

Partial Replication of “Economic predictions with big data: the illusion of sparsity”

Mahyar Habibi
Yekta Yazdanifard

January 2022

We tried to replicate the main estimation function in the Giannone et al. (2021). The replication was successful in modeling the simulated data. However, the results diverged from those in the paper when we used the paper’s datasets.

1 Summary of the paper

Up until recently, constructing a model for predicting an outcome of interest involved often studying the pattern of changes of only a handful of explanatory variables. However, the availability of large datasets and developments in statistical and machine learning methods have opened the way to use high dimensional data in predictive problems. These methods can be generally divided to *Sparse* and *Dense* classes of models. Sparse models such as Lasso, revolve around finding a small subset of predictors with highest explanatory power and dense-modelling techniques aim to use all predictors in explaining the variation in the outcome. But to do so, they *shrink* (toward zero) the estimates predictor so that they can perform in high dimensional setting.

Giannone et al. study six widely used datasets in economics, to find out which class of the models above is probably more fitted in tasks that involve making predictions of economic outcomes. To this end, they specify a “spike-and-slab” prior that can be informative of the degree of sparsity in a linear model. In such a framework, there is a

hyper-parameter q which denotes the probability of regression coefficients being non-zero. Consequently, small q 's lead to sparse representations and large q 's create dense ones. If the coefficient is not zero, Giannone et al. model it to be a draw from the Gaussian distribution with variance γ^2 . The hyper-parameter γ^2 controls the degree of shrinkage in the model. The lower the variance of the Gaussian distribution is, the shrinkage is stronger, and vice versa. Overall, q and γ^2 taken together determine the degree of sparsity and shrinkage in the model.

Giannone and co-authors conduct Bayesian inference on the hyper-parameters of their model and consider a joint prior on q and γ^2 to not pre-impose the assumption of sparsity/density on the data. Then they use six widely used datasets with application in macroeconomics, microeconomics and finance to estimate their model. The posterior results suggest that prediction in economic is typically not characterized by sparsity. Out of six applications, only the posterior result of the first microeconomics dataset, containing the crime rates in a cross-section of US states, favors an sparse representation.

Giannone, Lenza and Primiceri also observe that the data are not very informative on the identity of predictors that must be included or disregarded in the model. Therefore, they conclude that ignoring model uncertainty and naively assuming sparsity in predictive problems might mislead the researchers by creating an “illusion of sparsity”.

2 Replication

We tried to replicate the main estimation function of the paper, called *SpikeSlabGLP* in *Julia*. Although the replicated function was able to produce appropriate results when tested on simulated data, it produced results which were substantially different from the paper when actual datasets were used. We spent considerable time trying to figure the source behind this malfunctioning. But, unfortunately we have been unsuccessful in identifying the problem. The replicated function's results is very close to those of the original function in the first iteration of the model. But, quickly diverges after the next few iterations.

References

Giannone, D., Lenza, M., Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity.