Nowcasting US GDP: A Machine Learning Approach

Davis Berlind, Alex Marsh, Stella McMullen

April 25, 2019

Abstract

As many key economic variables are released with considerable delay, there is imperfect information about the current state of the economy. 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, called 'nowcasts'. This paper considers the implementation and performance of machine learning algorithms to create an automated nowcasting model for United States real GDP growth. We apply Elastic Net, Ridge Regression, Bayesian linear regression, and Gaussian process regression algorithms to a large set of quarterly, monthly, and daily maroeconomic indicators over the 2009Q2-2018Q4 period. We compare the predictive accuracy of these nowcasts with that of an autoregressive model, the St. Louis Fed Real GDP Nowcast, and the New York Fed Staff Nowcast. We find that Bayesian Linear regression outperforms the autoregressive model and other algorithms considered, and conclude that machine learning algorithms have great potential to improve early estimates of GDP.

1 Introduction

Due to delays in the release of information by governmental organizations, there is imperfect information about the current state of the economy. The final estimate for US Gross Domestic Product (GDP) for the first quarter of 2019 will not be released until June 27, 2019, three months after the end of the quarter (BEA, 2019). Nowcasting models attempt to address the problem of delayed information by predicting the state of economic variables in the recent past, the present, and the near future. These models have become increasingly popular tools for forecasters and institutions when attempting to provide real-time information on current economic conditions, and have been implemented in many countries including New Zealand, the United States, Germany, Mexico, Philippines, Spain, the United Kingdom, and Germany (Richardson et al., 2018; Giannone et al., 2008; Jung et al., 2018).

Using methods from machine learning in conjunction with high-dimensional data sets has proven to be a viable approach for generating nowcasts (Murphy, 2013). The nowcasting literature features the use of clustering algorithms, regularized regression, Support Vector Machines, and Neural Networks (Richardson et al., 2018). Many of these approaches to forecasting have delivered promising results, with Richardson et al. (2018) and Jung et al. (2018) both finding that machine learning methods can outperform the traditional statistical models used to forecast economic data.

In this paper, we investigate the performance of various regression models within the scope of nowcasting US GDP, including Elastic Net, Ridge, Gaussian process, and Bayesian linear regression. We use a wide data set of domestic and international economic variables covering the period of 2009Q1-2018Q4 in order train our models. Furthermore, we look at a simple autoregressive model as a baseline for comparing the results of the machine learning approaches we have specified.

We find that Bayesian Linear regression outperforms the autoregressive model and other regression methods considered, and conclude that machine learning approaches have the potential to improve early estimates of GDP.

This paper follows Richardson et al. (2018), who use a large macro data set to nowcast New Zealand GDP and Jung et al. (2018), who use a large macro data set to nowcast GDP in seven broadly representative advanced and emerging economies. We add to this literature i) by comparing the performance of machine learning algorithms to the performance of nowcasting models implemented by the Federal Reserve and other government organizations in the US, and ii) by exploring the use of Bayesian linear regression and Gaussian process regression, which allows us to provide predictive distributions of unobserved GDP growth. The incorporation of such variability is beneficial to policy makers attempting to quantify and plan for uncertainty in economic forecasts (Christensen et al., 2018).

This paper is structured as follows. Section 2 outlines our nowcasting model methodology, Section 3 outlines our results, Section 4 concludes, and Section 5 discusses future research opportunities.

2 Methods

2.1 Data

We use a large data set of 677 economic and financial variables beginning in April 1990 and ending in October 2018. We gathered the data from the Federal Reserve Economic Data (FRED) database maintained by the St. Louis Fed using the FRED API (so as to automate data collection and cleaning). The data occurs at daily, weekly, bi-weekly, monthly, and quarterly frequencies. Additionally, we include a one quarter lag for all variables in the training data set and two lags of quarterly GDP growth, making the total width of our design matrix 1,356 variables. We chose the variables to include based on Stock and Watson (2002), Richardson et al. (2018), Jung et al. (2018), and Bańbura et al. (2010). Following Stock and Watson (2002), we classify macroeconomic variables into a number of categories such as consumption and price indexes; these categories and examples can be found in Table 2 of Appendix A.

2.2 Nowcasting Model

After downloading the data, we perform transformations to achieve stationarity as is recommended in the literature (see Bańbura et al. (2010)). This mostly involves calculating the percent changes between observations in each data series. We use a Dickey-Fuller test to confirm stationarity.

Data series occurring at higher than quarterly frequencies are projected out to the end of the quarter using an AR(2) model. Both the actual and projected high frequency data is then aggregated up to the quarterly level. This method of *bridge equations* is widely used in the literature to forecast quarterly GDP from higher frequency data (Rünstler et al., 2009).

We then train each of the models on an expanding window of training data that is explained in-depth in the following section. For the purpose of model validation, we predict GDP growth for the the first quarter out of the training set using each model. The models we train include:

- **Autoregression (baseline):** an AR(2) model.
- **Bayesian Linear Regression:** with non-informative hyperparameters for the prior gamma distributions on α (the penalty term) and λ (the model precision) (i.e. $\alpha_1 = \alpha_2 = \lambda_1 = \lambda_2 = 10^{-6}$). During fitting, the parameters are estimated by maximizing the marginal log likelihood.
- **Ridge regression:** with penalty $\alpha = 1$.
- **Elastic Net regression**: with the ℓ_1 and ℓ_2 penalties optimized using grid-search cross-validation on the full data set.
- **Gaussian Process regression**: with a sum-kernel composed of a constant kernel (for numerical stability), scaled Gaussian kernel, and White kernel (to explain noise in the data). This kernel was chosen using grid-search cross-validation on the full data set, and the hyperparameters are selected during fitting by maximizing the marginal log likelihood.

Please see Appendix B and Murphy (2013) for further discussion of these models. Note that in the case of Ridge and Elastic Net regression, standard errors for the out-of-sample predictions are estimated using a bootstrap procedure (which also explains why the confidence intervals for these predictions appear to be too tight).

2.3 Forecast evaluation methodology

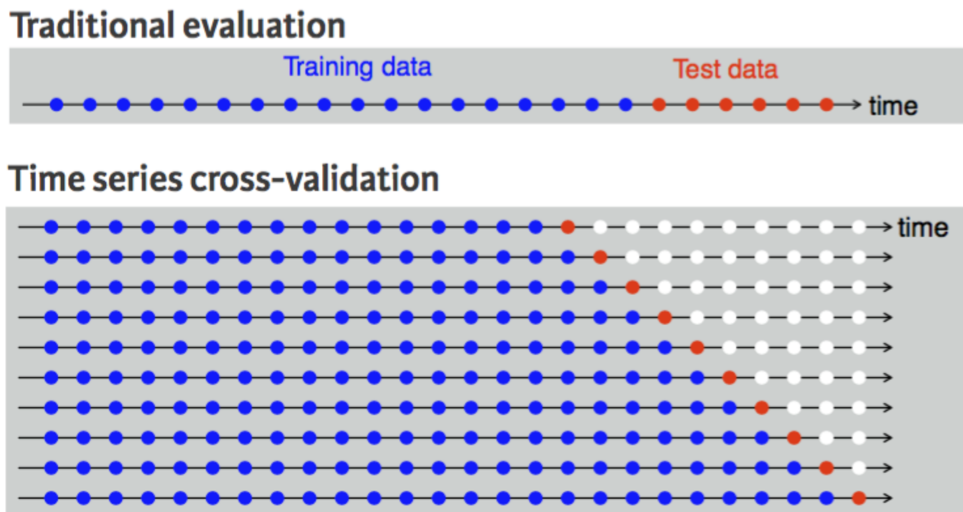

Figure 1: Representation of Time Series Cross Validation

In this section, we discuss our forecast evaluation methodology. We use one-step ahead time series cross-validation to evaluate the performance of our forecasts. We train each model over an expanding window, starting in 1990Q2. Each test set consists of one observation. The corresponding training set consists of all observations prior to the observation that forms the test set, with the first test set being 2008Q4. This method of time series cross validation is depicted in Figure 1.

The forecast accuracy is computed by averaging over the test sets. We use the mean-square error (MSE) root-mean-square error (RMSE) and the mean absolute error (MAE) as measures of forecast accuracy.

3 Results

In this section, we discuss the main results of our analysis. Table 1 shows the performance of the machine learning models and autoregressive model over the 2008Q4-2018Q3 period. We also present the nowcasting predictions of the models and the corresponding 95% prediction intervals in Figure 2. Our results indicate that Bayesian Linear Regression produces nowcasts that outperform the autoregressive model and the other machine learning models in every measure of error considered. Although the AR model generates nowcasts that produce lower MSE and RMSE than Elastic Net, Ridge regression, and Gaussian Process regression over our sample period, the AR ultimately results in a higher MAE.

