causalweight: An R Package for Causal Inference and Mediation Analysis

11 January 2019

Introduction

Researchers in epidemiology, economics, political sciences, or other social sciences frequently aim at evaluating the causal effect of some intervention or treatment, as well as learning about the mechanisms through which a causal effect operates. This paper introduces the R package `causalweight` for analyzing the causal effect of a treatment as well as its mechanisms (based on mediation analysis that incorporates intermediate outcomes called mediators) under various identifying assumptions. All estimators rely on some form of inverse probability weighting (IPW), by weighing outcomes by the inverse of a specific conditional probability or propensity score. The `causalweight` package includes treatment evaluation under treatment selection on observables with and without controlling for non-random outcome attrition or sample selection (Huber 2012, 2014b), instrumental variable-based estimation of local average treatment effects when controlling for observed covariates (Froelich 2007), and mediation analysis for investigating causal mechanisms with selection on observables or instrumental variable assumptions (Huber 2014a; Froelich and Huber 2017). The nonparametric identification strategies underlying the estimators avoid imposing strong functional form restrictions in the structural models considered. The estimation of the propensity scores relies on probit or logit specifications.

Overview of the core functions

The core of `causalweight` consists of four functions aimed at user-friendly treatment evaluation and mediation analysis. The following table illustrates the structure of `causalweight` by assigning to each of the main functions the corresponding treatment effect/mediation model.

Table 1: Main functions of the `causalweight` package

Functions in R	Treatment effect models
treatweight	Treatment evaluation with sample selection correction
medweight	Causal mediation analysis with a binary treatment
medweightcont	Causal mediation analysis with a continuous treatment
lateweight	Local average treatment effect with covariates
medlateweight	Causal mediation analysis with instrumental variables

The function `treatweight` implements treatment evaluation under treatment selection on observables, optionally with correcting for sample selection or non-ignorable outcome attrition based on either a selection on observables/missing at random assumption or an instrument. To tackle the double selection problem into the treatment and into the subpopulation with non-missing outcomes, it makes use of both treatment and selection propensity scores to appropriately reweigh observations by IPW, see (Huber 2012, 2014b). The function `treatweight` allows computing the average treatment effect in the total population (ATE) and on the treated (ATET).

The function `medweight` implements mediation analysis to investigate the causal mechanisms of a binary treatment under selection on observables based on IPW. More specifically, it computes (i) the (total) average treatment effect, (ii) the average natural *indirect* effect, which operates through an intermediate outcome (or mediator) situated on the causal path between the treatment and the outcome, and (iii) the (unmediated) average natural *direct* effect, see (Huber 2014a). The *indirect* and *direct* effect estimates are returned under either potential treatment state. The function `medweight` allows computing the effects for both the total population and the subpopulation of the treated.

`medweightcont` estimates causal mechanisms (natural *direct* and *indirect* effects) of a continuous treatment under a selection on observables assumption assuming that all confounders of the treatment and the mediator, the treatment and the outcome, or the mediator and the outcome are observed. Units are weighted by the inverse of their conditional treatment densities (known as generalized propensity scores) given the mediator and/or observed confounders, which are estimated by linear or loglinear regression, see (Hsu et al. 2018).

The function `lateweight` returns the local average treatment effect (LATE) of a binary endogenous treatment based on IPW using a binary endogenous instrument that is conditionally valid given observed covariates, see (Froelich 2007). In addition, it returns the intention-to-treat effect of the instrument on the outcome, as well as the first-stage effect of the instrument on the treatment. The function `lateweight` permits estimating the local average treatment effect among all subjects whose treatment complies with the instrument (LATE) and among treated compliers (LATTs) by weighing units by the inverse of their instrument propensity scores.

The function `medlateweight` computes the causal mechanisms (natural direct and indirect effects) of a binary treatment among treatment compliers based on distinct instrumental variables (IVs) for the treatment and the mediator, which are assumed to be conditionally valid given a set of observed covariates. The treatment and its instrument are assumed to be binary while the mediator and its instrument are assumed to be continuous. This motivates combining the LATE approach with a control function approach for tackling mediator endogeneity, see Theorem 1 in (Froelich and Huber 2017). The function `medlateweight` yields (i) the (total) local average treatment effect (LATE) among compliers based on IPW, (ii) the average natural *direct* and *indirect* effects under either potential treatment state among compliers based on IPW, and (iii) parametric direct and indirect effect estimates (imposing effect homogeneity across treatment states) based on regression.

The vignettes from the R package `causalweight` provide details on the models and the implementation of the corresponding estimators. In addition, the vignettes give illustrative examples in R.

Summary

`causalweight` is a comprehensive software package having an active user base. Its strength lies in its functionality (it runs on all standard operating systems), diversity of different estimation methods, and simple handling that does not require deep programming knowledge. It has been applied in the area of causal analysis for treatment effect evaluation in graduate courses and recent papers, for example . . . . The source code uploaded on CRAN (The Comprehensive R Archive Network) is available on https://github.com/hbodory/causalweight.git.

References

Froelich, M. 2007. "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates." *Economics Letters* 139: 35–75. https://doi.org/10.1016/j.jeconom.2006.06.004.

Froelich, M., and M. Huber. 2017. "Direct and Indirect Treatment Effects - Causal Chains and Mediation Analysis with Instrumental Variables." *Journal of the Royal Statistical Society: Series B* 79 (5): 1645–66. https://doi.org/10.1111/rssb.12232.

Hsu, Y.-C., M. Huber, Y.-Y. Lee, and L. Pipoz. 2018. "Direct and Indirect Effects of Continuous Treatments Based on Generalized Propensity Score Weighting." *SES working paper 495, University of Fribourg.*

Huber, M. 2012. “Identification of Average Treatment Effects in Social Experiments Under Alternative Forms of Attrition.” *Journal of Educational and Behavioral Statistics* 37 (3): 443–74. https://doi.org/10.3102/1076998611411917.

———. 2014a. “Identifying Causal Mechanisms (Primarily) Based on Inverse Probability Weighting.” *Journal of Applied Econometrics* 29 (6): 920–43. https://doi.org/10.1002/jae.2341.

———. 2014b. “Treatment Evaluation in the Presence of Sample Selection.” *Econometric Reviews* 33 (8): 869–905. https://doi.org/10.1080/07474938.2013.806197.