Risk Factors for Oil Prices Before and After Covid-19

The task given was to explore the determinants of US oil price changes. As we saw with the Persian Gulf wars in the 1990's, oil can been seen as one of the most important factors to the world economy. In the Gulf wars, we saw that countries will fight over who controls oil production and refining. They will literally take over other countries just to possess the power to determine who will buy the oil commodities and for what price. According to the literature mentioned in our given task, "Literature argues that oil prices should show sensitivity" to the variables selected in our given data sets.

We are given two data sets: one is dated from September 17, 2014, to December 31, 2019, with 1,332 daily usable entries and the other is dated from January 20, 2020, to April 23, 2021 with 326 daily usable entries. The first data set specifically contains data before the COVID-19 world pandemic. The second one specifically contains data which incorporates the generally agreed upon date in which we saw COVID-19 affect the world with major countries forcing their citizens to be locked down at home and many economies ground to a halt with the exception of healthcare and safety.

Both data sets include the same variables which can be broken into 3 categories:

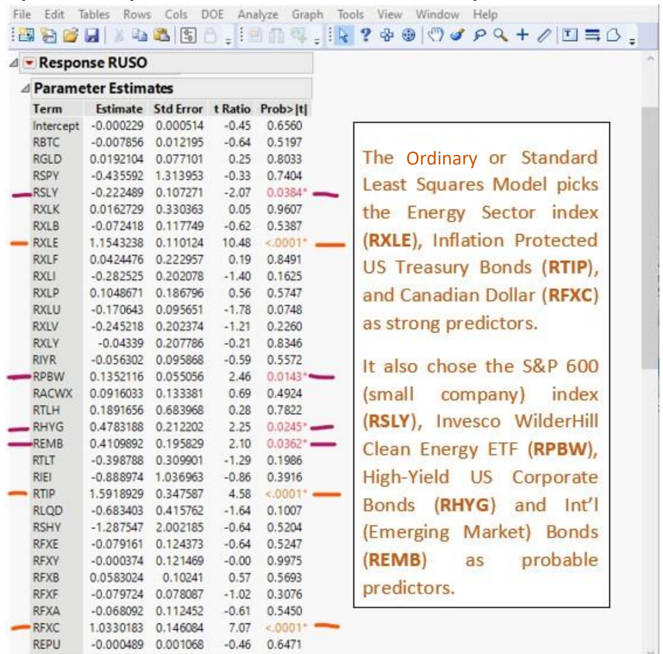
- Oil and Stock Prices – Economic Conditions seem to have a direct proportionate effect on both oil and stock prices; when one goes up, so does the other.
- Oil and Bond Prices – Bond rates also seem to have a direct variation with the price of US oil.
- Oil and Exchange Rates – The exchange value of the US Dollar relative to the exchange value of other world currencies tends to have an inverse relationship with US Oil rates.

In preparation of the Data Analysis, each of these two data sets had a cross-validation column added to the end of the data set. The observations were broken up into a group with 60% Training Data, 20% Validation data, and 20% Testing Data, otherwise known as a 60/20/20 split. SAS's JMP software was then able to calculate Statistics to be analyzed so that the most important variables could be chosen from the model to find predictive variables that would estimate how US Oil stocks would act in the future. Three Statistical models were compared: the **Ordinary or Least Squares Method (OLS)** which finds the values of the intercept and slope coefficient that minimizes the sum of the squared errors, the **Forward Stepwise Method** which adds one variable at a time to determine the most important variables, and the **Backward Stepwise Method** which starts with all the variables and takes away one variable at a time in order to find the best predictive variables.

Risk Factors for Oil Prices Before and After Covid-19

<u>Pre-Covid Data</u>: The Three Models used to find the most important predictive variable of oil prices:

1. Ordinary or Least Squares Method (OLS) – This method was set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Standard Least Squares Personality was run. This method found 3 strong predictors **RXLE**, **RTIP**, and **RFXC** plus 4 possible predictors as shown in the picture on the right. Using these 7 variables, a new model was created using the OLS Method which resulted in a **Training set R^2 value of 67.8%** and a **Test set R^2 value of 46.1%**, which are not close enough to say that this is a good model for predicting future oil prices.

