Referee report for ECOSTA-D-22-00151

"A Computationally Efficient Mixture Innovation Model for Time-Varying Parameter Regressions"

The manuscript considers prior specification and computing for Bayesian time-varying parameter regression. In particular, the author proposes a continuous analog of a spike-and-slab prior on the dynamic coefficient innovations, which maintains shrinkage toward locally static coefficients while reducing computational cost. The method is compared to a spike-and-slab version in the simulation study and applied to forecasting examples with further comparisons.

There are several questions and concerns regarding the manuscript:

1. The manuscript repeatedly advertises $\mathcal{O}(n)$ scalability instead of $\mathcal{O}(2^K n)$ for certain spike-and-slab alternatives. First, I do not see how this is possible: the computing time must depend on the number of covariates K, even if it is linear. Second, it is widely common in spike-and-slab models to update one indicator at a time from its full conditional distribution (i.e., given the remaining indicators), which is linear in K. Such a stochastic search must still explore the 2^K space, but would seem to achieve the same computational scalability as the proposed approach. Why is this not considered or discussed as an alternative here?

2. The simulation study compares the proposed approach (LMI) to a spike-and-slab analog (MI). However, the forecasting comparison in the next section instead uses a restricted version of the MI (RMI). I don't see why this restricted version is necessary, especially in light of the comments above.

3. Previous work on dynamic shrinkage has shown that gains from these efforts are often in uncertainty quantification (i.e., narrower credible intervals that maintain nominal coverage), rather than point estimation or forecasting. Given that the proposed method does advertise dynamic shrinkage, it would improve the simulation study to report mean credible interval widths and empirical coverage for the regression coefficients, and to compare them to dynamic shrinkage alternatives.

4. The forecasting results seem to favor the proposed approach over the competing methods. However, it is not clear to me why the results should be so significant: these methods are estimating very similar models, and the main difference is one of computing. What is the explanation for such dramatic improvements in performance? Can this be explained better through the simulation study?

5. A related concern is the sensitivity to hyperparameters, which is a notorious challenge for spike-and-slab priors. Does that issue arise for this prior?

6. Given the lack of dynamics in Figure 6, it would be useful to include a non-dynamic Bayesian regression model for benchmarking.
7. Figure 1 shows an unusual trend in that the mixture indicator is trimodal: zero, one, and some other number. This behavior seems to differ from usual spike-and-slab structures. How should one interpret the peak between zero and one, and is it useful for modeling?
8. Inefficiency factors (Figure 5) are largest when t is small. Is there a way to improve the MCMC efficiency for these time points?
9. It is noted that three predictors "appear significant" (page 18). Can this be made more precise?