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-’18

Using higher frequency data, we are also able produce updated estimates of US GDP as additional data is released over the quarter. The evolution of nowcasts for US GDP for 2018Q3 from Gaussian Process Regression and Bayesian Linear Regression are depicted in Figure 3. This is beneficial to decision makers, as nowcasts harness all currently available information.

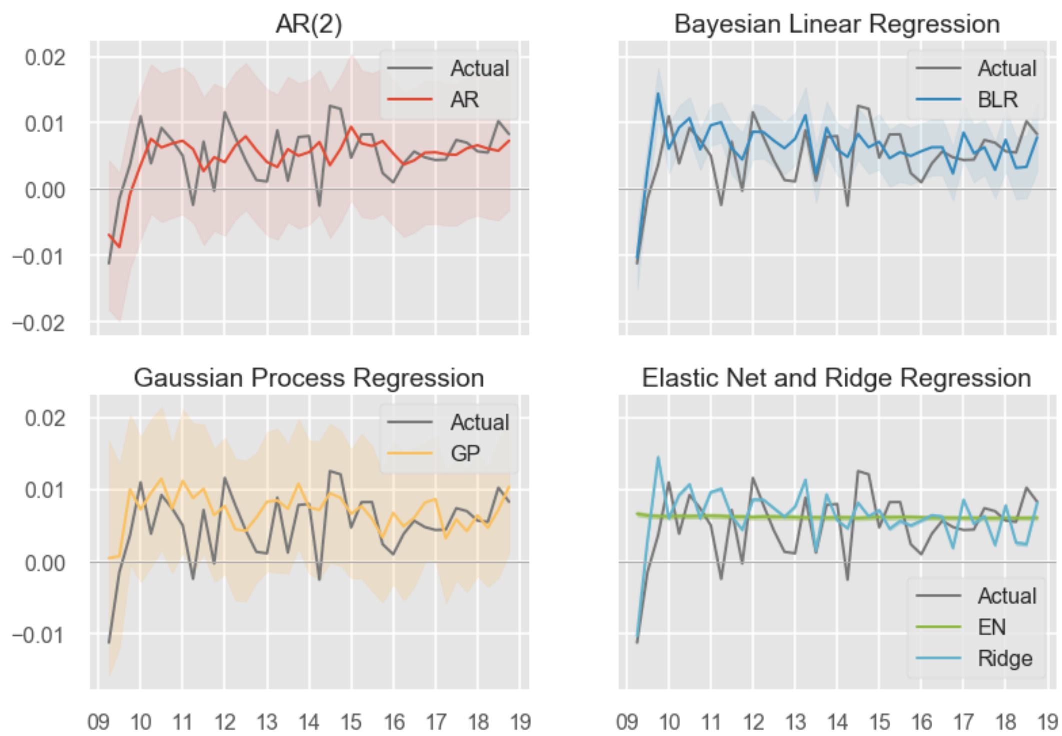

Figure 2: Quarterly US GDP growth and its nowcasts (2009Q1:2018Q3)