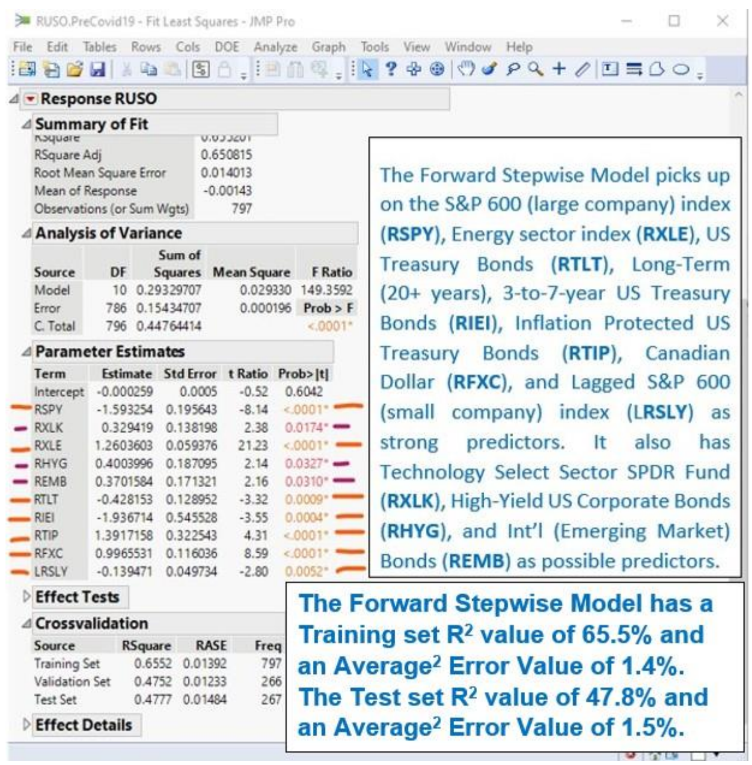


Response RUSO

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000229	0.000514	-0.45	0.6560
RBTC	-0.007856	0.012195	-0.64	0.5197
RGLD	0.0192104	0.077101	0.25	0.8033
RSPY	-0.435592	1.313953	-0.33	0.7404
RSLY	-0.222489	0.107271	-2.07	0.0384*
RXLK	0.0162729	0.330363	0.05	0.9607
RXLB	-0.072418	0.117749	-0.62	0.5387
RXLE	1.1543238	0.110124	10.48	<.0001*
RXLF	0.0424476	0.222957	0.19	0.8491
RXLI	-0.282525	0.202078	-1.40	0.1625
RXLP	0.1048671	0.186796	0.56	0.5747
RXLU	-0.170643	0.095651	-1.78	0.0748
RXLV	-0.245218	0.202374	-1.21	0.2260
RXLY	-0.04339	0.207786	-0.21	0.8346
RIYR	-0.056302	0.095868	-0.59	0.5572
RPBW	0.1352116	0.055056	2.46	0.0143*
RACWX	0.0916033	0.133381	0.69	0.4924
RTLH	0.1891656	0.683968	0.28	0.7822
RHYG	0.4783188	0.212202	2.25	0.0245*
REMB	0.4109892	0.195829	2.10	0.0362*
RTLT	-0.398788	0.309901	-1.29	0.1986
RIEI	-0.888974	1.036963	-0.86	0.3916
RTIP	1.5918929	0.347587	4.58	<.0001*
RLQD	-0.683403	0.415762	-1.64	0.1007
RSHY	-1.287547	2.002185	-0.64	0.5204
RFXE	-0.079161	0.124373	-0.64	0.5247
RFXY	-0.000374	0.121469	-0.00	0.9975
RFXB	0.0583024	0.10241	0.57	0.5693
RFXF	-0.079724	0.078087	-1.02	0.3076
RFXA	-0.068092	0.112452	-0.61	0.5450
RFXC	1.0330183	0.146084	7.07	<.0001*
REPU	-0.000489	0.001068	-0.46	0.6471

2. Forward Stepwise Method- This method was again set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Stepwise Personality was run. In the Stepwise Regression Control Window, the Stopping Rule that was chosen was the "Max Validation RSquare" and the "Forward" Direction was chosen. This variable-picking model returned 10 possible predictors as seen on the right. These variables were then included in a new OLS model which resulted in a **Training set R^2 value of 65.5%** and a **Test set R^2 value of 47.8%**. This model showed **RXLK**, **RHYG**, and **REMB** as very significant variables compared to the other 7 **RSPY**, **RXLE**, **RTLT**, **RIEI**, **RTIP**, **RFXC**, and **LRSLY**.



