

Review of: "Variable selection in generalized extreme value regression model using Bootstrap method"

Luca De Angelis¹

¹ University of Bologna

Potential competing interests: No potential competing interests to declare.

The paper deals with model selection in Generalized Extreme Value regression models as proposed by Calabrese and Osmetti (2013). The Authors identify the bootstrap algorithm proposed by Austin and Tu (2004) as the method to tackle the issue of the selection of regressors in such models.

Albeit being not more than an empirical exercise using the medical data from one case study, the paper is interesting and addresses an important issue.

Below are some comments that I think could improve the manuscript.

1. The title sounds too general and a bit misleading. As only one method and one regression model are considered, the title should reflect what is the true contribution of the paper. Statisticians could find this contribution too limited and too applied for their tastes, but the paper could be useful to practitioners and, e.g., researchers in medical fields. Therefore, a title that directs towards the actual contribution and value of the manuscript could significantly help the reader.
2. While I am happy with the choice of the GEV model specification, I do not understand (i) why considering only one bootstrap technique, and (ii) why specifically the one proposed by Austin and Tu (2004). Please add a justification to these choices or consider other bootstrap techniques for comparison.
3. It is not clear to me whether the issue at stake is related to high-dimensional data or, more in general, to model selection in regression models. If the former, then the Authors should state it clearly; if the latter, then please cite also "traditional" model selection methods (e.g. information criteria, maximum likelihood method, etc.).
4. The sentence about the fact that the bootstrap can handle complications in regression models, such as "model selection, heteroscedasticity of variances, nonlinearity in the model parameters, and bias due to transformation" should be discussed more in details. Is this true for any bootstrap techniques? Is this the case for the bootstrap algorithm considered in this paper?
5. To "validate" the results obtained using the bootstrap, the Authors consider information criteria, showing that AIC and BIC lead to the same outcome. To play devil's advocate, why is the bootstrap preferable in such context, although is less intuitive, more involved, and more time-consuming than standard IC?
6. I have found many typos. See, for instance, the Authors' affiliations, and at pages 2-3 of the paper: "proposed propose", "fi t", "infl uence", "confi dence", "accordind", "approch", etc.

