

AMATH 521 Final Project

Topic: Sector Outperformance Analysis

Contributors:

- 1.) Luke Lee
- 2.) Wipada Wannasiwaporn

Summary:

The goal of this project was to build a model that predicts the sectors that tend to outperform the market return using logistic regression with different penalties and try out different models for comparison. In order to confine our scope of the project, we decided to focus on the two largest sector in S&P500, Technology and Financial.

Technology Sector

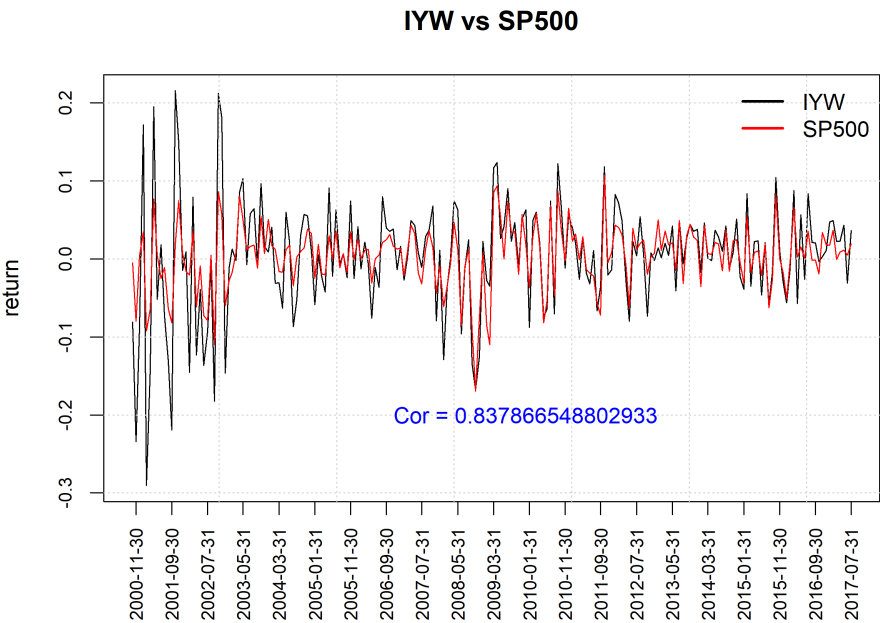
Data:

Our independent variables are basically the monthly macro economics data retrieved from FRED(<https://fred.stlouisfed.org/>). Some examples of these data are unemployment rate, FED funds rate, CPI, SP500 monthly return etc. We also added some sector related data such as Telecommunication export to possibly add prediction power for a specific sector as well. We convert these raw data into a return space using monthly simple return for an Index-like data and use percentage conversion for a probability data and rate data. Here are some samples of our data.

TABLE 1. Data Example

Date	IYW return	GDP	CSUSHPINSA	DGS10	TEDRATE	FEDFUNDS
10/31/2000	-0.080775444	0.011088	0.005507	0.0577	0.0057	0.0651
11/30/2000	-0.234329233	0.011088	0.005198	0.0548	0.0068	0.0651
12/31/2000	-0.087222647	0.011088	0.004617	0.0512	0.0067	0.064
01/31/2001	0.172170997	0.003422	0.003953	0.0519	0.0056	0.0598

For Technology Sector, we use BlackRock’s ETF(IYW) that tracks the Technology sector performance using Dow Jones as a benchmark. Overall, the returns of Technology sector moves strongly correlated with the return of S&P500 as seen in the high correlation value.



Methods & Algorithms:

