the United States and China to intervene, which would affect the war costs for both sides. We should therefore expect allies to be an important factor in predicting the costs of war and the effort to predict war costs should be able to account for alliances. Fortunately, the modeling approach I have used is flexible enough to incorporate the features of allies.

Return quickly to the set up of state features described earlier. Consider a scenario where states A and B are in an alliance and fight an alliance of states C and D. I can model each state's battle deaths in the following way:

Modeling Ally Features:
A Battle Deaths = f(A Features, Ally (B) Features, (C+D Features), Dyad Features, Year)
B Battle Deaths = f(A Features, Ally (A) Features, (C+D Features), Dyad Features, Year)
C Battle Deaths = f(A Features, Ally (C) Features, (A+B Features), Dyad Features, Year)
D Battle Deaths = f(C Features, Ally (D) Features, (A+B Features), Dyad Features, Year)

This again poses the question of how to incorporate features from additional states. As evidenced by the prior results, aggregating capabilities provided the best out of sample performance, so I aggregate ally features and incorporate their features directly. Here it is important to note that the original approach I adopt may be failing to account for alliance features. Presumably, if A and B fight alongside each other, their costs of conflict will be affected by each other's presence. That is, a state fighting alongside an ally should reasonably face different costs than if they fought alone. The original set up does not explicitly allow for this by not directly modeling ally features. With this in mind, we might expect incorporating ally features to improve our predictions. This is ultimately an empirical question: does modeling ally features lead to better out of sample predictions?

To answer this, I re ran the same methodology presented so far while including aggregated ally capabilities. I then ensembled these models to produce an another ensemble out of sample prediction for battle deaths. The end result produces very similar to that of the original ensemble (RMSE = 2.818 compared to RMSE 2.795). Some models perform better with the addition of ally features - the neural networks markedly improve over their performance in the original aggregated modeling approach. The tree based models continue to be the best performers, with similar results to that of the original results. At a glance, the results here do not demonstrate the utility of adding more features to accomodate allies. However, for the purpose of predicting multilateral conflicts, this approach may still be preferred for scholars seeking an estimate of costs. As there are too many hpothetical multilateral conflicts to predict them all, for the applied researcher I have written a function which permits the user to estimate the expected casualties for hypothetical pairings using the methodology here, and the function is available with instructions on GitHub. One extension in this area would be to predict alliance fulfillment and add a dimension of uncertainty to whether alliances will come to aid.

	RMSE	SD	Weight
Null	3.816	0.116	0.000
CINC+Year	3.434	0.077	0.000
OLS	3.027	0.134	0.000
PLS	3.016	0.147	0.074
Elastic Net	3.014	0.140	0.000
KNN	3.102	0.090	0.000
Cart	3.189	0.057	0.000
Random Forests	2.843	0.075	0.556
Boosted Trees	2.895	0.081	0.047
Cubist	2.889	0.087	0.198
SVM - Radial	3.000	0.086	0.000
Neural Nets	2.965	0.085	0.125
Ensemble RMSE: 2.818			

Table 1: Test performance of the candidate models having incorporated ally features more directly.

1.2 Political Institutions and War Costs

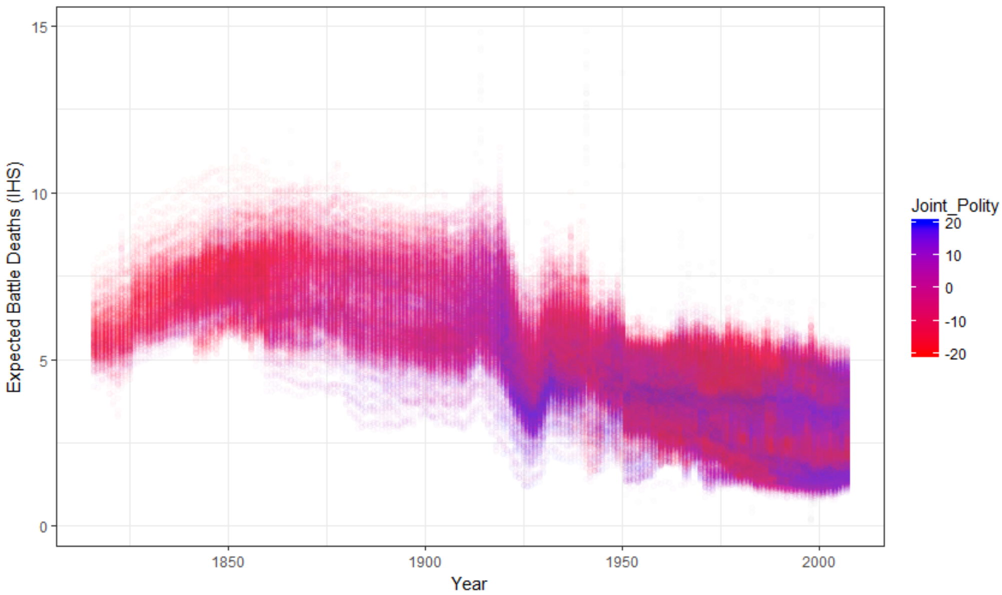
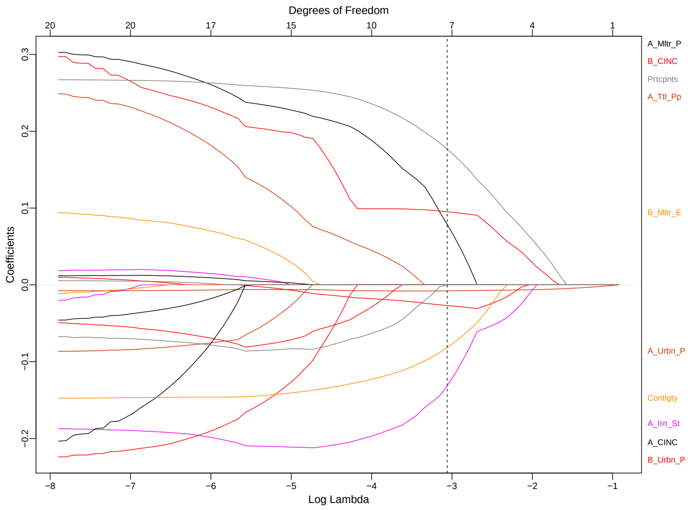


