Referee's Report on: Bayesian Restricted Likelihood Methods

The central observation in this paper is given in expression (3) p. 6. Using the notation in the paper, suppose a likelihood $L(\theta|y)$ can be factored as

$$L(\theta|y) = f(T(y)|\theta)f(y|T(y),\theta). \tag{1}$$

Then, if T is well chosen, it makes sense to assume

$$L(\theta|y) \approx f(T(y)|\theta).$$

As long as the ratio of the two sides is close to one, the factor $f(y|T(y),\theta)$ will not affect inference on θ or, arguably more important, prediction of y_{n+1} very much. In these cases, it makes sense to use posteriors based on $f(T(y)|\theta)$ as noted on p. 4. However, more generally, $f(y|T(y),\theta)$ affects $L(\theta|y)$ so it must be estimated as in Sec. 3.2.0 and Sec. 3.2.1.

This idea is developed in Sec. 3.2.1 in the context of linear models. The authors use the familiar frequentist Huber robust estimator in (6) for β and a similar robust estimator for σ.

As I understand it, since $f(T(y)|\theta)$ is known, the way to use (1) is to get an approximation to the full likelihood by estimating $f(y|T(y),\theta)$. The authors' technique is given by (7) and (8) p. 12. The idea is that a 'proposal density' must be found so that y_p represents outcomes of $f(y|T(y_{obs}),\theta)$ for some θ. The y_p's for each value of θ can be used as a sample for estimating the density $f(\cdot|T(y_{obs}),\theta)$. Given that this collection of densities is estimated, they can be used in (3) to form a posterior $w(\theta|y)$ for inference.

The balance of Sec. 3 gives some formalities to justify the procedure and some intuition.

Sec. 4 develops a useful and convincing application of the methodology developed. There are two parts to this: The first is a direct application of the technique in Sec. 3 for a standard normal theory regression model comparing five different instantiations of it based on five different base models. The second extends the first analysis to a hierarchical regression. The evaluation of the new method is done predictively – a major plus and unquestionably the right way to think about this problem.

Overall, this paper explores a reasonable setting – incompletely specified models – and offers a solution that is valid and useful in some generality. The paper is timely, novel, and the contribution is appropriate for *Bayes Analysis* in quality and importance; there is no call for further research. On the other hand, the paper could use some improvements in the exposition. I had a lot of difficulty in understanding the details and I'm not at all sure I do even now – after several readings. While some of my difficulties are due to weaknesses in my background, the authors should do a better job of explaining themselves.

Below are several points, major and minor, that the authors should address in a revision.

Major Points:

The writing in this paper is very smooth; it reads easily. This is good but I'm worried this is achieved by not providing enough detail and reasoning. Hence the bulk of my comments below are efforts to quantify what further details the authors almost certainly know and would be well advised to include.

1) *The Introduction and Sec. 2 are too long and detract from the focus of the paper:* The authors basically want to give (3) and two examples of it, namely (1) and (2). Then they want to explain how they generate $(y|T(y), \theta)$. The central idea is strong enough to stand on its own without the elaborate discussion. The discussion points seem more appropriate for a concluding section on the conceptual implications of using a restricted likelihood. I suggest this because the conceptual implications are not the focus of the paper, they result from the methodology. The discussion material in the early part of the paper could be moved to the last section and made more focused. There are also a good number of references that the authors might want to read and possibly cite in the general area of model uncertainty outside the across-model prior setting. In no particular order, some are: Zhu, Ibrahim, Tang 2011 BKA, Barnardo and Smith 2000 (for M-closed, complete, and open), Gustafson and Clarke 2004 CJS, Clarke 2010 BA, Draper 1996 JRSSB, Yuan and Clarke 1999 CJS, Clyde and Iverson 2013 (M-open, proceedings volume) and Clarke, Clarke and Yu 2013 (M-complete) BA.

2) *The paper should be more self-contained:* Even after several readings I do not understand the procedure on p. 12 satisfactorily. Details are left out, e.g., how (7) follows from (8); the one sentence explanation on l. 8-9 is not enough. Why is it OK to drop the conditioning on $T(y_p) = T(y_c) = T(y_{obs})$? Some notation seems odd e.g., what is the relationship amongst y_p, y_c, and y_{obs}? Where is the expression $p(\cdot)$? Or do the authors mean $p(\cdot|\theta)$? Also, I thought the idea of Metropolis-Hastings was to generate a sample from a distribution for which direct generation was difficult. If evaluation of $f(y|\theta)$ is straighforward (as stated on p. 12, l. 10) and $f(y|T(y), \theta)$ is available, why is the author's procedure important? I think the answer is that the authors want to compare $f(y|\theta)$ to the use of a $f(T(y)|\theta)$ corrected by an estimate of $f(y|T(y), \theta)$ so that one can use the authors procedure without having $f(y|\theta)$ but this should be stated explicitly.

