
DISCUSSION CONTRIBUTIONS

Proposer of the vote of thanks to Vansteelandt and Dukes and contribution to the Discussion of ‘Assumption-lean inference for generalised linear model parameters’

Rhian M. Daniel

Division of Population Medicine, Cardiff University, Cardiff, UK

Correspondence

Rhian M. Daniel, Division of Population Medicine, Cardiff University, Cardiff, UK.

Email: danielr8@cardiff.ac.uk

1 | TWO CONTRASTING PHILOSOPHIES

Traditional statistical modelling starts from a family \mathcal{F} of observed data laws indexed by unknown parameters of interest β . The goal is to make inference about β under the assumption that \mathcal{F} contains the true law. By labelling β ‘of interest’, it is implied that \mathcal{F} can be expressed such that β naturally encompasses the main scientific goal, which is not always the case. Furthermore (e.g. ch.1 Cox & Hinkley, 1979) if \mathcal{F} does not contain the truth then the inferential theory loses relevance and the interpretation of β is obscure. Such concerns sensibly lead to model checking procedures, which themselves raise further concerns, as VD describe.

The *causal inference* and *targeted learning* schools (Hernán & Robins, 2020; van der Laan & Rose, 2011) start instead from an estimand, chosen to reflect the scientific question, without reference to any statistical model. Subsequent estimation and inference are tailored to this estimand, sometimes using a parametric model \mathcal{F} , but not to define the estimand. The targeted learning framework advocates replacing \mathcal{F} with machine learning algorithms, using the estimand’s influence function and accompanying theory to derive estimators with well-understood asymptotic behaviour.

Vansteelandt and Dukes (henceforth VD) propose a practical resolution to an important tension between two philosophies of statistical inference. I summarise these aspects before discussing how we might revise our understanding of ‘bias–variance trade-off’ in statistical modelling in the light of VD’s work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Author. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

Although the hygiene of the latter approach is eminently attractive, its implementation requires statistical expertise. In principle, each bespoke estimand demands that all subsequent steps be derived afresh, with no guarantees that the resulting estimator has good properties (e.g. when the estimand is too ambitious given the available data). Practical applications of targeted learning thus tend to focus on simple estimands (e.g. the marginal effect of a binary exposure) where off-the-shelf implementations are readily available. This leaves users in a quandary when their scientific question is more complex, for example when the exposure is continuous, as in the settings considered by VD.

2 | THE BEST OF BOTH WORLDS

VD start, as in the traditional approach, from a generalised linear model \mathcal{F} indexed by β . This has the advantage of restricting attention to quantities that are plausibly reliably estimable from the data. For any estimator $\hat{\beta}$, consistent under \mathcal{F} , their philosophy is to consider its probability limit β^* under \mathcal{A} , the set of *all* possible data laws. The honest estimand β^* is only considered acceptable if it corresponds (under \mathcal{A}) to a weighted average of parameters β_l , where each β_l has the interpretation of β restricted to levels l of some covariates L , not of primary interest.

VD set high standards for making inference about β^* , namely consistent estimation and parametric convergence rates under ‘lean’ regularity assumptions, honest inference after algorithm/variable selection, and no density estimation for continuous variables. They argue convincingly that such demands are necessary for the data to speak for themselves about β^* , and describe a general procedure that meets these standards in the case of any parameter motivated by a GLM.

3 | PRECISION IS BOUGHT WITH BLUNTNESS NOT BIAS

That VD propose essentially non-parametric estimation may seem alarming in the light of the curse of dimensionality (Stone, 1985). Indeed, the traditional approach based on one simple model is often justified on the grounds of a bias–variance trade-off: we assume a simple (‘wrong but useful’) model since it buys precision in modest-sized data sets. The simulation studies presented by VD illustrate that this intuition is faulty: their assumption-lean estimators are also relatively precise, but then at what cost?

Our intuition was developed in the context of the traditional approach in which \mathcal{F} plays the two roles described by VD: (i) estimand definition, and (ii) representing the set of possible data laws. Traditionally, choosing a more complex model \mathcal{F} leads simultaneously to a less parsimonious estimand and a larger set of possible data laws. VD, on the other hand, propose a parsimonious estimand, coupled with only very lean restrictions on the set of data laws: parsimony in the first sense but not the second. Figure 1 gives a simple illustration of how parsimony in both senses increases efficiency, but with parsimony of type (i) having a greater impact than type (ii).

Since consistent estimation is guaranteed under very lean assumptions, and thus bias essentially avoided, the sacrifice made by VD’s parsimony (in the first sense) with which they buy precision is, I believe, not bias but bluntness. A more nuanced (less blunt) understanding of, say, a continuous exposure’s effect on an outcome, could be gained by choosing a less parsimonious summary, for example one that separately summarises the effect in more sub-groups, but at the cost of increased variance.

