Introduction to the Virtual Issue:
Past and Future Research Agenda on Causal Inference

Kosuke Imai

Department of Politics, Princeton University, Princeton NJ 08544
email: kimai@princeton.edu

If you ask political scientists whether a major goal of their empirical research is to test causal relationships, I suspect that most of them answer "yes." In contrast, many statisticians would say that their methods may improve predictions but have little to do with causal inference. As a result, even as an increasing number of social scientists use statistical methods to infer causal effects in their empirical research, many of these methods were not originally designed for causal inference.

This intellectual gap has been rapidly filled by two "revolutions." In the community of empirical researchers, the *identification revolution* has occurred where the key assumptions for identifying causal quantities are taken much more seriously than before. Problems related to measurement and selection bias are no longer brushed under the table. The discipline has recently witnessed a rise in the use of randomized experiments, and an increasing number of political scientists now choose research designs to improve the credibility of identification assumptions (for example, see the 2009 *Political Analysis* special issue on "Natural Experiments in Political Science").

Simultaneously, among methodologists in various disciplines, the *potential outcomes revolution* has taken place where traditional statistical methods are reframed within the formal framework of causal inference. The potential outcomes framework, which originates in the classic works of Neyman and Rubin, makes explicit the fundamental problem of causal inference: to infer causal effects one must estimate counterfactual outcomes from observed data. Along with the development of new methodologies, this revolution has led to re-interpretations of familiar statistical methods such as regression and instrumental variables methods, and clarified the exact conditions under which they can be used to infer causal relationships from observed data.

Many political methodologists contributed to these new developments. Here, I selected six articles published in *Political Analysis* between 2005 and 2010 that represent the leadership taken by political methodologists to contribute to advancement of causal inference. These articles have influenced the way in which political scientists and other social scientists conduct their empirical research, and three of them won the Warren Miller best paper award for 2005, 2007, and 2009. After briefly discussing each article with references to other relevant *Political Analysis* articles, I identify further research challenges where unresolved and yet important methodological problems await the contributions from political methodologists.

1. Articles in this Virtual Issue

Over the last decade, an increasing number of political scientists have used field experiments, and this led to methodological work on how to efficiently design and analyze such experiments. **David Nickerson**'s (2005, Vol 13(3), pp. 233–252) work shows how to efficiently design field experiments in the presence of noncompliance by adjusting the treatment/control ratio, matching subjects prior to random assignment of treatment, and using placebos in the control group. He derives explicit analytical expressions for efficiency gains of these designs and illustrates their use by applying them to get-out-the-vote field experiments. Field experiments differ from laboratory experiments because of the potential deviations from standard experimental protocols such as noncompliance and attrition. In addition, field experiments can be costly, which necessitates efficient experimental designs. Nickerson's work offers a promising first step towards methodological sophistication when designing field experiments and is being followed up by others (e.g., Gerber *et al.*, 2010).

In addition to field experiments, the availability of inexpensive online surveys led to an increasing use of survey experiments where respondents are exposed to randomly selected survey questions, images, or even audio and video clips. **Daniel Corstange**'s (2009, Vol. 17(1), pp. 45–63) article demonstrates how to analyze list experiments, which is an increasingly popular survey methodology to elicit truthful responses to socially sensitive questions such as racial prejudice. While many applied researchers used a simple difference-in-means estimator, Corstange shows how to build a multivariate regression model so that researchers can infer not only the proportion of those who have racial prejudice but also what respondent characteristics are associated with racial prejudice. Following his work, a number of new statistical methods are being developed for various survey experiment techniques including randomized response and endorsement experiments (e.g., Gingerich, 2010; Bullock *et al.*, 2011; Imai, 2011).

For most political science questions of interest, conducting randomized experiments is simply not an option and researchers must overcome the methodological challenges of inferring causal relationships from observational data. As a research design strategy to minimize selection bias, researchers often use natural experiments in which they find instances where the treatment variables of interest can be plausibly assumed to be randomly assigned. While such research design strategies can increase internal validity of findings, their results may not generalize to a greater population of interest. Thus, one important methodological challenge is the question of how and when one can improve external validity of empirical findings from natural experiments.

Donald Green, Terence Leong, Holger Kern, and Alan Gerber (2009, Vol. 17(4), pp. 400–417) address this point by examining the external validity of regression discontinuity design, which is a popular natural experiment research design strategy. Using the results from a randomized experiment as a benchmark, they show how the estimates from the regression discontinuity design can be sensitive to model specification and bandwidth selection. Following their lead, others have also shown that care must be taken when generalizing the results from natural experiments (Dunning, 2008; Caughey and Sekhon, 2011). As such research designs and statistical methods become more popular, their misuse may also become more common. An important role of political methodologists is to clarify proper ways in which these powerful methodological tools can be used to draw causal conclusions.

One major methodological challenge of observational studies is the issue of model dependence. Sensitivity to minor changes in model specification can significantly decrease the credibility of empirical findings. Randomized experiments reduce model dependence because randomization of the treatment guarantees the similarity of treatment and control groups. In this case, adjusting for covariates makes little difference for estimating treatment effects. **Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart** (2007, Vol. 15(3), pp. 199–236) show that matching methods can reduce model dependence in observational studies by making the distribution of observed covariates similar between the treatment and control groups. By preprocessing the data via matching, researchers can make their empirical findings robust to model misspecification. In their proposed framework, matching methods are not the replacement of traditional regression modeling. Rather, they improve the performance of regression models by making empirical results less dependent on model specification.

As another set of methods to achieve covariate balance between treatment and control groups, **Adam Glynn and Kevin Quinn** (2010, Vol. 18(1), 36–56) introduce weighting techniques that have become part of standard methodological toolkit in biostatistics and epidemiology. In some sense, weighting methods can be seen as a generalization of matching methods and they can be effective even in the situations where matching methods fail to achieve covariate balance. As explained by the authors, weighting can also make regression models robust to possible misspecification. This article will facilitate their use in political science.

Finally, in observational studies, a major source of model dependence may come from a particular form of covariate imbalance called extrapolation, which results from the fact that some treated units simply lack comparable control units. **Gary King and Langche Zeng** (2006, Vol. 14(2), pp. 131–159) propose a statistical method to diagnose the possible existence of extrapolation in a high-dimensional covariate space. The method highlights the fundamental difficulty of covariate adjustment in observational studies due to the curse of dimensionality – in a high dimensional covariate space, data points are much further away from each other. The dilemma is that while researchers wish to adjust for many covariates in order to reduce selection bias, this often leads to a greater degree of model dependence in the process of covariate adjustments.

2. Future Research Agenda

The articles in this virtual issue have made significant contributions to the methodological literature on causal inference. But at the same time, there remain a large number of important, unresolved issues. Below, I identify

four such challenges and briefly refer to ongoing work by political methodologists, some of which has appeared this year or is forthcoming in the pages of *Political Analysis*.

First, matching and weighting methods need to be improved so that they do not require informal and tedious process of balance checking by data analysts. Specifically, a measure of covariate balance must be explicitly determined a priori and a matching/weighting method needs to optimize this metric. Several political methodologists have already developed such matching methods, which I believe will significantly improve empirical research (e.g., Diamond and Sekhon, 2011; Hainmueller, 2011; Iacus *et al.*, 2011).

Second, while much existing research focuses on the estimation of average treatment effects, treatments typically do not affect different units in the same manner. Efficient estimation of heterogeneous treatment effects, therefore, is essential for testing scientific theories, understanding how treatments operate, and generalizing the results from experimental sample to a target population. It is well known that post-hoc subgroup analysis can yield false discovery of subpopulation-specific treatment effects. Therefore, we need principled ways to identify units who receive the most benefit (or harm) from treatment. New statistical techniques are being developed to address this possibility (e.g., Green and Kern, 2010; Hartman *et al.*, 2010; Imai and Strauss, 2011).

