

Connecting Theory to Policy with Forecasting

Presentation to the Decadal Survey of the Social and Behavioral Sciences

National Academy of Sciences

10/11/2017



Christopher Gelpi
Chair of Peace Studies & Conflict Resolution
Mershon Center for International Security Studies
Professor of Political Science

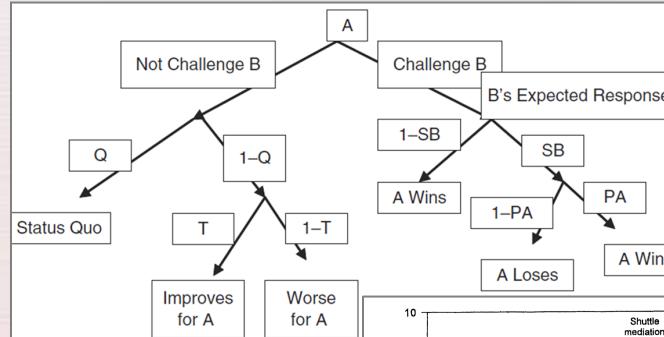
Ways to Use Forecasts of Violence

- Point predictions built from theory
- Evaluate policy relevance of theory



Three Kinds of Forecasts

- Game Theoretic

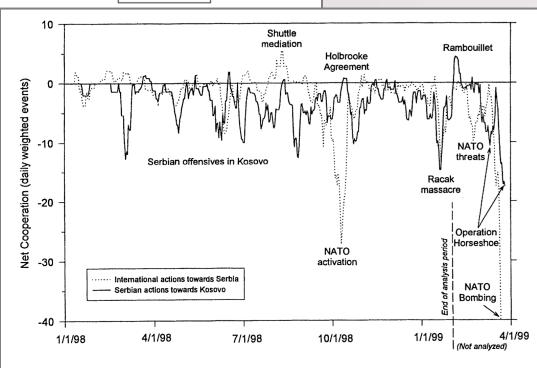


- Time Series

Table 1. Logit model of transnational terrorist incidents, 1968–2007

Polity Score Target	0.031***
Polity Score Origin	-0.024***
Target GDP (logged)	0.352***
Target Logged Population	0.267***
Target Major Power	-0.104
Origin Major Power	0.604***
Colonial Tie	1.248***
Ethnic Tie	0.152
Ethnic \times Post Cold War	-0.241
Ethnic \times Post-9/11	0.034
Dyadic Alliance	0.670***
Alliance \times Post Cold War	0.061
Dyadic Rivalry	0.708***
Post-Cold War Era	-0.577***
Post-9/11 Era	-1.550***
Log of Distance	-0.291***
Peace Years Spline 1	-0.293***
Peace Years Spline 2	2.005***
Peace Years Spline 3	-4.347***
Peace Years Spline 4	3.099***
Constant	-9.833***
Observations	632,058

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



- Structural

Blending Time-Series and Structural Approaches

- Ward et. al (2013) monthly civil war onset, 1997-2011
- Structure, behavior, and country intercepts

FIGURE 1. Separation plots for CRISP Model Prediction of UCDP Data

In-Sample: January 2001–December 2010



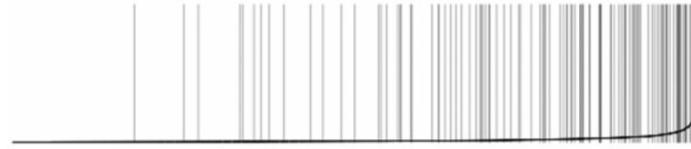
Out-of-Sample: January 2011–December 2011



Building Policy Relevant Structural Models

- Muchlinski et. al. (2016) on annual civil war onset 1945-2000
- Coefficients vs. random forests for rare events

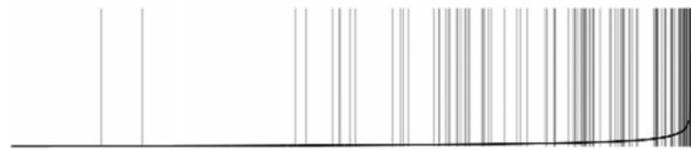
Fearon and Laitin (2003)



Collier and Hoeffler (2004)



Hegre and Sambanis (2006)



Random Forests



Fig. 1 Separation plots for all classifiers.