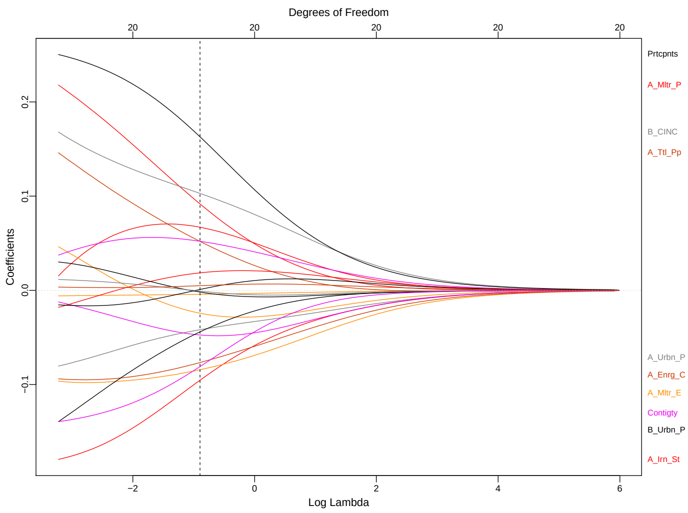
Figure 1: Expected battle deaths by dyad Polity scores

The enterprise of this paper is to provide a measure of war costs with which applied researchers can examine outcomes. Though I have primarily sought to develop the measure in this paper, one immediate topic to examine is the interaction of political institutions and expected war costs. One of the justifications for this enterprise is that of the democratic peace theory. One argument in this realm is that because democracies must win the wars they fight, democracies will fight harder, leading to a much more costly war for their opponent.(BDM et al., 1999) Because a war between two democracies would be prohibitively costly for both parties, we rarely if ever observe wars between democratic states. With this intuition we would expect to see very high expected battle deaths for democratic pairings relative to that of mixed or autocratic pairings. Figure 9 shows the DiCE estimates scaled by the combined polity score of the dyad. The main observation here is that democracies do not exhibit different patterns of war costs than do other country pairings. Indeed, if anything there is a small pattern in the opposite direction, that pairings of democracies have lower expected war costs. A simple t-test comparing the average costs of democratic dyads vs all others lends evidence to this proposition: democratic dyads have lower costs than others. This is in line with the evidence shown prior that institutional variables offered little value to the task of predicting battle deaths. But this result should be interpreted with some care. The costs of conflict produced here are conditional on states having entered into a dispute. That democracies experience lower expected costs might simply indicate that democracies are better at signaling information (cite audience costs literature) and thereby better able to avert costly conflicts.

1.3 Regularization and Feature Selection



(a) Lasso



(b) Ridge

Though I show the results of variable importance plots following (Hill and Jones, 2014) in the paper, I show here the results of a lasso and ridge regression to further identify the subset of predictors which offer improvements in out of sample performance. Using the one standard error rule advocated by (Hastie, Tibshirani and Friedman, 2009), I tuned a lasso and ridge using cross validation to minimize the RMSE. Table 6 shows the coefficients from these models, while FIgure 10 shows the coefficient paths with the a dotted line to indiciate the value of lambda which was selected via cross validation. The lasso shrinks the coefficients of the NMC components for each state, which should be expected as these predictors are highly correlated. Interestingly, the lasso retains different components for states A and B, with iron and steel, military personnel, and Polity 2 remaining for state A and only CINC remaining for side B. Taken as a whole, the relatively low number of predictors in the model does not fully illustrate the utility of penalized methods, but this offers another cut at determining the relevant subset of features used in modeling battle deaths. I place more emphasis on variable importance scores from the random forest and Cubist models largely because these models perform much better in cross validation. The elastic net, which combines the regularization of the ridge and the feature selection of the lasso, achieves largely the same performance as a standard linear model during cross validation, which the tree based models consistently outperform.

	OLS	Lasso	Ridge
(Intercept)	15.207	15.303	7.700
Year	−0.008	−0.008	−0.004
A Iron Steel	−0.185	−0.132	−0.083
A Military Expenditures	−0.042	–	−0.080
A Military Personnel	0.314	0.079	0.077
A Energy Consumption	−0.064	–	−0.072
A Total Population	0.266	–	0.043
A Urban Population	−0.086	–	−0.039
A CINC	−0.240	–	0.063
A Polity2	0.013	−0.027	−0.048
B Iron Steel	−0.054	–	0.020
B Military Expenditures	0.099	–	−0.027
B Military Personnel	−0.038	–	0.049
B Energy Consumption	0.015	–	−0.004
B Total Population	−0.020	–	0.005
B Urban Population	−0.231	–	−0.036
B CINC	0.336	0.095	0.096
B Polity2	0.011	–	−0.003
Contiguity	−0.148	−0.082	−0.069
Participants	0.268	0.177	0.146
ICOW Salience	0.006	–	0.006

Table 2: Coefficient estimates from an OLS compared to shrunken coefficients from the lasso and ridge. Predictors centered and scaled.

1.4 Tuning Parameters

I relied on the caret package (Kuhn, 2008) in R for estimation and cross validation. I present here the tuning parameters for each of the candidate models used in this paper. Though I separately tuned models for the strongest opponent, average, and aggregate models, I list only the aggregate tuning parameters here in order to save space.

1. Partial Least Squares (Wold, 1985)

    - Packages: pls

- Tuning Parameters: components = 10

2. Elastic Net (Zou and Hastie, 2005)
    - Packages: elasticnet
    - Tuning Parameters: fraction = 0.55, lambda = 0
3. k-Nearest Neighbors (Cover and Hart, 1967)
    - Packages: knn
    - Tuning Parameters: k = 11
4. Classification and regression trees (CART) (Breiman et al., 1984)
    - Packages: rpart
    - Tuning Parameters: maxdepth = 7
5. Random forests (Breiman, 2001)
    - Packages: randomForest
    - Tuning Parameters: mtry = 5
6. MARs (Friedman, 1991)
    - Packages: earth
    - Tuning Parameters: n.trees=50. interaction.depth=9, shrinkage=0.1, n.minobsinnode=10
7. Stochasic Gradient Boosted Trees (Friedman, Hastie and Tibshirani, 2001) (Elith, Leathwick and Hastie, 2008)
    - Packages: gbm
    - Tuning Parameters: mstop = 150, maxdepth = 3, nu = 0.1
8. Cubist (Kuhn et al., 2012)
    - Packages: Cubist
    - Tuning Parameters: committees = 20, neighbors = 9
9. Support vector machines with a radial kernel (Scholkopf et al., 1997)
    - Packages: kernlab
    - Tuning Parameters: sigma=0.045, C=1
10. Averaged Neural Networks (Scholkopf et al., 1997)
    - Packages: nnet
    - Tuning Parameters: size = 5, decay = 0.09, bag = T

References

BDM, James D Morrow, Randolph M Siverson and Alastair Smith. 1999. "An institutional explanation of the democratic peace." *American Political Science Review* pp. 791–807.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45(1):5–32.

Breiman, Leo, Jerome Friedman, Charles J Stone and Richard A Olshen. 1984. *Classification and regression trees.* CRC press.

Cover, Thomas and Peter Hart. 1967. "Nearest neighbor pattern classification." *IEEE transactions on information theory* 13(1):21–27.

Elith, Jane, John R Leathwick and Trevor Hastie. 2008. "A working guide to boosted regression trees." *Journal of Animal Ecology* 77(4):802–813.

Friedman, Jerome H. 1991. "Multivariate adaptive regression splines." *The annals of statistics* pp. 1–67.

Friedman, Jerome, Trevor Hastie and Robert Tibshirani. 2001. *The elements of statistical learning.* Vol. 1 Springer series in statistics Springer, Berlin.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. Unsupervised learning. In *The elements of statistical learning.* Springer pp. 485–585.

Hill, Daniel W and Zachary M Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(03):661–687.

Kuhn, Max. 2008. "Caret package." *Journal of Statistical Software* 28(5):1–26.

Kuhn, Max, Steve Weston, Chris Keefer and Nathan Coulter. 2012. "Cubist Models For Regression." *R package Vignette R package version 0.0* 18.

Scholkopf, Bernhard, Kah-Kay Sung, Christopher JC Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio and Vladimir Vapnik. 1997. "Comparing support vector machines with Gaussian kernels to radial basis function classifiers." *IEEE transactions on Signal Processing* 45(11):2758–2765.

Wold, Herman. 1985. "Partial least squares." *Encyclopedia of statistical sciences* .

Zou, Hui and Trevor Hastie. 2005. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(2):301–320.