i) Linear Regression

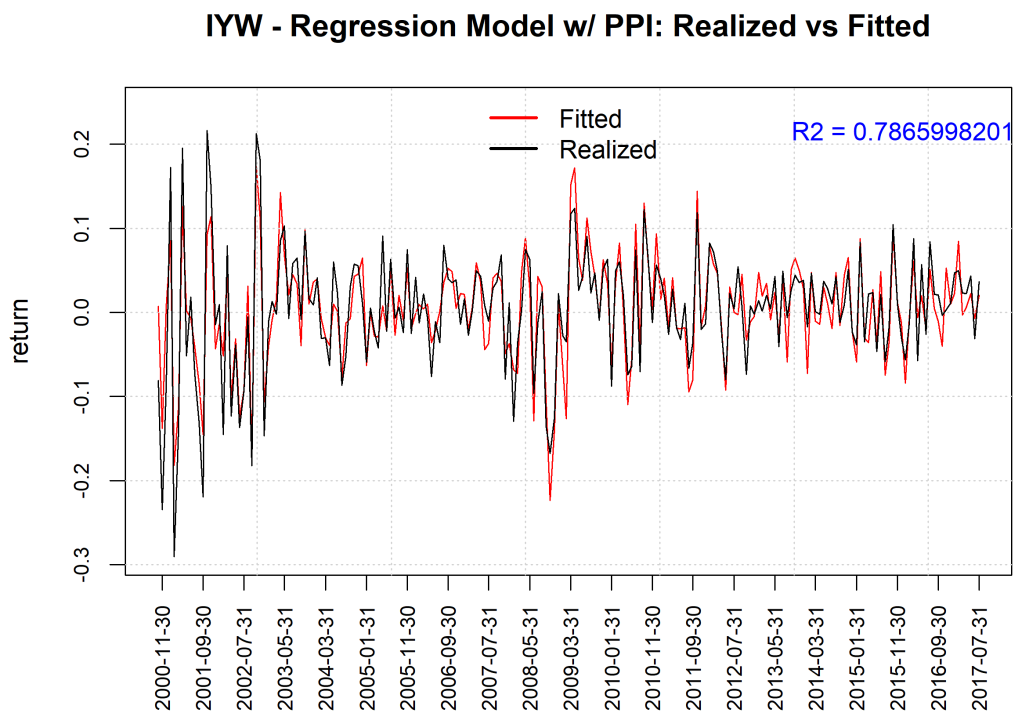
We first want to make sure that our data can explain the data well, so doing linear regression is one way of doing sanity check on our data. By using our whole data set, it turns out that the explained variance(R^2) is quite high for this data set, so we know our independent variables can somehow explain our dependent variable.

$$\mathbf{r} \in \mathbb{R}^T, \mathbf{F} \in \mathbb{R}^{T \times n}, \mathbf{x} \in \mathbb{R}^n$$

$$\mathbf{r} \sim F_1x_1 + F_2x_2 + \dots + F_nx_n$$

where \mathbf{r} represents our sector return(IYW - dependent variable) and \mathbf{x} represents our economics data(independent variables).

The results of this linear regression shown a high coefficient of determination R^2 .



After doing some sanity check, we try to predict whether IYW will beat the market in the next period by assuming we know the exact values of our independent variables. We use the combination of rolling windows of 5, 6, 7, ..., 10 years of training samples to predict the next 1, 2, 3, ..., 6 months return.

After we get our predicted return, we check whether this return is greater than the market and use this as a boolean result of our linear regression. Using this boolean output, we can assess our outperformance accuracy by using

$$Accuracy = \frac{\text{Numbers of correct predictions}}{\text{Numbers of total predictions}}.$$

Using accuracy as our selected criterion, we ran different combination of windows size and as a result, using 7 years of training data and 2 months of prediction period will give us the highest accuracy for linear regression method.

The plots below show the result of prediction using 7 years training data to predict the next 2 months value.

