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 has over 1,000 variables, reduced to over 600 over the January 1990 to April 2018 period.

Method

- Dataset is downloaded from FRED using an API, cleaned, and transformed so it is stationary (following Bańbura et al. (2010)) (all automated). Lags of variables are included.
- 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. Performance is measured by mean-square error, root-mean-square error and mean absolute error.

We train each algorithm over an expanding window from 2009Q1 to 2019Q1, with the training set beginning in 1990Q2.

Results

Model	MSE	RMSE	MAE
AR	1.972e-05	4.441e-03	3.644e-03
Elastic Net	2.335e-05	4.832e-03	3.580e-03
Ridge Reg.	2.005e-05	4.478e-03	3.5829e-03
Gaussian Process Reg.	2.078e-03	4.559e-03	3.508e-03
Bayesian Linear Reg.	1.929e-05	4.392e-03	3.517e-03

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

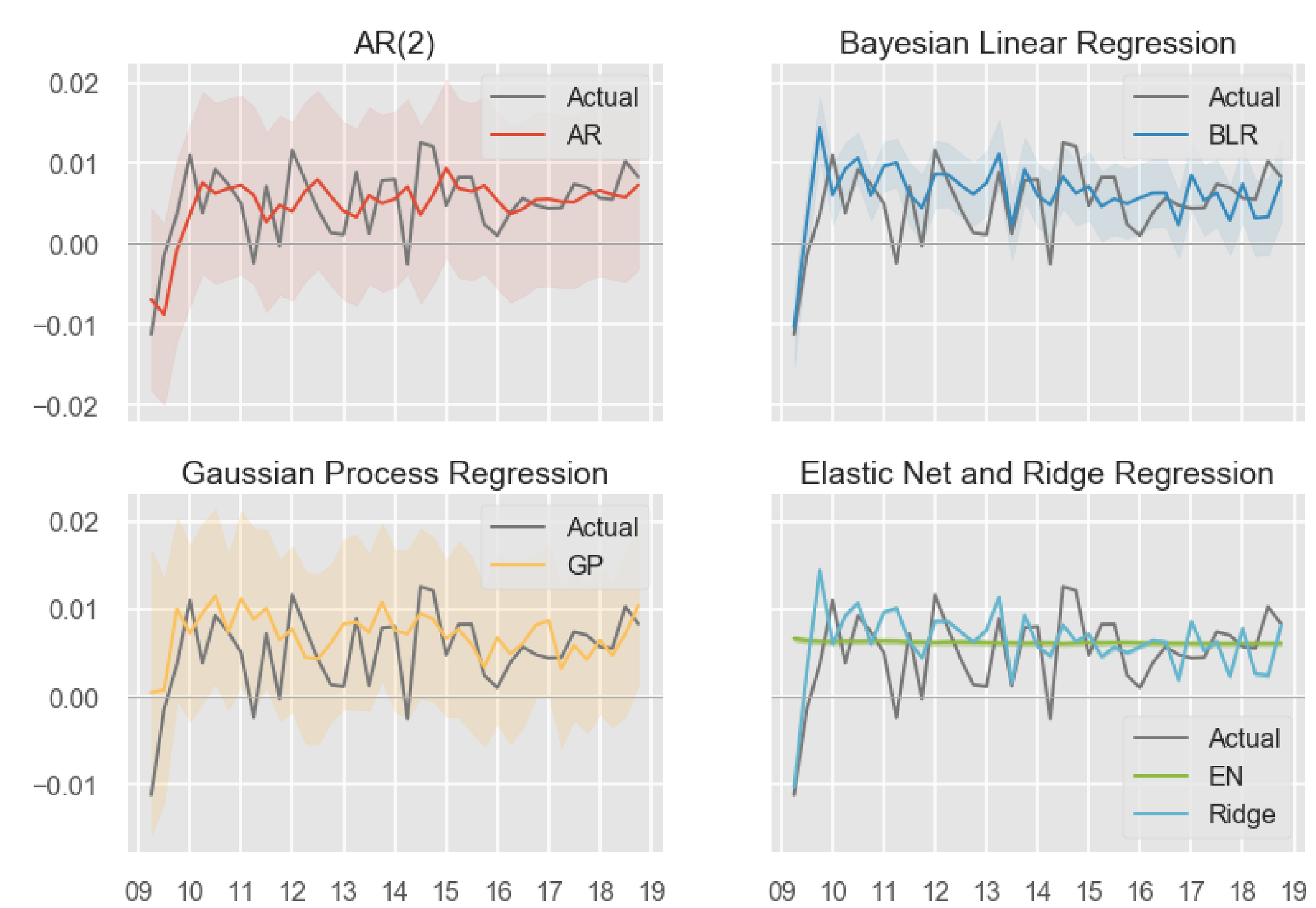

Figure 1: Quarterly US GDP growth and its nowcasts (2009Q1:2018Q4)

