croZackery Bradley

BAN 525

20220131

**Model 2 Assignment 2**

Introduction

This week’s assignment focuses on analyzing the determinants of renewable energy stock prices by utilized our newly learned penalized regression models. On behalf of the dataset, we will be working predominantly with the returns of the clean energy stock ETF (RPBW), which is considered to be one of the largest and most popular clean energy investment instruments in the US. Our goal is to examine what other variables may impact the performance of RPBW. As society transitions into the world of clean energy, it is reasonable to suggest that the stock market will also be in a transition as technology in the industry increases. Some of the factors that are thought to have a more significant impact on the continued development of this technology include interest rates (as new technology is created, this poses more risk to banks as it is essentially unknown, thus banks require significantly more compensation for their risk), and the price of oil (depending upon production levels, as well as how effective the transition to clean energy becomes, oil could still be a cheaper and more desirable energy producing method). This analysis can provide valuable information to investors as they continue to hunt for their next bullish investments. The statistical methods that will be utilized in this analysis include our standard OLS method, along with Lasso, Adaptive Lasso, Elastic Net, Adaptive Elastic Net, and Adaptive Lasso with t(5) and Cauchy distributions. All of these methods have distinct capabilities that can be highly desirable based on the type of analysis being performed. Lasso has the capability of shrinking uninformative coefficient all the way to zero, essentially eliminating their respective influences. Lasso typically bodes well in an environment where prediction accuracy is of the upmost importance. Elastic Net is a combination of the Ridge and Lasso methods in which a mixing function (usually fixed) is implied. Like Lasso, Elastic Net has also shown great prediction accuracy. We will also be utilization adaptation in regard to these two methods, as this strategy will give us a good idea about which variables are of the most importance. Employing adaptive strategies would allow us to penalize important variables at a much lesser rate, thus creating a more accurate model. Lastly, we will be implementing the Cauchy distributions in our model as well. This method, in lament terms acts as a strictly stable distribution.

Analysis and Model comparison

While all of our methods have been proven to be reliable in the right conditions, each method does not come complete without its own pros and cons. Lasso, for example, typically avoids over fitting, however it also has a tendency to be highly biased based on selections. Elastic Net can be especially valuable when more than one predictor is utilized but can also be detrimentally more expensive to compute than Lasso. Given the constraints of each method, the adaptive and Cauchy strategies are implemented as to enhance our overall capability. Procedurally, we will be employing a cross-validation analysis. Cross-validation techniques are built on the concept of leaving out a part of the data out of the estimation process as a buffer. As the predictions of the model stop improving as the data holdout process occurs, model growth then stops, and estimates are obtained. This creates a ‘split’ in the data. These splits are then broken down into three parts in the terms of our model. Those parts include training, validation, and testing splits respectfully. Training data is the portion that is used to estimate our model. Validation data is not used in estimate directly, but instead works in the background to determine the optimal point in which the model stops. Lastly, the test data is never actually used in the model estimation. Test data is used to represent “new observations” and assists in providing and unbiased analysis of the predictive ability of the model. With the validation method only being utilized indirectly, we choose our model that displays the best performance on the test data.

Upon closer analysis of our data, we can see that two of our models performed very close to each other. The Elastic Net method, much to my surprise is the model that performed the best in our analysis. With an RSquare of -0.065, followed by a RASE and AAE of 1.435 and 0.6629 respectively, the Elastic Net method proved to be an effective method of accuracy prediction on the basis of our dataset. Admittedly, I expected the Adaptive Lasso with t(5) distribution to perform best in our analysis. Given the outcome of this model, our data performance has shown that utilizing several different statistical methods in our models is of the upmost importance. Those small degrees of performance enhancement can cost companies millions if the wrong decision is made based on inferior data. Although we felt confident about our decision, we opted to perform additional analysis with the t(5) and Cauchy distributions. Upon analysis, we tested the distributions in a PP plot seen below. Although the data visuals looked promising, those looks were found to be deceiving as we performed a series of Anderson-Darling tests tests and determined that the distributions are better off as rejected.

**Graphical user interface

Description automatically generated**

**Fitted Cauchy Distribution**

**PP Plot**



**Goodness-of-Fit Test**

|  | **A2** | **Simulated p-Value** |
| --- | --- | --- |
| Anderson-Darling | 12.118114 | 0.0176\* |

**Fitted Student's t Distribution**

**PP Plot**



**Goodness-of-Fit Test**

|  | **A2** | **Simulated p-Value** |
| --- | --- | --- |
| Anderson-Darling | 2.6012304 | <.0001\* |

Interpretation

Our interpretation of our data analysis can simply not be complete without reporting variable importance. By definition, variable importance is essentially how important a singular variable is in the modeling process. These variables carry the most weight in the analysis of a dataset. By examining our methods, we can see that each of them weighed our variables quite differently. In our chosen model, Elastic Net weighed our RPBW variable as the most important. Elastic Net as having a total effect of just .831, proving that RPBW is a significant contributor in model performance.

**RPBW Variable Importance Elastic Net**

| **Column** | **Main Effect** | **Total Effect** |  |
| --- | --- | --- | --- |
| RPBW | 0.811 | 0.831 |  |
| XLK | 0.031 | 0.049 |  |
| SPY | 0.015 | 0.027 |  |

In closing of our analysis, we can say with certain confidence that Elastic Net performs adequately as opposed to other methods in regard to this particular data. This can be seen on our overall analysis of the other performance methods as Elastic Net has proven to be superior statistically speaking. Given the detrimental impact one variable can have on decision making, it is imperative that all companies study model performance in order to generate the best outcome for their futures.