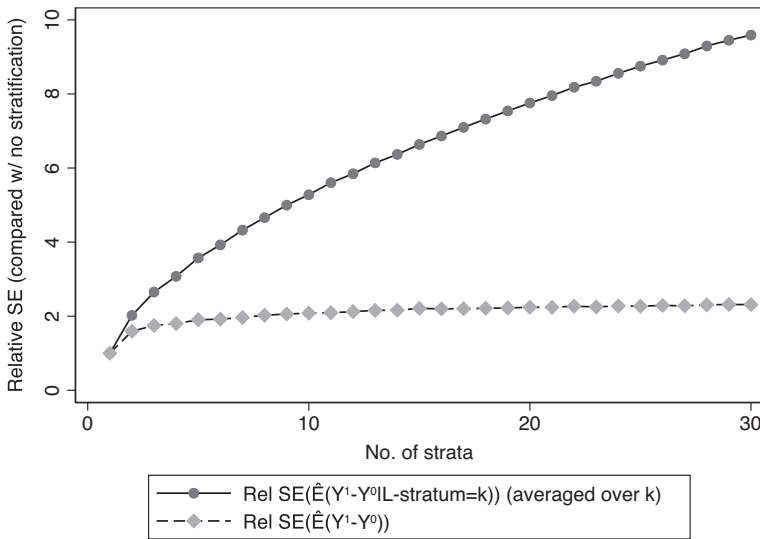


FIGURE 1 This graph shows the increase in relative standard error for the estimator of two different types of estimands after subdividing a covariate L into progressively more strata. The 3000 simulated datasets each with sample size 1000 are from a hypothetical observational study with a continuous confounder $L \sim N(0, 1)$, a binary exposure A with $\Pr(A = 1 | L) = \text{expit}(L)$ and a binary outcome Y with $\Pr(Y = 1 | A, L) = \text{expit}(-2 + 0.2AL^2)$. Each dataset is divided into an increasing number s of approximately equally-populated strata based on the observed quantiles of L . We first plot the empirical standard deviation of the stratum-specific estimator of the average causal effect in each stratum separately, when splitting into $s = 1, \dots, 30$ strata relative to 1 (i.e. no stratification). Since the SE varies by stratum, the plot in fact takes the average of the SE over the s strata. We then plot the relative empirical standard deviation of the estimator of the average causal effect (marginalised over the strata) when the data analysis model is stratified into s strata relative to 1. Since the true model for Y given A and L has the same form regardless of the value of L , the models with and without stratification are all correctly specified. This allows us to explore, on the one hand, the impact of needless flexibility in \mathcal{F} in the sense described in (ii) in the text (the slowly increasing lower line) compared with the additional impact of decreasing parsimony in the estimand of interest, i.e. sense (i) in the text (the more steeply increasing upper line)

4 | CONCLUDING REMARKS

VD start from the viewpoint that the two approaches in Section 1 are unsatisfactory. The traditional well-trodden path offers a comfortable ride but often to an unknown and uninteresting destination with a dishonest account of how we got there. On the other hand, the targeted learning path, in aiming admirably for the summit of a yet-to-be-conquered mountain, is often too perilous to navigate with our modest equipment and abilities. VD offer a third way, which feels on the surface much like the first, but leads to a well-defined destination that is both practically reachable and at least somewhere in the foothills of scientific interest. Beneath the surface lies much of the sophisticated technology from the targeted learning journey, but as passengers we need not necessarily know how to operate it, thanks to their general-purpose solution.

I conclude by congratulating Vansteelandt and Dukes on their innovative yet pragmatic proposal presented in a wonderfully didactic fashion that provokes us to rethink fundamental aspects of statistical modelling. I enthusiastically propose the vote of thanks.

REFERENCES

- Cox, D.R. & Hinkley, D.V. (1979) *Theoretical statistics*. Boca Raton: CRC Press.
- Hernán, M.A. & Robins, J.M. (2020) *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC.
- Stone, C.J. (1985) Additive regression and other nonparametric models. *The Annals of Statistics*, 13(2), 689–705.
- van der Laan, M.J. & Rose, S. (2011) *Targeted learning*. Berlin: Springer.

How to cite this article: Daniel, R.M. (2022) Proposer of the vote of thanks to Vansteelandt and Dukes and contribution to the Discussion of ‘Assumption-lean inference for generalised linear model parameters’. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84, 686–689. Available from: <https://doi.org/10.1111/rssb.12513>

DOI: 10.1111/rssb.12514

Seconder of the vote of thanks to Vansteelandt and Dukes and contribution to the Discussion of ‘Assumption-lean inference for generalised linear model parameters’

Vanessa Didelez^{1,2}

¹Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

²Department of Mathematics and Computer Science, University of Bremen, Bremen, Germany

Correspondence

Vanessa Didelez, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany.

Email: vdidelez@uni-bremen.de

In my view, one of the most important contributions of the field of causal inference has been to place the target of inference, the desired estimand, at the centre of the analysis. The estimand is chosen in view of the research question, and typically reflects what decision problem we need to solve or what our ideal (target) trial would be. Crucially, the (causal) estimand is *not* automatically a parameter that happens to parametrise a chosen model. The role of models is mostly as

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Author. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.