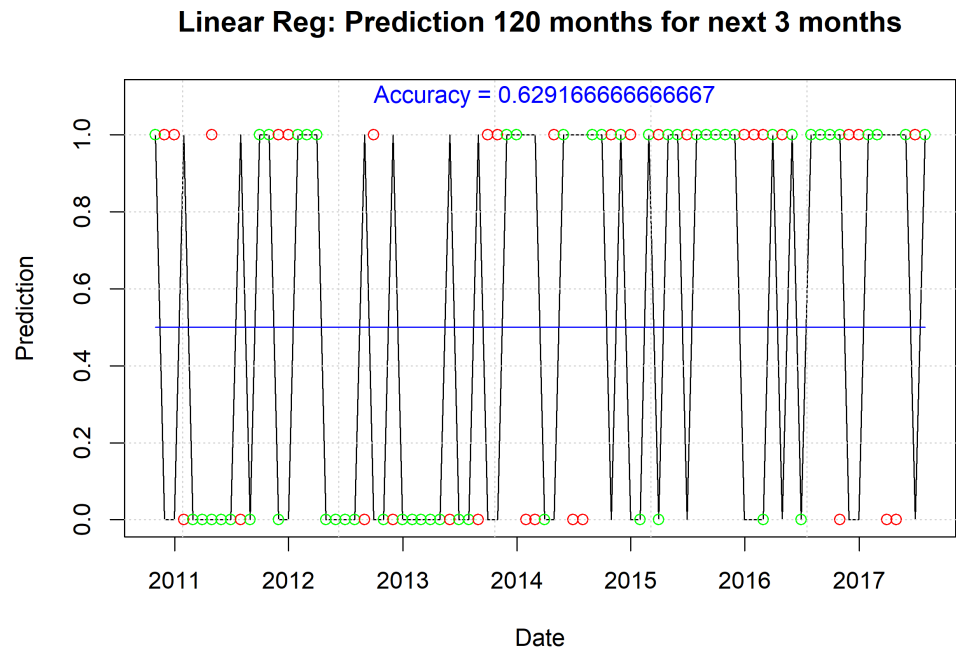


Figure 1: Linear Regression

ii) Linear Regression with Elastic Net

The plots below show the result of prediction using 9 years training data to predict the next 1 month value.

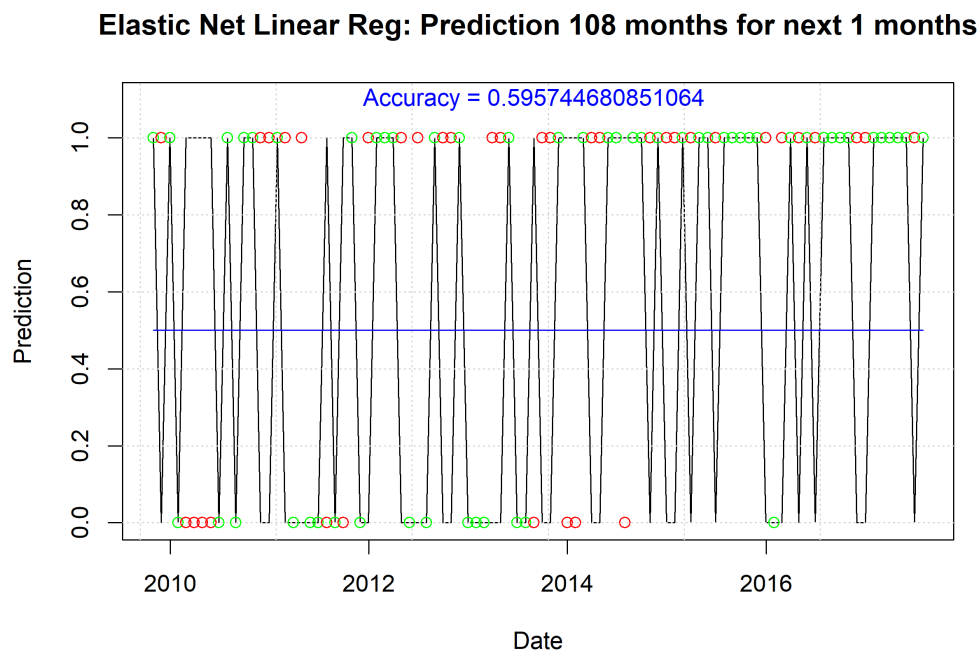


Figure 2: Linear Regression with Elastic Net

iii) Logistic Regression

The plots below show the result of prediction using 10 years training data to predict the next 2 months value.

