Overview

Do modern political science theories effectively explain the phenomena that are being studied? How can political scientists better assess the validity of their theories? I believe these two questions capture the driving force of today's research agendas. Each researcher wants to understand the phenomena they have dedicated themselves to, and as a result they also want to provide the best evidence that their theory and hypotheses are correct.

This dissertation will focus on proposing a new methodological framework for quantitative analysis in political science. The most common approach in political science and social science at large is null-hypothesis statistical testing (NHST). In this framework a researcher uses the following general structure:

1. Theorize and hypothesize about some phenomena.
2. Establish some a priori heuristic to judge empirical analysis of the theory. Most commonly this is setting an alpha level of .05.
3. Identify relevant data and collect it.
4. Estimate models to assess the relevant null hypotheses (H_0) identified in step one.
5. Using the heuristic established in step two infer what the models say about the validity of the theory and hypothesis of step one.

This framework in theory should provide sound empirical analysis of theories. However, throughout the recent history of political science the relatively simple steps of NHST have been abused. Schrodt (2014) elaborates on many of these abuses, including the use of 'garbage can models' (p. 288) and over-interpretation of coefficients as predictive despite the complete lack of predictive analysis. While his take is notably sharp-toned, the problems of NHST stem from the dependence on p-values as adjudicators of truth. A standard approach is to identify coefficients with p-values below or at the alpha level and use those p-values to validate the hypothetical story behind that variable.

Although contrived, this hypothetical story represents actual commonplace practice, especially outside of top journals. This approach is completely misplaced, and not at all interpreting p-values correctly. There is a long literature on the misuse of p-values and the NHST approach. Gill (1999) shows that one should not and cannot interpret a small p-value as indicating a variable is certainly non-randomly associated with the dependent variable. This conclusion not only misinterprets what the p-value represents, but it also is contingent on the alpha level which is arbitrarily defined.

Other critiques include Levine et al. (2008) who show that the p-value represents $P(H_0|Data)$, the probability that the null hypothesis is true given the observed data, but most quantitative research want to interpret it is $P(Data|H_0)$, the probability of observing this data (and therefore empirical relationship/coefficient) given the null hypothesis. One can only infer about the truth of the null hypothesis given the data at hand, one cannot infer responsibly or accurately about the validity of the alternative hypothesis based on coefficients from p-values.

Apart from these critiques about the foundation of NHST, I argue that this approach offers little to no leverage on actually predicting when phenomena occur. Most articles and books today will take significant coefficients as good 'predictors' of outcomes simply based on the p-value without ever considering how well the actual outcome is explained. A significant coefficient, even a significant and large coefficient, offer no insight into the added value of that variable in predicting the outcome of interest.

My proposed framework insists that research in political science that proports to care about explaining the outcome should assess how well it explains the outcome. One of the most cited terrorism articles in the last decade (225 citations according to Google Scholar 6/26/2019) looks at the effect of political competition on terrorism (Chenoweth 2010). Despite being published in a top journal (Journal of Politics) and being highly cited there is not even a flippant discussion of goodness of fit statistics or in-sample model prediction. Quite simply: one would be suprised to see model fit or prediction discussion in papers that are not primarily methodological.

Any theory that improves understanding of a phenomena, and therefor helps explain the phenomena, should also be helpful in predicting the phenomena. However, in-sample prediction is not particularly useful

for assessing the predictive ability of a theory. My new framework will propose a focus on a minimum of striving for some out-of-sample prediction and a maximum of full-scale cross-validation of models.

I propose that political science begins its transition into a focus on predictive accuracy rather than a focus on statistical significance. This can take several forms. One of the more simple approaches is to use least absolute shrinkage and selection operator (LASSO) regression. The model uses a user-specified parameter to simplify models by progressively setting variables' coefficients to zero if their correlation is sufficiently small that they do not predict the outcome well enough to merit their reduction in parsimony. This approach favors parsimony over all else, and is noteworthy for its ability to identify variables with meaningful correlations that stand up to multicollinearity.

LASSO regression estimated on a full sample provides a direct alternative to OLS, that is more parsimonious. One weakness is that it does not have standard errors and thus does not allow for statistical testing in the same way that NHST does. But, if the final coefficient of a variable is nonzero, then it can be concluded that it does have a stronger association with the dependent variable than other variables. The benefit of a more parsimonious model is that it allows for better prediction and less overfitting, and once this model is extended to cross-validation and out-of-sample prediction its advantages over standard OLS become more apparent.

My framework will outline the steps that should be taken to focus on how well models predict and how to assess the added value of a variable and associated theory/hypothesis to predicting the outcome of interest. I will also propose that research that focuses on subnational events such as terrorism, protests, and repression should focus on analyzing these events at a more local and disaggregated unit of anlaysis.

These settings (terrorism, protest, and repression) will provide an opportunity for me to show my framework in action, and also show how the standard research approach of country-year units of anlaysis drastically underrepresents the nuance of the issues being studied and also likely misrepresents which theories matter as a result. I will use a variety of machine learning approaches to try to predict when and where these events occur, and use these models to assess the theoreitcal literature as an example of my framework in action. Similar examples exist in the terrorism literature but at the country-year level of anlaysis (Basuchoudhary and Bang 2018).

References

Atin Basuchoudhary and James T. Bang. Predicting Terrorism with Machine Learning: Lessons from "Predicting Terrorism: A Machine Learning Approach". *Peace Economics, Peace Science and Public Policy*, 24(4), 2018. ISSN 1554-8597. doi: 10.1515/peps-2018-0040. URL http://www.degruyter.com/view/j/peps.2018.24.issue-4/peps-2018-0040/peps-2018-0040.xml.

Erica Chenoweth. Democratic Competition and Terrorist Activity. *The Journal of Politics*, 72(1):16–30, 2010. ISSN 0022-3816. doi: 10.1017/s0022381609990442. URL http://www.jstor.org/stable/10.1017/s0022381609990442.

Jeff Gill. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly*, 52(3): 647–674, September 1999. ISSN 1065-9129. doi: 10.1177/106591299905200309. URL https://doi.org/10.1177/106591299905200309.

Timothy R. Levine, René Weber, Craig Hullett, Hee Sun Park, and Lisa L. Massi Lindsey. A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research. *Human Communication Research*, 34(2):171–187, April 2008. ISSN 0360-3989, 1468-2958. doi: 10.1111/j.1468-2958.2008.00317.x. URL https://academic.oup.com/hcr/article/34/2/171-187/4210725.

Philip A Schrodt. Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research*, 51(2):287–300, March 2014. ISSN 0022-3433. doi: 10.1177/0022343313499597. URL https://doi.org/10.1177/0022343313499597.