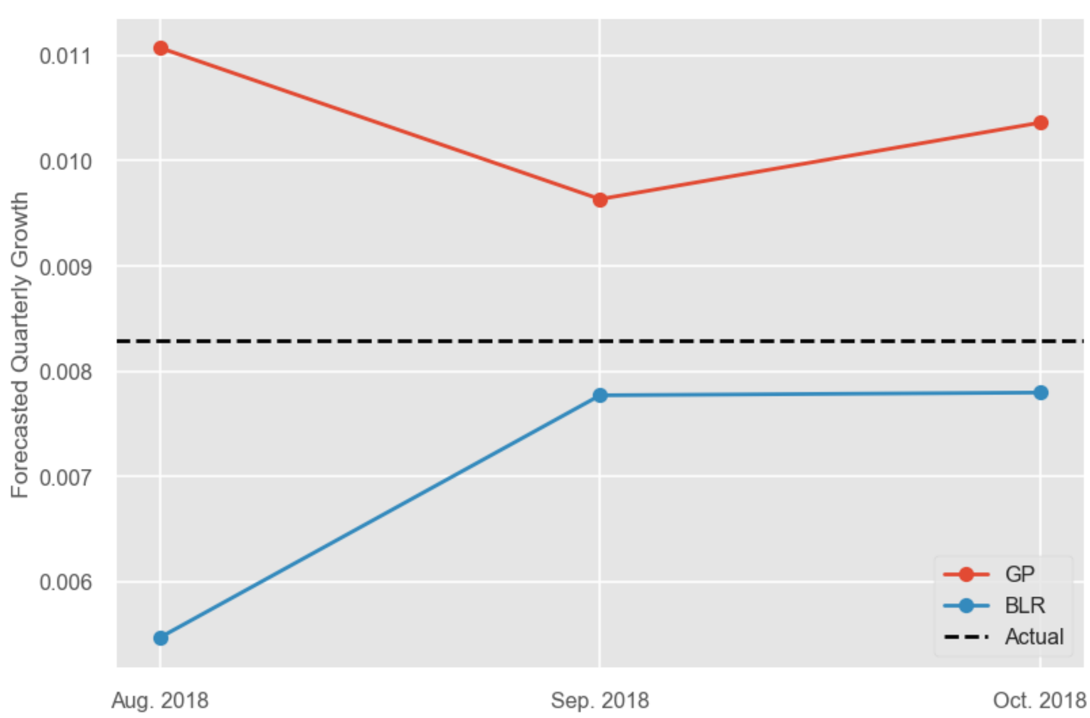

Figure 3: Progression of GDP nowcasts over 2018Q3

4 Conclusions

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 show that both Elastic Net and Ridge Regression (depicted in Figures 4 and 5 of Appendix A) favor industrial production variables.

5 Future Research

We would like to explore the use of additional machine learning techniques, including Bayesian Vector Autoregressive Models (discussed in detail in section 5.1) and neural networks. We would also like to investigate the use of machine learning algorithms such as AdaBoost to produce a combined weighted sum of the other algorithms that we have considered. This is expected to improve the performance of our nowcasts. Further, we would like to complete additional tests regarding the optimality of our forecasts (such as Mincer-Zarnowitz regression) and tests for comparing the predictive accuracy of the models we consider (such as the Diebold-Mariano test). 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.

5.1 Bayesian Vector Autoregressive Models

One method that we played with a bit but could not implement due to computational and time limitations were Bayesian Vector Autoregressive (VAR) models with mixed frequency data. VAR models allow the researcher to set up numerous interdependencies among variables in a system of equations which is ideal in the macroeconomic context where causality is not entirely understood and simultaneity across equations effects almost all observed data. Say there are m variables whose current value is believed to be structurally related to p lags of all the other variables in the system. Let $\boldsymbol{Y}_t' = (y_t^1, ..., y_t^m)$ and $\boldsymbol{Y}_{t-k}' = (y_{t-k}^1, ..., y_{t-k}^m)$ Then we can set up the system as follows:

$$\boldsymbol{Y}_t = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1}\boldsymbol{Y}_{t-1} + ... + \boldsymbol{\Phi_p}\boldsymbol{Y}_{t-p} + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\Phi_0}' = (\phi_0^1, ..., \phi_0^m)$, $\boldsymbol{\epsilon}_t' = (\epsilon_t^1, ..., \epsilon_t^m)$, and

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \phi_{11}^{(k)} & \phi_{12}^{(k)} & \cdots & \phi_{1m}^{(k)} \\ \phi_{21}^{(k)} & \phi_{22}^{(k)} & \cdots & \phi_{2m}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}^{(k)} & \phi_{m2}^{(k)} & \cdots & \phi_{mm}^{(k)} \end{bmatrix}$$

However, because VAR models are so flexible and allow for a rich parameterization of the structural model, they are very prone to overfitting and as such, it is very common to put a prior on the parameters and estimate the model within a Bayesian framework (Karlsson (2013)). This treatment is even popular among economists more prone to frequentist methods simply due to how easy it is to overfit VAR models of the macro-economy. A common prior used is called the Minnesotta prior first, introduced in Litterman (1979). This is a set of prior beliefs derived from the data that shrinks the parameters to reflect the macroeconomic data so that we have better informed forecasts and more certain parameter estimates.

