Nowcasting US GDP: A Machine Learning Approach

Davis Berlind, Alex Marsh and Stella McMullen

Department of Computer Science & Department of Economics, Duke University

Objective

To create an automated nowcasting model that produces real-time estimates of United States real GDP growth using a variety of machine learning algorithms, and evaluate their performance.

Motivation

Many key economic variables are released with considerable delay, with US GDP released 3 months after the end of the quarter. This means that decision makers rely on imperfect information about current economic conditions. Machine learning algorithms have great potential to be applied to large, high frequency data sets to obtain earlier estimates of economic variables such as GDP. This is called *nowcasting*.

Data set

We use a large data set of daily, monthly, and quarterly economic and financial statistics from Federal Reserve Economic Data (FRED). We choose variables to include based on Stock and Watson (2002), Richardson et al. (2018), Jung et al. (2018), and Bańbura et al. (2010). We classify variables following Stock et al. (2002). Our final data set had over 1,000 variables, over the January 1990 to April 2018 period.

Method

- Dataset is downloaded from FRED using an API, cleaned, and transformed (following Bańbura et al. (2010)) (all automated)
- Higher frequency variables (monthly, weekly, and daily) are projected out to the end of the quarter using an autoregressive model
- Higher frequency data (actual and projected) are aggregated to the quarterly level. This method of *bridge equations* is recommended in the literature and described in Barhoumi et al. (2008)
- Machine learning algorithms are applied to the aggregated data set to produce nowcasts of real GDP growth.
  - *Algorithms: Elastic net, Ridge regression, Bayesian vector autoregression, Gaussian process regression, and Bayesian linear regression (see Murphy (2013) and Ankargren et al. (2018) for an explanation of these models)*
- The performance of the nowcasts are compared to the performance of an autoregressive model

We train each algorithm over an expanding window from 2009Q1 to 2019Q1, with the training set beginning in 1990Q1.

Forecast Evaluation

The forecast accuracy of each model is calculated as the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_t = 1^T (y_t - \hat{y}_t)^2}{T}} \tag{1}$$

Results

Model	RMSE
AR	0.0003262
Elastic Net	0.0015681
Ridge Regression	0.0009271
Gaussian Process Regression	0.0009271
Bayesian Linear Regression	0.0009271
Bayesian Vector Autoregression	0.0009271

Table 1: Nowcast performances of models (RMSE) '09-'19

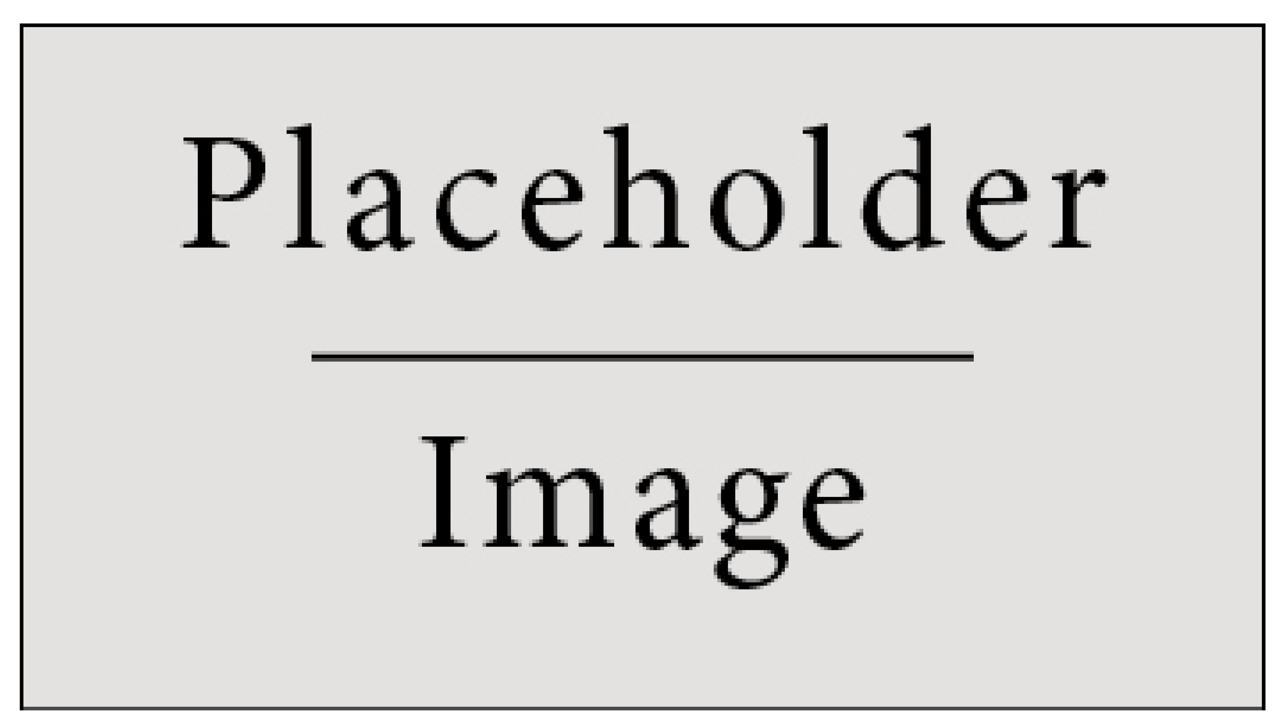

Figure 1: Progression of 2018Q4 GDP estimate over quarter

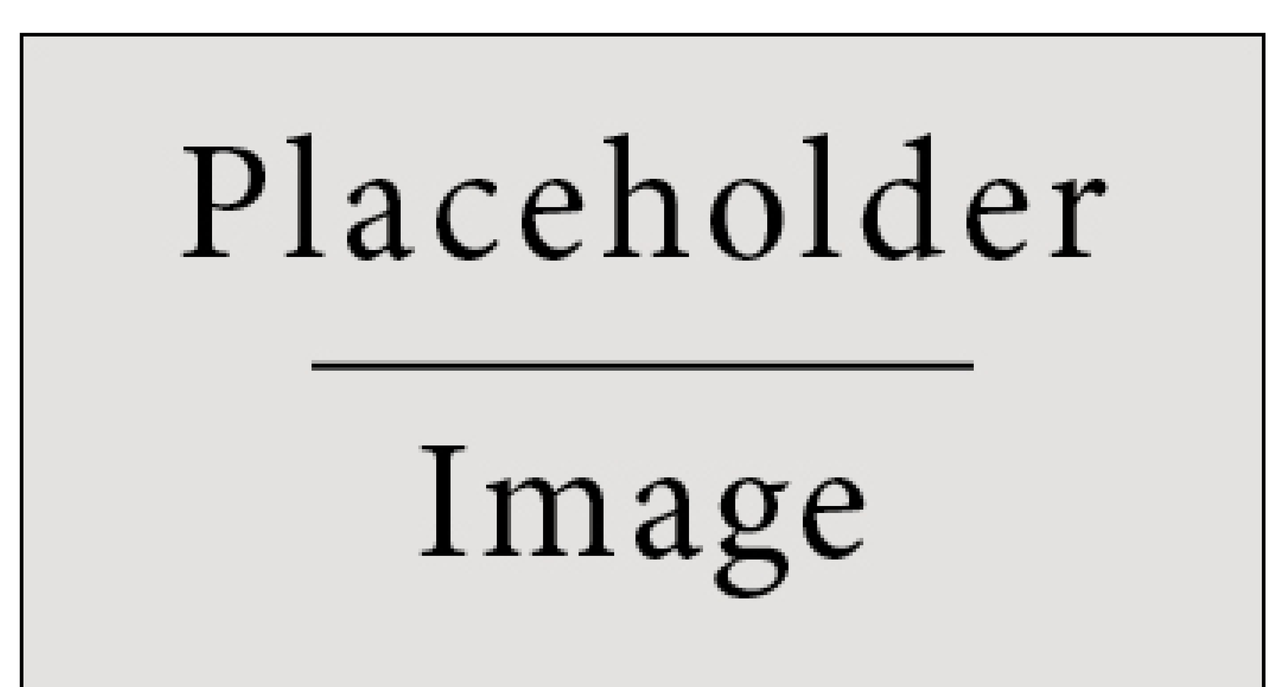

Figure 2: Quarterly GDP growth and its nowcasts (2009Q1:2019Q1)

Figure 3: Posterior distribution for Bayes - 2018Q4 GDP growth

Figure 4: Regularization paths

Discussion

We find that the majority of Machine Learning algorithms are able to produce more accurate forecasts than the AR model. Automated machine learning based nowcasting models are promising forecasting methods.

Conclusion

We find that the majority of Machine Learning algorithms are able to produce more accurate forecasts than the AR model. Automated machine learning based nowcasting models are promising forecasting methods.