Policy Relevant Structural Forecasts

- Muchlinski et. al. (2016)
- Random forest solves rare events problem

Table 1 Predicted probability of civil war onset: Logistic Regression and Random Forests

Civil war onset	<i>Models and predicted probability of civil war onset</i>			
	<i>Fearon and Laitin (2003)</i>	<i>Collier and Hoeffler (2004)</i>	<i>Hegre and Sambanis (2006)</i>	<i>Random Forests</i>
Afghanistan 2001	0.01	0.01	0.01	0.09
Angola 2001	0.04	0.01	0.01	0.13
Burundi 2001	0.00	0.00	0.00	0.05
Guinea 2001	0.00	0.00	0.01	0.22
Rwanda 2001	0.02	0.00	0.00	0.56
Uganda 2002	0.03	0.05	0.00	0.81
Liberia 2003	0.01	0.03	0.00	0.94
Iraq 2004	0.04	0.01	0.00	0.68
Uganda 2004	0.02	0.01	0.02	0.52

Identifying Policy Relevance of Theory

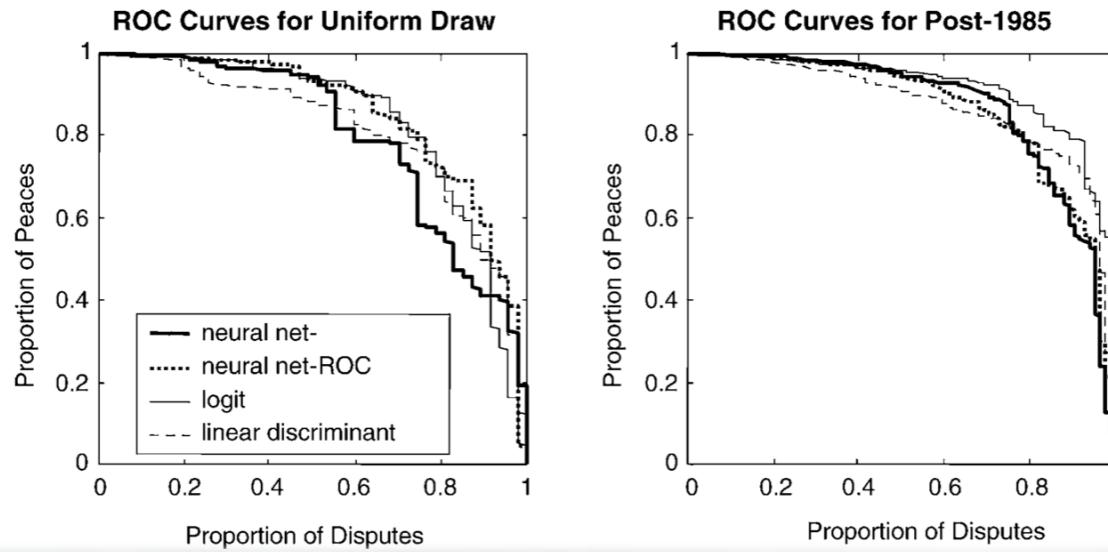
- Over-fitting & excessive complexity
- Over generalization of models
- Substantive significance
- Unrealistic marginal effects