While Bayesian VAR models are the most common methods of forecasting macroeconomic variables, they do not immediately lend themselves to nowcasting as all the variables still must be observed at the same frequency. To allow for mixed frequency data, we must treat the model as a state-space model where the unit of observation is the highest frequency data available and less frequent data are treated as a latent variable when missing. This has been the approach followed in Eraker et al. (2014), Schorfheide and Song (2015) and Ankargren et al. (2018). Furthermore, Ankargren et al. (2018) include a prior over the unconditional mean of the variables so that prior knowledge about the steady state of the economy can be incorporated to improve forecast accuracy. This is particularly useful as policy makers often “target” many macroeconomic variables e.g. the Federal Reserve typically target a 2% inflation rate, economists target a steady unemployment rate (the natural rate of unemployment) of 5%, etc.

Ankargren et al. (2018) published all the code to replicate their results on Zenodo and Github[1]; however, we could not get the code to work with our data. Nevertheless, the model laid out in Ankargren et al. (2018) is a promising method to achieving better nowcasts of GDP.

[1]Link: `https://zenodo.org/record/1145828#.XLj-L-tKjOR`

References

Ankargren, S., M. Unosson, and Y. Yang (2018, November). A mixed-frequency bayesian vector autoregression with a steady-state prior. Technical Report 2018:3, Uppsala University.

Bańbura, M., D. Giannone, and L. Reichlin (2010, December). Nowcasting. Working Paper Series 1275, European Central Bank.

BEA (2019, April). Release schedule: Upcoming releases 2019. `https://www.bea.gov/news/schedule`.

Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti (2018, January). Macroeconomic nowcasting and forecasting with big data. System Working Paper 18-04, Federal Reserve Bank of Minneapolis.

Christensen, P., K. Gillingham, and W. Nordhaus (2018). Uncertainty in forecasts of long-run economic growth. *Proceedings of the National Academy of Sciences of the United States of America 115*(21), 5409–5414.

Dahlhaus, T., J.-D. Guenette, and G. Vasishtha (2017). Nowcasting bric+m in real time. *International Journal of Forecasting 33*(1), 915–935.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*(1), 318–328.

Eraker, B., C. W. J. Chiu, A. T. Foerster, T. B. Kim, and H. D. Seoane (2014). Bayesian mixed frequency vars. *Journal of Financial Econometrics 13*(3), 698–721.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(1), 665–676.

Jung, J.-K., M. Patnam, and A. Ter-Martirosyan (2018, November). An algorithmic crystal ball: Forecasts-based on machine learning. IMF Working Paper 18/230, International Monetary Fund.

Kapetanios, G. and F. Papailias (2018, July). Big data & macroeconomic nowcasting: Methodological review. ESCoE Discussion Paper 2018-12, Office for National Statistics.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2B, Chapter 15, pp. 791–897. Radarweg 29, Amsterdam, 1043 NX, The Netherlands: North Holland.

Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics 186*(1), 294–316.

Kim, H. H. and N. R. Swanson (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics 178*(1), 352–367.

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting 29*(1), 43–59.

Li, J. and W. Chen (2014). Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting 30*(1), 996–1015.

Litterman, R. B. (1979, November). Techniques of forecasting using vector autoregressions. Technical Report 115, Federal Reserve Bank of Minneapolis.

Murphy, K. P. (2013). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

of St. Louis, F. R. B. and US (2019, April). Economic data. `https://fred.stlouisfed.org/`.

Richardson, A., T. v. F. Mulder, and T. Vehbi (2018, September). Nowcasting new zealand gdp using machine learning algorithms. CAMA Working Paper 47/2018, Australian National University: Crawford School of Public Policy.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze (2009). Short-term forecasting of gdp using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting 28*(7), 595–611.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics 33*(3), 366–380.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business Economic Statistics 20*(2), 147–162.

Appendix A: Figures

Variable Type	Examples
Real output and income	Industrial production, capacity utilization
Employment and hours	Unemployment rate, avg. weekly hrs worked
Real retail, manufacturing and trade sales	Total retail trade, total manufacturing
Consumption	Personal consumption expenditure
Housing starts and sales	Housing starts, mobile home construction
Real inventories and inventory-sales	Manufacturing inventories
Orders and unfilled orders	New orders index
Stock prices	S&P500 index
Interest rates	Effective federal funds rate
Money and credit quantity aggregates	M1 money stock
Price indexes	CPI all items
Average hourly earnings	Average hourly earnings of production workers
Miscellaneous	US exports

