Flexibly Detecting Effect Heterogeneity with an Application to the Effects of College on Reducing Poverty

Jiahui Xu [*] Jennie E. Brand [†] Tanvi Shinkre [‡] Nanum Jeon [§]

September, 2024

Abstract

Units of analysis in social research do not respond uniformly to events and interventions. Yet it is not always clear which axes of heterogeneity are most important to consider before data analysis. We use causal forests to nonparametrically uncover heterogeneous treatment effects. We then adapt causal forests and advance causal mediation forests to assess heterogeneous direct and indirect effects. This novel adaptation explores heterogeneity in the causal paths linking a treatment to an outcome through a binary, multinomial, or continuous mediator. Both causal forests and causal mediation forests robustly adjust for high-dimensional confounders, yielding asymptotically normal and $n^{1/2}$ consistent estimates. We show that forest-based approaches often outperform alternative methods in identifying effect heterogeneity. We apply the forest-based methods to study the heterogeneous effects of four-year college on reducing poverty with data from the

[*]Ph.D. Candidate in Sociology and Demography, the Pennsylvania State University, jiahuixu@psu.edu

[†]Professor of Sociology and Statistics & Data Science, UCLA; Co-Director, Center for Social Statistics, brand@soc.ucla.edu

[‡]Ph.D. Candidate in Statistics & Data Science, UCLA, tanvishinkre@g.ucla.edu

[§]Ph.D. Candidate in Sociology and M.S. Student in Statistics & Data Science, UCLA, njeon@g.ucla.edu

National Longitudinal Survey of Youth 1979 and 1997 cohorts and find large gains for disadvantaged youth.

Keywords: heterogeneity, moderated mediation, causal inference, causal mediation, machine learning, causal forests, causal mediation forests, generalized random forests, college education, poverty

1 Introduction

Effect heterogeneity is the norm. Units do not respond identically to different events, circumstances, or interventions (i.e., "treatments"). Effectively and flexibly uncovering heterogeneous returns to treatments across population subgroups deepens our scientific understanding of social processes (Athey and Imbens 2019; Brand et al. 2023; Manski and Garfinkel 1992; Xie et al. 2012; Zhou and Xie 2019, 2020). Response variation is also of significant interest to policymakers who seek to learn an optimal policy and develop effective interventions satisfying practical constraints (Athey and Wager 2021; Oprescu et al. 2019). Toward the same goals, it is useful to investigate the causal pathways linking a treatment to an outcome via a downstream intermediate factor, which occurs after the treatment and causally affects the outcome. Decomposing total effects into constituent direct and indirect effects via mediators provides sharper and clearer insights into the mechanisms underlying effect heterogeneity (Ting and Linero 2023). Sometimes referred to as *moderated mediation*, investigating heterogeneous direct and indirect effects allows us to detect and better understand the complex and variable causal effects of a treatment by highlighting variations in the causal paths through which the treatment exerts its effects (Muller et al. 2005).

In this paper, we introduce forest-based methods to nonparametrically estimate and infer heterogeneous patterns of the total treatment effect and the decomposed direct and indirect effects via a binary, multinomial, or continuous mediator. We first demonstrate the use of causal forests to uncover heterogeneous total treatment effects (Athey et al. 2019).

Then we propose causal mediation forests, an adaptation of causal forests, to estimate heterogeneous direct and indirect effects and enhance our understanding of variation in causal mechanisms. We focus on the natural direct effects (NDE)[1] and the natural indirect effects (NIE)[2], which sum to the total effect of a treatment (Pearl 2001; Tchetgen and VanderWeele 2014; Tchetgen Tchetgen and Shpitser 2012; Zheng and van der Laan 2012). Conceptually, NDE captures the effect of a treatment when the mediator is set to the level it would have been without the treatment or under a reference exposure level (Imai et al. 2010b; Pearl 2001; Zheng and van der Laan 2012). NIE captures the treatment effect transmitted via post-treatment and pre-outcome mediators. For example, in studying higher education, the total effect of college contrasts outcomes between people who attended college and those who never went to college, whereas NDE and NIE isolate different parts of the total college effect. NDE is the part *not* transmitted via downstream factors (e.g., attaining advanced degrees), and NIE is the so-called "continuation value" of college attendance (e.g., parts of the total college effects created by changing the probability of pursuing advanced degrees; Zhou 2022; Zhou and Pan 2023). Although NDE is referred to as a "direct" effect, it is quite likely that additional intermediate factors, those outside our model, may transmit part or all of these "direct" effects (Alwin and Hauser 1975).

