Daniel Sanchez
BAN 525
Dr. Cetin Ciner

Module 2 Assignment -- Penalized Regressions and Predicting Clean Energy Stock Prices

May 23rd, 2021

Introduction

On May 18, 2021, President Joe Biden sat inside the cabin of a Ford F-150, declaring emphatically, “This sucker’s quick!” The Ford F-150 is the most popular model in Ford’s iconic line of American-made full-size pickup trucks. But what was so special about this truck that the President (with his stone-faced Secret Service detail riding shotgun) would be taking it for a test-drive? The answer: this truck was an *electric* vehicle.

Clean energy has been touted as our only hope in combatting climate change. Low-carbon, renewable and sustainable energy usage appears to be gaining more traction in the world. Countries are now vying for position in a future that leverages clean energy technologies. To understand where this future is heading (and where we might invest in it), we can look for the determinants of clean energy stock returns.

The dataset for this analysis contains 1659 observations between September 17, 2014 and April 23, 2021. These values represent daily (trading day) returns from 18 variables from Stocks, Bonds, Currency Markets, Overall Market Sentiments, and a Cryptocurrency (Bitcoin). 18 corresponding lagged variables from one period prior to the contemporaneous data are also included. The inclusion of Bitcoin in the dataset is somewhat of a wildcard, due to its extreme volatility and reputation for consuming high amounts of electricity in its mining. RPBW, the response variable in this dataset, reflects the returns of the clean energy stock ETF, PBW.

The analysis in this report has been performed in JMP Pro 16, using OLS (Ordinary Least Squares) and the following penalized regression methods:

1.) Lasso
2.) Adaptive Lasso
3.) Elastic Net
4.) Adaptive Elastic Net and,
5.) Adaptive Lasso, with t(5) and Cauchy distributions

Analysis and Model Comparison

OLS: This is being used as a benchmark for the analysis. It is not expected to be the top performer in the group, since its results involve all of the variables and tends to overfit models to the training set.

Lasso: Using the absolute value of a determined penalty, this method shrinks the variables that contain little information, so that those are not included in the final model.

Adaptive Lasso: This method is similar to Normal Lasso. However, it first takes into consideration the results of OLS before applying its penalization factor to unimportant variables.

Elastic Net: This method applies both an absolute and squared penalization value to uninformative variables. Similar to Lasso, unimportant variables are eliminated in its model.

Adaptive Elastic Net: This method considers the results of OLS before applying it penalized regression approach.

Adaptive Lasso, with t(5) and Cauchy distributions: Since the other models assume normal distributions, running Adaptive Lasso with t(5) and Cauchy distributions allows the model to acknowledge a high frequency of outliers.

The cross-validation procedure applied to this analysis acknowledges that the data is time-series in nature. Hence, the data set is split into three sections chronologically (as opposed to being randomized). Implementing a 60-20-20 split, the first 995 rows are used to train the models, rows 996 - 1327 are tied to validation, and rows 1328 - 1659 are set aside for testing. This forward-looking approach to cross-validation helps to ensure that the model considers the sequential nature of the dataset.

Hoping to predict the price of clean energy stock returns, RPBW is selected as the response variable. All of the other variables in the dataset are included as predictor variables. Using JMP 16 Pro, this analysis was conducted with all six methods. As predictions from each model were generated, they were saved and stored in new columns in the dataset. The figure below shows the comparison of how all six models performed with the test data.

Predictor	Creator	RSquare	RASE	AAE
RPBW Prediction Formula OLS	Fit Least Squares	0.6203	0.0219	0.0169
RPBW Prediction Formula Lasso	Fit Generalized Lasso	0.6121	0.0222	0.0172
RPBW Prediction Formula ALasso	***Fit Generalized Adaptive Lasso***	**0.6260**	**0.0218**	**0.0168**
RPBW Prediction Formula Elastic Net	Fit Generalized Elastic Net	0.6121	0.0222	0.0172
RPBW Prediction Formula AElastic Net	***Fit Generalized Adaptive Elastic Net***	**0.6260**	**0.0218**	**0.0168**
RPBW Prediction Formula t(5) ALasso	Fit Generalized Adaptive Lasso	0.6229	0.0218	0.0169
RPBW Prediction Formula Cauchy ALasso	Fit Generalized Adaptive Lasso	0.6173	0.0220	0.0170

The Adaptive versions of Lasso and Elastic Net outperformed the other models, achieving identical metrics in terms of RSquare (0.6260), RASE (0.0218), and AAE (0.0168). This likely occurred due to none of the active predictors being highly correlated with each other. Both regression methods applied similar Lambda Penalties (0.0024794 for Adaptive Lasso, and 0.0025044 for Adaptive Elastic Net). Having two methods tie in this model comparison resulted in both being included in the next steps of the analysis.

Interpretation

Of the 36 total candidate variables, only 4 were selected in the Adaptive Lasso and Adaptive Elastic Net models: RSLY, RXLK, RACWX, and RUSO. The estimates for all other variables were reduced to zero, and not included in either of the final models. This is unlike the OLS benchmark, which included all variables in its final model. The parameter estimates for each of the final models are shown in the tables below.

Adaptive Lasso Parameter Estimates

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
RSLY	0.6815513	0.0656340	107.82982	<.0001*	0.5529109	0.8101916
RXLK	0.2181598	0.0564589	14.930877	0.0001*	0.1075025	0.3288171
RACWX	0.1918935	0.0657822	8.5095029	0.0035*	0.0629628	0.3208242
RUSO	0.0474635	0.0192966	6.0500319	0.0139*	0.0096429	0.0852842
Intercept	-0.000665	0.0003066	4.7085192	0.0300*	-0.001266	-6.437e-5

Adaptive Elastic Net Parameter Estimates

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
RSLY	0.6816127	0.0656419	107.82354	<.0001*	0.552957	0.8102684
RXLK	0.2180654	0.0564691	14.912542	0.0001*	0.107388	0.3287428
RACWX	0.1921981	0.0657966	8.5328046	0.0035*	0.0632392	0.321157
RUSO	0.047056	0.0193155	5.9349329	0.0148*	0.0091982	0.0849137
Intercept	-0.000666	0.0003067	4.7120786	0.0300*	-0.001267	-6.463e-5

Represented mathematically, this is the final model proposed by both of these methods:

$$RPBW_t = -.001 + (.682)RSLY_t + (.218)RXLK_t + (.192)RACWX_t + (.047)RUSO_t$$

The model characteristics show a relationship between clean energy stocks(RPBW) and these variables: RSLY(small companies), RXLK(technology companies), RACWX(international markets), and RUSO(oil prices). Notably, the features selected in the final model all consist of contemporaneous data; no lagged variables are included in either of the final models.

Variable importance analysis reveals that, in both Adaptive Lasso and Adaptive Elastic Net, RSLY accounts for approximately 80% of the fluctuation in RPBW. To put this into perspective, the OLS benchmark model predicted RSLY as having a total effect of only 53%. According to these penalized regression models, however, RSLY has by far the most impact on RPBW than any other variables in the dataset. RXLK accounts for around 13% of the fluctuation in the response variable. RACWX has a total effect of 5%. And, although RUSO is included in both models, its total effect is minimal (between 1-2%).

Adaptive Lasso Variable Importance

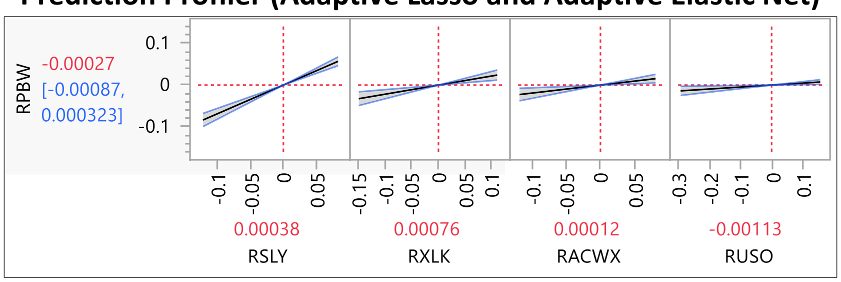
Column	Main Effect	Total Effect
RSLY	0.779	0.805
RXLK	0.103	0.129
RACWX	0.035	0.057
RUSO	0.01	0.02

Adaptive Elastic Net Variable Importance

Column	Main Effect	Total Effect
RSLY	0.776	0.797
RXLK	0.11	0.131
RACWX	0.04	0.058
RUSO	0.009	0.017

The prediction profiler for both models is shown below, revealing identical trends between the model variables. The steep slope of RSLY indicates a much higher correlation with RPBW, than that of RXLK, RACWX, and RUSO. Manipulating the value of RSLY in the prediction profiler reveals a large corresponding shift in the model's other variables. Such high correlation between RSLY and RPBW nearly allows RSLY to serve as its proxy.

Prediction Profiler (Adaptive Lasso and Adaptive Elastic Net)



It is worth examining the variables that were *not* included in the final model. No correlation was identified between bond markets, currency markets, or Bitcoin. Large companies (such as Ford Motor Company) were also not found in the final model. This may change as clean energy moves further into the mainstream, through natural market forces and/or government intervention. Similarly, lagged variables were not chosen to be in the final model. This makes it very difficult to predict the future of clean energy stocks, based on a one-period lag. This does not necessarily rule-out the predictive capability of using an alternative lag frequency. Additionally, analysis of future time-series data may also reveal relationships between clean energy stock returns and either the same variables or others outside of this dataset.