Table 2: Variable types included in model

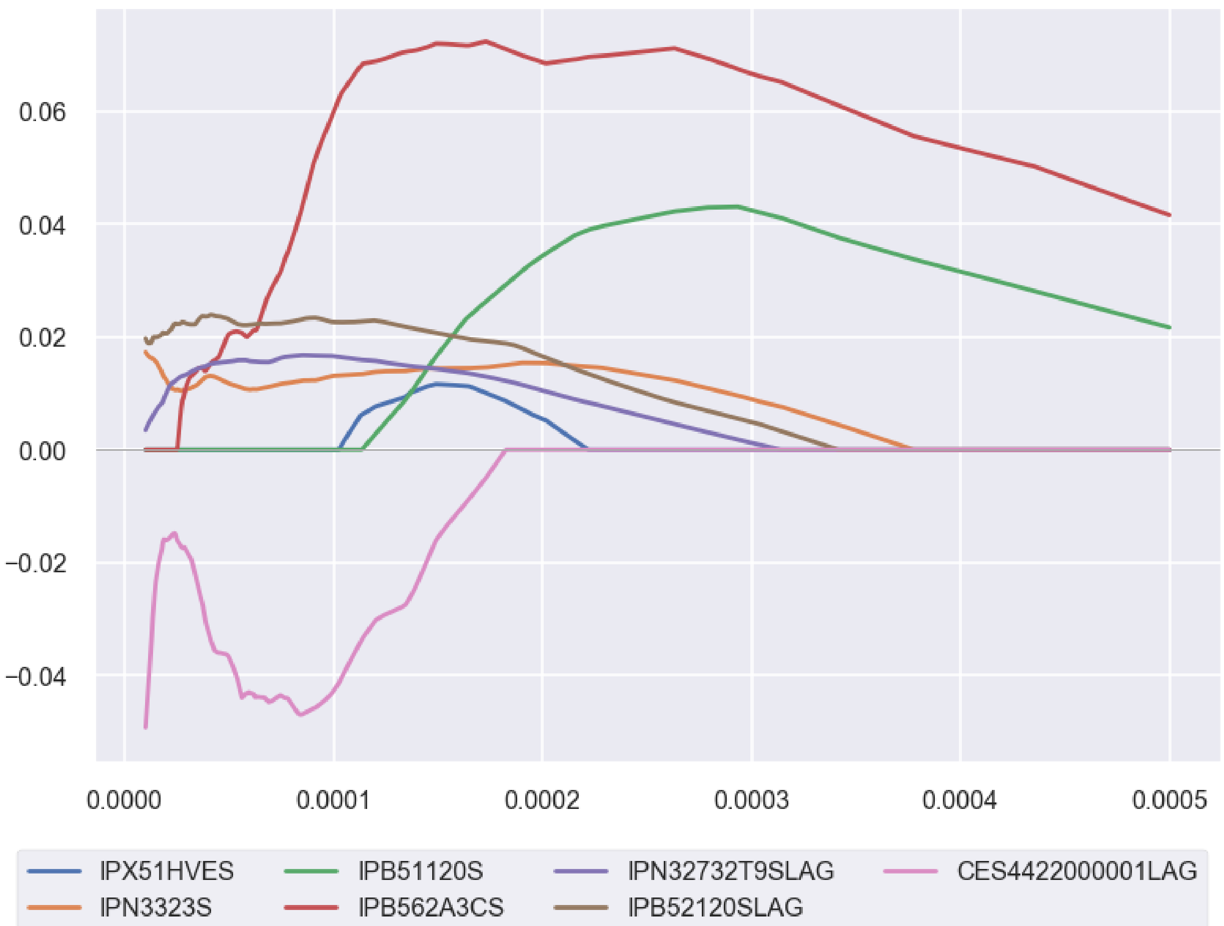

Figure 4: Regularization Path: Elastic Net Regression

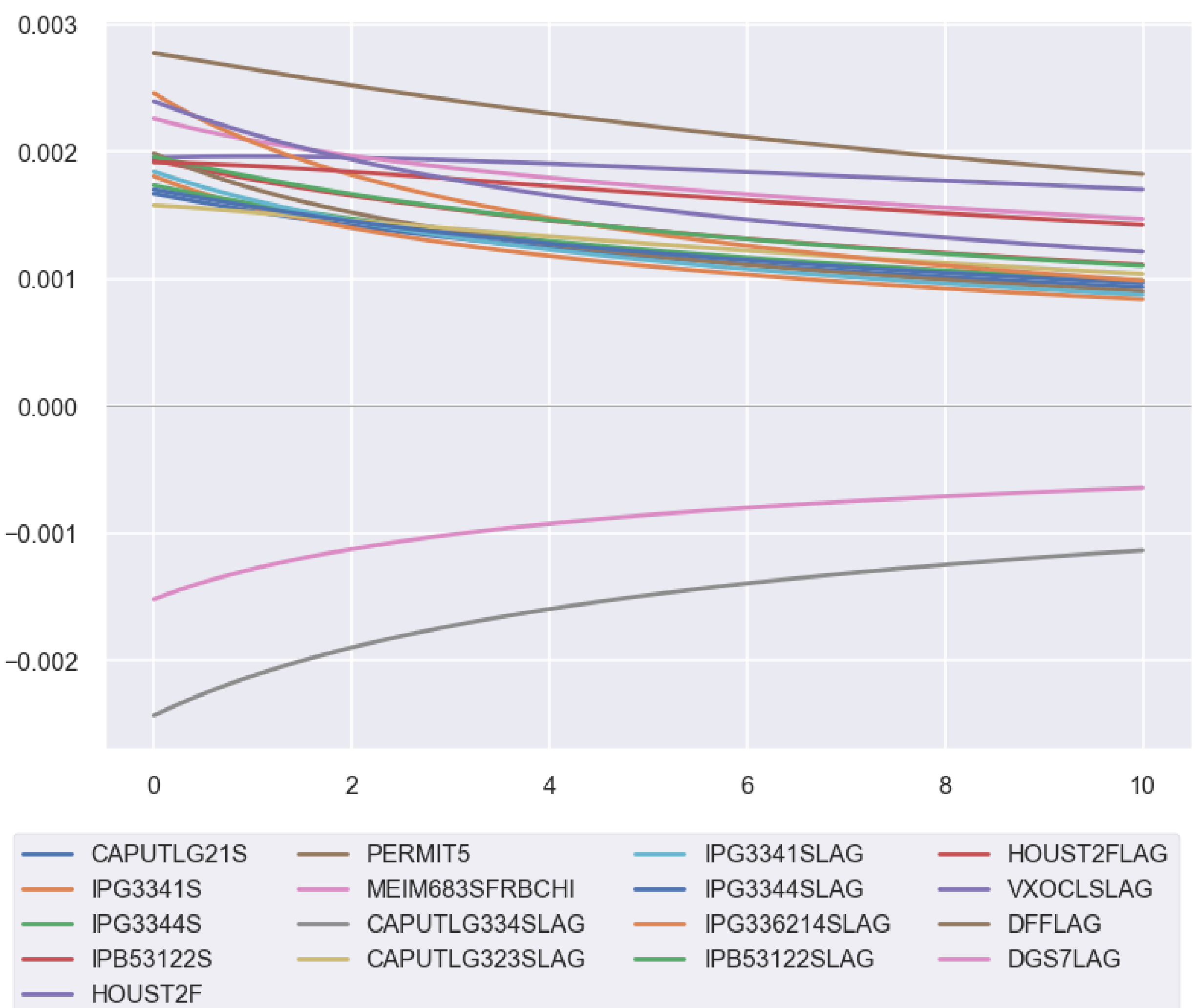

Figure 5: Regularization Path: Ridge Regression

Appendix B: Machine Learning Algorithms

5.2 Models

In this section, we provide a description of the Machine Learning and benchmark models we used for nowcasting GDP.

5.2.1 Autoregressive Model (AR(2))

We use an autoregressive process of order one as a benchmark model to compare the performance of the machine learning algorithms. An AR(1) process is defined as follows:

$$GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 GDP_{t-2} + \epsilon_t$$

where GDP is quarterly GDP growth.

5.2.2 Ridge Regression and Elastic Net

In Ridge Regression, a complexity penalty is added to the OLS loss function so that we minimize the sum of squared residuals and penalize the size of parameter estimates:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^{\,2} = ||y - X\hat{\beta}||^2 + \lambda||\hat{\beta}||^2$$

In Elastic Net, the regression penalty is a convex sum of the Ridge and Lasso penalties.

5.2.3 Bayesian Linear Regression

Bayesian linear regression is an approach to linear regression within the context of Bayesian inference, in which the linear regression is formulated using probability distributions rather than point estimates.

5.2.4 Gaussian process regression

In Gaussian process regression, learning and the kernel function are used to predict the value for an unseen point from training data. The prediction is a one-dimensional Gaussian distribution, rather than a single point.