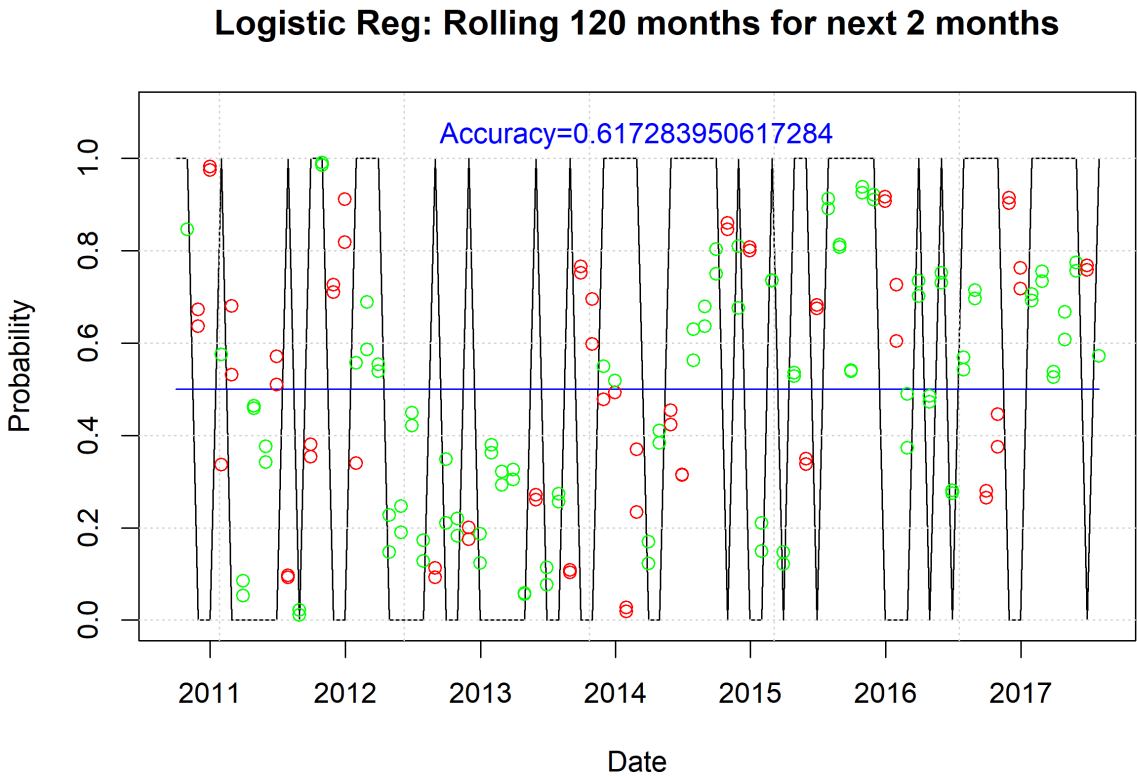
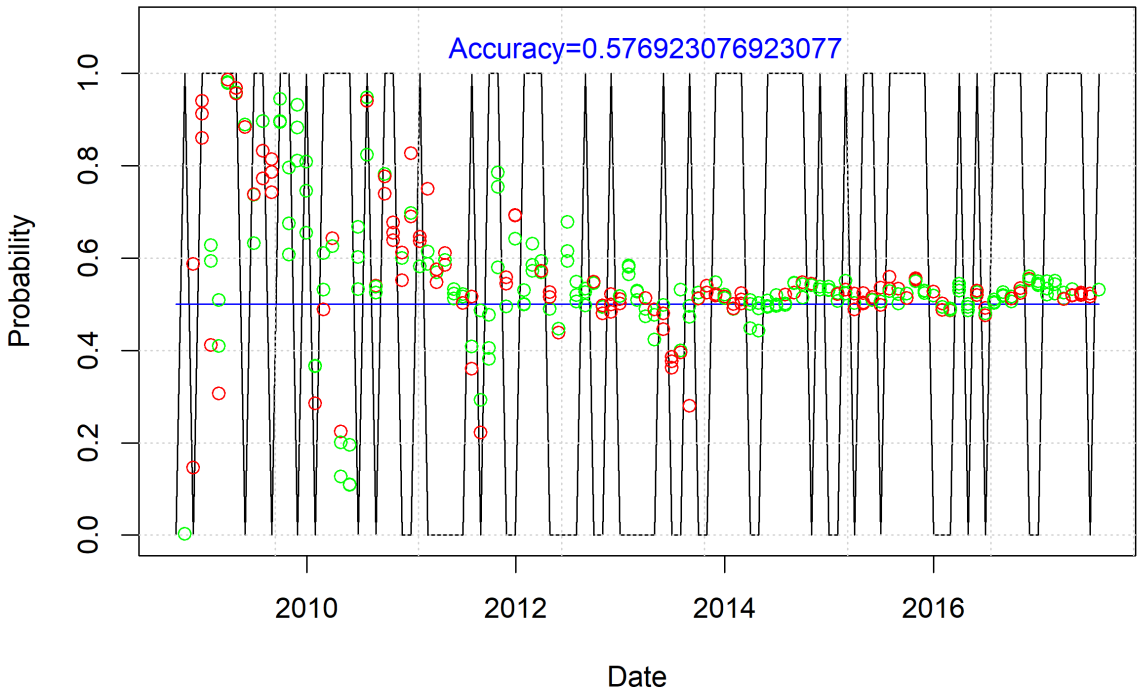


