

Anonymous - Referee report

April 08, 2019 - 09:20

Is the contribution of the paper potentially significant?

In my opinion, absolutely. The paper articulates a subtle and elegant observation about the nature of statistical modelling (and in fact all modelling): that error can be broken down into variance and bias. The point is, as I see it, that ...[more]

... a complex model may have high error under conditions of uncertainty, while misleadingly conveying low bias, which can be mistaken for low error. In situations of real, radical uncertainty about what the "correct" solution is, the "bias bias" the author speaks of is the tendency to prefer models with low bias over those with low variance, even though both alternatives can equally contribute to error. Note that simple models are (by definition) easier to understand, and that is (in my opinion) important to the author's point. Bias bias is implicitly preferring more complex models because they appear more mathematically consistent, even though their complexity belies variability in the answers they might provide. Moreover, that variability could relate to "bias" in another sense of that word.

A somewhat unfortunate aspect of the bias/variance breakdown is that "bias," in the casual meaning of that word (e.g., prejudice), may be worse and more hidden in complex models, precisely because of what this paper indicates. That is to say, in models like deep learning networks, biases (intentional or otherwise) in the selection of features, processing steps, training data, etc., can be hidden in model complexity. Low bias (in the sense used in the paper) only acts to further hide these effects in complex models, by conveying a sense of mathematical sophistication and reliability. A simpler model might be equally wrong, but the biases in the casual sense of the word will be less hidden there.

However, that is one amongst a broad variety of problems with language in the field of statistics, and the author need not address that in this paper. The paper is clear and internally consistent on this point, so that is not a complaint.

A valuable aspect of the paper is the offering of four common and understandable examples where a complex model misleads, and a simpler model is more appropriate under conditions of high uncertainty. In these examples, the author shows that technically the simpler models do better.

Is the analysis correct?

Yes, and given the gravity and importance of the issue, and the clarity of the paper's articulation of that issue, I most certainly feel it should be published and widely read.

I have a question I'd like to ask the author, but that is off point for the paper, per se. That question is whether the author feels there is an intrinsic value to simpler models because of their higher variance, in that such models may allow greater human adaptivity, for psychological reasons. That is to say, when one contemplates such models as a part of human reasoning, perhaps higher variance allows for psychologically easier switching of models when new data fails to agree with existing models. Lower bias models may provide a sense of confidence that is more difficult to shake as data fails to conform with current assumptions.

I've approached that question as a matter of psychology, but it would also be useful

to consider it as a technical question about algorithmic modelling. Do simpler (and other higher) bias models prove to be more adaptable (in terms of computational effort) when the data stream is from a non-stationary source?

As I said, those are questions I'd like to discuss with the author, but I don't feel they need addressing in this paper, which is quite informative and self-contained.

Henry Brighton - Response to reviewer

July 09, 2019 - 11:07

Dear reviewer,

Thanks for taking the time to read the paper and provide feedback. I think that it is first worth clarifying that we shouldn't assume that (1) simple models have low variance and high bias, and (2), complex models have low bias and high variance. Depending on the ...[more]

... problem, the opposite can also be true. And of course, the sample size is crucial.

I think your question raises the issue of strategy selection/meta-learning, and issues relating to how one switches or revises a strategy over time. I didn't tackle this issue in the article and you are right to bring it up. One view is that if we rely only on feedback to guide strategy selection then the complexity/simplicity of the strategy is not really an issue. Performance is all that matters. Your main point (I will assume) alludes to deliberation on the part of decision maker, and you raise the question of what role complexity plays in relation to psychological issues in strategy choice. From a technical standpoint, which you are curious about, I'm not sure what can be said without stating more precisely the nature of the problem. In short: I'm not entirely convinced by general claims about the benefits of "simple" models. It depends on the problem, and the nature of the "complex" models they compete with. I'd reiterate the point I made in the article: That variation among the strategies being considered is what is crucial.

I hope this is useful,

Henry Brighton

Mark Fenton-O'Creevy, The Open University - Commentary

June 17, 2019 - 07:26

In this nicely argued paper, Brighton builds on prior work on the 'bias bias' with Gigerenzer (Brighton & Gigerenzer, 2015; Gigerenzer & Brighton, 2009), in which they have argued for the power of simple heuristics in decision-making under conditions of uncertainty (defined here by sparse available observations and little knowledge ...[more]

... about causal processes). In this paper Brighton moves on from a focus on the conditions in which simple decision rules outperform more complex modelling to deploying similar arguments (about the relative contribution to prediction error of variance and bias) to question implicit and explicit assumptions about normative rationality that underpin much of modern economics and the relationship of rationality to optimisation.