Third, one of the common criticisms about randomized experiments and statistical methods in general is that they can show *whether* or not treatments causally affect outcomes but fail to ascertain *how* and *why* such causal effects arise. New statistical methods and research design strategies must be developed in order to effectively identify causal mechanisms that underlie causal effects of interest. This is an active area of interdisciplinary research that has also been of interest to political methodologists (e.g., Bullock *et al.*, 2010; Glynn and Quinn, 2011; Imai *et al.*, 2011).

Finally, a majority of existing statistical methods are only applicable in cross-section settings, and yet political science data often have time-series or time-series cross-section data structures. Methodological research has begun to address these extensions (e.g., Abadie *et al.*, 2010), and much work remains to be done. For example, carry-over and spill-over effects can further complicate causal analysis in these settings. Where current methodological approaches often rely on the assumption of independence, many treatments exercise influence into the future or across multiple individuals/countries.

3. Concluding Remarks

Political scientists joined the identification and potential outcomes revolutions more than a decade after these movements took place in other disciplines. Indeed, methodological articles that utilize the modern framework of causal inference only appeared in political science journals within the last several years. And yet, as the articles of this virtual issue demonstrate, political methodologists have begun to make serious contributions to the interdisciplinary literature on causal inference methodology.

Funding

About the Author

Kosuke Imai is an Assistant Professor of Politics at Princeton University. He has worked on various causal inference topics including causal mediation analysis, propensity scores, matching, and survey experiments. His published and working papers are available at http://imai.princeton.edu

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American Statistical Association* **105**, 490, 493–505.

Bullock, J. G., Green, D. P., and Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology* **98**, 4, 550–558.

Bullock, W., Imai, K., and Shapiro, J. N. (2011). Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan. *Political Analysis* Forthcoming.

Caughey, D. and Sekhon, J. (2011). Elections and the regression-discontinuity design: Lessons from close U.S. House races, 1942–2008. *Political Analysis* Forthcoming.

Corstange, D. (2009). Sensitive questions, truthful answers?: Modeling the list experiment with LISTIT. *Political Analysis* **17**, 1, 45–63.

Diamond, A. and Sekhon, J. (2011). Genetic matching for estimating causal effects: A new method of achieving balance in observational studies. Working Paper, Department of Political Science, University of California, Berkeley.

Dunning, T. (2008). Model specification in instrumental variables regression. *Political Analysis* **16**, 3, 290–302.

Gerber, A. S., Green, D. P., Kaplan, E. H., and Kern, H. L. (2010). Baseline, placebo, and treatment: Efficient estimation for three-group experiments. *Political Analysis* **18**, 3, 297–315.

Gingerich, D. W. (2010). Understanding off-the-books politics: Conducting inference on the determinants of sensitive behavior with randomized response surveys. *Political Analysis* **18**, 3, 349–380.

Glynn, A. N. and Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis* **18**, 1, 36–56.

Glynn, A. N. and Quinn, K. M. (2011). Why process matters for causal inference. *Political Analysis* **19**, 3, 273–286.

Green, D. P. and Kern, H. L. (2010). Detecting heterogenous treatment effects in large-scale experiments using Bayesian additive regression trees. *The Annual Summer Meeting of the Society of Political Methodology, University of Iowa* .

Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., and Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis* **17**, 4, 400–417.

Hainmueller, J. (2011). Entropy balancing for causal effects: Multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* Forthcoming.

Hartman, E., Grieve, R., and Sekhon, J. S. (2010). From SATE to PATT: The essential role of placebo test combining experimental and observational studies. *The Annual Meeting of the American Political Science Association, Washington D.C.* .

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* **15**, 3, 199–236.

Iacus, S. M., King, G., and Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. *Political Analysis* Forthcoming.

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* **106**, 494, 407–416.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* **105**, 4, Forthcoming.

Imai, K. and Strauss, A. (2011). Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. *Political Analysis* **19**, 1, 1–19.

King, G. and Zeng, L. (2006). The danger of extreme counterfactuals. *Political Analysis* **14**, 2, 131–159.

Nickerson, D. W. (2005). Scalable protocols offer efficient design for field experiments. *Political Analysis* **13**, 3, 233–252.