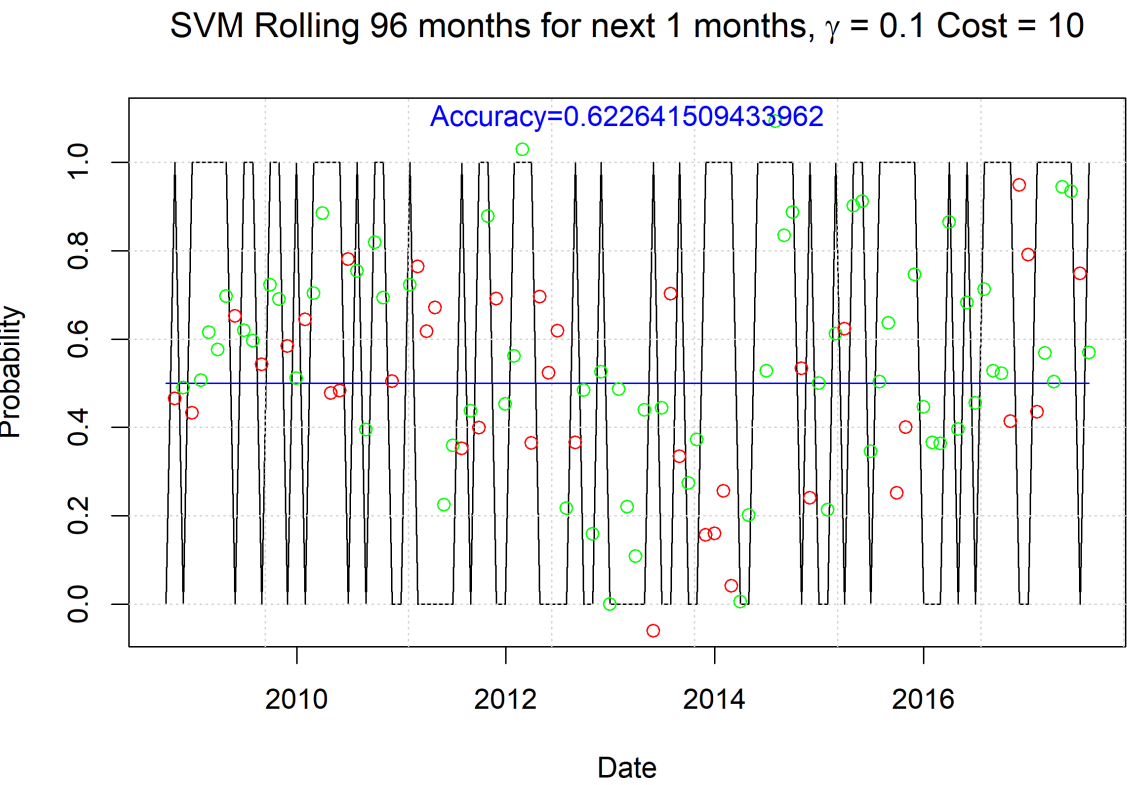
Figure 3: A simple caption

iv) Logistic Regression with Elastic Net

Elastic Net Logistic: Rolling 96 months for next 3 months with $\alpha = 0.8$



v) Support Vector Machine



gfjSGGJneklfniiqngienrhrbreqng

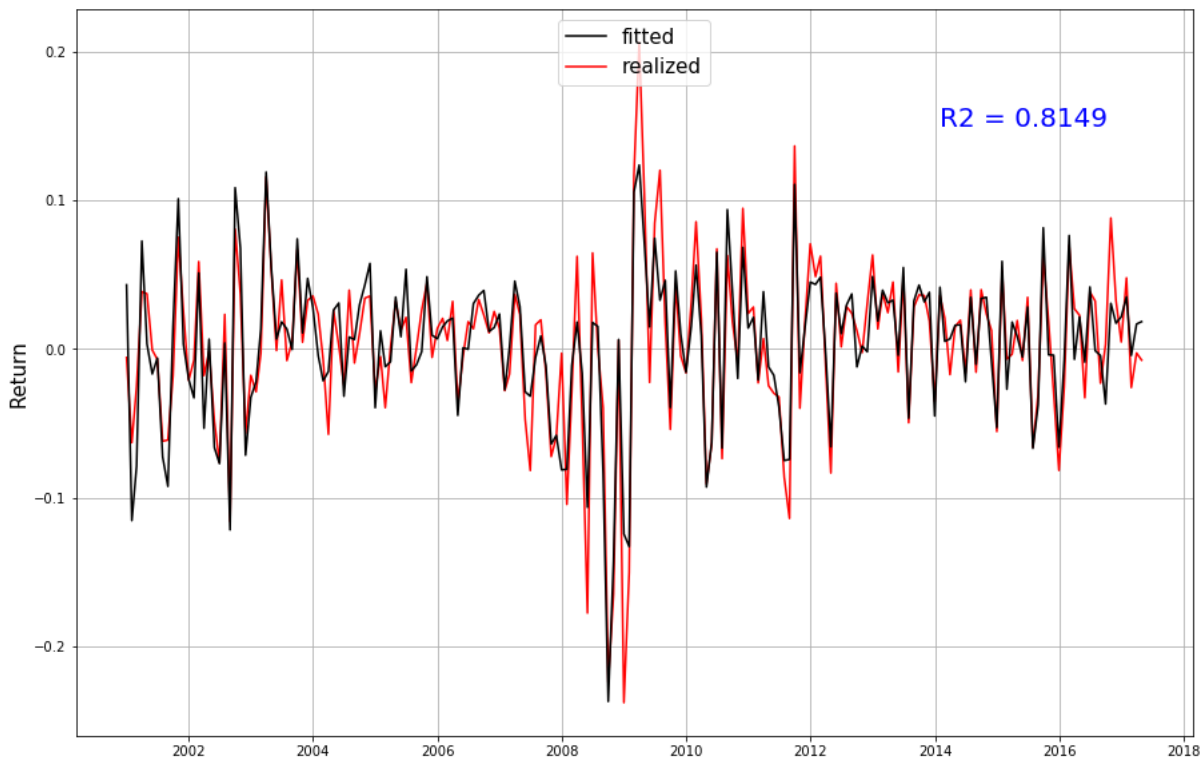
Financial Sector

Linear regression:

Data selection

To verify that the data chosen for the financial sector(IYF iShare) we fit a linear regression and see if our underlying data can explain the sector returns. As shown in the figure below the selected data have a high R^2 value of 81.5%

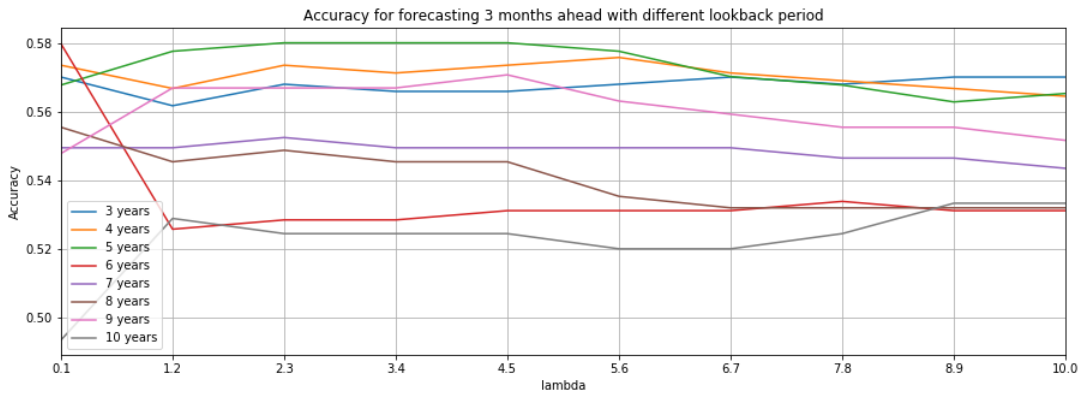
IYF - Linear Regression: Realized vs Fitted