At the heart of this critique is a) the insight that variance affects prediction error more than bias in conditions of uncertainty (as defined above); and, b) drawing attention to the large approximations involved in making small world representations of large-world models (in particular though quantifying unquantifiable uncertainty), relative to the smaller approximation involved in producing a particular solution from a given model.

In reviewing this paper, I have the advantage of having been present at a presentation of an earlier version at the Rebuilding Macroeconomics Conference referenced in the paper acknowledgements. I mention this since the combination of the paper and the audience discussion led me to an important insight about economic constructions of rationality. In response to the paper a number of economists made a particular critique (perhaps mistaking the main thrust of the argument, but in a revealing way). The critique was essentially ‘you seem to be implying that economists don’t understand the problems that you are describing and that we produce over-fitted models. You are wrong we understand the importance of parsimony in the performance of models out of sample and for this reason in constructing macro-economic models we are very focussed on developing models with only a restricted set of predictor variables’ (I paraphrase for brevity). My insight: Macro-economic models are a form of simple heuristic and this behaviour is regarded as normatively rational by many economists. However, by contrast, much economic theory regards failure to act on full information by ordinary mortals as a significant breach of normative rationality. For me, this paradox deserves greater scrutiny.

Discussion of the trade-offs between variance and bias and the implications of simple versus more complex models is not new. However, the significance of the paper and its major contribution lies in the way in which these arguments are deployed to question standard understandings of normative rationality in economic thought.

Whilst genuinely admiring of this paper and the potential insights it offers for rethinking rationality in economic theory and modelling, I have some concerns. First, I feel the paper would benefit from more explicit discussion of the nature of uncertainty. Whilst the explicit operating definition of uncertainty offered early in the paper focuses on sparse observations and poor knowledge of causal processes, other tacit definitions seem apparent in the paper. These include the uncertainty of approximation of small world models to large world problems and, e.g. in discussion of the $1/N$ heuristic, the non-ergodicity of market price behaviour over extended time spans. Second, the paper glides over the question of causal theorising without discussion of the role of such theorising in variable selection and model construction. For example, both models presented in the first burglary example rest on a common theoretical assumption that burglars tend to commit crimes within some (perhaps idiosyncratic) radius of their home address. In practice, all models rest on assumptions and all assumptions rest on some form of theoretical reasoning whether naïve or sophisticated. This bears on the question of fitting simple rules to context, which is prominent in the work the author builds on.

Brighton, H., & Gigerenzer, G. (2015). The bias bias. *Journal of Business Research*, 68(8), 1772-1784.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in cognitive science*, 1(1), 107-143.

Henry Brighton - Response to reviewer

July 09, 2019 - 11:08

Dear reviewer,

Yes, I fully agree with your point about the interpretation of the presentation by some of the audience. My point was not to claim that nobody has thought about bias/variance in relation to predictive models. My point is that these insights, and their implications, are neglected when ...[more]

... it comes to thinking about rationality. In this article, I tried to make this point clear and unambiguous.

I also agree that the various categories of uncertainty could be set out more clearly, along with how they impact on different facets of the argument. I've attempted this elsewhere, and perhaps I should elaborate on these issues in the revised version of this article. Your point about models and their assumptions is good one. To what extent do models always involve causal, or some form reasoned assumptions? I would respond by saying that the criminal profiling example you use illustrates your point, but for many problems a different perspective is needed because we may know next to nothing about the underlying causal processes. I think the point here is that I'm focusing on these tricky problems where, in practice, the critical issue tends to be choosing the right features rather than the right causal model. Nevertheless, the criminal profiling example illustrates that even when we do have some idea about the causal factors relating to the problem, we still need to reduce variance somehow. So, in short, yes, what knowledge we have of the causal processes giving rise to observations should guide model development when it can. However, such models need to be evaluated relative to others which don't attempt to model these processes. The proof is in the prediction error.

I hope this response addresses your point,
Henry Brighton