Identifying Excessive Complexity

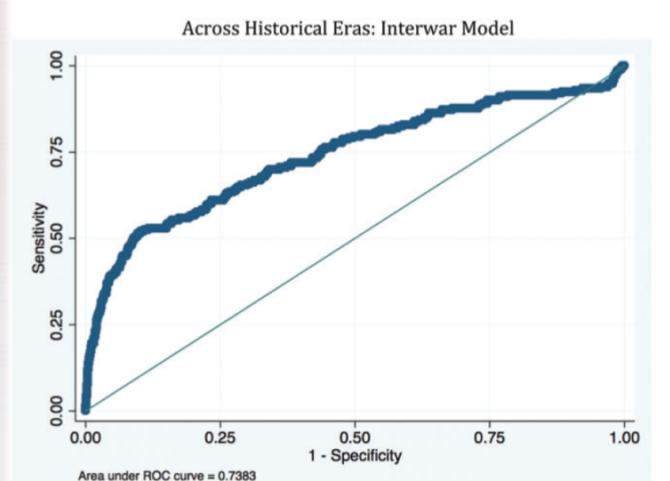
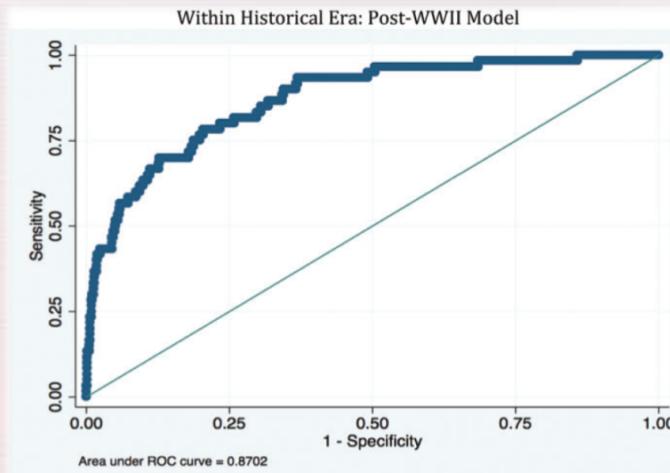
- deMarchi et. al. (2004) MID onset 1945-1989
- Simple logit vs. complex neural network
 - What about a random forest?

FIGURE 1. ROC Curves for Neural Network, Logit, and Linear Discriminant Estimators across Uniform Draw and Post-1985 Test Sets



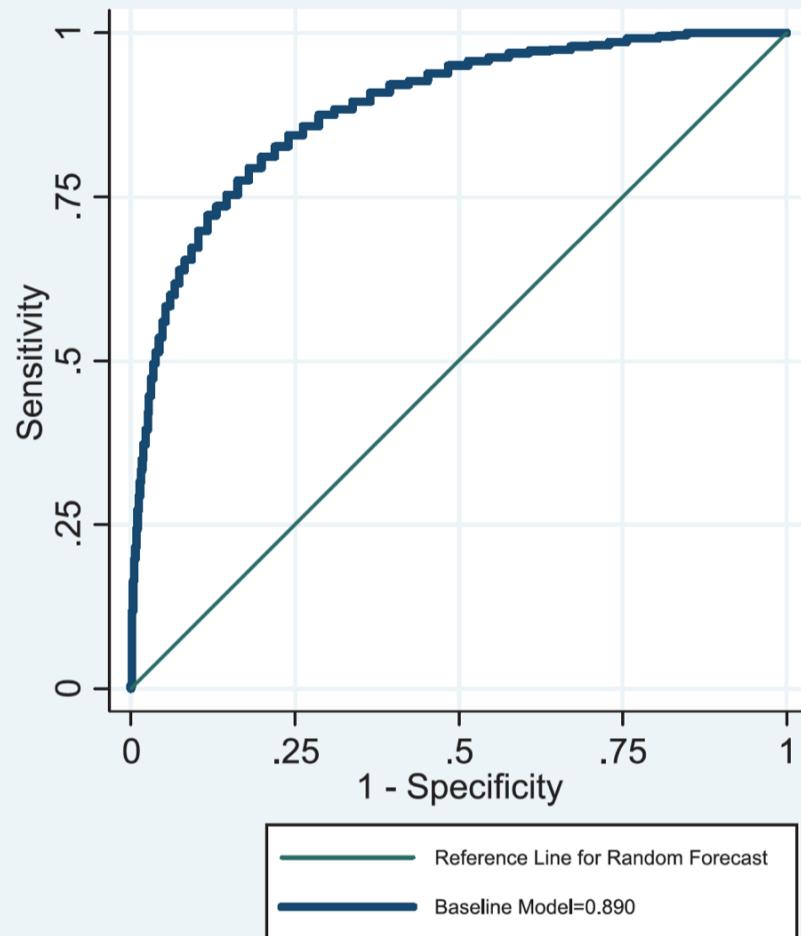
Over Generalization of Effects

- Jenke and Gelpi (2016) MID onset 1945-1992
- Forecasting models trained on 1945-1992 vs. 1914-1945
- AUC 0.87 vs. 0.74