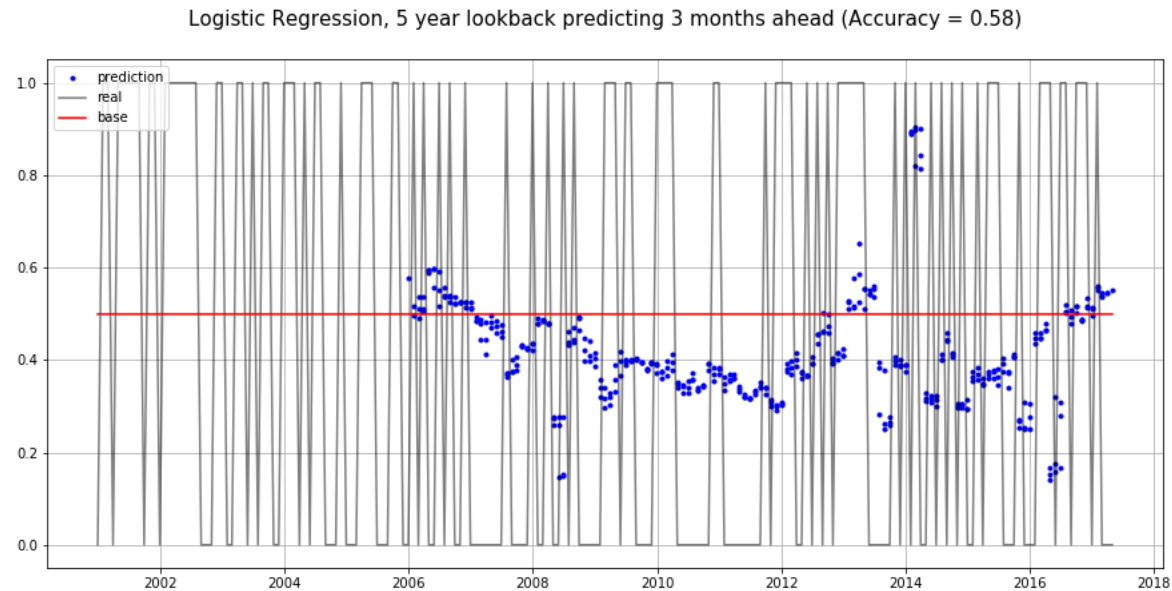
Now that I am satisfied with the data selection, we can move on to testing out more complicated methods that could help us predict this sector out-performing the benchmark. We calculated our binary predictor variable by taking the excess return of our sector and assigned 1 when it was positive and 0 otherwise.

LOGISTIC REGRESSION

As mentioned before we are applying a robust method where we use past n years of data to predict certain months ahead and applying this on a rolling basis. In addition to these two parameters, we need to select the penalty type (l1 or l2) and the penalty value(λ). After trying numerous combinations of these parameter's the optimal parameters selected were 5 year looking window, forecasting 3 months ahead, penalty of l2, with lambda of 3.4.



Here is a comparison of predicted probability against the realized values. The results were not as good as we have hoped for and partly because the probabilities we predict does not seem to be sparse.



