Prediction Can Be Safely Used as a Proxy for Explanation in Causally Consistent Bayesian Generalized Linear Models

Maximilian Scholz[1], Paul-Christian Bürkner[1]

[1]) Cluster of Excellence SimTech, University of Stuttgart, Germany
maximilian.scholz@simtech.uni-stuttgart.de

Bayesian modeling provides a principled approach to quantifying uncertainty in model parameters and structure and has seen a surge of applications in recent years. Despite a lot of existing work on an overarching Bayesian workflow, many individual steps still require more research to optimize the related decision processes. One such practice is the use of prediction as a proxy for explanation in the context of latent inferential goals. However, there is an apparent gap in the joint consideration of causality, predictive performance, and parameter recoverability with the goal of optimizing the latter. Here, we approach it from the perspective of model selection in a Bayesian workflow. When PR is the goal, predictive performance reduces to a conveniently available supporting utility that ideally helps to select models with better PR at real data inference time [1]. In practice at least, and despite the theoretical arguments for caution, this assumption is very commonly (and often implicitly) made whenever explanatory model choices are based on out-of-sample posterior predictive metrics or their approximations, such as AIC [2, 3], DIC [4], WAIC [5, 6] or ELPD-LOO [7, 6]. As we know from counter examples [8, 9], this assumption cannot hold in general, but it remains unclear under which conditions it is actually justified. To answer the question of "Can predictive performance be reliably used as a proxy for (causal) parameter recoverability, within a set of statistical models that all share the same latent causal model and does the answer to this question depend on whether or not the latent model is causally consistent?" we conducted a large simulation study of Bayesian generalized linear models. The results indicate that for causally consistent models predictive performance can be safely used as a proxy for parameter recoverability as exemplarily shown in Figure 1 where better relative predictive performance is predictive of a lower parameter RMSE.

References

[1] Bürkner, P.C., Scholz, M. and Radev, S.T., 2022. Some models are useful, but how do we know which ones? Towards a unified Bayesian model taxonomy. arXiv preprint arXiv:2209.02439.

[2] Watanabe, S., 2009. Algebraic geometry and statistical learning theory (Vol. 25). Cambridge university press.

[3] McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.

[4] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A., 1998. Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models (pp. 98-009). Research Report, 98-009.

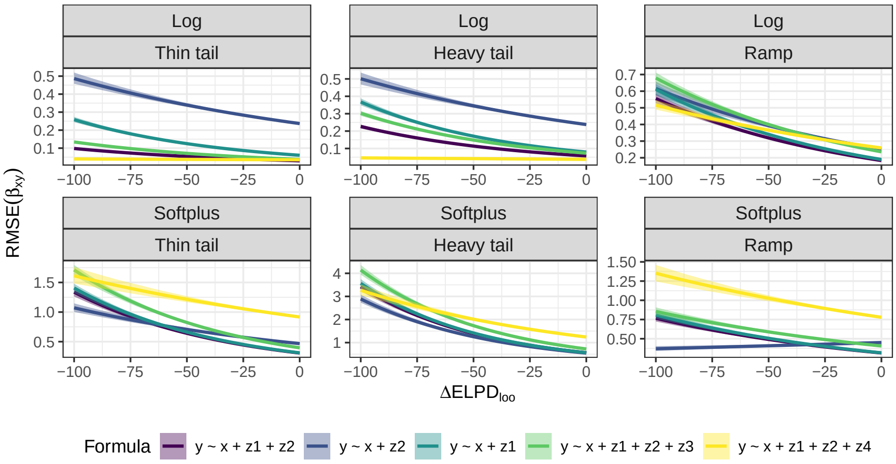

Figure 1: Conditional effects of $\Delta\text{ELPD}_{\text{loo}}$ on RMSE(β_{xy}) for double-bounded data and models. Split by data generating link and shape. A negative slope implies better parameter recovery with better predictive performance.

[5] Watanabe, S. and Opper, M., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research, 11(12).

[6] Vehtari, A., Gelman, A. and Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing, 27, pp.1413-1432.

[7] Vehtari, A. and Ojanen, J., 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison.

[8] Betancourt, Michael (2020). Towards A Principled Bayesian Workflow (RStan). Retrieved from `https://github.com/betanalpha/knitr_case_studies/tree/master/principled_bayesian_workflow`, commit aeab31509b8e37ff05b0828f87a3018b1799b401

[9] Hagerty, M.R. and Srinivasan, V., 1991. Comparing the predictive powers of alternative multiple regression models. Psychometrika, 56, pp.77-85.