Bond University

Research Repository

Assumption-lean methods are not necessarily superior to assumption-rich methods, but they are surely complimentary. In future applied work, I look forward to *adding* some of these generic estimand-centric methods to my usual, honestly-reported, model comparisons and tentative inductive leaps.

REFERENCES

Cartwright, N. (1994) *Nature's capacities and their measurement*. Ch2. Oxford: Oxford University Press.

Fisher, R.A. (1959) *Statistical methods, experimental design and scientific inference*, 1990 edition. Oxford: Oxford University Press.

How to cite this article: Hunt, I. (2022) Ian Hunt'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, 711–712. Available from: https://doi.org/10.1111/rssb.12525

Kuldeep Kumar's contribution to the Discussion of 'Assumption-lean inference for generalised linear model parameters' by Vansteelandt and Dukes

Kuldeep Kumar

Bond University, Gold Coast, Queensland, Australia

Correspondence
Kuldeep Kumar, Bond University, Gold Coast, Queensland, Australia.
Email: kkumar@bond.edu.au

According to Leo Breiman (2001), there are two broad cultures for analysing and modelling to reach conclusions from the data. The first one is data modelling culture where the value of the parameters are estimated from the data and then the model is used for information and/or prediction. The second one is algorithmic modelling culture, where the approach is to find a function f(x) using an algorithm that operates on x to predict y. In the first case, statistical tools like OLS are used whereas machine learning tools perform much better in the second case. I should

congratulate the authors for developing a hybrid model using machine learning and traditional statistical techniques like OLS. However, I do have few queries on the real-life applications of the proposed model. Most of the simulation results for main effects, effect modification and misspecified link function are valid for sample size $n = 500$ or more. Also, the data analysis example presented in section 7 has a sample size of 2500. How is the empirical bias affected when the sample size is small? Bayesian models incorporate prior information, which can be specified quantitatively in the form of a distribution. Do the authors think a Bayesian parameter generalised linear model can perform better if some prior information is available? Also, I am not sure how this model tackles the problem of multicollinearity? Finally, in the context of high dimensional variable selection, what will happen if the covariates are not distributed normally?

REFERENCE

Breiman, L. (2001) Statistical modeling: the two cultures. *Statistical Science*, 16(3), 199–231.

Michael Lavine and James Hodges' contribution to the Discussion of 'Assumption-lean inference for generalised linear model parameters' by Vansteelandt and Dukes

Michael Lavine[1] | **James Hodges**[2]

[1]Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts, USA

[2]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
James Hodges, Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA.
Email: hodge003@umn.edu

The authors advocate a style and rhetoric of statistical analysis that begins with specifying an estimand and ends with an estimate and interval. They say, 'The starting point... is to come up