KIEP Opinions

www.kiep.go.kr

Machine Learning in Macroeconomics: Forecasting in the Presence of Instabilities

Yaein BAEK
Ph.D., Associate Research Fellow, International Macroeconomics Team
Korea Institute for International Economic Policy

In the presence of instabilities in economics, such as the global financial crisis in 2008 and the recent COVID-19 pandemic, we face challenges in forecasting macroeconomic and financial variables. Predictors that have been useful in normal times would be different from those at the time of crisis, which implies that the performance of traditional forecasting methods is likely to fail.

Recently forecasters have turned their attention to Machine Learning (ML), which refers to automated predictive algorithms that are able to deal with a large number of models and predictors and/or describe nonlinear mappings nonparametrically. ML has become an important estimation and forecasting tool due to availability of “big data” in economics, and an increasing number of macroeconomic studies show successful forecasting performance applying ML methods.

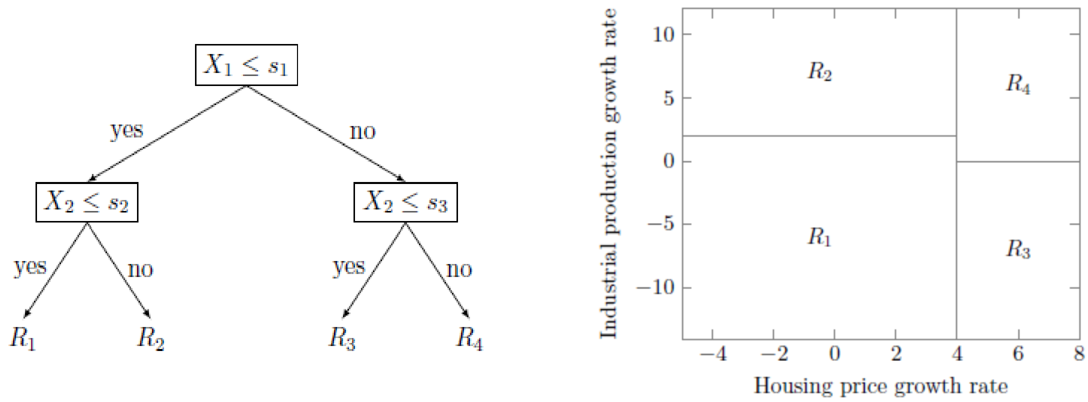
Which ML method do we use for macroeconomic forecasting? ML methods can be divided into three groups: supervised, unsupervised, and reinforcement learning. Most studies in economics apply supervised ML methods, where the task is to learn a function that maps an input (independent variables) to an output (dependent variable) based on data organized as input-output pairs. Supervised ML methods can be categorized into three groups: linear, nonlinear, and ensemble-based methods (Masini et al. 2021).

Linear ML methods are largely composed by penalized regression, such as Ridge (Hoerl and Kennard 1970), LASSO (Tibshirani 1996), adaptive LASSO (Zou 2006), and elastic net (Zou and Hastie 2005). Studies have applied penalized regression to forecast macroeconomic variables such as inflation rate and GDP growth, and show that these methods can outperform conventional forecasting tools. There are also applications on forecasting financial variables such as the equity premium and realized volatility. Penalized regression is a way to aggregate all available data while performing shrinkage – model parameters are replaced with smaller or zero values to handle the large dimension of the data. The method shrinks unimportant parameters toward zero to induce parsimony in the model. Hence, forecasters do not have to worry about selecting predictors in a large data set, which is particularly useful in periods of economic uncertainty.

The second group of supervised ML are nonlinear methods such as neural networks and random forest (RF). Most of the early literature focus on neural network methods, and with the recent availability of large data sets, RF has gained attention. Medeiros et al. (2021) show that ML models with a large number of covariates are systematically more accurate than benchmarks in forecasting US inflation, and among them RF outperform all others. In particular, the authors show that the gains of the RF model are larger during recessions and periods of instability, especially during and after the Great Recession in 2007-2009. To go into detail, RF was proposed by Breiman (2001) to reduce the variance of regression trees by bootstrap aggregation (bagging) of randomly constructed regression trees. A regression tree is a nonparametric model that approximates an unknown nonlinear function with local predictions using recursive partitioning of the space of the covariates (Breiman 1996). Consider a simple example: we are interested in predicting Y based on two variables X_1 and X_2. As illustrated in the left side of Figure 1, the first node of the tree splits the sample according to whether X_1 is larger than s_1 or not. The second node in the left (right) takes the sub-sample with low (high) X_1 values and splits it by X_2. The final result is partitioning into four groups, each corresponding to a terminal node of the tree, and the sample average of Y is computed for each group. A prediction is made by following a path in the tree, where the optimal splitting point in each node ($s_i, i = 1,2,3$) is determined to minimize the sum of squared errors. For instance, suppose we are interested in forecasting U.S. GDP growth rate (Y) one-year-ahead based on housing price growth rate