A related point is that the discussion in Sec. 3.2.1 was also hard to follow because too many details were left out. The role of z^* needs to be explained better. While the words in the last two lines of p. 12 may be obvious to those who know the methodology, giving the mathematical formalities and the motivations for the various steps would make them and p. 13, 14 easier to follow. In particular, where does y^* come from? Why would one want to transform y^* into y as in Theorem 1? How does the expression in Theorem 1 lead to the set in (9)? I have no doubts that the authors can provide more formality and motivation for each step in the procedure; it's just a matter of doing it.

3) *The intuition should be better explicated:* The authors have clearly thought carefully about the intuition behind their procedure, but some how it doesn't

come across in Figs. 1 and 2 which I am unable to link to the formalities they do present. The authors have not yet found a good way to express their ideas so that they can be readily assimilated by some one who does not already understand. I don't understand the steps in the reasoning in Sec. 3.2.2. It's OK to relegate proofs to an Appendix, but the reasoning should be clearer in the main text.

4) *The results for the application need better labeling/exposition:* It would be helpful if the authors could write out exactly and explicitly what cross-validation type of evaluation is being calculated. This is described at the bottom of p. 24, top of p. 25 in words, but the mathematical expressions should be given fully as well. This would clarify the role of the base method and the comment on p. 27, l. 3-4. The same sort of comment applies to p. 27-28: Please give the mathematical expressions rather than just the words. (This would make the last lines on p. 29 easier to follow.) Maybe it's just the way that my copy printed, but in Figs. 3, 4, and 5 the ranges/dots are hard to distinguish from each other so it's hard to tell which is what. Could the different cases have different symbols that would be easier to tell apart? Even making the figures larger would help. To make it easier for the reader could the discussion of Table 1 p. 31 refer to the rows in the table? That would make it easier for the reader to tell which comparisons the discussion was commenting on. A related question: What does Fig. 1 show that is essential to the discussion?

5) *Gather and organize your thoughts on model mis-specification in Sec. 5:* This is a sort of repeat of item #1, but there are many places where the authors comment on various aspects of their analysis, the comments are valid and helpful, but they detract from the flow of the paper. If these were gathered and organized in the Discussion section the authors points would be more effectively made. For instance, the paragraph beginning In addition to... on p. 4, the first part of the paragraph beginning Further examples... on p. 6, middle paragraph on p.8, the discussion of instability in the first two paragraphs of Sec. 3.2, the paragraph beginning We digress...on p. 28. There are other cases where remarks are valid but maybe they are not in the right place and would be more effectively made in the Discussion.

Minor Points:

Here are a few specific points. Some are picky, some are more substantive, but they are all smaller than those above. If I had more time, I'd probably find more suggestions, but I'm sure the authors can see the general tenor of my points is to encourage them to be clearer about the details.

1) p. 3, para. beginning The focus of this work...: State this more formally with mathematical notation.

2) p. 3, l. -2: The word restricted likelihood is OK, but if the authors could use some other word it would be better. The term restricted is already used in the sense of restricted MLE and that's not the same as the authors mean here. This would help avoid possible confusion.

3) p. 4, l. 11: Wong and Clarke CJS 2004 shows a case where using less but well chosen data gives better prediction.

4) p. 4, l. -8: The authors might want to have a look at Bernardo and Smith 2000 for their discussion on M-closed, -complete, -open problems. Maybe mention variance-bias decompositions for MSE?

5) p. 4, l. -4,-5: Yes, mis-specification and outliers somehow represent the same problem...Could this point be made with more clarity?

6) p. 5, l. -4: term should be factor.

7.p. 6, l. 7: Alternatively should be Hence (for instance).

8) Keep the ordering of conditioning variables the same e.g., (3) p. 6 uses $(y|T(y), \theta)$ while p. 1, l. -2 uses $(y|\theta, T(y))$.

9) p. 23, l. -1: 'tunning' should be 'tuning'