RUSO.PreCovid19 - Fit Least Squares - JMP Pro

Response RUSO

Summary of Fit

RSquare	[illegible]
RSquare Adj	0.650815
Root Mean Square Error	0.014013
Mean of Response	-0.00143
Observations (or Sum Wgts)	797

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	10	0.29329707	0.029330	149.3592
Error	786	0.15434707	0.000196	Prob > F
C. Total	796	0.44764414		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000259	0.0005	-0.52	0.6042
RSPY	-1.593254	0.195643	-8.14	<.0001*
RXLK	0.329419	0.138198	2.38	0.0174*
RXLE	1.2603603	0.059376	21.23	<.0001*
RHYG	0.4003996	0.187095	2.14	0.0327*
REMB	0.3701584	0.171321	2.16	0.0310*
RTLT	-0.428153	0.128952	-3.32	0.0009*
RIEI	-1.936714	0.545528	-3.55	0.0004*
RTIP	1.3917158	0.322543	4.31	<.0001*
RFXC	0.9965531	0.116036	8.59	<.0001*
LRSLY	-0.139471	0.049734	-2.80	0.0052*

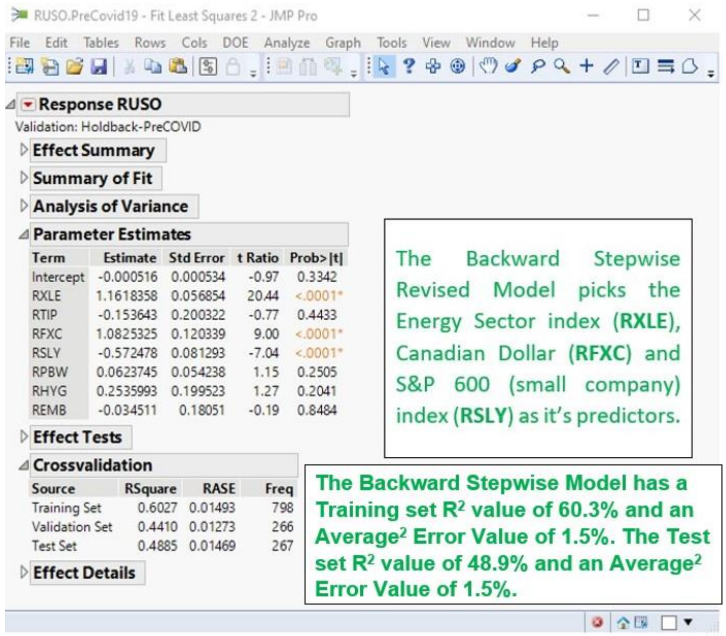
Effect Tests

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.6552	0.01392	797
Validation Set	0.4752	0.01233	266
Test Set	0.4777	0.01484	267

Effect Details

3. Backward Stepwise - This method was again set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Stepwise Personality was run. In the Stepwise Regression Control Window, the Stopping Rule that was chosen was the "Max Validation RSquare" and the "Backward" Direction was chosen. This model initially has ALL the variables and returned 10 possible predictors as seen on the right. This method found the same 7 variables as the OLS Method as strong or possible. Using these 7 variables, a new model was created using the OLS Method which resulted in a **Training set R^2 value of 60.3%** and a **Test set R^2 value of 48.9%**, which are not close enough to say that this is a good model for predicting future oil prices.



RUSO.PreCovid19 - Fit Least Squares 2 - JMP Pro

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

Response RUSO

Validation: Holdback-PreCOVID

Effect Summary

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000516	0.000534	-0.97	0.3342
RXLE	1.1618358	0.056854	20.44	<.0001*
RTIP	-0.153643	0.200322	-0.77	0.4433
RFXC	1.0825325	0.120339	9.00	<.0001*
RSLY	-0.572478	0.081293	-7.04	<.0001*
RPBW	0.0623745	0.054238	1.15	0.2505
RHYG	0.2535993	0.199523	1.27	0.2041
REMB	-0.034511	0.18051	-0.19	0.8484

The Backward Stepwise Revised Model picks the Energy Sector index (**RXLE**), Canadian Dollar (**RFXC**) and S&P 600 (small company) index (**RSLY**) as it's predictors.

Effect Tests

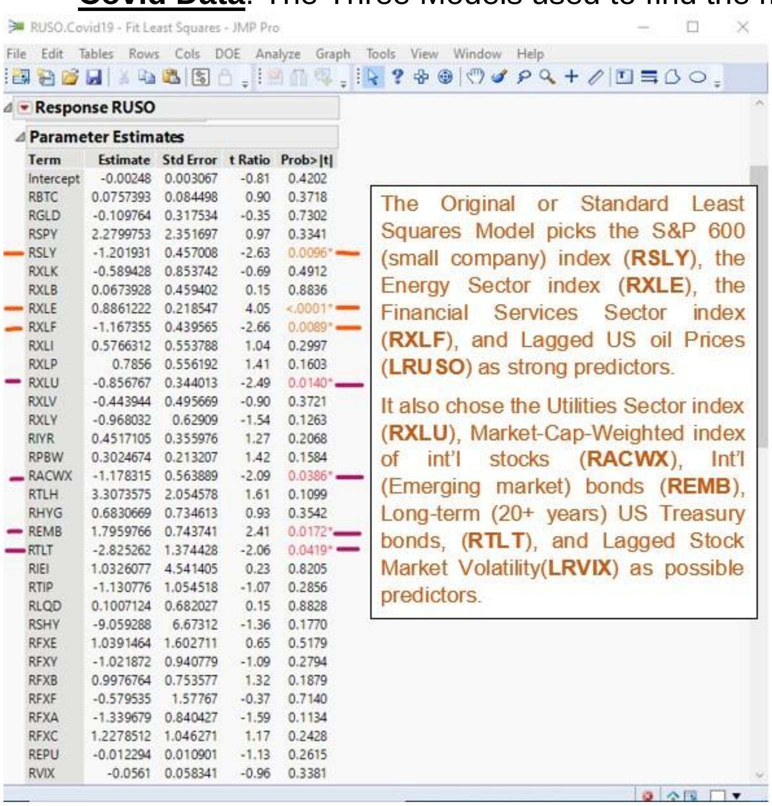
Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.6027	0.01493	798
Validation Set	0.4410	0.01273	266
Test Set	0.4885	0.01469	267

The Backward Stepwise Model has a Training set R^2 value of 60.3% and an Average^2 Error Value of 1.5%. The Test set R^2 value of 48.9% and an Average^2 Error Value of 1.5%.

Effect Details

<u>**Covid Data**</u>: The Three Models used to find the most important predictive variable of oil prices:



RUSO.Covid19 - Fit Least Squares - JMP Pro

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

Response RUSO

Parameter Estimates

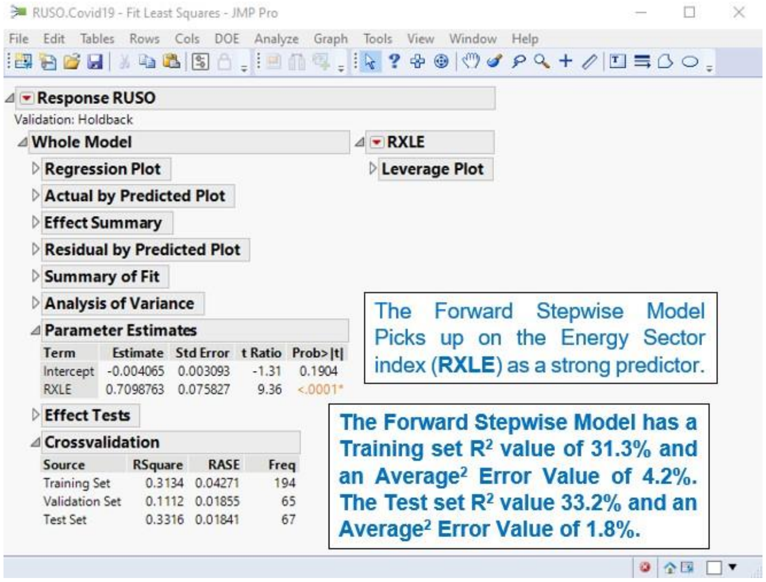
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.00248	0.003067	-0.81	0.4202
RBTC	0.0757393	0.084498	0.90	0.3718
RGLD	-0.109764	0.317534	-0.35	0.7302
RSPY	2.2799753	2.351697	0.97	0.3341
RSLY	-1.201931	0.457008	-2.63	0.0096*
RXLK	-0.589428	0.853742	-0.69	0.4912
RXLB	0.0673928	0.459402	0.15	0.8836
RXLE	0.8861222	0.218547	4.05	<.0001*
RXLF	-1.167355	0.439565	-2.66	0.0089*
RXLI	0.5766312	0.553788	1.04	0.2997
RXLP	0.7856	0.556192	1.41	0.1603
RXLU	-0.856767	0.344013	-2.49	0.0140*
RXLV	-0.443944	0.495669	-0.90	0.3721
RXLY	-0.968032	0.62909	-1.54	0.1263
RIYR	0.4517105	0.355976	1.27	0.2068
RPBW	0.3024674	0.213207	1.42	0.1584
RACWX	-1.178315	0.563889	-2.09	0.0386*
RTLH	3.3073575	2.054578	1.61	0.1099
RHYG	0.6830669	0.734613	0.93	0.3542
REMB	1.7959766	0.743741	2.41	0.0172*
RTLT	-2.825262	1.374428	-2.06	0.0419*
RIEI	1.0326077	4.541405	0.23	0.8205
RTIP	-1.130776	1.054518	-1.07	0.2856
RLQD	0.1007124	0.682027	0.15	0.8828
RSHY	-9.059288	6.67312	-1.36	0.1770
RFXE	1.0391464	1.602711	0.65	0.5179
RFXY	-1.021872	0.940779	-1.09	0.2794
RFXB	0.9976764	0.753577	1.32	0.1879
RFXF	-0.579535	1.57767	-0.37	0.7140
RFXA	-1.339679	0.840427	-1.59	0.1134
RFXC	1.2278512	1.046271	1.17	0.2428
REPU	-0.012294	0.010901	-1.13	0.2615
RVIX	-0.0561	0.058341	-0.96	0.3381

The Original or Standard Least Squares Model picks the S&P 600 (small company) index (**RSLY**), the Energy Sector index (**RXLE**), the Financial Services Sector index (**RXLF**), and Lagged US oil Prices (**LRUSO**) as strong predictors.

It also chose the Utilities Sector index (**RXLU**), Market-Cap-Weighted index of int'l stocks (**RACWX**), Int'l (Emerging market) bonds (**REMB**), Long-term (20+ years) US Treasury bonds, (**RTLT**), and Lagged Stock Market Volatility(**LRVIX**) as possible predictors.

1. Ordinary or Least Squares Method (OLS) – This method was set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Standard Least Squares Personality was run. This method found 4 strong predictors **RSLY, RXLE**, **RXLF**, and **LRUSO** plus 5 possible predictors as shown in the picture on the left. Using these 9 variables, a new model was created using the OLS Method which resulted in a **Training set R^2 value of 70.6%** and a **Test set R^2 value of negative 90.8%**, which was reflected noise and showed a poor model.

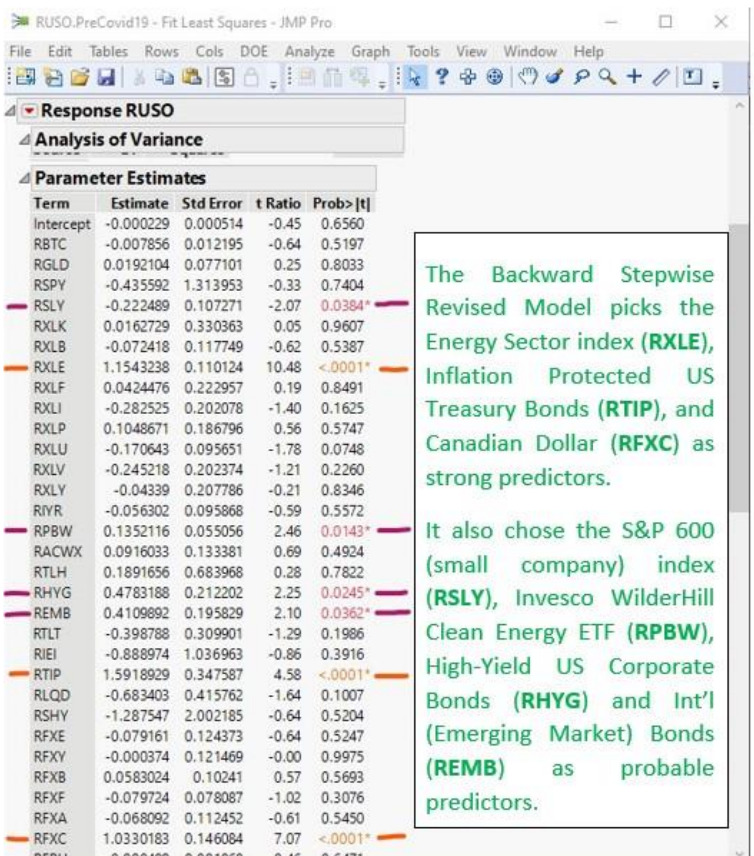
2. Forward Stepwise Method- This method was again set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Stepwise Personality was run. In the Stepwise Regression Control Window, the Stopping Rule that was chosen was the "Max Validation RSquare" and the "Forward" Direction was chosen. This variable-picking model returned 1 possible predictor as seen on the right. This variable was then included in a new OLS model which resulted in a **Training set R^2 value of 31.3%** and a **Test set R^2 value of 33.2%**. Having these two results similar, showed a good model.



Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.004065	0.003093	-1.31	0.1904
RXLE	0.7098763	0.075827	9.36	<.0001*

Source	RSquare	RASE	Freq
Training Set	0.3134	0.04271	194
Validation Set	0.1112	0.01855	65
Test Set	0.3316	0.01841	67

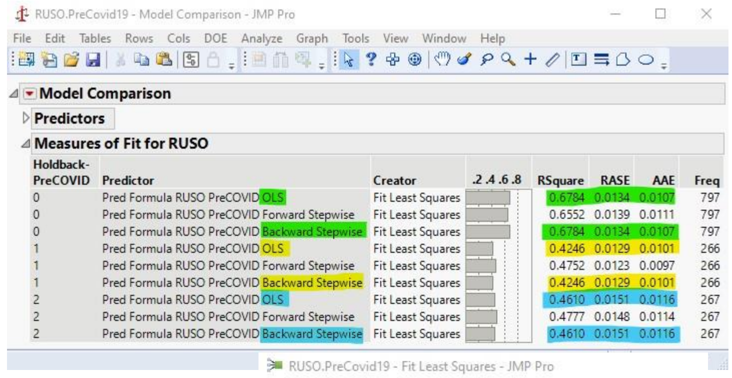
3. Backward Stepwise - This method was again set-up using the SAS JMP software to have a Y Variable of the concurrent returns of US Oil (RUSO) with a Validation Column. All other concurrent return or lagged variables were initially included and the Stepwise Personality was run. In the Stepwise Regression Control Window, the Stopping Rule that was chosen was the "Max Validation RSquare" and the "Backward" Direction was chosen. This model initially has ALL the variables and returned 10 possible predictors as seen on the right. This method found the same 9 variables as the OLS Method as strong or possible. Using these 9 variables, a new model was created using the OLS Method which resulted in a **Training set R^2 value of 70.6%** and a **Test set R^2 value of 90.8%**, which are not similar enough to call this a good model.



Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000229	0.000514	-0.45	0.6560
RBTC	-0.007856	0.012195	-0.64	0.5197
RGLD	0.0192104	0.077101	0.25	0.8033
RSPY	-0.435592	1.313953	-0.33	0.7404
RSLY	-0.222489	0.107271	-2.07	0.0384*
RXLK	0.0162729	0.330363	0.05	0.9607
RXLB	-0.072418	0.117749	-0.62	0.5387
RXLE	1.1543238	0.110124	10.48	<.0001*
RXLF	0.0424476	0.222957	0.19	0.8491
RXLI	-0.282525	0.202078	-1.40	0.1625
RXLP	0.1048671	0.186796	0.56	0.5747
RXLU	-0.170643	0.095651	-1.78	0.0748
RXLV	-0.245218	0.202374	-1.21	0.2260
RXLY	-0.04339	0.207786	-0.21	0.8346
RIYR	-0.056302	0.095868	-0.59	0.5572
RPBW	0.1352116	0.055056	2.46	0.0143*
RACWX	0.0916033	0.133381	0.69	0.4924
RTLH	0.1891656	0.683968	0.28	0.7822
RHYG	0.4783188	0.212202	2.25	0.0245*
REMB	0.4109892	0.195829	2.10	0.0362*
RTLT	-0.398788	0.309901	-1.29	0.1986
RIEI	-0.888974	1.036963	-0.86	0.3916
RTIP	1.5918929	0.347587	4.58	<.0001*
RLQD	-0.683403	0.415762	-1.64	0.1007
RSHY	-1.287547	2.002185	-0.64	0.5204
RFXE	-0.079161	0.124373	-0.64	0.5247
RFXY	-0.000374	0.121469	-0.00	0.9975
RFXB	0.0583024	0.10241	0.57	0.5693
RFXF	-0.079724	0.078087	-1.02	0.3076
RFXA	-0.068092	0.112452	-0.61	0.5450
RFXC	1.0330183	0.146084	7.07	<.0001*
REPU	-0.000489	0.001068	-0.46	0.6471

