Penalized Regressions and Predicting Clean Energy Stocks during Covid-19

1. INTRODUCTION

With global policies such as the 2015 Paris Agreement, there is a growing focus on renewable energy sources. There is an increase in investment in Solar Farms, Wind Turbines, and Electric Transportation in order to reduce the large amount of greenhouse gases contributing to our Global Warming Issues. It is a fair exploration to analyze the effects of the Price of Oil, the price of High Technology Shares and General Interest Rates on Renewable Energy Stock Prices in the U.S., specifically, Invesco WilderHill Clean Energy ETF (RPBW).

The data set to be considered consists of 611 daily recordings of twenty different stocks in these three sectors. The dates range from January 2, 2020 to June 3, 2022 which include data from the time of the COVID world pandemic. With the inclusion of these turbulent times, we can possibly see unexpected influences on our Renewable Energy Stock Prices. The data set itself does not contain a Date, so it is a classification data set, not a time-series data set.

These twenty different stocks from these three different financial sectors were transformed into returns by taking log differences. Then one period of lag results was also recorded. The data analysis used both the returns and the lagged results to determine what was the strongest predictor on the renewable energy stock prices.

The method of exploration began by creating a cross-validation column in which was actually already provided in the given data set. This cross-validation allowed three randomized sections of data to be compared for best estimation: the Training subset, the Validation (Holdback) subset and the Test subset. As the seven estimation models were created in JMP, there was a comparison of their Training and Test subset estimation and error values. Then, the models were combined into a complete Model Comparison using a feature in the JMP software. The comparison of these results was narrowed down to show that there were three models that performed the best. A final comparison of these led to an overall best performer and the data was applied to this estimation model to find the best predictor variable for the Response Variable: Invesco WilderHill Clean Energy ETF (RPBW).

2. ANALYSIS AND ESTIMATION MODEL COMPARISON

The estimation methods chosen were the Ordinary (or Standard) Least Squares Regression, the Lasso Generalized Regression (Adaptive and non-Adaptive) and the Elastic Net General Regression (Adaptive and non-Adaptive). Methods have advanced so that the use of Adaptive methods helps avoid overfitting and penalizing large coefficients. Also, as a default, JMP assumes errors are normally distributed so this analysis additionally, looked at the normalized regression as well as the t(5) and Cauchy Distributions of the data with Adaptive Lasso Regression.

Type of Method	**Advantages**	**Disadvantages**
Ordinary / Standard Least Squares	It is well-known, easy to explain and easy to understand.	It is sensitive to outliers and does not deal well with singularities.
Lasso Regression	Avoids overfitting models and removes variables that do not affect the response variable	Shrinks uninformative coefficients all the way to zero. Arbitrarily picks variables that are correlated
Adaptive Lasso Regression	Consistent variable selection where adaptive weights are used for penalizing different coefficients in the L1 penalty	Like Lasso, it cannot do group selection with highly correlated variables.
Elastic Net Regression	Keeps correlated variables and adds weighted penalties to them.	Takes more time to compute since it combines Ridge and Lasso.
Adaptive Elastic Net Regression	Encourages a grouping effect: either selects the correlated group or omits them	Increase in computational time.
Adaptive Lasso w/ T5	Works with distributions that assumes thick tails; large skewness or kurtosis	Same as Lasso above
Cauchy Regression	It addresses continuous data describing resonance behavior	Same as Lasso above

Each model was created in JMP using either the default Standard Least Squares or the Generalized Regression Personality, each of which were eventually saved to the Data Table for future use. From the Generalized Regression Method, options could be chosen for the type of Distribution and the Type of Estimation Method Preferred. The option to make the Estimation Method Adaptive was also a choice.

RPBW.Covid - Generalized Regression - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help

