DOI: 10.1111/rssb.12528

Elizabeth L. Ogburn, Junhui Cai, Arun K. Kuchibhotla, Richard A. Berk and Andreas Buja's contribution to the Discussion of 'Assumption-lean inference for generalised linear model parameters' by Vansteelandt and Dukes

Elizabeth L. Ogburn[1] | **Junhui Cai**[2] | **Arun K. Kuchibhotla**[3] | **Richard A. Berk**[2,4] | **Andreas Buja**[5]

[1]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

[2]Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[4]Department of Criminology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[5]Flatiron Institute, Simons Foundation, New York, New York, USA

Correspondence
Elizabeth L. Ogburn, John Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.
Email: eogburn@jhsph.edu

Not all conditional associations between outcomes and exposures are of interest. Those that are tend to be directional: up or down. The simplest way to assess directionality is to fit a confounder-adjusted linear exposure term, as the authors propose. We agree with this approach as some of us have argued that linear slopes are meaningful and interpretable even if the directional association is not linear (Buja et al., 2019, section 10). The authors, and Whitney et al. (2019), remind us that severely misspecified adjustment can result in distortions of linear exposure slopes. In their examples, the $A-L$ distributions have U-shaped nonlinearities and, as a result, naive linear adjustment produces a biased estimate of the true slope. Thorough data analysis could unearth such exposure-confounder structure if present in real data. A greater worry for practitioners is missing an essential confounder that biases or reverses the direction of association. The authors' inferential framework does not require L to control for all $A-Y$ confounding, but meaningful use

We congratulate the authors on their excellent article (Vansteelandt and Dukes, 2021). In this comment, we highlight a few practical issues related to their proposal.

of the estimand likely does—and therefore practitioners should select L with care and interpret estimates in conjunction with sensitivity analyses.

The authors' project of assumption lean inference rests on the assumption that nuisance parameters can be estimated nonparametrically at rate $n^{1/4}$. It is surprising to us that this property is widely assumed to hold for machine learning methods. The authors are in good company with this assumption, but, for example, the random forests included in the authors' analyses can have large bias if a tuning parameter is chosen badly (Olson, 2018), and as far as we know cross-validation has not been shown to reliably choose good tuning parameters. Even if $n^{1/4}$ rates are achieved asymptotically, slower rates of convergence may require large samples before asymptotic approximations are useful. This points to the importance of methods to test or help ensure that the required rates are achieved (Liu et al., 2020; Robins et al., 2008; van der Laan et al., 2021), or to perform valid inference under slower rates (Cattaneo and Jansson, 2018; Kuchibhotla et al., 2021).

We re-ran the authors' code and applied HulC, a new method for the construction of assumption—lean confidence intervals (Kuchibhotla et al., 2021).[1] We found that the point estimates are indeed sensitive to choice of tuning parameters. Although HulC intervals are wider, they are valid even if approximate normality does not hold, as would be the case if the nuisance estimators converge slower than $n^{-1/4}$, as long as the estimator satisfies a weaker median unbiasedness property (Kuchibhotla et al., 2021).

How to cite this article: Ogburn, E.L., Cai, J., Kuchibhotla, A.K., Berk, R.A. & Buja, A. (2022) Elizabeth L. Ogburn, Junhui Cai, Arun K. Kuchibhotla, Richard A. Berk and Andreas Buja's contribution to the Discussion of 'Assumption-lean inference for generalised linear model parameters' by Vansteelandt and Dukes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84, 715–716. Available from: https://doi.org/10.1111/rssb.12528

[1]The data and code were provided by the authors. We modified the code slightly, removing the support vector machine method from the SuperLearner library because of an error message. Because of this, our point estimates are close, but not identical, to those reported by the authors. The code to produce all tables is available at https://github.com/cccfran/HulC-on-VD.