Risk Factors for Oil Prices Before and After Covid-19

Using JMP's Model Comparison Tool, plus the three models all three models in both data sets could be compared.



RUSO.PreCovid19 - Model Comparison - JMP Pro

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

Model Comparison

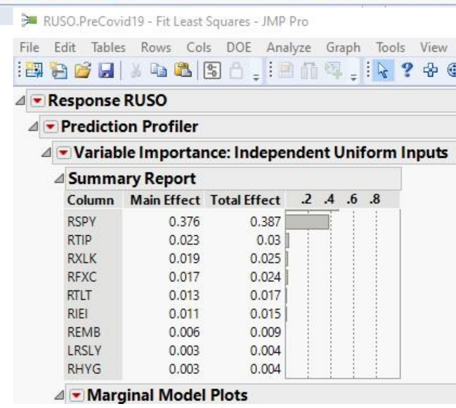
Predictors

Measures of Fit for RUSO

Holdback-PreCOVID	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
0	Pred Formula RUSO PreCOVID OLS	Fit Least Squares		0.6784	0.0134	0.0107	797
0	Pred Formula RUSO PreCOVID Forward Stepwise	Fit Least Squares		0.6552	0.0139	0.0111	797
0	Pred Formula RUSO PreCOVID Backward Stepwise	Fit Least Squares		0.6784	0.0134	0.0107	797
1	Pred Formula RUSO PreCOVID OLS	Fit Least Squares		0.4246	0.0129	0.0101	266
1	Pred Formula RUSO PreCOVID Forward Stepwise	Fit Least Squares		0.4752	0.0123	0.0097	266
1	Pred Formula RUSO PreCOVID Backward Stepwise	Fit Least Squares		0.4246	0.0129	0.0101	266
2	Pred Formula RUSO PreCOVID OLS	Fit Least Squares		0.4610	0.0151	0.0116	267
2	Pred Formula RUSO PreCOVID Forward Stepwise	Fit Least Squares		0.4777	0.0148	0.0114	267
2	Pred Formula RUSO PreCOVID Backward Stepwise	Fit Least Squares		0.4610	0.0151	0.0116	267

<u>Pre-Covid Data</u>:

The Forward Stepwise Model was chosen as the better model as it had 17.8% difference when comparing the Training R^2 values and 0.09% different with their Root Average Squared Error (RASE) values. The other two were 21.7% different when comparing the Training R^2 values and 0.17% different their Root Average Squared Error (RASE) values. Using this model and JMP's Variable Importance Tool, <u>the S&P 600 (large company) index (**RSPY**) stood out as the most important factor</u> when predicting oil prices prior to the COVID-19 Pandemic. This correlates to the literature which states that Economic Conditions seem to have a direct proportionate effect on both oil and stock prices; when one goes up, so does the other.