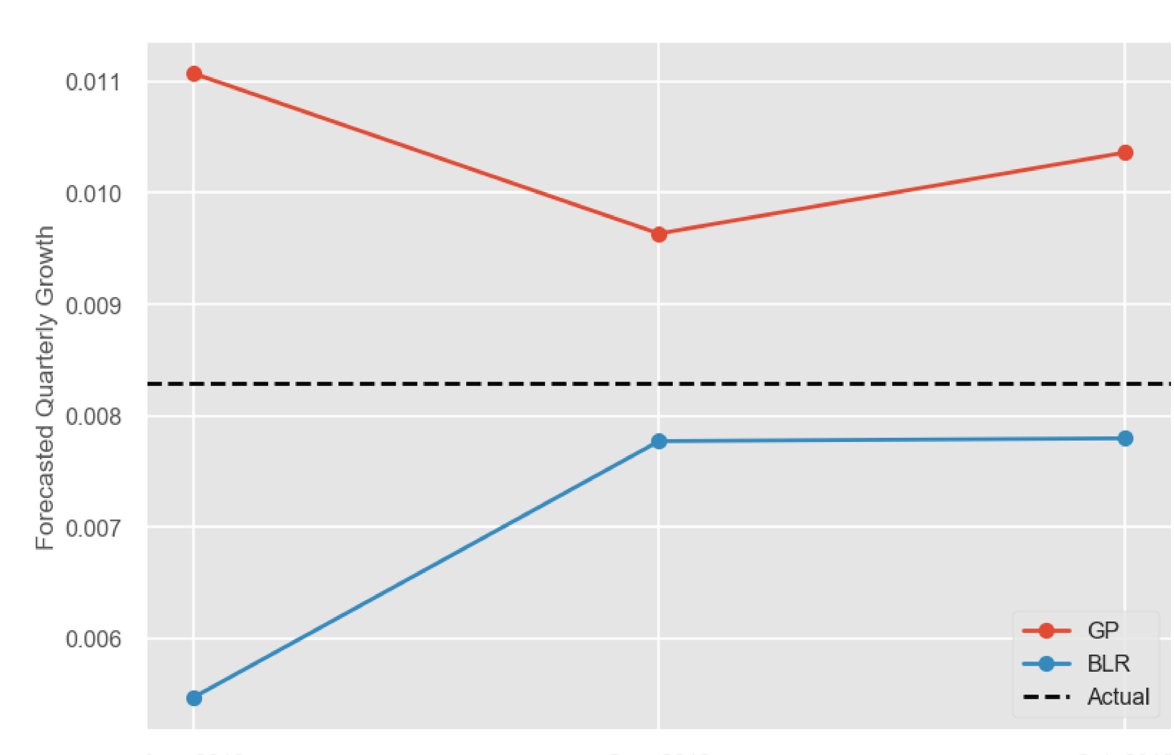

Figure 2: Progression of GDP nowcasts over 2018Q4

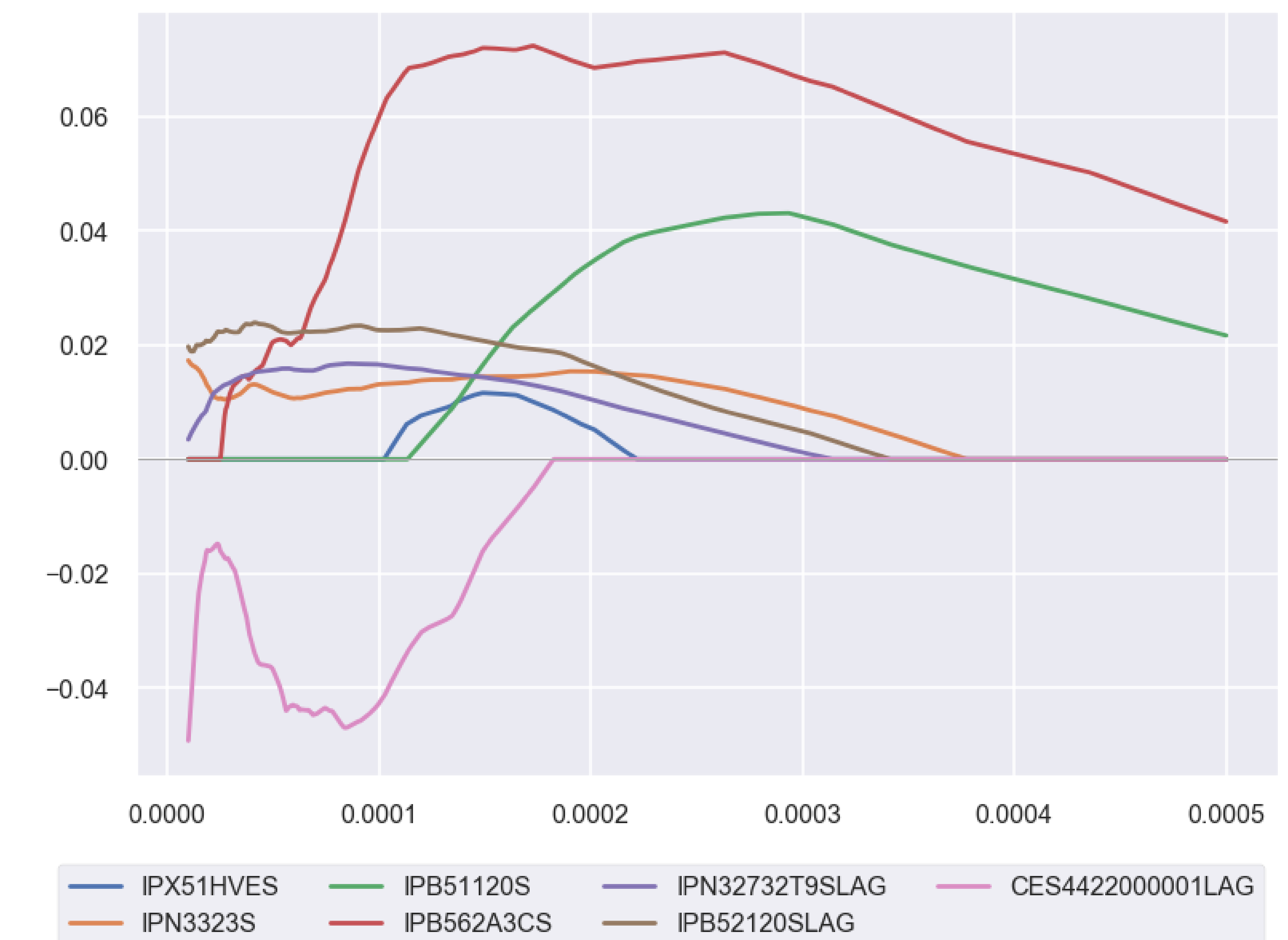

Figure 3: Regularization Path: Elastic Net Regression

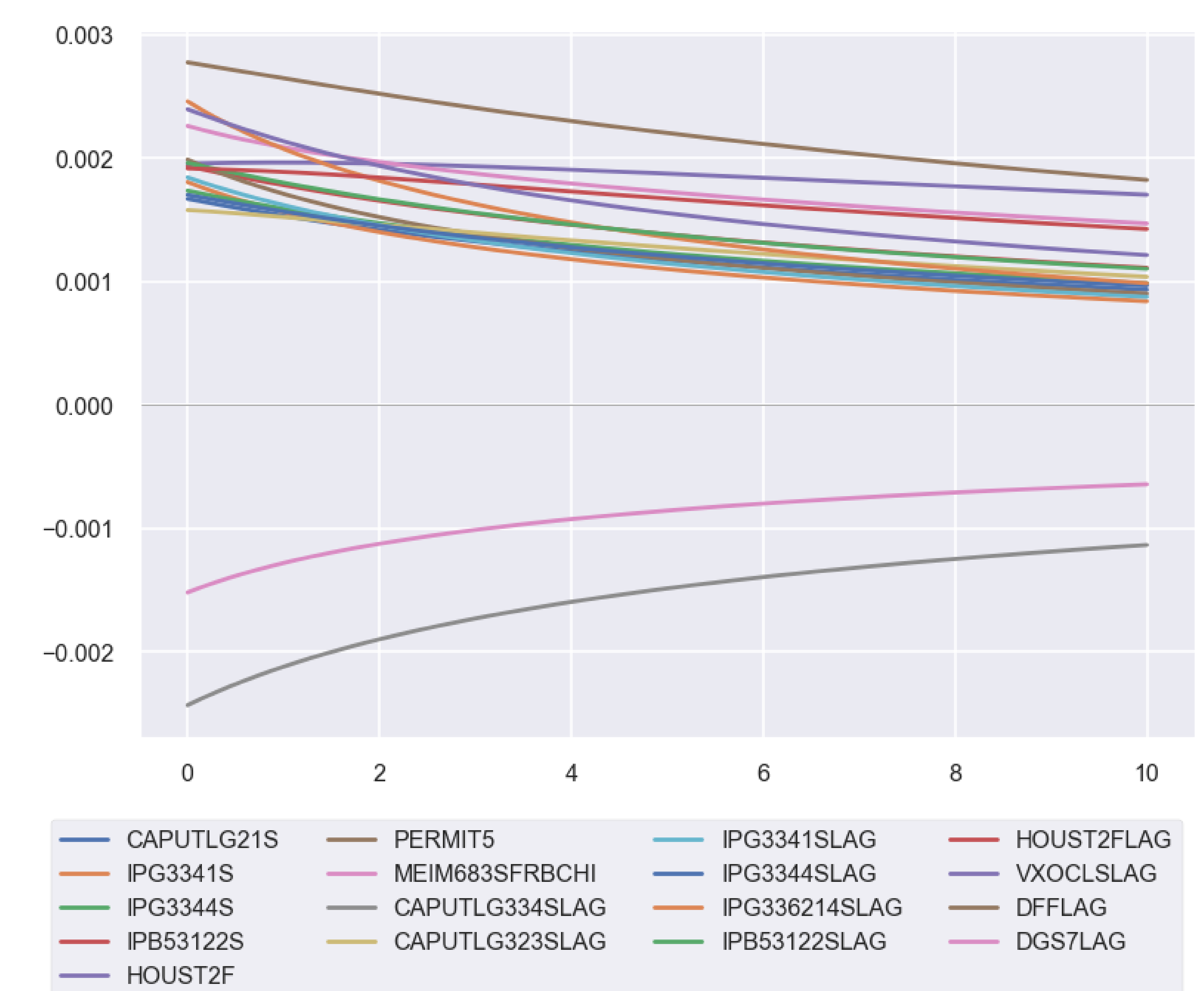

Figure 4: Regularization Path: Ridge Regression

Conclusion

We find that the Bayesian Linear Regression outperforms the AR model. Automated machine learning based nowcasting models are promising forecasting methods. The regularization paths both seem to favor industrial production variables. In future research, we would like to explore the use of additional machine learning techniques, including neural networks. We would also like to investigate the use of data from additional sources, and potentially refining our choices of variables to include in our data set.