

Forecasting US recessions using large scale data sets and machine learning algorithms

Barış Soybilgen

Istanbul Bilgi University, Turkey
baris.soybilgen@bilgi.edu.tr

Abstract:

Identifying the start of a US recession as early as possible is crucial for both market participants and policymakers. However, it is a well-known fact that it is hard to forecast US recessions accurately in a timely fashion. Usually, the best models to predict US recessions are simple probit models including yield spread and stock market returns. In this study, we analyze whether using large scale data sets with state art of machine learning algorithms improve over basic models. We use three main types of machine algorithms: tree-based algorithms, neural network-based algorithms, and support vector machine variants. As the large-scale data set, we use the FRED-MD data set including more than 100 financial and macro variables. We further use dimension reduction techniques such as factor models, autoencoders, and lasso models before feeding large scale data sets into the machine learning algorithms. Results show that neural network models usually have the highest prediction performance. Support vector machines and tree-based models don't yield desired results. De-noising data set and feature extraction don't improve results in general. In many cases, it is hard to beat simple models.

Keywords: Machine Learning, Recessions, Forecasting, Dimension Reduction, USA

JEL Codes: E37, E32, C38