Referee Report on A Computationally Efficient Mixture Innovation Model for Time-Varying Parameter Regressions

Summary

This paper develops a dynamic shrinkage prior for time-varying parameter regressions. This is motivated by the desire to allow for time-varying parameters in settings with many predictors, while overcoming the computational burden that arises in a standard mixture innovation model. The idea is to shrink the time-varying parameters approximately to zero via the logistic function (rather than exactly to zero in a mixture innovation model). This new dynamic shrinkage prior is illustrated with two empirical applications on equity premium and inflation forecasting.

Comments

1. Time-varying parameter regressions are widely used, and increasingly in applications with a large number of predictors. Therefore, I think there's value in developing dynamic shrinkage priors that are not too computationally intensive to implement. The new prior is motivated as a computationally feasible alternative to the mixture innovation model by using a continuous mixture rather than a discrete mixture. As such, this new prior is closer to the class of continuous scale mixtures shrinkage priors than the mixture innovation model. (For example, the new prior does not retain the nice interpretation of the mixture innovation model as Bayesian model averaging over a large number of specifications). It would therefore be better to reposition the new prior as a member of the class of continuous mixtures shrinkage prior.

2. The idea of using a continuous mixture rather than a discrete mixture to speed up computation has been around for more than a decade (e.g., Griffin and Brown, 2010; Polson and Scott, 2010). And there are by now a large literature on dynamic shrinkage in TVP settings using continuous scale mixture priors. Other continuous scale mixture priors are certainly computationally feasible for settings considered in this paper. Hence, the paper would need to do a better job to motivate

the new prior. For instance, I'm not able to find any theoretical motivation/discussion on why the new prior is better than other continuous dynamic shrinkage priors. Such information would help the readers better appreciate the contribution of the paper.

3. In addition to the dynamic horseshoe prior cited in the paper, the new prior is also similar to the normal-gamma prior for TVP-VARs in Pruser (2021). In both cases the error variance in the state equation, v_j, follows a gamma distribution. It would therefore be useful to discuss the theoretical differences between the two priors and compare them empirically.

4. There's some discussion on a certain conditional distribution of d_{it} in Section 2.2 and I appreciate that. But the case is too special ($\rho = 0$ and so there's no dynamics). It would be useful to work out, for example, the conditional distribution of d_t given d_{t-1}, so that it can be compared to other dynamic shrinkage priors (e.g., dynamic horseshoe). Even though this conditional distribution probably can't be derived analytically, it would still be possible to plot the density for some selected values of d_{t-1} (and other parameters) using numerical integration.

References

Griffin, J. and P. Brown (2010). Inference with normal-gamma prior distributions in regression problems, Bayesian Analysis, 5(1), 171–188

Polson, N. and J. Scott (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction, Bayesian statistics, 9, 501–538.

Pruser, J. (2021). The horseshoe prior for time-varying parameter VARs and monetary policy, Journal of Economic Dynamics and Control, 104188.