Sufficient Forecasting Using Factor Models

J. Fan[1], **L. Xue**[2*] and J. Yao[3]

[1] *Princeton University; jqfan@princeton.edu*
[2] *The Pennsylvania State University; lzxue@psu.edu*
[3] *Citadel LLC; jiaweiy@princeton.edu*
[*]*Presenting author*

We consider forecasting a single time series when there is a large number of predictors and a possible nonlinear effect. The dimensionality was first reduced via a high-dimensional approximate factor model implemented by the principal component analysis. Using the extracted factors, we develop a novel forecasting method called the sufficient forecasting, which provides a set of sufficient predictive indices, inferred from high-dimensional predictors, to deliver additional predictive power. The projected principal component analysis will be employed to enhance the accuracy of inferred factors when a semi-parametric approximate factor model is assumed. Our method is also applicable to cross-sectional sufficient regression using extracted factors. The connection between the sufficient forecasting and the deep learning architecture is explicitly stated. The sufficient forecasting correctly estimates projection indices of the underlying factors even in the presence of a nonparametric forecasting function. The proposed method extends the sufficient dimension reduction to high-dimensional regimes by condensing the cross-sectional information through factor models. We derive asymptotic properties for the estimate of the central subspace spanned by these projection directions as well as the estimates of the sufficient predictive indices. We further show that the natural method of running multiple regression of target on estimated factors yields a linear estimate that actually falls into this central subspace. Our method and theory allow the number of predictors to be larger than the number of observations. We finally demonstrate that the sufficient forecasting improves upon the linear forecasting in both simulation studies and an empirical study of forecasting macroeconomic variables.