RUSO.PreCovid19 - Fit Least Squares - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools View

Response RUSO

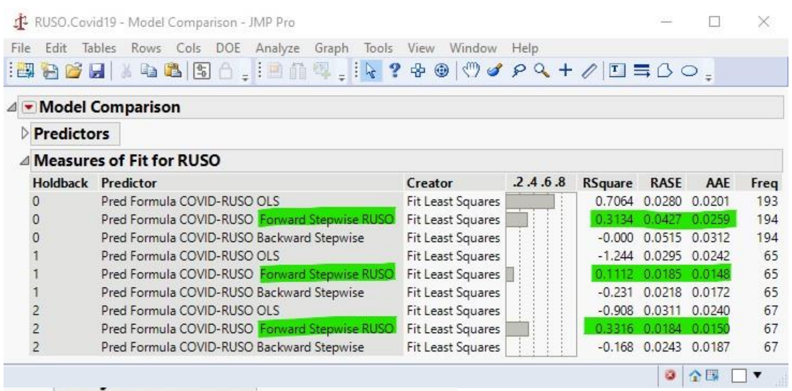
Prediction Profiler

Variable Importance: Independent Uniform Inputs

Summary Report

Column	Main Effect	Total Effect	.2 .4 .6 .8
RSPY	0.376	0.387	
RTIP	0.023	0.03	
RXLK	0.019	0.025	
RFXC	0.017	0.024	
RTLT	0.013	0.017	
RIEI	0.011	0.015	
REMB	0.006	0.009	
LRSLY	0.003	0.004	
RHYG	0.003	0.004	

Marginal Model Plots



RUSO.Covid19 - Model Comparison - JMP Pro

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

Model Comparison

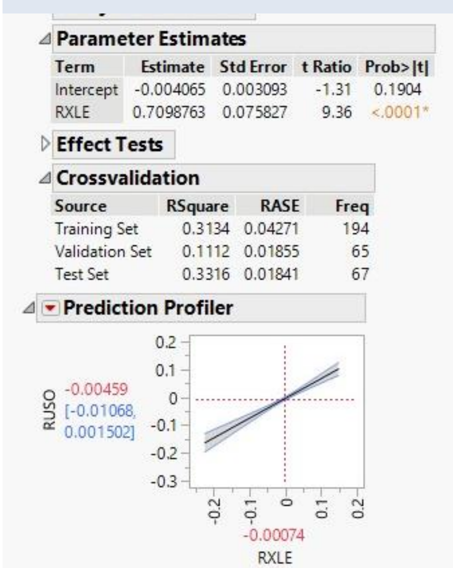
Predictors

Measures of Fit for RUSO

Holdback	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
0	Pred Formula COVID-RUSO OLS	Fit Least Squares		0.7064	0.0280	0.0201	193
0	Pred Formula COVID-RUSO Forward Stepwise RUSO	Fit Least Squares		0.3134	0.0427	0.0259	194
0	Pred Formula COVID-RUSO Backward Stepwise	Fit Least Squares		-0.000	0.0515	0.0312	194
1	Pred Formula COVID-RUSO OLS	Fit Least Squares		-1.244	0.0295	0.0242	65
1	Pred Formula COVID-RUSO Forward Stepwise RUSO	Fit Least Squares		0.1112	0.0185	0.0148	65
1	Pred Formula COVID-RUSO Backward Stepwise	Fit Least Squares		-0.231	0.0218	0.0172	65
2	Pred Formula COVID-RUSO OLS	Fit Least Squares		-0.908	0.0311	0.0240	67
2	Pred Formula COVID-RUSO Forward Stepwise RUSO	Fit Least Squares		0.3316	0.0184	0.0150	67
2	Pred Formula COVID-RUSO Backward Stepwise	Fit Least Squares		-0.168	0.0243	0.0187	67

<u>Covid Data</u>:

The Forward Stepwise Model was chosen as the better model as it had a positive R2 value compared with the OLS Method and the Backward Stepwise Method having negative R2 values for their testing data, reflecting a large amount of noise in the data. Using this model and JMP's Variable Importance Tool, <u>the Energy Sector index (**RXLE**) stood out as the most important factor</u> when predicting oil prices during the COVID-19 Pandemic. This correlates to the literature which states that Economic Conditions seem to have a direct proportionate effect on both oil and stock prices; when one goes up, so does the other.



Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.004065	0.003093	-1.31	0.1904
RXLE	0.7098763	0.075827	9.36	<.0001*

Effect Tests

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.3134	0.04271	194
Validation Set	0.1112	0.01855	65
Test Set	0.3316	0.01841	67

Prediction Profiler

When comparing the two data sets, it definitely appears that more variables could impact the prediction of oil prices prior to COVID. During COVID, each countries' reaction to the pandemic would have more impact on the price of oil, individually. The Energy Sector stocks since COVID have declined quite a bit as the world economy has stabilized.