Double-robust and assumption-lean methods for confounding adjustment in generalised partially linear models

Stijn Vansteelandt[1,2]; Oliver Dukes[1]; David Whitney[3]

1 Ghent University, Belgium
2 London School of Hygiene and Tropical Medicine, U.K.
3 Imperial College, U.K.

Abstract:
Due to concerns about parametric model misspecification, there is interest in using machine learning to adjust for confounding when evaluating the causal effect of an exposure on an outcome. Unfortunately, exposure effect estimators that rely on machine learning predictions are generally subject to so-called plug-in bias, which can render naive p-values and confidence intervals invalid. Progress has been made via proposals like targeted maximum likelihood estimation and more recently double machine learning, which rely on learning the conditional mean of both the outcome and exposure. Valid inference can then be obtained so long as both predictions converge (sufficiently fast) to the truth. Focusing on (generalised) partially linear regression models, we show that a specific implementation of the machine learning techniques can yield exposure effect estimators that have small bias even when one of the first-stage predictions does not converge to the truth. The resulting tests and confidence intervals are doubly robust. We moreover discuss a further extension that aims to be completely assumption-lean. It does this by building on the efficient influence function of a well-chosen estimand, which reduces to the same effect previously targeted when the (generalised) partially linear regression model holds, but continues to capture the conditional association of interest otherwise.

Keywords:
Causal inference; Double robustness; Estimand; Machine learning; Variable selection

References:

1. Dukes, O., & Vansteelandt, S. (2021). Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika*, *108*(2), 321-334.

2. Vansteelandt, S., & Dukes, O. (2020). Assumption-lean inference for generalised linear model parameters. *arXiv preprint arXiv:2006.08402*. Forthcoming in JRSS-B.