Statistical Analysis for Industrials components in the S&P 500

United Parcel Service (UPS)

Introduction

United Parcel Service is an American shipping and logistics company and a leading provider of global supply chain management solutions, founded in 1907. As a pivotal player in the transportation industry, UPS landed a spot in the S&P 500, a stock market index. The S&P 500 is one of the most widely known stock market indices in the world, tracking the stock performance of 500 of the largest companies in the United States. The abundance of daily stock data from the S&P 500 provides an opportunity to conduct a detailed analysis of the factors influencing UPS's stock performance. To gain a deeper understanding of the UPS stock, it is crucial to identify influencing factors that contribute to the daily fluctuations in the stock's price. This study aims to develop a robust multiple regression model by leveraging the stock prices of other companies within the S&P 500 with the goal of predicting daily stock prices for United Parcel Service (UPS).

Abstract Executive Summary

The goal of this analysis is to determine the best model for predicting the daily stock prices for United Parcel Service (UPS). To do this, the study began by analyzing the correlation between UPS and all other stocks within the S&P 500 in order to identify

those with the most significant influence on UPS's daily stock performance. From this initial correlation, several stocks were chosen based on their strong associations with the prices for UPS. The several chosen stocks will be correlated alongside UPS in order to determine the best model. The stocks selected, including HWM, RTX, CTAS, and more, are used as independent variables. In total, this analysis will consider eight different stocks within the S&P 500 that may impact the daily stock prices for UPS.

The full list of independent variables analyzed are below:

- Howmet Aerospace (HWM)
- RTX Corporation (RTX)
- Cintas (CTAS)
- Lockheed Martin (LMT)
- 3M (MMM)
- Rockwell Automation (ROK)
- Veralto (VLTO)
- Verisk Analytics (VRSK)

These independent variables are used in order to identify the most significant association with the daily stock price UPS. To conduct the analysis, Multiple Regression and several other statistical tools will be utilized with the goal of understanding influencing factors that contribute to the daily stock prices for UPS.

The data collected focuses on daily stock prices of all companies within the S&P 500 from November 1st, 2023 to October 31st, 2024. This dataset includes 252 observations and 77 variables. From the 77 variables, the analysis will only utilize eight, Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) , 3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK). These are the eight variables that were found to be most strongly associated with UPS daily stock prices and will be further correlated for our multiple regression models. Determining the best multiple regression model involves a multi-step process of implementing several statistical regression techniques that test different combinations of variables with UPS, the dependent variable. The process requires a stepwise model for the primary selection of variables, followed by four other major selection criteria methods including: R-square Criterion, Adjusted R-square Criterion, CP Criterion and Press Criterion. Following the creation of these models, their output is analyzed and compared based on several specific statistical criteria and the best model is revealed. To further enhance the analysis, additional models incorporating interaction terms and second-order terms are generated based on the best initial model. The purpose of this is to assess whether these additional variables have an impact on the model's overall efficiency and predictive accuracy, or if other models are best. Additionally, a final model is analyzed in order to test the presence of multicollinearity. All of these models are implemented in order to find the best overall model for predicting the daily stock prices for UPS.

After diving into each of the Multiple Regression, Stepwise, R-square Criterion, Adjusted R-square Criterion, CP Criterion and Press Criterion, Interaction, Second Order, and Multicollinearity models, the findings indicate that the Second Order Model (Model 3: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2) is the most significant model in predicting predicting the daily stock prices for UPS. Our analysis revealed that each model generated and analyzed appeared to be statistically significant and close in performance, although the Second Order Model was the most suitable choice for predicting UPS prices. In order to better understand these results, further explanation regarding the process of this analysis will be displayed through the exploratory analysis, correlation analysis, methodology and Multiple Regression techniques used.

Research Question

The analysis of this study focuses on the following research question: What is the best model for predicting the daily stock prices for United Parcel Service (UPS) based on the eight chosen independent variables that are most associated with daily stock prices for United Parcel Service? (Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) , 3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK))

Data

The dataset used within this study was retrieved from the S&P 500, classified by the Global Industry Classification Standard (GICS). It contains only data gathered from November 1st, 2023 to October 31st, 2024. Therefore, the dataset focuses on roughly 1 year of stock pricing statistics for stocks within the S&P 500. As stated above, this dataset includes 252 observations and 77 variables, or stocks. For the purpose of this analysis, eight specific stocks variables are examined along with UPS:

HWM - Howmet Aerospace, Aerospace & Defense

RTX - RTX Corporation, Aerospace & Defense

CTAS - Cintas, Diversified Support Services

LMT - Lockheed Martin, Aerospace & Defense

MMM - 3M, Industrial Conglomerates

ROK - Rockwell Automation, Electrical Components & Equipment

VLTO - Veralto, Environmental & Facilities Services

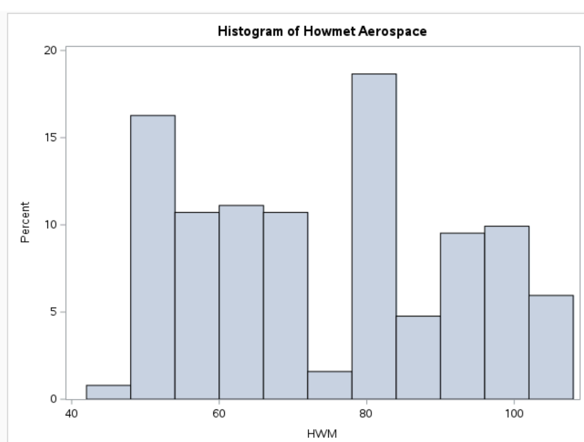
VRSK - Verisk Analytics, Research & Consulting Services

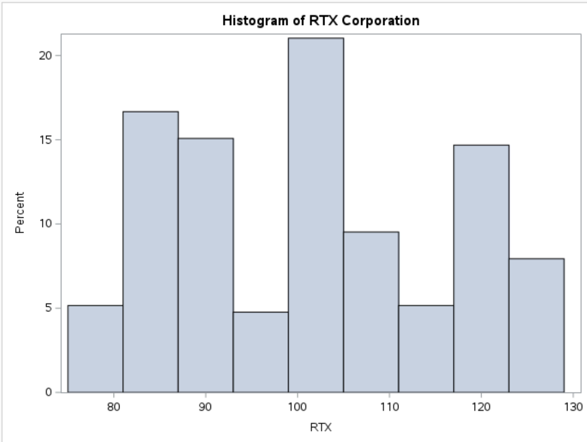
These variables were chosen because they were found to represent the eight variables, or stocks, most strongly associated with the daily stock prices for UPS. The data gathered from these eight variables in correlation to UPS will help to identify which regression model has the most significant impact on predicting daily stock prices for United Parcel Service.

Exploratory Data Analysis

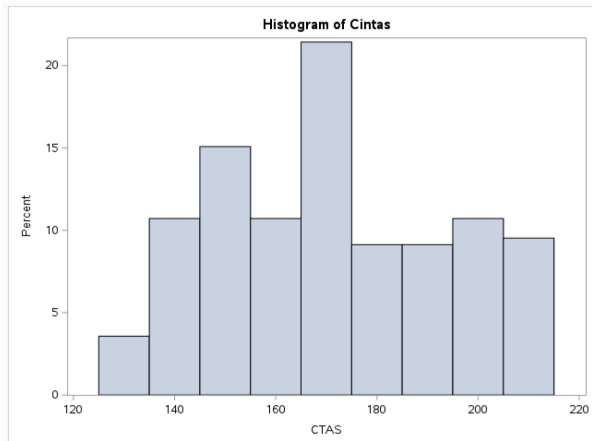
The very first step of the analysis involved conducting Exploratory Data Analysis (EDA), which is used to analyze and investigate datasets and summarize main characteristics. To better understand the distribution of the variables, each component is visualized and displayed with a histogram. The eight important independent variables are to be displayed in histogram form (Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) , 3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK)). Each histogram displays a unique distribution of data, and provides a solid foundation for evaluating stocks variability.

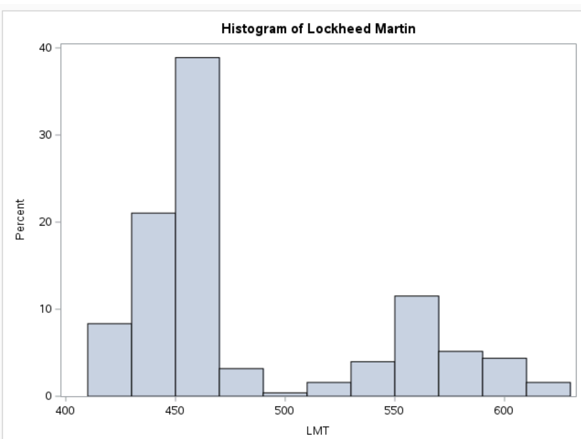
To begin, the histogram for Howmet Aerospace (HWM) displays moderate spread fluctuations, with prices ranging from around 50 to around 100. There appears to be a potential outlier towards