Generalized Regression for RPBW

Normal Elastic Net with Validation Column

Effect Tests

Source	Nparm	DF	Wald ChiSquare	Prob > ChiSquare	
RVWO	1	1	26.781856	<.0001*	
RXLK	1	1	14.10706	0.0002*	
RSPY	1	1	7.1151723	0.0076*	
LRBWX	1	1	5.7821888	0.0162*	
LRIBND	1	1	5.6690729	0.0173*	
ROVX	1	1	3.0863455	0.0790	
RVIX	1	1	2.7590655	0.0967	
LRLQD	1	1	2.3535358	0.1250	
RVNQ	1	1	1.6826656	0.1946	
LRHYG	1	1	1.2757563	0.2587	
LRVIX	1	1	1.2159642	0.2702	
RGLD	1	1	1.1743309	0.2785	
LRSPY	1	1	1.0901011	0.2964	
LRUSO	1	1	0.8740277	0.3498	
RTLH	1	1	0.86938	0.3511	
LRGLD	1	1	0.6416994	0.4231	
LRTIP	1	1	0.335176	0.5626	
LRPBW	1	1	0.2840588	0.5941	
RBWX	1	1	0.2805621	0.5963	
LRVWO	1	1	0.2127096	0.6447	
RUSO	1	1	0.155583	0.6933	
LRVNQL	1	1	0.1554288	0.6934	
RLQD	1	1	0.1167329	0.7326	
RHYG	1	1	0.0775459	0.7807	
LRSLY	1	1	0.0496515	0.8237	
LROVX	1	1	0.0341543	0.8534	
LRFXE	1	1	0.0021269	0.9632	
RTIP	1	0	0	1.0000	Removed
RFXE	1	0	0	1.0000	Removed
REMB	1	0	0	1.0000	Removed

Generalized Normalized Elastic Net Regression model picks Vanguard Emerging Markets Stock Index Fund (**RVWO**), Technology Select Sector SPDR Fund (**RXLK**), and S&P 500 (large Company) index (**RSPY**) as strong predictors for the Response Variable Invesco WilderHill Clean Energy ETF (**RPBW**).

It also chose the Lagged SPDR Bloomberg International Treasury Bond (**LRBWX**) and Lagged SPDR Bloomberg International Corporate Bond (**LRIBND**) as possible predictors for the Response Variable Invesco WilderHill Clean Energy ETF (**RPBW**).

This model kept only 27 variables of the original 40 Return and Lagged variables.

3. MODEL COMPARISON

After the seven Estimation Methods listed in the previous chart were created and saved as a column, a Model Comparison was produced using those seven columns to determine the best Predictor variable for the chosen Response Variable. The measures of error that were compared included the R^2 value (the higher the number, the better) as well as the Residual Standard Error (RASE) and the Average Absolute Error (AAE), both of which have better results with the lowest results. Additionally, these values were compared the results from the training subsets and the Testing subsets to make sure that there were no very high differences in these values, resulting in overfitting of the data.

Looking at Test Set Values to determine the Best of 7 Models

- **Adaptive Lasso** has the highest **R^2** value with 0.6301.
- **Adaptive Elastic Net** has the 2nd highest **R^2** value with 0.6300.
- **Adaptive Lasso t (5)** and **Adaptive Lasso Cauchy** share the lowest Residual Standard Error (**RASE**) with 0.0215.
- **Adaptive Lasso t (5)** has the lowest Average Absolute Error (**AAE**) with 0.0215.

After getting mixed results in the seven Model Comparison chart, the top three performers were put into a model of their own. Again, the measures of error that were compared included the R^2 value (the higher the number, the better) as well as the Residual Standard Error (RASE) and the Average Absolute Error (AAE), both of which have better results with the lowest results. The Adaptive Lasso t (5) outperformed the Adapted Lasso with normal distribution and the Adapted Lasso with the Cauchy distribution.

RPBW.Covid - Model Comparison 3 - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help

Model Comparison

Predictors

Measures of Fit for RPBW

Holdback	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
0	RPBW Prediction Formula AdaptiveLassoT5	Fit Generalized Adaptive Lasso		0.7285	0.0181	0.0137	364
0	RPBW Prediction Formula AptiveLassoCauchy	Fit Generalized Adaptive Lasso		0.7059	0.0189	0.0140	364
0	RPBW Prediction Formula AdaptiveLasso	Fit Generalized Adaptive Lasso		0.7171	0.0185	0.0142	364
1	RPBW Prediction Formula AdaptiveLassoT5	Fit Generalized Adaptive Lasso		0.4604	0.0171	0.0135	122
1	RPBW Prediction Formula AptiveLassoCauchy	Fit Generalized Adaptive Lasso		0.4604	0.0171	0.0135	122
1	RPBW Prediction Formula AdaptiveLasso	Fit Generalized Adaptive Lasso		0.4564	0.0172	0.0134	122
2	RPBW Prediction Formula AdaptiveLassoT5	Fit Generalized Adaptive Lasso		0.6426	0.0215	0.0164	123
2	RPBW Prediction Formula AptiveLassoCauchy	Fit Generalized Adaptive Lasso		0.6418	0.0215	0.0166	123
2	RPBW Prediction Formula AdaptiveLasso	Fit Generalized Adaptive Lasso		0.6301	0.0219	0.0166	123

RPBW.Covid - Generalized Regression - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools

Generalized Regression for RPBW

t(5) Adaptive Lasso with Validation Column

Model Summary

Measure	Training	Validation	Test
Generalized RSquare	0.7120478	0.4617273	0.6386236
Lambda Penalty	0.8395614	.	.

Solution Path

Parameter Estimates for Original Predictors

Effect Tests

Source	Nparm	DF	Wald ChiSquare	Prob > ChiSquare	
RSLY	1	1	92.864885	<.0001*	
RXLK	1	1	33.847371	<.0001*	
RSPY	1	1	23.536541	<.0001*	
RVWO	1	1	11.556272	0.0007*	
LRBWX	1	1	5.2285416	0.0222*	
LRIBND	1	1	4.1687956	0.0412*	
LRSPY	1	1	2.5543541	0.1100	
ROVX	1	1	2.3500379	0.1253	
LRLQD	1	1	1.8402766	0.1749	
LRVWO	1	1	1.5192751	0.2177	
LRVIX	1	1	1.4038956	0.2361	
RVSS	1	1	1.0492442	0.3057	
RVNQL	1	1	0.9846348	0.3211	
LRHYG	1	1	0.9664575	0.3256	
LRVNQL	1	1	0.6916334	0.4056	
LRVSS	1	1	0.5780408	0.4471	
RTIP	1	1	0.1838694	0.6681	
LROVX	1	1	0.1819877	0.6697	
RVEU	1	1	0.1736959	0.6768	
LREMB	1	1	0.0670517	0.7957	
LRSLY	1	1	0.03522	0.8511	
RUSO	1	1	0.0229273	0.8796	
RHYG	1	1	0.0037583	0.9511	
RFXE	1	1	5.2084e-5	0.9942	
RLQD	1	0	0	1.0000	Removed

Generalized t (5) Adaptive Lasso Regression model picks S&P 600 (small company) index (**RSLY**), Technology Select Sector SPDR Fund (**RXLK**), S&P 500 (large Company) index (**RSPY**), and Vanguard Emerging Markets Stock Index Fund (**RVWO**) as strong predictors for the Response Variable Invesco WilderHill Clean Energy ETF (**RPBW**).

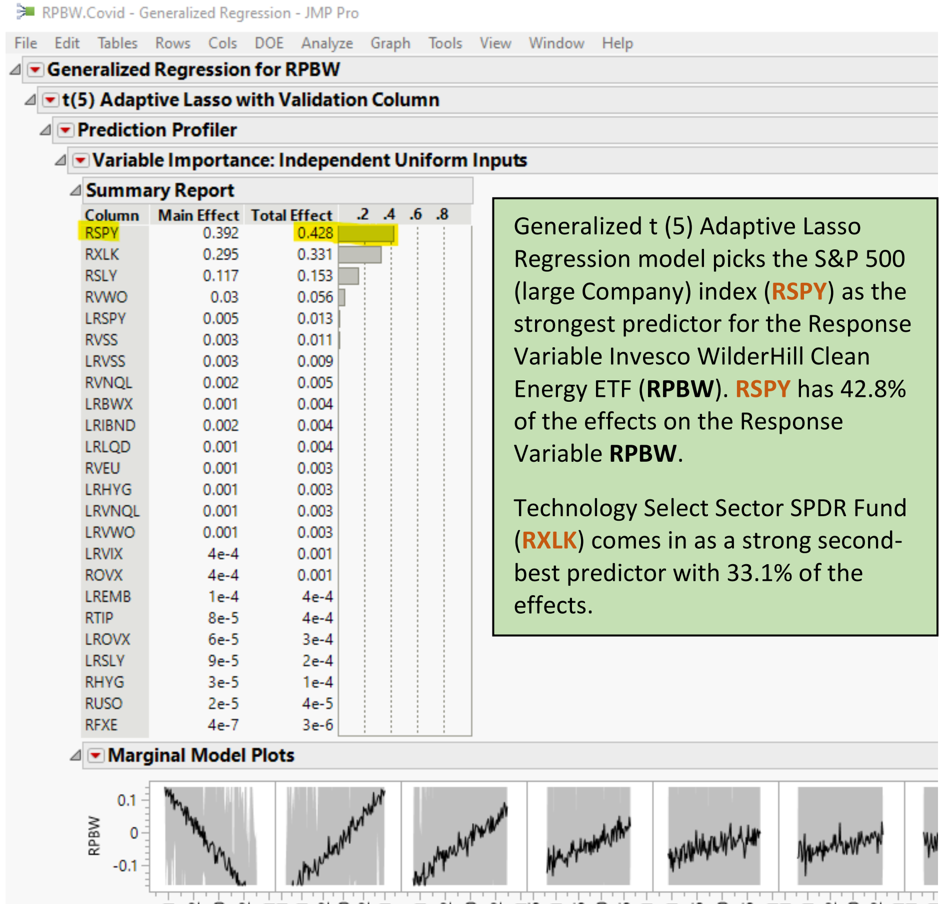
It also chose the Lagged SPDR Bloomberg International Treasury Bond (**LRBWX**) and Lagged SPDR Bloomberg International Corporate Bond (**LRIBND**)as possible predictors for the Response Variable Invesco WilderHill Clean Energy ETF (**RPBW**).