Unrealistic Extrapolation of Effects: Democracy and Terrorism

- Gelpi and Avdan (2015) transnational terrorist attacks 1968-2007
- Baseline model out of sample AUC = 0.89



Democracy, Wealth & Terrorism

- Gelpi and Avdan (2015)
- With democracy AUC = 0.89, without = 0.886

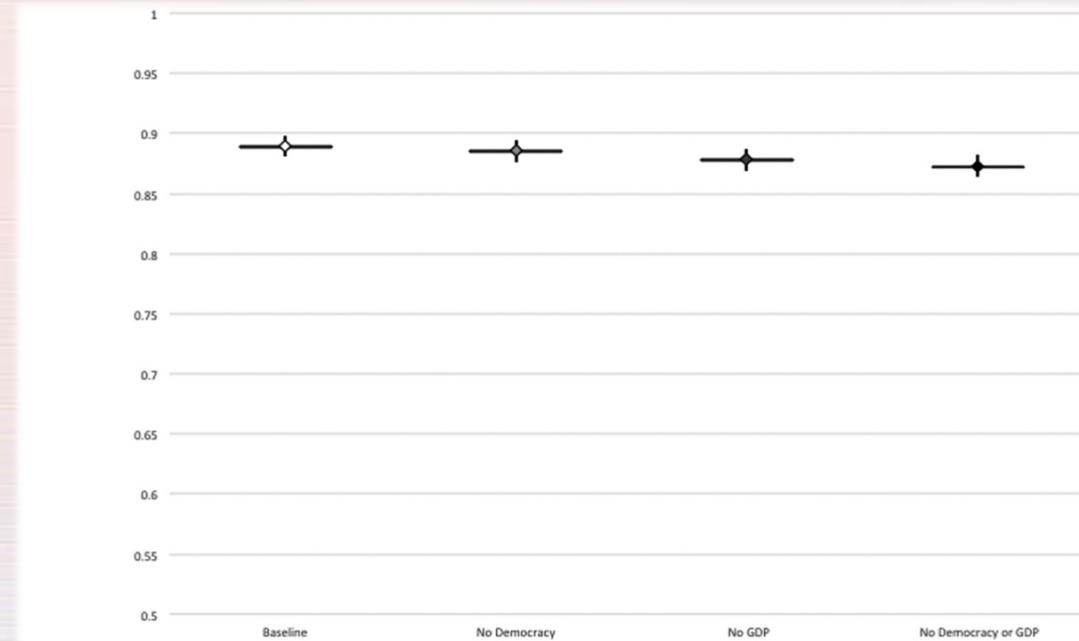


Figure 2. Estimated areas under the ROC curve for models deleting democracy and wealth.

Forecasting Integrates Theory and Policy

- Structural models make relevant predictions
- Forecasting disciplines policy relevance of theory
- Limitation: causal impact
- Complement with causal inference designs



References

- De Marchi, Scott, Christopher Gelpi, and Jeffrey D. Grynaviski. "Untangling neural nets." *American Political Science Review* 98.2 (2004): 371-378.
- Gelpi, Christopher and Nazli Avdan. "Democracies at Risk? A Forecasting Analysis of Regime Type and the Risk of Terrorist Attack." *Conflict Management and Peace Science*. (2015) Online First 11/3/2015
- Jenke, Libby and Christopher Gelpi. "Theme and Variations: Historical Contingencies in the Causal Model of Inter-State Conflict." *Journal of Conflict Resolution*. (2016) Online First 1/6/2016
- Muchlinski, David, et al. "Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data." *Political Analysis* 24.1 (2015): 87-103.
- Ward, Michael D., et al. "Learning from the past and stepping into the future: Toward a new generation of conflict prediction." *International Studies Review* 15.4 (2013): 473-490.