Some recent work considers methods for assessing effect heterogeneity (Brand et al. 2023), such as propensity score methods (e.g., Xie et al. 2012), outcome-model-based (g-formula) methods (e.g., Robertson et al. 2021), tree-based machine learning methods (e.g., Athey and Imbens 2016; Brand et al. 2021), and double/debiased machine learning methods (e.g., Díaz et al. 2021; Farbmacher et al. 2022; Zhou 2022). The causal forests and causal mediation forests we describe extend the existing methodological tools for sociologists in several ways. First, causal forests and causal mediation forests can tackle a rich set of high-dimensional characteristics associated with the treatment, mediators, and outcomes in identifying total, direct, and indirect effects. In social science, individual decisions (e.g., attending college) are affected by many factors at individual-, family-, and contextual levels. When we aim to consider many factors, traditional regression methods suffer from the "curse of dimensionality." For instance, controlling for interaction effects between years of schooling and family income measured at each life stage becomes problematic when these variables contain numerous categories. Suppose years of schooling includes 15 categories and we divide family income into 1000 values. The resulting interaction terms' dimension (15,000 = 15×1000) may exceed the number of observations, leading to inconsistent estimates from traditional regression methods. However, forest-based methods appropriately regulate the estimation in high-dimensional settings and reach the fundamental limits of structural-agnostic functional estimation in causal inference (Balakrishnan et al. 2023).

Second, causal forests and causal mediation forests rely on orthogonal estimating equations to provide doubly and triply robust estimation, yielding asymptotically normal and $n^{1/2}$ consistent estimates with theoretical guarantees. Causal forests generate doubly robust estimates for total treatment effects. That is, if either the propensity score model or outcome model is specified correctly, the estimate of total effects will approach the ground truth of parameter values at a $n^{1/2}$ rate. Causal mediation forests generate triply robust estimates for NDE and NIE, such that we will have consistent estimates for the (in)direct effects if any two of the propensity score, mediator, and outcome models are valid, where mediators can be either binary, multinomial, or continuous.

Third, causal forests and causal mediation forests nonparametrically identify patterns of effect heterogeneity. Different from alternative double/debiased machine learning approaches we consider, which often run a low-dimensional linear regression at the final step for estimating heterogeneous effects (Farbmacher et al. 2022; Zhou 2022), causal forests and causal mediation forests provide fully nonparametric estimation to capture effect heterogeneity in a flexible manner. The forest approach is thus particularly beneficial for nonlinear forms of heterogeneity (Athey et al. 2019; Oprescu et al. 2019). Although the final step in double/debiased machine learning methods can be performed nonparametrically to capture nonlinear heterogeneity, the estimates typically have higher variance than those from causal forests and causal mediation forests. The higher variance of the double/debiased machine learning methods suggests that forest-based methods may offer better extrapolation of results in out-of-sample settings. In addition, before estimating heterogeneous treatment effects across subpopulations, we can use a rank average treatment effect as a metric to test for the presence of heterogeneity and calibrate the overall performance of the forest-based methods in capturing effect heterogeneity (Yadlowsky et al. 2021).

Fourth, causal forests and causal mediation forests produce more stable effect estimates than single tree-based methods (e.g., causal trees), and they are more computationally efficient than alternative double/debiased machine learning methods. Moreover, we can pair forest-based methods with decision trees to effectively and intuitively visualize patterns of heterogeneity (Athey and Wager 2021). The interpretability of these methods makes them less like black boxes than other machine learning techniques.

We conduct simulations to compare causal forests and causal mediation forests with existing causal machine learning approaches. Forest-based approaches often outperform alternative methods in identifying effect heterogeneity. To illustrate their empirical use, we assess college effects on reducing poverty using data from the National Longitudinal Survey of Youth 1979 and 1997 cohorts. Improving access to higher education is a critical factor in equalizing socioeconomic attainment (e.g., Brand 2023; Hout 1988; Zhou 2022). Research suggests that socioeconomic returns to college are heterogeneous – some youth benefit more from a college education than others. Some research shows that youth who are least likely to attend and complete college benefit most (e.g., Brand 2023; Brand and Xie 2010; Giani et al. 2020), and in particular, that college reduces socioeconomic disadvantage most for disadvantaged students (Brand 2023). More research is needed, however, to assess variation in the effects of college pathways on reductions in socioeconomic disadvantage.[3]