This model kept only 24 variables of the original 40 Return and Lagged variables.

Finally, as a clear best estimation model, the model was put to use to estimate the parameters, their effects on the response variable and to find the strongest predictor for the Response Variable Invesco WilderHill Clean Energy ETF (RPBW). As seen in the green chart to the left, 4 strong predictors were chosen from the original list of 40 potential predictor variables. The Lasso Method is known to be able to estimate and predict simultaneously and therefore, drops uninformative variables. In this data set, sixteen variables were dropped and not considered in the last step of Predictor Variable Prediction.

Risk Factors for Oil Prices Before and After Covid-19

The generalized t (5) Adaptive Lasso Regression model picks the S&P 500 (large Company) index (RSPY) as the strongest predictor for the Response Variable Invesco WilderHill Clean Energy ETF (RPBW). JMP's Variable Importance: Independent Uniform Inputs Tool states that RSPY is responsible for 42.8% of the effect on RPBW. It also showed that these sixteen predictor variables effected RPBW, with nineteen of them effecting it 0.5% or less.



Column	Main Effect	Total Effect
RSPY	0.392	0.428
RXLK	0.295	0.331
RSLY	0.117	0.153
RVWO	0.03	0.056
LRSPY	0.005	0.013
RVSS	0.003	0.011
LRVSS	0.003	0.009
RVNQL	0.002	0.005
LRBWX	0.001	0.004
LRIBND	0.002	0.004
LRLQD	0.001	0.004
RVEU	0.001	0.003
LRHYG	0.001	0.003
LRVNQL	0.001	0.003
LRVWO	0.001	0.003
LRVIX	4e-4	0.001
ROVX	4e-4	0.001
LREMB	1e-4	4e-4
RTIP	8e-5	4e-4
LROVX	6e-5	3e-4
LRSLY	9e-5	2e-4
RHYG	3e-5	1e-4
RUSO	2e-5	4e-5
RFXE	4e-7	3e-6

Generalized t (5) Adaptive Lasso Regression model picks the S&P 500 (large Company) index (**RSPY**) as the strongest predictor for the Response Variable Invesco WilderHill Clean Energy ETF (**RPBW**). **RSPY** has 42.8% of the effects on the Response Variable **RPBW**.

Technology Select Sector SPDR Fund (**RXLK**) comes in as a strong second-best predictor with 33.1% of the effects.

4. MODEL INTERPRETATION

On the right, The Marginal Model Plots the Predictor Variables in the same order presented in the Summary Report and it shows an Inverse Relationship to our Response variable in the bottom left graph. This means that when the S&P 500 (large Company) index (RSPY) goes down, the Invesco WilderHill Clean Energy ETF (RPBW) increases. However, the second most important variable, Technology Select Sector SPDR Fund (RXLK) is responsible for 33.1% of the effect on RPBW. When RXLK increases, so does the Invesco WilderHill Clean Energy ETF (RPBW). These two Predictor Variables amount to 75% of the effect these predictor variables have on the US Clean Energy Stocks. In conclusion, Oil Prices did not have much effect on renewable energy stock prices in the U.S., but High Technology Shares and General Interest Rates do influence Clean Energy Stocks.