Executive Summary

The Stock Market is known for being extremely unpredictable. With so many companies and potential relationships, to predict the daily closing price of a stock for a company may take considerable time and effort. A share of F5's stock price is around $200. Over the range of time this data covers, it has fluctuated considerably; this project aims to predict F5's daily closing stock price with other companies' prices. Firstly, the entire stock market will not be used but only the S&P 500. Secondly, A subset of the S&P 500, the Information Technology GICS sector, will be analyzed. All analysis will be done with SAS.

F5's daily closing stock price is the dependent variable. F5 is a company that focuses on online security. They provide services to many large companies, with Tesla and Microsoft being notable customers. They have competitors in other parts of the stock market, but none are in the data collected. Because they are focused on security, the demand for their services has been high because of the switch to more online learning classes. However, recently their stock has been doing slightly worse than usual; the price per share at the end of the data's range is actually the same as it was in the beginning despite F5 showing increases in their stock for most of the range.

It was suspected that any of the information technology daily stock prices could relate to F5's daily closing price significantly. Performing a variable screening method reduced the number of variables from 74 to 32. Multicollinearity was highly present in this stage; many independent variables had to be removed since they correlated with each other and not strictly with F5's daily stock price. A correlation matrix was performed to see if each variable was at least moderately correlated with F5. If this condition wasn't satisfied, they were removed. The final model included no higher order terms. Adding higher order terms may make the model too complex, leading to high variance for out of sample data, so opting for a simpler model, but not too simple, is the method of approach that was taken. An inclusion of an interaction between companies ANET and MSI made the prediction equation more accurate. The R-squared adjusted for the final model with variables VRSN, TER, CDW, ANET, MSI, and the interaction between ANET and MSI was .9080. All of the variables, bar the interaction term and its components, have variance inflation factors less than 10. Every variable was at least moderately correlated with F5. The root mean squared error and coefficient of variation were low at 5.05 and 2.45, respectively. All variables and the overall utility of the model were extremely significant, each with a p-value less than .05. The fit diagnostics were all satisfactory; this includes the residual plots, the Q-Q plot, and the Cook's D plot. Only 5% of the observations were outliers, and very few were extreme outliers. The model is excellent overall, with all of the components meeting or exceeding the minimum requirements for a good model.

Data

The data spans from April 1, 2021 to March 31, 2022. The main interest was F5's daily closing stock prices and predicting that based on other variables. The analysis left only these variables' daily closing stock prices in the final model: Verisign (VRSN), who mainly handles arrays of network infrastructure with a slight focus in online security services; Teradyne (TER), who designs automatic test equipment to test materials from their clients; CDW Corporation (CDW), who provides products such as laptops, servers, and other software for governments and businesses; and the interaction between Arista Networks (ANET) and Motorola Solutions (MSI), the former focusing on providing network operating systems and network switches, while the latter provides software, telecommunications equipment, and more.

When examining the relationship between F5's daily stock price and one of the parameters here, the trends were similar, especially in the final fiscal business quarter in 2021 (Figure 1.1 - Figure 1.5). F5's stock prices were low in the summer months and kept increasing until January. F5 switching to a software-based business model and schools implementing more online learning may be some causes of this increase. Schools require security in online environments; when school is out of session, the demand for F5's license is low, as seen from the trends in the line graph.

Because MSI and ANET interacted closely with one another, a line graph was made to examine the relationship between them (Figure 1.5). Although they behaved similarly over the time span, it is somewhat difficult to determine that they did have an interaction. An interaction plot supplement showed it existed (Figure 2.7).

A Cook's D plot for the residuals shows that only 15 out of the 274 observations are outliers. Out of these 15, three of them appear to be major outliers (Figure 1.6 - Figure 1.8). The value for an outlier was if the residual was over a cent and a half from the predicted line. The maximum was 18 cents, and the next two highest were 4 cents. The rest of the outliers were about two cents away from the predicted line. With only about 5% of the data being outliers, with many of them not being egregious values, I believe they would have little influence over the final model's predicting power especially with how consistent the rest of the observations are.

Methodology

All the companies in the Information Technology GICS sector for the S&P 500, along with the closing stock price for each day within the time span, are included. Because so many variables were present, a variable screening method had to be performed so that only the potentially most significant terms remain. This was done by stating "selection=stepwise" in the "proc reg" model statement in SAS. Stepwise selection searches through all the variables and picks the most significant variable for the best one-variable model; then, it searches again for the next variable that would make for the best two-variable model. The procedure constantly performs t-tests in order to ensure the most statistically significant variables are there; consequently, some variables that are included can be abruptly taken out if they are no longer deemed significant by the process. This is done until the highest variable first-order model that has all statistically significant terms is created.

The final result was a 32 variable first-order model. However, with so many variables, multicollinearity is inevitable. By observing the variance inflation factor “/vif” in the “proc reg” model statement, we can check if there is evidence for multicollinearity for each variable. A high VIF means one of the independent variables is highly correlated with another independent variable. The goal is to have the independent variables only correlate well with the dependent variable. Any VIF value of above 10 shows evidence for multicollinearity. A value close to 0 is ideal, but a good threshold is 10 and below. Many of the variables were removed, usually the ones with the highest vif taken out one by one, until all the variables had acceptable VIF values.

The next step involved creating separate data in SAS since we were interested in higher order terms. Scatterplots for all of the variables that passed the screening selection but were removed because of their VIF were created. Many showed evidence for slight curvature The t-tests revealed that many of them scatterplots actually had a better fit if the variable was transformed in this way, by squaring it. The notable ones were marked as to be added later to the model after testing to see if an interaction between some variables was present first.

F5 is a company specializing in communications equipment. The four other companies who share this niche are Arista Networks (ANET), Cisco Systems (CSCO), Juniper Networks (JNPR), and Motorola Solutions Inc (MSI). It was hypothesized that there would be some interaction between a combination of these companies with F5, so tests for each combination were conducted, and new data was created for these potential interaction terms.

Two correlation matrices were created; the first to see if each of the independent variables correlated at least moderately strongly with F5, and the other to see if any of the variables taken out had strong correlations with F5. This was performed with the “proc corr” statement.

After all these steps, the final model was created. Some exploratory data analysis was done afterwards, such as highlighting the differences in stock prices over the time span between F5 and one of the independent variables, examining the fit plots in the output, and observing the outliers shown by the Cook’s D plot.

Results

After cleaning up some data, stepwise regression was performed in the model statement for all variables (Figure 2.8). The default slentry of .15 was used. That is, any variable over this limit was excluded since their t-test was deduced to be insignificant by the process. Although the alpha value, the level of significance, is .05, there might be potential variables that may have been excluded that just weren’t significant when so many variables were included. This is the rationale behind using .15 as the slentry.

Many of the variables were highly correlated with one another, so manual removal of the highest VIF observations were performed. Additionally, some of the removed variables were added again to see if they would be significant with less values in the model, though none of them ended up being significant. This was done until all the parameters had acceptable VIF values. The first order model at the end of this process included the variables INTC, VRSN, TER, WDC, PAYX, SEDG, and CDW. The specifics of this first order model were satisfactory (Figure 1.9).

It’s important to also test higher order or interaction terms to determine if a better model is achievable. The next step was creating and observing the scatterplots and seeing if there was a higher order trend between F5’s stock price and a specific company’s. The variables observed were those that were taken out during the VIF removal process. Many of the 2nd order variables had better correlations with F5’s closing stock price. They were implemented one at a time into the first order model. However, the squared terms were always insignificant when this was done. Replacing one higher order term with another always led to the same result: an extremely high p-value. Although not all the higher order variables were tested, it was inferred that if even the most significant higher order terms were incompatible with the model, then the rest wouldn’t be any better. Ultimately, no higher order terms were implemented. Other variable transformations were thought about, but the relationship in the scatterplots seemed too complex to model accurately with simple transformations. Log transformations and reciprocal transformations would not suffice, along with others, so experimentation was at a minimum for this aspect.