(X_1) and industrial production growth rate (X_2). The second node of the tree in Figure 1 takes the low housing price growth sub-sample and splits them by industrial production growth, and the second node in the right does the same to the high housing price growth sub-sample. Suppose the optimal splitting points are estimated as $s_1 = 4\%$, $s_2 = 2\%$, and $s_3 = 0\%$ based on data spanning from 1975 to 2021, so that there are four groups as in the right side of Figure 1. The U.S. housing price growth at 2021Q1 is 5.5% and the industrial production growth rate is 1.6%, which fall in R_4. Hence, the forecast of the 2022Q1 GDP growth rate is equal to the average GDP growth rate of 1975-2021 that falls within R_4.

Figure 1. Example of a Regression Tree



Although regression trees are intuitive and capture nonlinearity, they have a tenancy to overfit. The RF can overcome this by growing many trees on bootstrap samples of the original data, with further randomization obtained by randomly selecting a subset of original covariates. The RF forecast is the average of the forecasts of each tree applied to the original data. Medeiros et al. (2021) provide evidence that both nonlinearity and variable selection play a key role in the superiority of the RF, in which instabilities in economics are an important source of nonlinearity. These properties are again emphasized in Borup and Schütte (2022), which forecast U.S. employment growth using Google Trends data. ML takes advantage of the data-rich environment formed by Google search activity, which results in good performance of forecasts in non-normal times.

The third group is ensemble-based methods that include Bagging (Breiman 1996; Inoue and Kilian 2008) and complete subset regression (Elliott et al. 2013). There are also other ML "hybrid" methods that combine ideas from linear and nonlinear ML models to generate forecasts.

Likewise, there are numerous ML models that are not introduced here, and further developments are ongoing. However, it is important to keep in mind that the theoretical properties for most ML methods are not well known for time-series, which is the context that is relevant for macroeconomic forecasting. The majority of the theory has been developed only to independent and identically distributed data, and a large component of it assumes orthogonal regressors. The theoretical properties of ML that makes it useful in economic forecasting in the presence of instabilities are yet to be uncovered. Nonetheless, we have witnessed the recent empirical evidence of successful macroeconomic forecasting performed by nonlinear ML combined with large data. As much as we expect to see further developments in the field of ML for macroeconomic forecasting, it is important that forecasters can make the most of ML techniques. KIEP

References

Borup, D. and E. Schütte. 2022. "In search of job: Forecasting employment growth using Google trends." *Journal of Business & Economic Statistics,* 40, 186-200.

Breiman, L. 1996. "Bagging predictors." *Machine Learning,* 24, 123-140.

Breiman, L. 2001. "Random forests." *Machine Learning,* 45, 5-32.

Elliott, G., A. Gargano, and A. Timmermann. 2013. "Complete subset regressions." *Journal of Econometrics,* 177(2), 357-373.

Hoerl, A. and R. Kennard. 1970. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics,* 12(1), 55-67.

Inoue A. and L. Kilian. 2008. "How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation." *Journal of the American Statistical Association,* 103, 511-522.

Masini, R. P., M. C. Marcelo, and E. F. Edurado. 2021. "Machine learning advances for time series forecasting." *Journal of Economic Surveys*, 1-36.

Medeiros, M. C., G. F. Vasconcelos, Á. Veiga and E. Zilberman. 2021. "Forecasting inflation in a data-rich environment: the benefits of machine learning methods." *Journal of Business & Economic Statistics,* 39(1), 98-119.

Tibshirani, R. 1996. "Regression shrinkage and selection via the LASSO." *Journal of the Royal Statistical Society: Series B(Methodological),* 58(1), 267-288.

Zou, H. 2006. "The adaptive LASSO and its oracle properties." *Journal of the American Statistical Association,* 101, 1418-1429.

Zou, H. and T. Hastie. 2005. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B(Statistical Methodology),* 67(2), 301-320.