

Statistical Methods for Evidence-Based National Security Policy

Professor Bear F. Braumoeller
Department of Political Science
The Ohio State University
braumoeller.1@osu.edu

Despite the importance of government policy, there are surprisingly few formal studies of the likely causal impact of a given policy prior to the decision to implement it. The *evidence-based policy* movement seeks to fill this need by creating new statistical methods and utilizing existing methods and research designs to draw reliable causal inferences from both experimental and observational data. Unfortunately, this movement has made relatively few inroads into national and international security analysis in the United States. The goal of this line of research would be to develop and apply tools for evidence-based national security policy in the hopes of providing timely, actionable assessments of the likely impact of a policy before it is implemented.

Evidence-Based Policy

Evidence-based policy refers to policy that is informed by rigorous statistical analysis of available evidence. More specifically, evidence-based policy is a form of *causal inference*, a subset of statistical analysis that seeks to make valid causal claims based on either experimental or observational evidence. Unfortunately, it is challenging to speak to the question of group behavior based on individual-level experiments, so national security analysts are typically left with observational data.

Fortunately, the field of causal inference focuses on research designs and statistical estimators that allow researchers to measure causal impact based on observational data. In short, these techniques allow us to approximate the results of a randomized, controlled experiment—the gold standard for measuring causal effects—in non-experimental, naturally occurring settings (see, e.g., [1, 2, 3]).

The goal of evidence-based policy analysis is to provide a neutral assessment of the likely impact of a given policy. The neutrality of the assessment is crucial for its credibility. Perhaps the most widely-known practitioners of evidence-based policy in the American government are the staff of the Congressional Budget Office (CBO), whose dispassionate estimates of the likely economic impact of economic policies are both respected and resented precisely because they are arrived at in a neutral, scientific manner.

At present, despite the inroads that evidence-based policy has made into a variety of fields, including public policy, [4, 5, 6] I am aware of no applications of evidence-based policy to national security policy in the United States. In other words, despite the importance of such questions as whether international institutions and treaties actually change state behavior, whether sanctions are effective in achieving their stated goals,

whether particular counter-insurgent programs have the desired effect, and whether third-party interventions actually improve the outcome of conflicts, the tools of evidence-based policy have yet to be applied to them in a systematic fashion. The development of these tools is particularly important for evaluating complex security situations in which the actions of each party affect both the outcome and the actions of other involved parties. And the potential costliness (in terms of both money and lives) of many national security decisions makes correct assessment of their effects critical.

Existing Solutions and Their Problems

While the majority of scholars in the social sciences are aware of the problem of deriving valid causal inferences from observational data (i.e., endogeneity bias), at least in principle, most seem to believe that it can be taken care of simply by adding observed confounding variables to the right-hand side of a linear, logit, or probit regression. Unfortunately, doing so generally does not resolve the problem. While it is technically not impossible to address endogeneity bias in this manner, these techniques assume both that all possible confounders have been measured and included in the equation *and* that the functional form of their relationship to the outcome of interest has been correctly specified—assumptions that will almost certainly not be met in practice.

For many years, the standard method in economics for deriving valid causal inferences from observational data has been to utilize an *instrumental variable*. Instrumental variables are variables that have an impact on the outcome under study only by virtue of their impact on the policy that may or may not have influenced it. To take a simple example, although it might be difficult to gauge the impact of exercise on health in purely observational data because poor health can make it more difficult to exercise, an employer program that incentivizes exercise could serve as an instrument for exercise because it is not conditional on the health of the employee and could only have an impact on health via the exercise that it promotes. While instruments have a long history in the social and behavioral sciences, more recent research has emphasized that they can be highly problematic if they are invalid, in that they are correlated with the error term in the outcome equation, or if they are only weakly correlated with the endogenous (policy) variable [7].

Matching methods are a compelling way to cut through the thicket of problems surrounding instrumental variable approaches. They are not without problems of their own, however. Foremost among them is the assumption that all confounders have been measured and incorporated into the analysis—an assumption known as “selection on observables.” To the extent that this assumption has been recognized as being potentially problematic, the solution is typically to measure unmeasured confounders and include them in the analysis [8, 9]. As one prominent student of causal inference notes in a recent review, however, “selection on observables is a very strong assumption. It is often difficult to imagine that selection on observables is plausible in many contexts.” [10]

A more promising approach is the use of *simultaneous likelihood methods*—in particular, multiple equation probit models with endogenous dummy regressors, also known as recursive models. These models address endogeneity directly by estimating the coefficients in two (or more) equations simultaneously. They allow one to capture the impact of unobserved confounders by modeling the correlation between the error terms of the equations. Unfortunately, such methods rely heavily on distributional and functional form assumptions and can be very sensitive to violations of those assumptions [11, 12, 8]. Given the number of distributional assumptions in standard simultaneous likelihood methods, this sensitivity can be nontrivial. Mismodeled dependencies that appear, for instance, in the tails of the distribution linking the two equations

(that a linear measure of association can not fully capture), undetected nonlinear covariate-response relationships and mismodeled probabilities related to the outcomes of two equations can have severe consequences for parameter estimation [13, 14, 15, 16, 17].

Finally, *natural experiments*—sets of cases in which the only causally relevant difference is the policy, or “treatment”—can provide valid causal inferences if the implementation of the policy is effectively random. Fortunately for citizens but unfortunately for researchers, national security policy is rarely randomized.

It seems extremely unlikely that there will ever be a perfect method of deriving causal inferences from observational data. At the same time, there has been very significant progress in the area of causal inference in the last 20 years. Spurring the development of even better methods and applying them to national security analysis could provide a substantially better quality of information for policy makers facing difficult policy choices.

Goals

The first goal of research in this area should be to import existing statistical tools and develop and refine new statistical ones for use in evidence-based national security analysis. Such tools should meet the following desiderata:

- They should be well-designed for policies that are endogenous—that is, those policies that are prompted by the same conditions that they hope to ameliorate (e.g. food aid, peacekeeping, etc.) The impact of endogenous policies is especially hard to assess because the conditions that prompt the policy *confound*, or distort, our estimate of the causal impact of the policy. Peacekeeping missions can look very ineffective, for example, unless we account for the fact that peacekeepers are only sent to places where they are needed, and those places are far less likely to remain peaceful to begin with.
- Given the complexity of world politics and the practical challenges inherent in obtaining data in a timely manner for analysis, they should ideally not require that all possible confounding variables be measured and included in the analysis.
- They should be as precise and accurate as possible in realistic benchmark tests while relying as little as possible on fragile and restrictive assumptions.

Put more succinctly, while it is generally impossible to estimate causal effects with any meaningful degree of confidence without making *some* assumptions [18], the goal would be to make the fewest and most reasonable assumptions possible.

The second goal of research in this area should be to estimate the causal impact of foreign and defense policies of interest to policy makers. Examples might include:

- the impact of policies designed to limit the severity of terror attacks
- the impact of service provision on the success of insurgencies
- the impact of third-party intervention on state stability
- the impact of treaties banning or restricting the use of certain categories of weapons on state behavior, with an eye toward likely impact of treaties on unmanned weapons and cyberweapons

Ideally, an effort of this nature would be administered by a national-level funding organization such as the National Science Foundation and executed at both publicly and privately funded research centers. Priority should ideally be given to interdisciplinary research teams that can credibly produce research that is of more value than the sum of its parts, but funding should not exclude individual scholars or intradisciplinary teams that can produce deeper and more focused contributions.

Summary

Measuring the causal impact of state policy on outcomes of interest is one of the biggest methodological challenges in the social and behavioral sciences, for two reasons: behavior is generally endogenous, and the threat of unobserved variables that confound the relationship between behavior and outcomes is pervasive. As a result, policy is too often based on evidence that is either anecdotal or descriptive rather than causal in nature.

The evidence-based policy movement is a concerted effort to rectify this problem. Existing tools for causal inference in observational settings, while a vast improvement over simple correlational methods, are not without problems of their own: some are ill-suited to inference in the presence of unobserved confounders, which abound in national security settings and cannot typically be measured in a timeframe that is suited to policy analysis, while others rely on functional form assumptions that can produce substantial bias in estimates of average treatment effects.

Simply importing existing tools to produce assessments of likely policy impact would be a valuable contribution to national security. Developing newer and better ones that provide valid causal inferences under a wider range of challenging real-world situations would be an even greater contribution. The foundations of this research already exist and attempts to improve on it are ongoing; promoting and incentivizing such attempts with an eye toward applications in national security would yield a substantial payoff.

References

- [1] R. J. LaLonde. Evaluating the Econometric Evaluations of Training Programs. *The American Economic Review*, 76(4):604–620, 1986.
- [2] J. D. Angrist and A. B. Krueger. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106:979–1014, 1991.
- [3] J. Lyall. Does Indiscriminate Violence Incite Insurgent Attacks?: Evidence from Chechnya. *Journal of Conflict Resolution*, 53(3):331–362, June 2009.
- [4] K. Bogenschneider and T. Corbett. *Evidence-based policymaking: insights from policy-minded researchers and research-minded policymakers*. Routledge Academic, New York, 2010.
- [5] N. Cartwright and J. Hardie. *Evidence-based policy: a practical guide to doing it better*. Oxford Univ. Press, Oxford, 2012.
- [6] C. F. Manski. *Public policy in an uncertain world: analysis and decisions*. Harvard Univ. Press, Cambridge, Mass., 2013.
- [7] M. P. Murray. Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives*, 20(4):111–132, 2006.
- [8] B. A. Simmons and D. J. Hopkins. The Constraining Power of International Treaties: Theory and Methods. *American Political Science Review*, 99(04), November 2005.
- [9] Y. Lupu. The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects. *American Journal of Political Science*, pages 912–925, 2013.
- [10] L. Keele. The Statistics of Causal Inference: A View from Political Methodology. *Political Analysis*, 23(3):313–335, July 2015.
- [11] C. Winship and R. D. Mare. Models for Sample Selection Bias. *Annual Review of Sociology*, 18:327–350, 1992.
- [12] A. E. Sartori. An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions. *Political Analysis*, 11(2):111–138, 2003.
- [13] S. Chib and E. Greenberg. Semiparametric modeling and estimation of instrumental variable models. *Journal of Computational and Graphical Statistics*, 16:86–114, 2007.
- [14] R. J. A. Little. A note about models for selectivity bias. *Econometrica*, 53:1469–1474, 1985.
- [15] C. Monfardini and R. Radice. Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study. *Oxford Bulletin of Economics and Statistics*, 70:271–282, 2008.
- [16] G. Marra and R. Radice. Estimation of a semiparametric recursive bivariate probit model in the presence of endogeneity. *Canadian Journal of Statistics*, 39:259–279, 2011.
- [17] G. Marra and R. Radice. Flexible Bivariate Binary Models for Estimating the Efficacy of Phototherapy for Newborns with Jaundice. *International Journal of Statistics and Probability*, 4:46–58, 2015.

- [18] C. F. Manski. Nonparametric Bounds on Treatment Effects. *The American Economic Review*, pages 319–323, 1990.