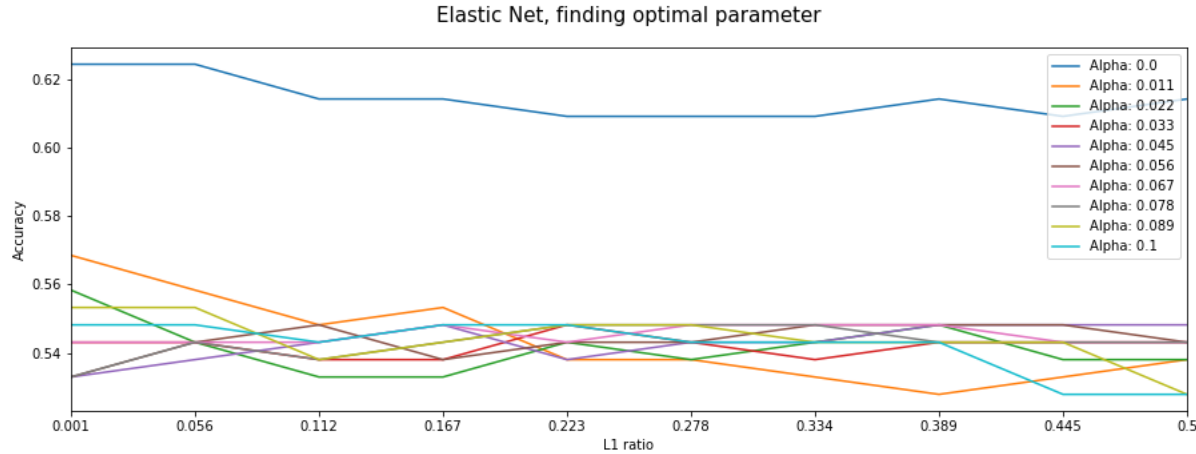
ELASTIC NET

Another method we considered was Elastic Net. Elastic Net’s objective function is an extension of linear regression it adds two additional terms, a L1 norm and a L2 norm.

$$\min_w \frac{1}{2n} ||y - Xw||_2^2 + \alpha \cdot \text{l1_ratio} \cdot ||w||_1 + \frac{1}{2} \cdot \alpha \cdot (1 - \text{l1_ratio}) \cdot ||w||_2^2$$

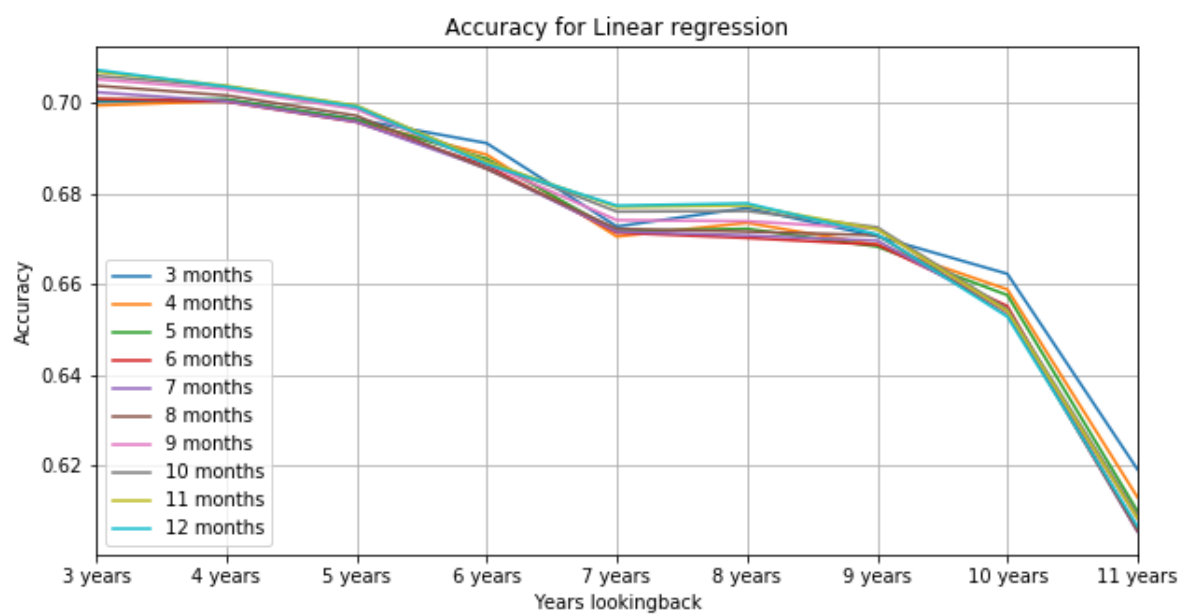
As seen in the objective function, there are two parameters to choose from. α is the penalty parameter and L1_ratio chooses the ratio of the L1 and L2 penalties.

Unfortunately, the optimal parameter chosen for this model seem to be basically linear regression. The combination with the highest accuracy was when α , the penalty parameter, was zero. This simply equates to linear regression.



BACK TO LINEAR REGRESSION

Applying the same rolling basis method, linear regression still outperforms the other methods with 70.72% accuracy. Interestingly, linear regression seem to do better with shorter training set data. The optimal parameter for this model was looking back 3 years to predict 3 months return ahead.



Summary & Comments: