

MPMC Proposal: The Compatibility of Causal Inference and Bayesian Analysis in Political Science

Michael G. DeCrescenzo^{*}

Submitted January 18, 2019

Introductions to Bayesian inference often use experiments to exemplify the intuition of Bayesian updating.¹ These narratives often go something like...

Suppose that a research literature has some hypotheses about the causal effect of X on Y . Researchers conduct experiments, each time updating the information they have about their hypotheses, conditional on the experiments they have conducted to date.

Despite the regularity with which professors tell students this story, the discipline has not internalized the lesson. Studies that employ *both* a causal inference framework and Bayesian estimation are exceedingly rare in political science. Why is this the case, and should it be changed? In this proposal, I outline an essay that discusses points of contention between these two methodological camps, makes an argument for viewing causal inference and Bayesian analysis as having fundamentally compatible goals, and exemplifies the benefits of combining both methodological traditions in practical examples.

Throughout this proposal, I refer to “causal inference” as research designs derived from explicit causal models and assumptions that allow the estimation of causal parameters from data (which include experiments as well as difference-in-differences, instrumental variables, synthetic control, matching, regression discontinuity, and others). I refer to *Bayesian analysis* as statistical models that include models for unobserved parameters (prior distributions) alongside models for observed data (likelihood functions). I refer to “causal inference with Bayesian estimation” (CIBE for short) as causal inference designs whose parameters are estimated using priors. In a Bayesian causal model (Rubin 1978; Imbens and Rubin 1997), an observed outcome is compared to a posterior distribution of potential outcomes that reflects joint uncertainty over all model parameters. In an applied setting, it should be sufficient to think about a modeling scenario where a causal effect β is given a prior distribution $\beta \sim \mathcal{D}(\cdot)$.

^{*}Ph.D. Candidate, Political Science, University of Wisconsin–Madison. Thanks to William T. Christiansen, Andrew Heiss, Anna Meier, Laura Meinzen-Dick, Daniel Putman, and Zach Warner for helpful feedback.

¹The “Eight Schools” SAT coaching experiment and meta-analyses of experiments are conventional examples in many introductions to Bayesian hierarchical modeling (see e.g. Gelman et al. 2013).

Compatible Goals

The main argument of the paper is that causal inference and Bayesian estimation, despite their recent siloing in political science, are compatible with one another because they have the same underlying goal: getting the best possible parameter estimates. They pursue this goal using different tactics—causal inference identifies causal parameters by “de-biasing” the research design, while Bayesian analysis improves parameter estimates by incorporating prior information—but the tactics themselves are subordinate to the broader goal.

What would be the advantages of CIBE? How can we do better by combining Bayesian analysis into an already-strong causal inference design? There are a number of ways to answer this question. First, there are practical benefits. The data used in causal inference designs often contain features for which Bayesian models provide practical and intuitive solutions. Multilevel models are an efficient tool to account for clustered randomization, repeated within-subject measures, and other sources of correlated error across observations (see e.g. Green and Vavreck 2008) by modeling multiple sources of variation directly. Because Bayesian models estimate the distribution of all parameters simultaneously, they can naturally proliferate uncertainty through multi-equation modeling routines such as sequential g-estimation, instrumental variables, and other multi-equation designs for issues such as treatment non-compliance (Acharya, Blackwell, and Sen 2016; Horiuchi, Imai, and Taniguchi 2007). Hierarchical models allow researchers to pool information across similar treatments, which is especially applicable to the analysis of conjoint experiments. This proliferated uncertainty also means that multiple hypothesis testing is second nature, since the analyst can calculate any quantity they desire using the MCMC samples for an arbitrary number of parameters.

Another way to phrase the benefits of CIBE is philosophical. What *is* the treatment effect? In Bayesian analysis, the posterior distribution directly characterizes the plausibility of treatment effects given the data. This allows researchers to say (on firm philosophical ground) that the treatment effect probably falls within some specified range. Being able to make precise statements about well-identified treatment effects is the goal of causal inference, and CIBE facilitates that goal.

Sources of Tension

Although there are advantages of CIBE, scholars in each camp have distinct norms of analysis that would need to be adjudicated in order to make space for CIBE. Primarily, the objective in causal inference to de-bias the research design extends into other areas of research decision-making; causal inference researchers tend also to prefer unbiased estimators, reflecting their desire to purge their work of any researcher intervention that does not “let the data speak.” I describe this ethic in causal inference as a “conservative” preference for simple models and minimal assumptions. Researchers want to be led by the data, not by the model. This fosters a subtle skepticism that Bayesian analysis

is used to *defy* data—if the researcher employs enough subjective assumptions, they can conclude anything.

Is there a way of doing Bayesian analysis that fits the conservative norms of causal inference work without sacrificing the advantages of the Bayesian model? The paper focuses on *weakly informative priors* (WIPs) and *regularizing priors* as two practices that enable researchers to engage in CIBE in harmony with the norms of causal inference. Gelman describes a prior as weakly informative if it is “proper but is set up so that the information it does provide is intentionally weaker than whatever actual prior knowledge is available” (2006: p. 517). WIPs are informative enough to rule out estimates that make no sense for the problem at hand (e.g. that exceed the range of your dependent variable), but not so informative that the data can’t speak. Regularizing priors, on the other hand, are designed to prevent overfitting by shrinking effects toward zero. These priors introduce a little more bias into the analysis, but the bias is unmistakably conservative. Rhetorically, the purpose of a regularizing prior is to ensure that whatever effects the researcher finds are of a high enough statistical power to overcome a conservative prior.

Proposed Examples

I propose to walk through two examples of CIBE. First is a re-analysis of a regression discontinuity design by Hall (2015), using WIPs to show how weak but sensible prior information can contain and stabilize causal effect estimates. Second will be a simulation of data from a conjoint design, which demonstrates how multilevel modeling can be used to account for multiple sources of variance, pool information from similar treatments, and increase the generalizability of causal effect estimates when marginalizing over independent treatment dimensions.

What the Paper is not

The purpose of this essay is *not* to convince all practitioners of causal inference methods to use Bayesian methods, nor is it to convince anybody of the “superiority” of Bayesian methods. In fact, the paper recognizes and celebrates the skepticism that causal inference holds toward *any* method that infringes on the methodological conservatism and importance of letting data speak. Rather, the purpose of the paper is to make some space for Bayesian methods in causal inference work. These two camps can make each other better, but they need to talk to one another. Bayesian methods are not inconsistent with the goals, nor are they necessarily inconsistent with the informal norms, of causal inference practitioners. But there does need to be some adjudication of potential disagreements and areas of misunderstanding in order for each camp to benefit from the strengths of the other.

References

- Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110 (3): 512–529.
- Gelman, Andrew. 2006. "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)." *Bayesian analysis* 1 (3): 515–534.
- Gelman, Andrew, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian data analysis*. Chapman and Hall/CRC.
- Green, Donald P, and Lynn Vavreck. 2008. "Analysis of cluster-randomized experiments: A comparison of alternative estimation approaches." *Political Analysis* 16 (2): 138–152.
- Hall, Andrew B. 2015. "What happens when extremists win primaries?" *American Political Science Review* 109 (01): 18–42.
- Horiuchi, Yusaku, Kosuke Imai, and Naoko Taniguchi. 2007. "Designing and analyzing randomized experiments: Application to a Japanese election survey experiment." *American Journal of Political Science* 51 (3): 669–687.
- Imbens, Guido W, and Donald B Rubin. 1997. "Bayesian inference for causal effects in randomized experiments with noncompliance." *The annals of statistics*: 305–327.
- Rubin, Donald B. 1978. "Bayesian inference for causal effects: The role of randomization." *The Annals of statistics*: 34–58.