We suspected that an interaction term would most likely have a positive effect on the model. Specifically, this is concerning the companies in IT who have the same niche as F5. The individual components of the interaction term are included in the model out of necessity; the significant term is the interaction between those two variables. There is some sense to why these would be strongly correlated with F5. Some of these companies offer products like laptops and phones, while others are in charge of domain names on the internet. F5 is a company focused on security, so it could be reasonable to assume that if one of these company’s stock prices behave a certain way, then F5’s would also behave similarly. Perhaps there is something faulty in CDW Corporation's servers and it is deemed unsafe. There could be positive or negative relationships that could result.

Each combination was tested. The interaction between ANET and MSI had the best effect on the model. However, the VIF factor is extremely high at 643. This is inevitable; by definition, the three terms are all correlated with another in some way, so it is not unnatural to see a value so high. Because the t-test for this term is very significant and the sign of the parameter is positive – which is what one would expect – for the final model, I believe it is okay to include this term despite the VIF being so high. The interaction plot supplement showed that an interaction did exist since the lines were flared out, indicating dependence between the variables (Figure 2.7).

The final step is to create correlation matrices. First, one was made to see which of the variables didn’t at least moderately correlate with F5’s stock price (Figure 2.2). INTC, WDC, and SEDG were removed for this reason; they were weakly correlated with correlation coefficients below .05. Then, the second correlation matrix examined the relationship between F5 and all variables that were deemed either insignificant or had a high VIF (Figure 2.0 and Figure 2.1). These were the variables after the stepwise selection process. Some of the companies, like AMD, ADP, and KLAC had extremely high correlation coefficients from .8 to .9 with F5. When trying to include variables with coefficients this high into the current best model up to this point, they were either not significant or their VIF value was too high. None of these variables gathered from the 2nd matrix were used.

The best and final model included the variables VRSN, TER, CDW, ANET, MSI, and the interaction between ANET and MSI (Figure 2.4). The correlation matrix shows all the variables at least moderately correlated with F5 (Figure 2.3). The equation F5 = 342.518 - .696(VRSN) + .573(TER) + .422(CDW) - 1.943(ANET) -.701(MSI) + .009(ANET)(MSI) is the prediction equation corresponding to F5's daily closing stock price. When all variables have a closing stock price of $0, F5 will have one of $342.52. Holding all other variables constant, for every 1 unit increase in VRSN, F5's stock price will decrease by $0.69. This interpretation can be fit to VRSN, TER, and CDW. For ANET, -1.943 + .009(MSI) represents the change in F5's stock price for every $1 increase in ANET while holding all variables and MSI fixed. Likewise, -0.701 + .009(ANET) represents the change in F5's stock price for every $1 increase in MSI while holding all variables and ANET fixed. The root mean squared error, or standard deviation of residuals, is $5.05. We can expect 95% of the observations to fall within two standard deviations, or $10.10, from the predicted regression line. The coefficient of variation, the ratio between the standard deviation and the mean, is $2.45. The coefficient of determination, adjusted, is .9081, meaning 90.8% of the variation in F5 can be explained by the parameters in this model. That is, the model is about 90% accurate. The global F-test was extremely significant, as were the rest of the parameters with their t-tests with p-values less than .0001. The residual histogram is approximately normally distributed, so transformations are unneeded. The residual plots for each variable are randomly scattered as well, a satisfactory result (Figure 2.6). The fit-mean is larger than the residual plot in range, another sign of a great model. The quantile plot shows many of the observations along the line, a trend that is close to ideal. The Cook's D plot only shows a couple of potential outliers and one major outlier, but the rest of the residuals are satisfactory (Figure 2.5).

The parameters for the best model at predicting F5's stock price make sense. For example, some companies create laptops and phones. With products like this, there will always be a demand for network security. One would expect F5's stock price to be at least slightly influenced by some of these companies, whether that be negatively or positively.

One of the only potential problems is how VRSN is positively correlated with F5, yet it has a negative value when multiple regression is performed. I believe it is because this is also something that is inevitable; one of the variables, or perhaps more, are influencing this result. Because the variables have low VIF values, this does not appear to be multicollinearity but rather just a product of having multiple variables. Furthermore, this is the only variable in the model in which the sign was contrary to what was expected. Overall, the model is extremely strong when it comes to prediction power.

Conclusions

The final model is extremely good at predicting the stock price for F5. It has a 90% accuracy rate, significant values for each parameter, excellent fits for graphs, and small spread values. Though it is a great model, it is only good at predicting in-sample data. If we wish to predict out-of-sample data, we must either collect more data or perform a test and train split. About 25% of the data would have to be test data to simulate out-of-sample data, while the remaining would be the training, in-sample data. Using the training data's model to predict the testing model's observations should yield more accurate results as to whether the model is good for predicting out-of-sample data. For the next project, it may be useful to expand the range of companies from more than just the Information Technology GICS sector. Perhaps looking at the competitors' daily stock closing price from this timespan would be beneficial for the model.

Appendix


250
200
150
100
FFIV
Apr 2021
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2022
Feb
Mar
Apr
Date
FFIV
ANET

Figure 1.0

Figure 1.1

Figure 1.2

Figure 1.3

Figure 1.4

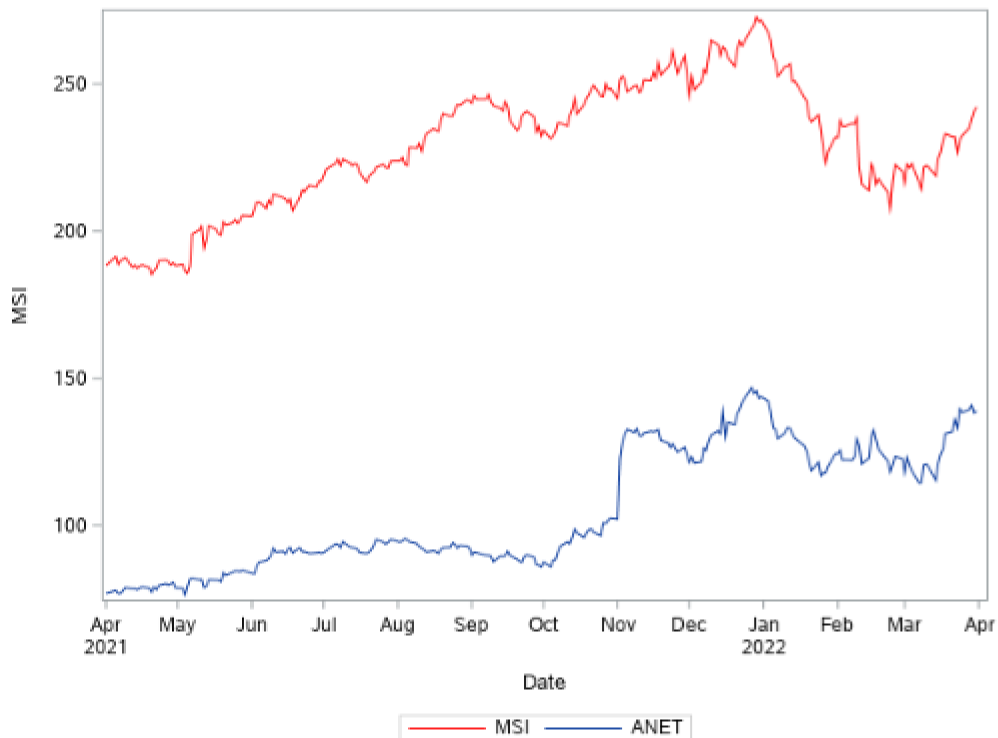
250
200
150
100
MSI
Apr 2021
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2022
Feb
Mar
Apr
Date
MSI
ANET

Figure 1.5

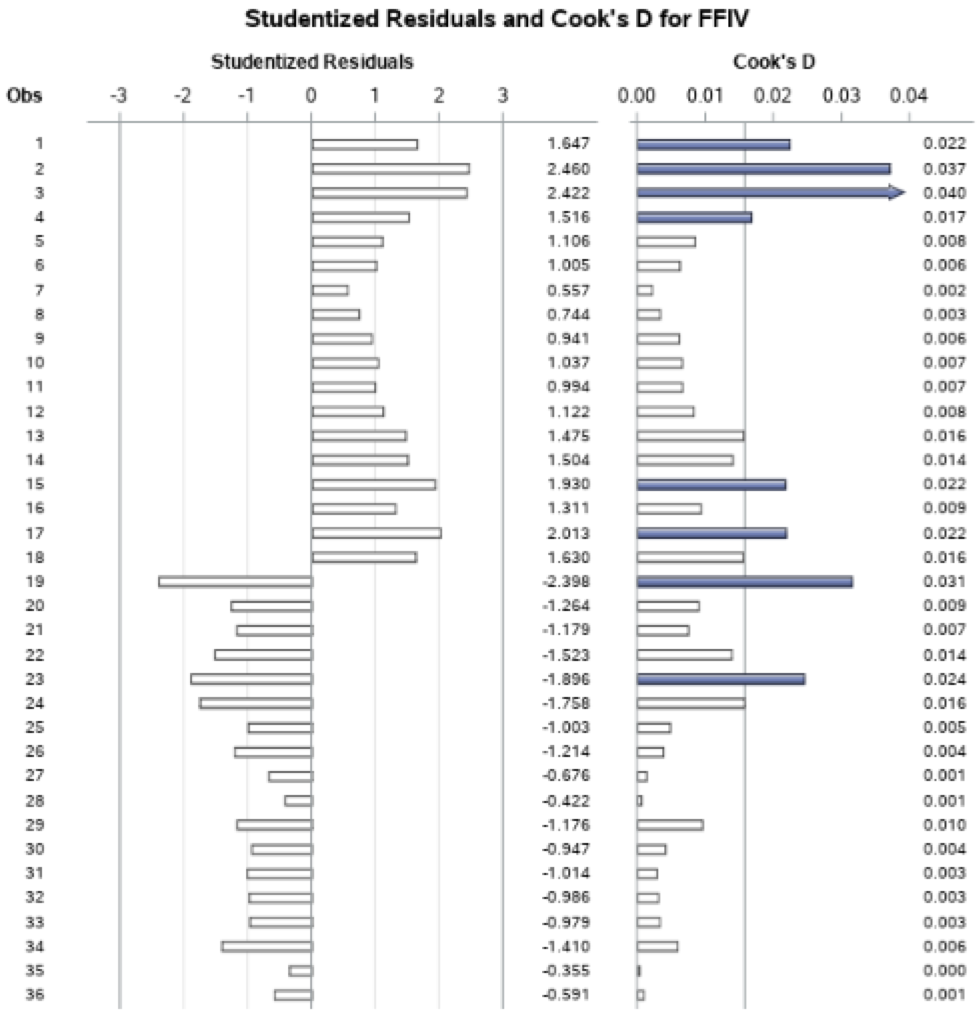
Studentized Residuals and Cook's D for FFIV
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03 0.04
1 1.647 0.022
2 2.460 0.037
3 2.422 0.040
4 1.516 0.017
5 1.106 0.008
6 1.005 0.006
7 0.557 0.002
8 0.744 0.003
9 0.941 0.006
10 1.037 0.007
11 0.994 0.007
12 1.122 0.008
13 1.475 0.016
14 1.504 0.014
15 1.930 0.022
16 1.311 0.009
17 2.013 0.022
18 1.630 0.016
19 -2.398 0.031
20 -1.264 0.009
21 -1.179 0.007
22 -1.523 0.014
23 -1.896 0.024
24 -1.758 0.016
25 -1.003 0.005
26 -1.214 0.004
27 -0.676 0.001
28 -0.422 0.001
29 -1.176 0.010
30 -0.947 0.004
31 -1.014 0.003
32 -0.986 0.003
33 -0.979 0.003
34 -1.410 0.006
35 -0.355 0.000
36 -0.591 0.001

Figure 1.6

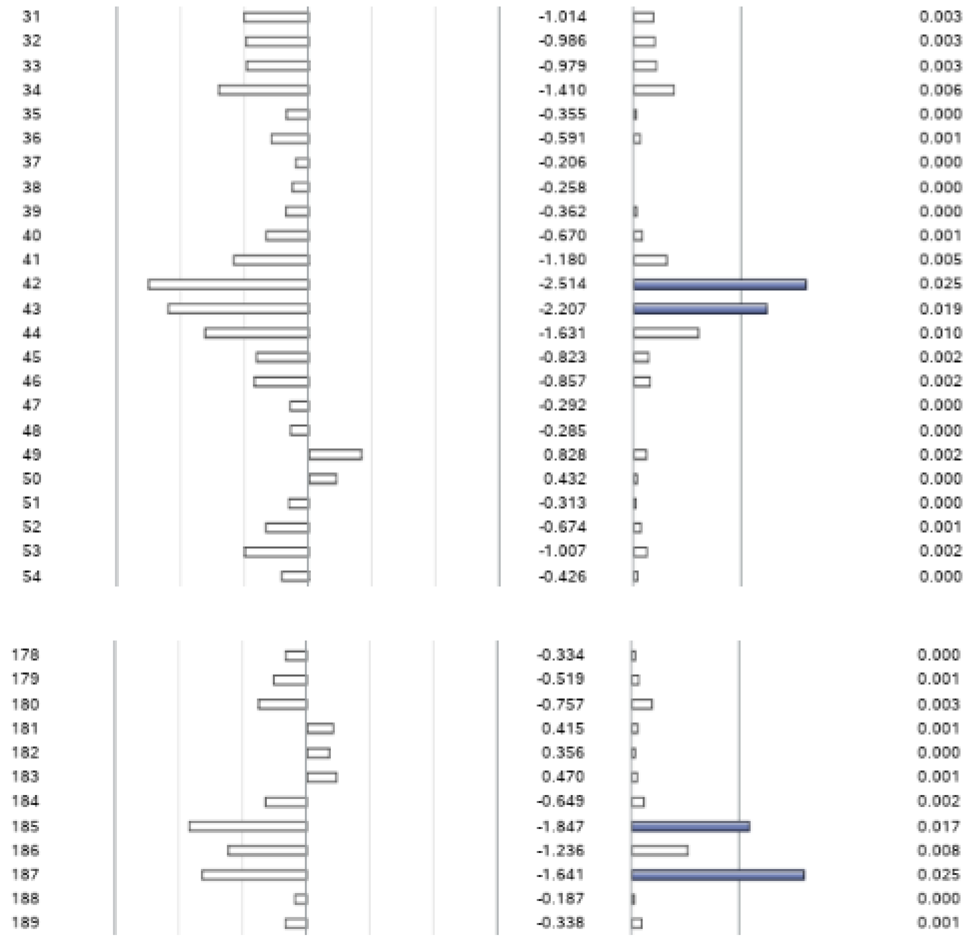
31 -1.014 0.003
32 -0.986 0.003
33 -0.979 0.003
34 -1.410 0.006
35 -0.355 0.000
36 -0.591 0.001
37 -0.206 0.000
38 -0.258 0.000
39 -0.362 0.000
40 -0.670 0.001
41 -1.180 0.005
42 -2.514 0.025
43 -2.207 0.019
44 -1.631 0.010
45 -0.823 0.002
46 -0.857 0.002
47 -0.292 0.000
48 -0.285 0.000
49 0.828 0.002
50 0.432 0.000
51 -0.313 0.000
52 -0.674 0.001
53 -1.007 0.002
54 -0.426 0.000
178 -0.334 0.000
179 -0.519 0.001
180 -0.757 0.003
181 0.415 0.001
182 0.356 0.000
183 0.470 0.001
184 -0.649 0.002
185 -1.847 0.017
186 -1.236 0.008
187 -1.641 0.025
188 -0.187 0.000
189 -0.338 0.001

Figure 1.7

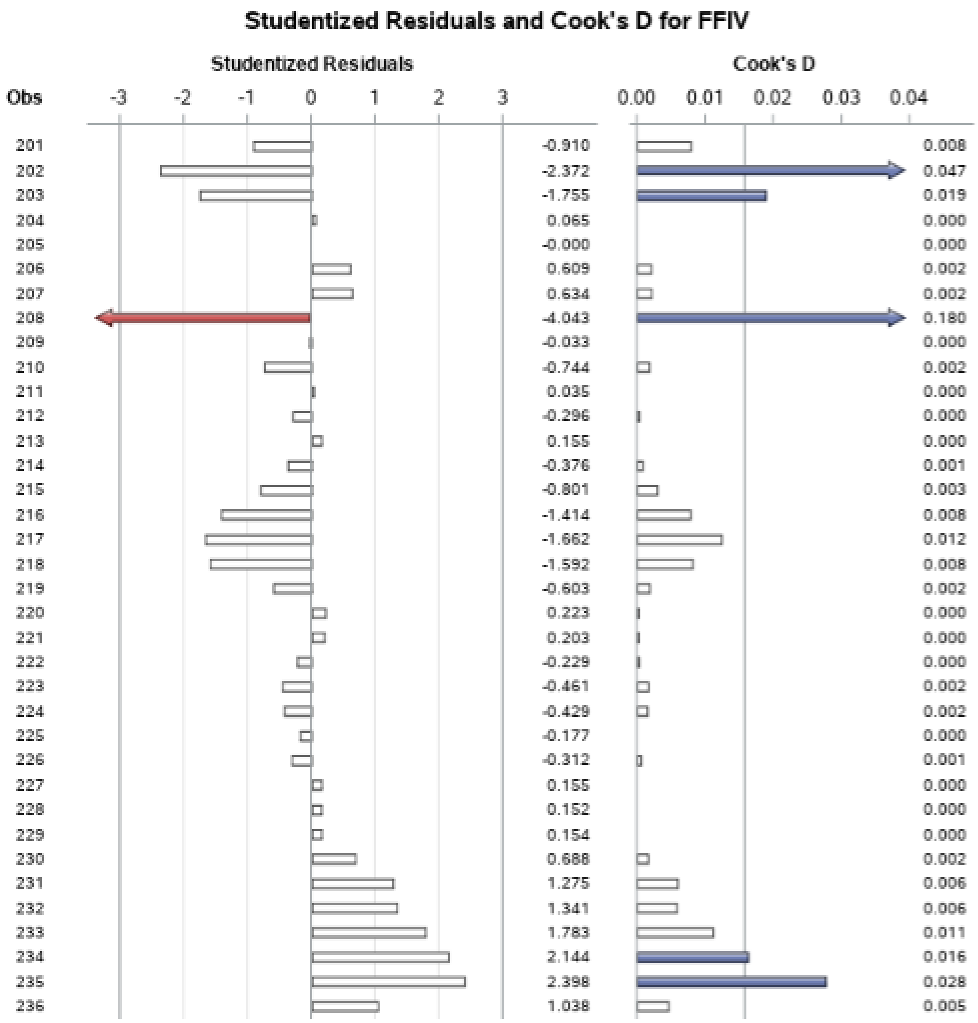
Studentized Residuals and Cook's D for FFIV
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03 0.04
201 -0.910 0.008
202 -2.372 0.047
203 -1.755 0.019
204 0.065 0.000
205 -0.000 0.000
206 0.609 0.002
207 0.634 0.002
208 -4.043 0.180
209 -0.033 0.000
210 -0.744 0.002
211 0.035 0.000
212 -0.296 0.000
213 0.155 0.000
214 -0.376 0.001
215 -0.801 0.003
216 -1.414 0.008
217 -1.662 0.012
218 -1.592 0.008
219 -0.603 0.002
220 0.223 0.000
221 0.203 0.000
222 -0.229 0.000
223 -0.461 0.002
224 -0.429 0.002
225 -0.177 0.000
226 -0.312 0.001
227 0.155 0.000
228 0.152 0.000
229 0.154 0.000
230 0.688 0.002
231 1.275 0.006
232 1.341 0.006
233 1.783 0.011
234 2.144 0.016
235 2.398 0.028
236 1.038 0.005

Figure 1.8

The REG Procedure
Model: MODEL1
Dependent Variable: FFIV FFIV

Number of Observations Read	254
Number of Observations Used	253
Number of Observations with Missing Values	1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	64355	9193.50444	404.53	<.0001
Error	245	5567.99438	22.72651		
Corrected Total	252	69923			

Root MSE	4.76723	R-Square	0.9204
Dependent Mean	206.06375	Adj R-Sq	0.9181
Coeff Var	2.31347		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	38.91673	12.51172	3.11	0.0021	0
INTC	INTC	1	0.88376	0.13775	6.42	<.0001	4.04263
VRSN	VRSN	1	-0.43277	0.04885	-8.86	<.0001	3.89770
TER	TER	1	0.86012	0.03810	22.57	<.0001	3.88876
WDC	WDC	1	-0.65479	0.08643	-7.58	<.0001	5.03379
PAYX	PAYX	1	0.46448	0.06398	7.26	<.0001	4.82434
SEDG	SEDG	1	0.03945	0.01015	3.89	0.0001	1.55876
CDW	CDW	1	0.43938	0.03535	12.43	<.0001	1.70944

Figure 1.9

Pearson Correlation Coefficients, N = 253
Prob > |r| under H0: Rho=0

	KLAC	CDNS	INTC	ADP	VRSN	TER	QCOM	NVDA	NXPI	KEYS	WDC
FFIV FFIV	0.90881 <.0001	0.87256 <.0001	-0.23015 0.0002	0.85650 <.0001	0.54937 <.0001	0.75193 <.0001	0.74441 <.0001	0.77500 <.0001	0.69297 <.0001	0.84669 <.0001	-0.29208 <.0001

	FLT	SNPS	ANET	CTXS	PAYX	AMD	NTAP	FISV	ACN	ADI
FFIV FFIV	-0.57117 <.0001	0.79913 <.0001	0.70110 <.0001	-0.55834 <.0001	0.69590 <.0001	0.85035 <.0001	0.59006 <.0001	-0.28184 <.0001	0.83614 <.0001	0.65159 <.0001

Figure 2.0

CSCO	ENPH	ZBRA	AVGO	EPAM	ANSS	MSI	JNPR	SEDG	CDW	BR
0.71680 <.0001	0.54824 <.0001	0.50571 <.0001	0.71632 <.0001	0.42184 <.0001	0.65294 <.0001	0.72702 <.0001	0.57887 <.0001	0.47758 <.0001	0.72475 <.0001	0.45263 <.0001

Figure 2.1

Pearson Correlation Coefficients, N = 253 Prob > |r| under H0: Rho=0									
	INTC	SEDG	WDC	VRSN	TER	CDW	ANET	MSI	anet_msi
FFIV FFIV	-0.23015 0.0002	0.47758 <.0001	-0.29208 <.0001	0.54937 <.0001	0.75193 <.0001	0.72475 <.0001	0.70110 <.0001	0.72702 <.0001	0.78626 <.0001

Figure 2.2

Pearson Correlation Coefficients, N = 253 Prob > |r| under H0: Rho=0						
	VRSN	TER	CDW	ANET	MSI	anet_msi
FFIV FFIV	0.54937 <.0001	0.75193 <.0001	0.72475 <.0001	0.70110 <.0001	0.72702 <.0001	0.78626 <.0001

Figure 2.3

The REG Procedure
Model: MODEL1
Dependent Variable: FFIV FFIV

Number of Observations Read	254
Number of Observations Used	253
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	63643	10607	415.55	<.0001
Error	246	6279.29311	25.52558		
Corrected Total	252	69923			

Root MSE	5.05228	R-Square	0.9102
Dependent Mean	206.06375	Adj R-Sq	0.9080
Coeff Var	2.45181		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	342.51768	33.36601	10.27	<.0001	0
VRSN	VRSN	1	-0.69579	0.05021	-13.86	<.0001	3.66594
TER	TER	1	0.57259	0.04645	12.33	<.0001	5.14649
CDW	CDW	1	0.42173	0.04836	8.72	<.0001	2.84766
ANET	ANET	1	-1.94344	0.28468	-6.83	<.0001	343.95875
MSI	MSI	1	-0.70098	0.11230	-6.24	<.0001	58.66988
anet_msi		1	0.00948	0.00123	7.72	<.0001	643.35634

Figure 2.4

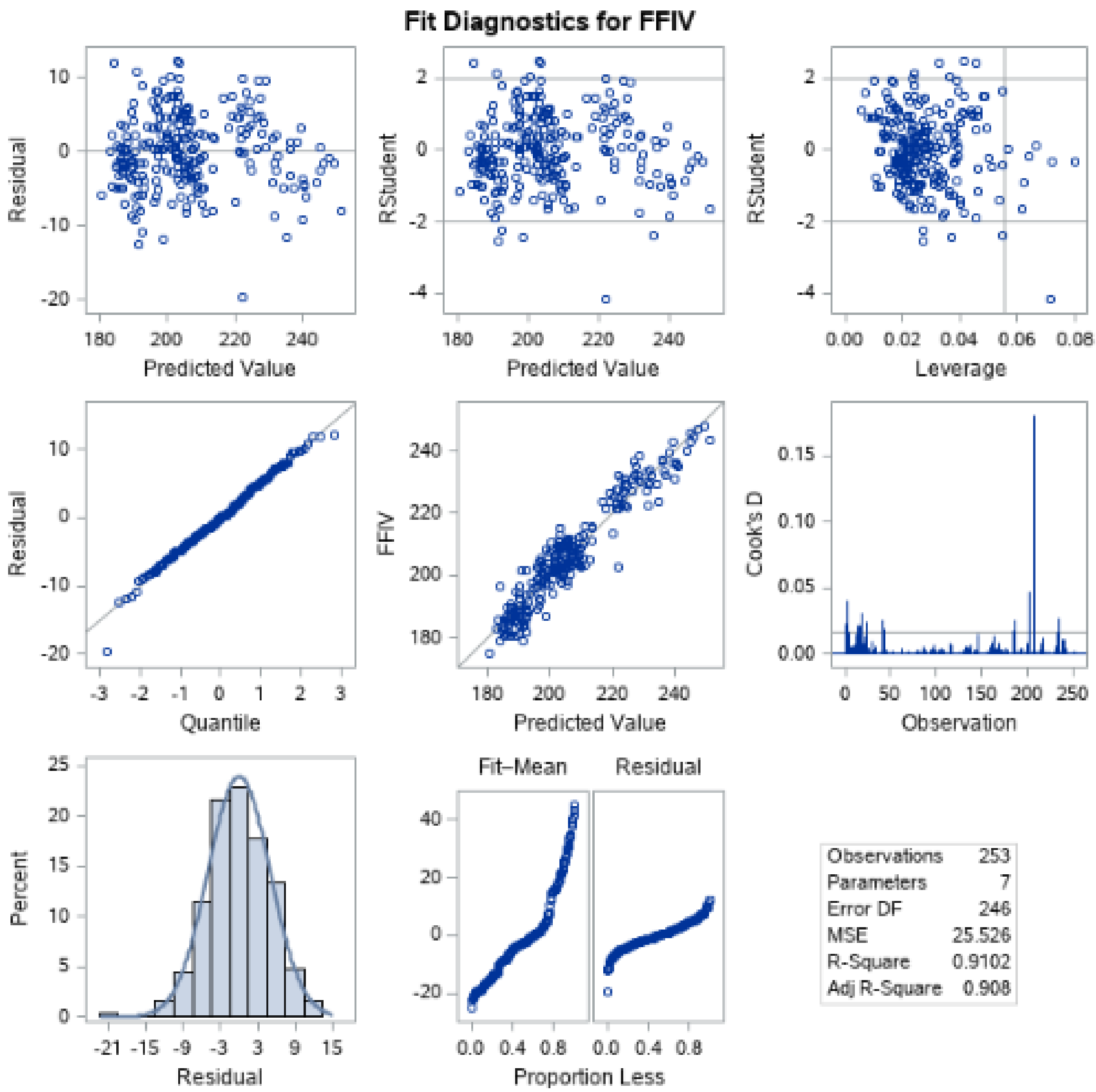
Fit Diagnostics for FFIV
Residual
RStudent
Predicted Value
Leverage
Quantile
FFIV
Cook's D
Observation
Percent
Fit–Mean
Residual
Proportion Less
Observations 253
Parameters 7
Error DF 246
MSE 25.526
R-Square 0.9102
Adj R-Square 0.908

Figure 2.5

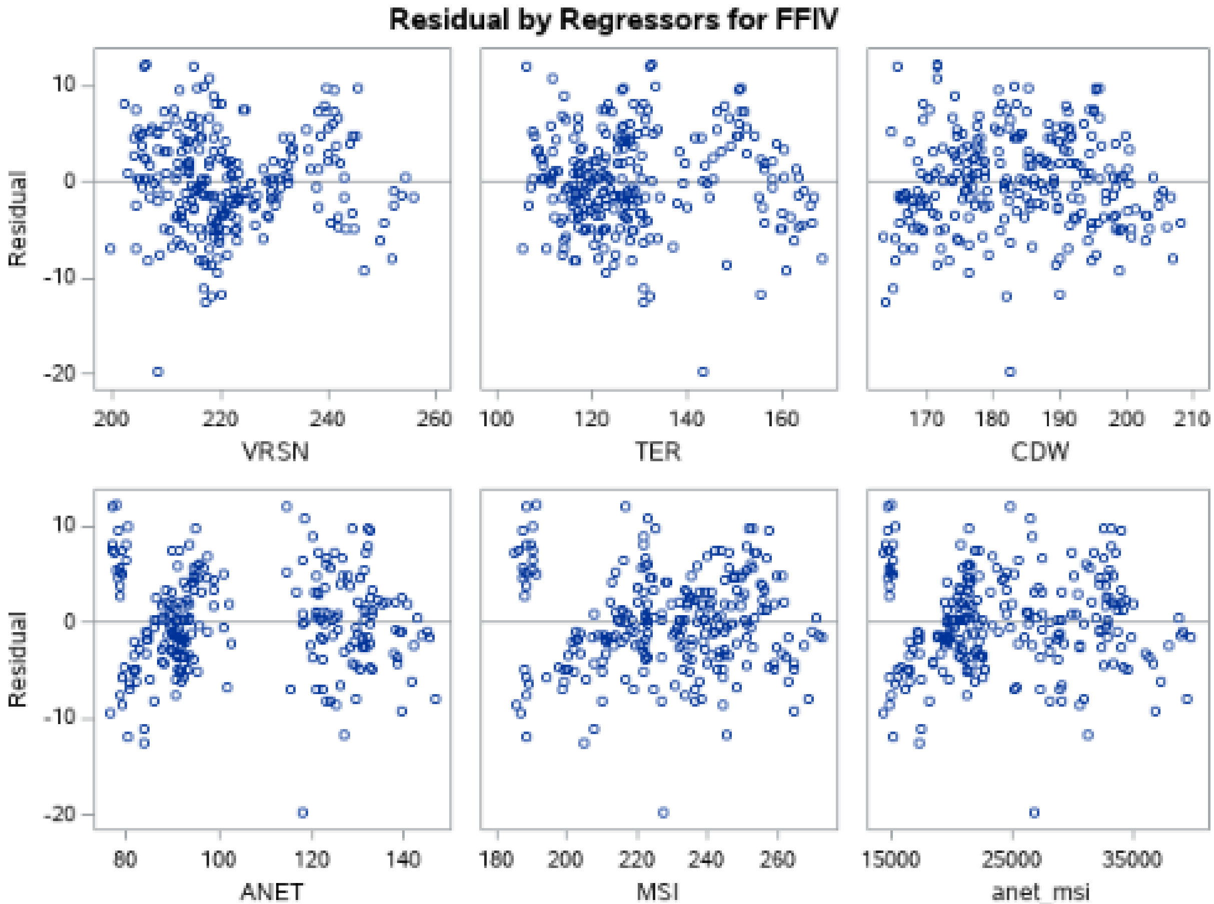
Residual by Regressors for FFIV
Residual
10
0
-10
-20
200
220
240
260
VRSN
100
120
140
160
TER
170
180
190
200
210
CDW
80
100
120
140
ANET
180
200
220
240
260
MSI
15000
25000
35000
anet_msi

Figure 2.6

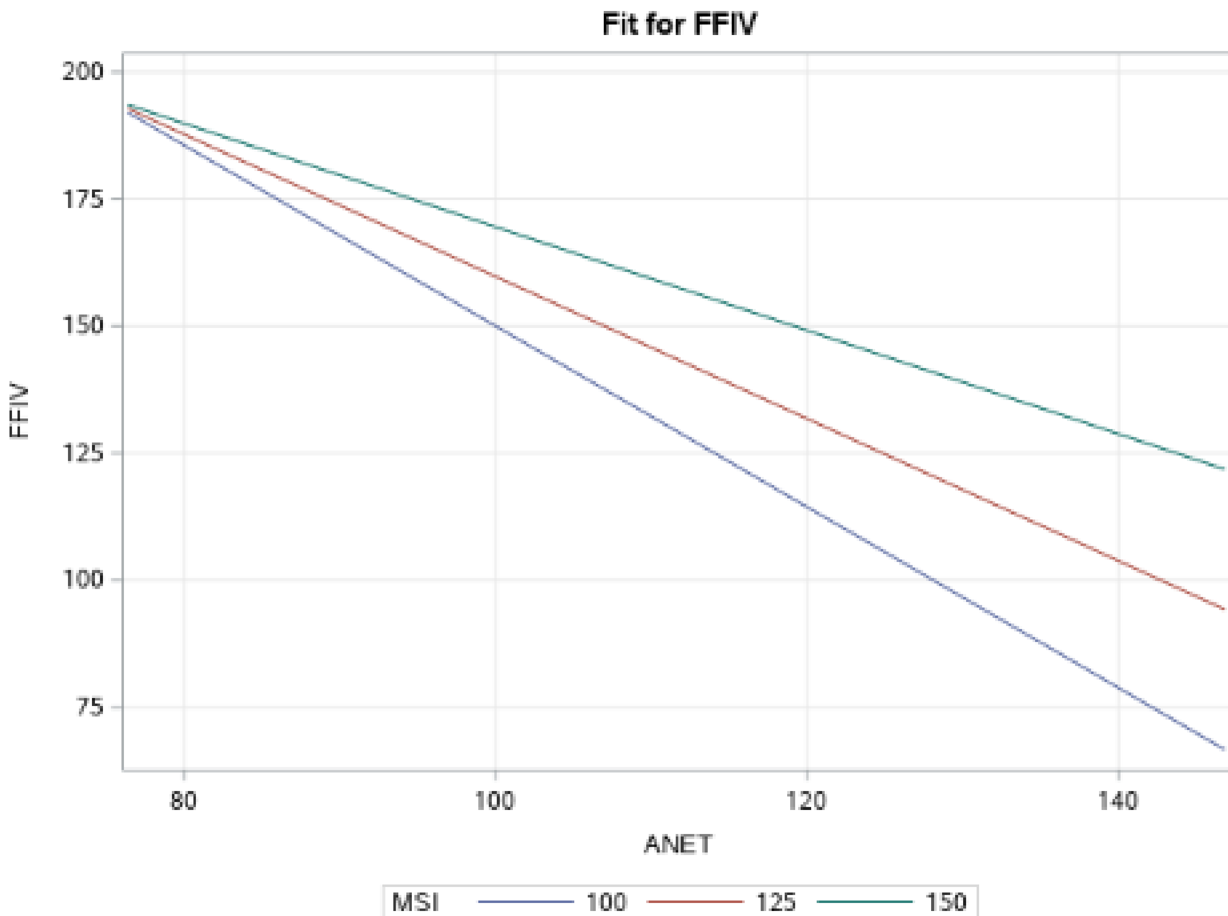
Fit for FFIV
200
175
150
125
100
75
FFIV
80
100
120
140
ANET
MSI
100
125
150

Figure 2.7

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	KLAC		KLAC	1	0.8259	0.8259	1595.02	1191.04	<.0001
2	CDNS		CDNS	2	0.0469	0.8728	1100.39	92.15	<.0001
3	INTC		INTC	3	0.0290	0.9018	795.501	73.44	<.0001
4	ADP		ADP	4	0.0209	0.9227	575.899	67.11	<.0001
5	VRSN		VRSN	5	0.0174	0.9401	393.797	71.63	<.0001
6	IT_new			6	0.0105	0.9506	284.352	52.38	<.0001
7	KEYS		KEYS	7	0.0087	0.9593	194.523	52.14	<.0001
8	NVDA		NVDA	8	0.0028	0.9620	167.352	17.69	<.0001
9	ANET		ANET	9	0.0026	0.9646	141.757	17.89	<.0001
10	AMD		AMD	10	0.0023	0.9670	118.905	17.19	<.0001
11	AVGO		AVGO	11	0.0012	0.9682	107.807	9.37	0.0025
12	FLT		FLT	12	0.0011	0.9693	98.3040	8.49	0.0039
13	NTAP		NTAP	13	0.0015	0.9708	84.7142	12.03	0.0006
14	CTXS		CTXS	14	0.0009	0.9717	76.7564	7.91	0.0053
15	WDC		WDC	15	0.0006	0.9723	72.5516	5.01	0.0261
16	APH		APH	16	0.0006	0.9729	68.2594	5.17	0.0239
17	MPWR		MPWR	17	0.0006	0.9734	64.3674	4.92	0.0275
18	PAYX		PAYX	18	0.0004	0.9739	61.6333	4.00	0.0465
19	NXPI		NXPI	19	0.0004	0.9743	59.1634	3.83	0.0516
20	IBM		IBM	20	0.0005	0.9749	55.4436	4.98	0.0266
21	AAPL		AAPL	21	0.0003	0.9752	54.0071	3.02	0.0837
22	ADI		ADI	22	0.0005	0.9756	51.1239	4.35	0.0381
23	EPAM		EPAM	23	0.0004	0.9761	48.5991	4.09	0.0444
24	FISV		FISV	24	0.0003	0.9763	47.8279	2.52	0.1139
25		APH	APH	23	0.0002	0.9761	48.0203	1.99	0.1594
26		AAPL	AAPL	22	0.0002	0.9759	48.1611	1.94	0.1653
27	MA		MA	23	0.0003	0.9762	46.6842	3.16	0.0766
28	V		V	24	0.0006	0.9769	41.8127	6.40	0.0121
29	FTNT		FTNT	25	0.0003	0.9772	40.6626	2.96	0.0868
30	NLOK		NLOK	26	0.0004	0.9776	38.4595	4.00	0.0467
31		INTC	INTC	25	0.0002	0.9774	38.3536	1.80	0.1807
32		MPWR	MPWR	24	0.0001	0.9773	37.7942	1.37	0.2437
33	GPN_new			25	0.0003	0.9776	36.2897	3.35	0.0684
34	ENPH		ENPH	26	0.0002	0.9778	35.9769	2.22	0.1372
35	SEDG		SEDG	27	0.0003	0.9781	35.0599	2.83	0.0940
36		AVGO	AVGO	26	0.0001	0.9780	34.3582	1.26	0.2631
37	CSCO		CSCO	27	0.0004	0.9784	32.0599	4.22	0.0411
38	PAYC		PAYC	28	0.0004	0.9788	29.8790	4.16	0.0425

Figure 2.8

```
/* Generated Code (IMPORT) */
/* Source File: Spring 2022 Lab #3 Closing Prices.xlsx */
/* Source Path: /home/u60677641/sasuser.v94/SAS Homework */
/* Code generated on: 4/16/22, 8:56 AM */

%web_drop_table(WORK.IMPORT);


FILENAME REFFILE '/home/u60677641/sasuser.v94/SAS Homework/Spring 2022 Lab #3 Closing Prices.xlsx';

PROC IMPORT DATAFILE=REFFILE
    DBMS=XLSX
    OUT=WORK.stocks;
    GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.stocks; RUN;


%web_open_table(WORK.IMPORT);

proc print data=stocks(obs=10); run;

data stocks_date;
    set stocks;

data stocks;
    set stocks;
    drop Date;
    IT_new = input(IT, comma9.);
    drop IT;
    GPN_new = input(GPN, comma9.);
    drop GPN;

proc contents data=stocks; run;

proc reg data=stocks;
    model FFIV = ACN
ADBE
ADP
AKAM
AMD
APH
```

Figure 2.9

```
APH
ADI
ANSS
AAPL
AMAT
ANET
ADSK
AVGO
BR
CDNS
CDW
CDAY
CSCO
CTXS
CTSH
GLW
DXC
ENPH
EPAM
FIS
FISV
FLT
FTNT
IT_new
GPN_new
HPE
HPQ
INTC
IBM
INTU
IPGP
JKHY
JNPR
KEYS
KLAC
LRCX
MA
MCHP
MU
MSFT
MPWR
MSI
NTAP
```

Figure 3.0

```
NTAP
NLOK
NVDA
NXPI
ORCL
PAYX
PAYC
PYPL
PTC
QRVO
QCOM
CRM
STX
NOW
SWKS
SEDG
SNPS
TEL
TER
TXN
TRMB
TYL
VRSN
V
WDC
ZBRA /selection=stepwise;
    run;

*variables we're interested in
KLAC CDNS INTC ADP VRSN TER QCOM NVDA NXPI KEYS WDC FLT SNPS ANET CTXS PAYX AMD NTAP FISV ACN ADI CSCO ENPH ZBRA AVGO EPAM ANSS MSI JNPR SEDG CD


proc corr data=stocks;
    var KLAC CDNS INTC ADP VRSN TER QCOM NVDA NXPI KEYS WDC FLT SNPS ANET CTXS PAYX AMD NTAP FISV ACN ADI CSCO ENPH ZBRA AVGO EPAM ANSS MSI JNPR
    with ffiv;
    run;
*remove keys, snps, cdns;

proc reg data=stocks;
    model FFIV = KLAC INTC ADP VRSN TER QCOM NVDA NXPI WDC FLT ANET CTXS PAYX AMD NTAP FISV ACN ADI CSCO ENPH ZBRA AVGO EPAM ANSS MSI JNPR SEDG
    run;
*remove MSI NVDA ACN;
```

Figure 3.1

```
proc reg data=stocks;
    model FFIV = KLAC INTC ADP VRSN TER QCOM NXPI WDC FLT ANET CTXS PAYX AMD NTAP FISV ADI CSCO ENPH ZBRA AVGO EPAM ANSS JNPR SEDG CDW BR /vif;
    run;
*remove AMD Juniper adp;

proc reg data=stocks;
    model FFIV = KLAC INTC VRSN TER QCOM NXPI WDC FLT ANET CTXS PAYX NTAP FISV ADI CSCO ENPH ZBRA AVGO EPAM ANSS SEDG CDW BR /vif;
    run;
*remove ANET and AVGO;

proc reg data=stocks;
    model FFIV = KLAC INTC VRSN TER QCOM NXPI WDC FLT CTXS PAYX NTAP FISV ADI CSCO ENPH ZBRA EPAM ANSS SEDG CDW BR /vif;
    run;
*remove ZBRA;

proc reg data=stocks;
    model FFIV = KLAC INTC VRSN TER QCOM NXPI WDC FLT CTXS PAYX NTAP FISV ADI CSCO ENPH EPAM ANSS SEDG CDW BR /vif;
    run;
*remove ANSS and KLAC;

proc reg data=stocks;
    model FFIV = INTC VRSN TER QCOM NXPI WDC FLT CTXS PAYX NTAP FISV ADI CSCO ENPH EPAM SEDG CDW BR /vif;
    run;
*remove ENPH and FISV;

proc reg data=stocks;
    model FFIV = INTC VRSN TER QCOM NXPI WDC FLT CTXS PAYX NTAP ADI CSCO EPAM SEDG CDW BR /vif;
    run;
*remove BR and NXPI;

proc reg data=stocks;
    model FFIV = INTC VRSN TER QCOM WDC FLT CTXS PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove CTXS;

proc reg data=stocks;
    model FFIV = INTC VRSN TER QCOM WDC FLT PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove FLT;

*now lets add some that we removed;
```

Figure 3.2

```
proc reg data=stocks;
    model FFIV = INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;

*add adp amd MSI NVDA cdns keys snps anet avgo zbra anss klac fisv br nxpi ctxs;

proc reg data=stocks;
    model FFIV = adp amd MSI NVDA cdns keys snps anet avgo zbra anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vi
    run;
*remove snps;

proc reg data=stocks;
    model FFIV = adp amd MSI NVDA cdns keys anet avgo zbra anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove keys;

proc reg data=stocks;
    model FFIV = adp amd MSI NVDA cdns anet avgo zbra anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove anet;

proc reg data=stocks;
    model FFIV = anet adp amd MSI NVDA cdns avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*added back anet and removed zbra, take out nvda;

proc reg data=stocks;
    model FFIV = anet adp amd MSI cdns avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove AMD;

proc reg data=stocks;
    model FFIV = anet adp MSI cdns avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove ANET;

proc reg data=stocks;
    model FFIV = adp MSI cdns avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove CDNS;
```

Figure 3.3

```
proc reg data=stocks;
    model FFIV = adp MSI avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove ADP;

proc reg data=stocks;
    model FFIV = MSI avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;


proc corr data=stocks;
    var MSI avgo anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW;
    with ffiv;
    run;
*remove AVGO;

proc reg data=stocks;
    model FFIV = MSI anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*remove MSI;

proc reg data=stocks;
    model FFIV = anss klac fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;

proc reg data=stocks;
    model FFIV = anss fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;

proc reg data=stocks;
    model FFIV = fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*removed anss, klac;

proc corr data=stocks;
    var fisv br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW;
    run;

proc reg data=stocks;
    model FFIV = fisv nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
```

Figure 3.4

```
proc reg data=stocks;
    model FFIV = br nxpi ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*getting rid of fisv instead and adding br back;

proc reg data=stocks;
    model FFIV = br ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO EPAM SEDG CDW /vif;
    run;
*got rid of nxpi
*get rid of epam;

proc reg data=stocks;
    model FFIV = br ctxs INTC VRSN TER QCOM WDC PAYX NTAP ADI CSCO SEDG CDW /vif;
    run;
*get rid of adi (terrible t value);

proc reg data=stocks;
    model FFIV = br ctxs INTC VRSN TER QCOM WDC PAYX NTAP CSCO SEDG CDW /vif;
    run;
*get rid of qcom;

proc reg data=stocks;
    model FFIV = br ctxs INTC VRSN TER WDC PAYX NTAP CSCO SEDG CDW /vif;
    run;
*remove ctxs;

proc reg data=stocks;
    model FFIV = br INTC VRSN TER WDC PAYX NTAP CSCO SEDG CDW /vif;
    run;
*remove ntap;

proc reg data=stocks;
    model FFIV = br INTC VRSN TER WDC PAYX CSCO SEDG CDW /vif;
    run;
*remove csco and br;

proc sgplot data=stocks;
    scatter y=ffiv x=ctxs;
    run; *might be useful doing some variable transformations, come back to this later;
```

Figure 3.5

```
proc reg data=stocks;
    model FFIV = INTC VRSN TER WDC PAYX SEDG CDW /vif;
    run; *best model so far, but we should test for interaction/squared terms for other variables we excluded.
    line 211 may be a good place to start;



data stocks_2;
    set stocks;
    ctxs_sq = ctxs*ctxs;
    csco_sq = csco*csco;
    br_sq = br*br;
    ntap_sq = ntap*ntap;
    qcom_sq = qcom*qcom;
    adi_sq = adi*adi;
    nxpi_sq = nxpi*nxpi;
    epam_sq = epam*epam;
    fisv_sq = fisv*fisv;
    anss_sq = anss*anss;
    klac_sq = klac*klac;
    msi_sq = msi*msi;
    avgo_sq = avgo*avgo;
    adp_sq = adp*adp;
    ctns_sq = ctns*ctns;
    cdns_sq = cdns*cdns;
    adp_sq = adp*adp;
    amd_sq = amd*amd;
    keys_sq = keys*keys;
    snps_sq = snps*snps;
    anet_sq = anet*anet;


proc sgplot data=stocks;
    scatter y=ffiv x=ctxs;
    run; *might be useful doing some variable transormations, come back to this later;
proc reg data=stocks_2;
    model ffiv = ctxs ctxs_sq;
    run;
```

Figure 3.6

```
proc sgplot data=stocks;
    scatter y=ffiv x=csco;
    run;
proc reg data=stocks_2;
    model ffiv = csco csco_sq;
    run;
*use as sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=br;
    run;
proc reg data=stocks_2;
    model ffiv = br br_sq;
    run;


proc sgplot data=stocks;
    scatter y=ffiv x=ntap;
    run;
proc reg data=stocks_2;
    model ffiv = ntap ntap_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=qcom;
    run;
proc reg data=stocks_2;
    model ffiv = qcom qcom_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=ctxs;
    run;
proc reg data=stocks_2;
    model ffiv = ctxs ctxs_sq;
    run;
```

Figure 3.7

```
proc sgplot data=stocks;
    scatter y=ffiv x=adi;
    run;
proc reg data=stocks_2;
    model ffiv = adi adi_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=nxpi;
    run;
proc reg data=stocks_2;
    model ffiv = nxpi nxpi_sq;
    run;
*debatable but use sq term;

proc sgplot data=stocks;
    scatter y=ffiv x=epam;
    run;
proc reg data=stocks_2;
    model ffiv = epam epam_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=fisv;
    run;
proc reg data=stocks_2;
    model ffiv = fisv fisv_sq;
    run;
*terrible predictor in general, DONT use this. maybe log transform?;


proc sgplot data=stocks;
    scatter y=ffiv x=anss;
    run;
proc reg data=stocks_2;
    model ffiv = anss_sq;
    run;
*use sq term, though not very good;
```

Figure 3.8

```
proc sgplot data=stocks;
    scatter y=ffiv x=klac;
    run;
proc reg data=stocks_2;
    model ffiv = klac_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=msi;
    run;
proc reg data=stocks_2;
    model ffiv = msi msi_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=avgo;
    run;
proc reg data=stocks_2;
    model ffiv = avgo avgo_sq;
    run;


proc sgplot data=stocks;
    scatter y=ffiv x=adp;
    run;
proc reg data=stocks_2;
    model ffiv = adp adp_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=cdns;
    run;
proc reg data=stocks_2;
    model ffiv = cdns cdns_sq;
    run;
*USE SQ TERM ESPECIALLY;
```

Figure 3.9

```
proc sgplot data=stocks;
    scatter y=ffiv x=amd;
    run;
proc reg data=stocks_2;
    model ffiv = amd amd_sq;
    run;
*use sq term for sure;


proc sgplot data=stocks;
    scatter y=ffiv x=adp;
    run;
proc reg data=stocks_2;
    model ffiv = adp adp_sq;
    run;
*better off not using sq term, the increase seems negligible;



proc sgplot data=stocks;
    scatter y=ffiv x=keys;
    run;
proc reg data=stocks_2;
    model ffiv = keys keys_sq;
    run;
*use sq term;


proc sgplot data=stocks;
    scatter y=ffiv x=snps;
    run;
proc reg data=stocks_2;
    model ffiv = snps snps_sq;
    run;
*use sq term;



proc reg data=stocks_2;
    model FFIV = INTC VRSN TER WDC PAYX SEDG CDW
     /vif;
    run;
```

Figure 4.0

```
Data stocks_3;
    set stocks_2;
    anet_csco = anet*csco;
    anet_jnpr = anet*jnpr;
    anet_msi = anet*msi;
    csco_jnpr = csco*jnpr;
    csco_msi = csco*msi;
    jnpr_msi = jnpr*msi;


proc reg data=stocks_3;
    model ffiv = anet csco anet_csco;
    run;
    *this looks good!;

proc reg data=stocks_3;
    model ffiv = anet jnpr anet_jnpr;
    run;
    *maybe?;

proc reg data=stocks_3;
    model ffiv = anet msi anet_msi;
    run;
*should use this probably;

proc reg data=stocks_3;
    model ffiv = csco_jnpr csco jnpr;
    run;

*use;

proc reg data=stocks_3;
    model ffiv = csco_msi csco msi;
    run;
*use;

proc reg data=stocks_3;
    model ffiv = jnpr_msi jnpr msi;
    run;
*use;
```

Figure 4.1

```
proc glm data=stocks_3;
 model ffiv=anet | msi / solution; *dependent variables separated by bar;
 ods select ParameterEstimates ;
 store GLMModel;
 run;

proc plm restore=GLMModel noinfo;
   effectplot slicefit(x=anet sliceby=msi = 100 125 150); *can specify by age, remove numbers
run;


proc reg data=stocks_3;
    model FFIV = INTC VRSN TER WDC PAYX SEDG CDW /vif;
    run;*.9181;


proc corr data=stocks;
    var KLAC CDNS INTC ADP VRSN TER QCOM NVDA NXPI KEYS WDC FLT SNPS ANET CTXS PAYX AMD NTAP
    with ffiv;
    run;


proc reg data=stocks_3;
    model FFIV = INTC VRSN TER WDC SEDG CDW anet msi anet_msi /vif;
    run;*.9247;

proc reg data=stocks_3;
    model FFIV = VRSN TER CDW anet msi anet_msi /vif;
    run; *absolute best model;

proc corr data=stocks_3;
    var INTC SEDG WDC VRSN TER CDW anet msi anet_msi;
    with FFIV;
    run;

proc corr data=stocks_3;
    var VRSN TER CDW anet msi anet_msi;
    with FFIV;
    run;
```

Figure 4.2

```
proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=FFIV / lineattrs=(color=red);
    series x=Date y = CDW;
    run;

proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=FFIV / lineattrs=(color=red);
    series x=Date y = TER;
    run;

proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=FFIV / lineattrs=(color=red);
    series x=Date y = VRSN;
    run;

proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=FFIV / lineattrs=(color=red);
    series x=Date y = ANET;
    run;

proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=FFIV / lineattrs=(color=red);
    series x=Date y = MSI;
    run;

proc sgplot data=stocks_date nocycleattrs;
    series x=Date y=MSI / lineattrs=(color=red);
    series x=Date y = ANET;
    run;



proc reg data=stocks_3;
    model FFIV = VRSN TER CDW anet msi anet_msi /all;
    run;
```

Figure 4.3