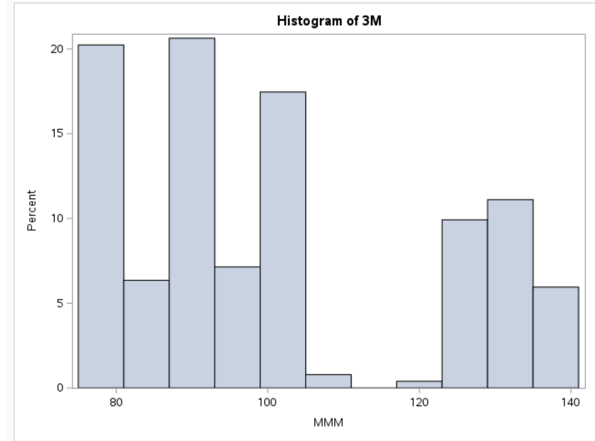
the lower end, around 40. There does not appear to be a strong skew. As for the histogram for RTX Corporation (RTX), the distribution seems to be somewhat symmetric with no extreme outliers. Prices range from around 70 to 130 suggests there is normal fluctuation over time. The histogram for Cintas (CTAS), displays a much wider and more symmetrical distribution. With prices ranging from 120 to 220, and a peak at around 170. This peak may indicate an outlier but not a very

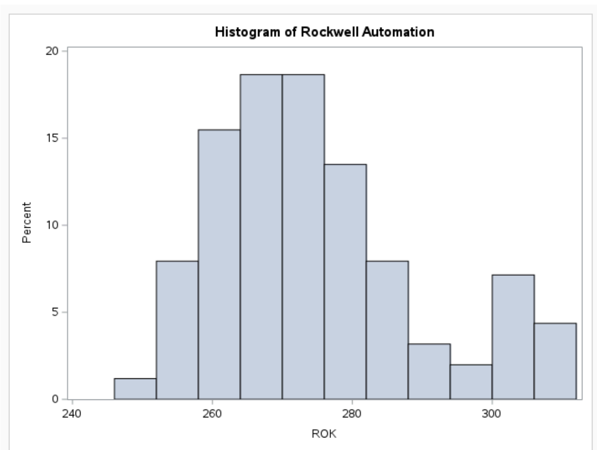




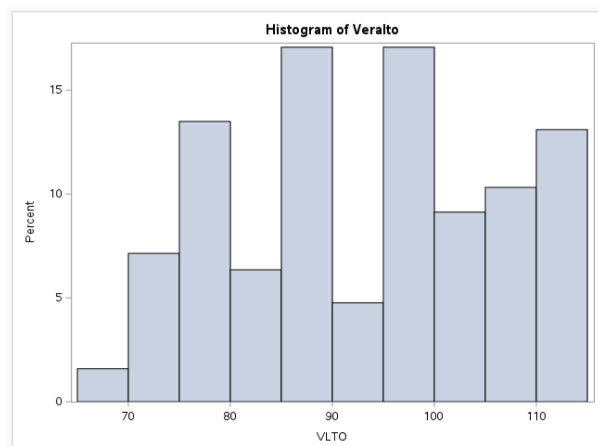
extreme one. The next histogram, for Lockheed Martin (LMT), shows a strong right skew, as most of the observations are seen to be around 450. The graph shows that there are more observations towards the lower end. This suggests that there is minimal fluctuation in the stock's prices. Next up is the histogram for 3M (MMM), it is clear to see that there are two distinct clusters, one ranging from 80 to 100 and the other from 120 to 140. This could imply that there were two separate ranges during different periods of time. The histogram for

Rockwell Automation (ROK), displays a somewhat symmetrical distribution with most values lying between 260 and 290. Additionally, there does not appear to be any outliers, which could suggest that pricing for this stock may be more

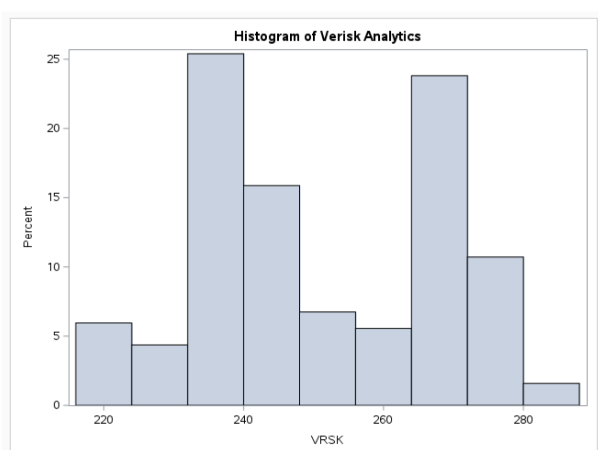




consistent. The next histogram, Veralto (VLTO), displays a range between 80 to over 110. It is clear to see a drop off towards the lower end, near 70, which may indicate an outlier. Additionally, there is another drop in

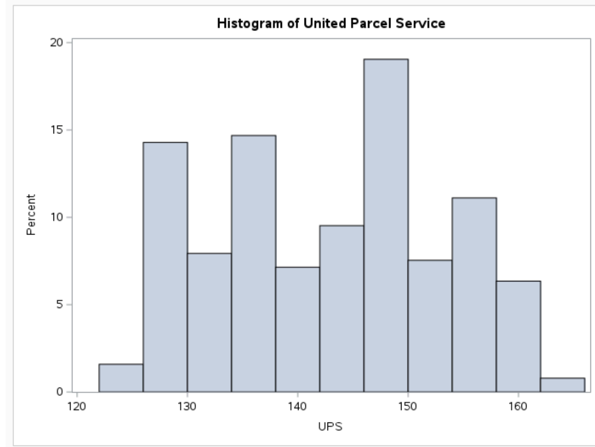


between two very high values around 90, which is important to be aware of. The last of our eight stock variables is Verisk Analytics (VRSK), which appears to display



a range between 220 to 280. Although there seems to be low values on both the lower end of the graph and the higher end. Because of the distribution of data, it is hard to tell if there are any outliers.

Another important histogram to visualize is the dependent variable of the analysis, United Parcel Service (UPS). In this histogram it is clear to see a fairly symmetrical spread. There does not appear to be any clear skew, although there is a slight central tendency around 150. Prices range from around 120 to 160, with lower frequencies at both the lower and higher end, which may indicate the presence of outliers. Overall, the balanced distribution of UPS suggests stability and consistency in the stock's daily prices.



The visualization of each of these variables displayed unique distributions. Some variables revealed a more symmetrical distribution, including RTX Corporation, Cintas (CTAS), Rockwell Automation and the dependent variable, United Parcel Service (UPS). Through the exploration and visualization of histograms key trends were identified and provided a foundational understanding of the dataset as a whole as well as aided in the overall prediction of daily stock prices for UPS.

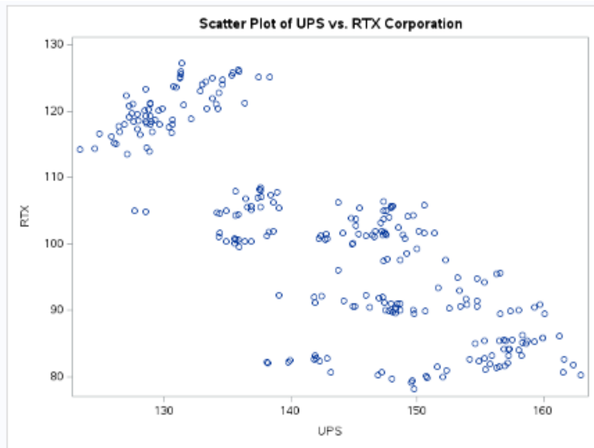
Correlation Analysis

In order to determine the eight variables that are most strongly associated with the daily stock prices for UPS, a correlation analysis was implemented. This is done by analyzing the output of a ‘proc corr’ procedure in SAS, which calculates the correlation coefficients between specific stocks listed within the S&P 500 (all 77 variables previously mentioned) and UPS. The correlation coefficient measures the strength and direction of the linear relationship between two variables. This means that the eight variables with the highest correlation coefficients represent the strongest associations. As stated above, the process began by running a ‘proc corr’ that included all 77 variables or stocks listed in the S&P 500 alongside our dependent variable, UPS. This output displayed that there were several highly correlated stocks in regards to the UPS stock. In order to ensure that the model would capture the most relevant predictors and avoid weaker, less impactful variables, the variables with an absolute correlation coefficient of 0.75 or greater were selected (a reasonable cutoff for strong predictors). This would narrow down variables that are both statistically significant and have a logical connection to UPS’s business operations, while still ensuring their predictive reliability. This process revealed that there were eight statistically significant variables that displayed a correlation coefficient greater than 0.75. The eight variables, or stocks, that were found to have the highest correlation coefficient and therefore have the strongest relationship to UPS are:

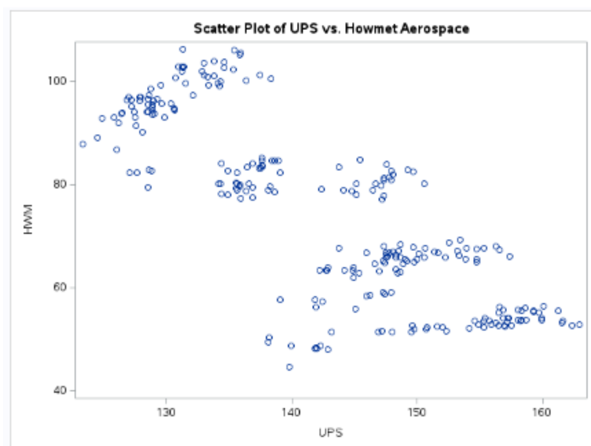
RTX (RTX Corporation): -0.81707	CTAS (Cintas): -0.77796
HWM (Howmet Aerospace): -0.80339	MMM (3M): -0.77309
ROK (Rockwell Automation): 0.79696	LMT (Lockheed Martin): -0.75456
VLTO (Veralto): -0.78342	VRSK (Verisk Analytics): -0.75028

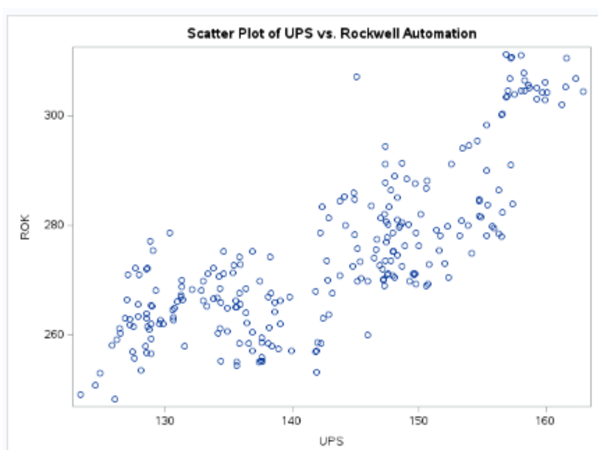
By identifying the eight variables with the strongest correlation to UPS, we can now proceed to the analysis and implementation of our regression models. But before that, visualizations for each of these variables are displayed to effectively assess these relationships between each of the eight selected variables and UPS. This step is done to spot any outliers or trends in the data.

Since the correlation between RTX and UPS is the strongest, let's start there. The scatter plot for RTX Corporation (RTX) displays a clear negative direction.

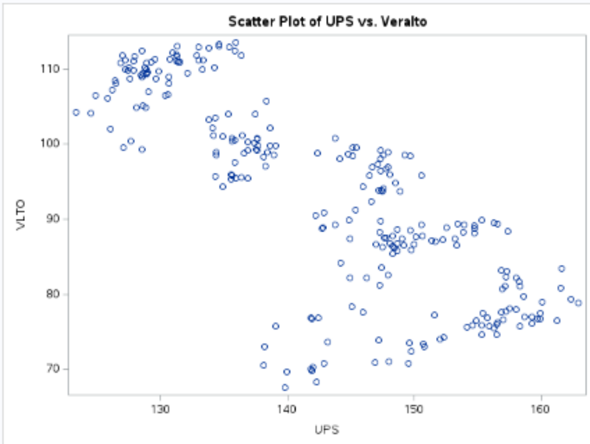


Additionally, this plot displays a linear trend within the data. The distribution of points displays a strong relationship, considering the points follow a linear path, as stated above. There does not appear to be any outliers. As for the scatterplot of HWM (Howmet Aerospace), similarly to the previous plot, there is a negative direction. Additionally, this plot also displays a linear trend. There is a moderate relationship, as the points somewhat gather around a line. Again there does not appear to be any obvious outliers. The scatter plot for ROK (Rockwell Automation) displays a positive

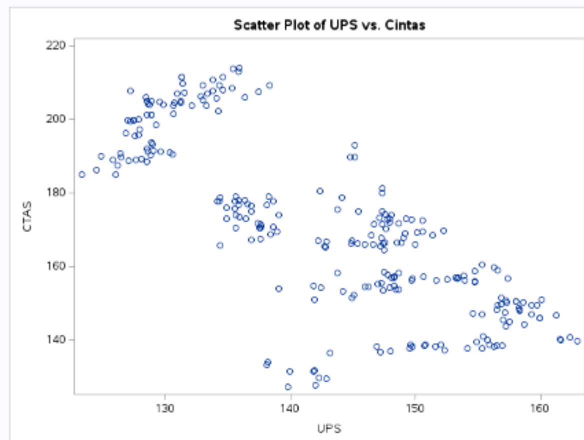




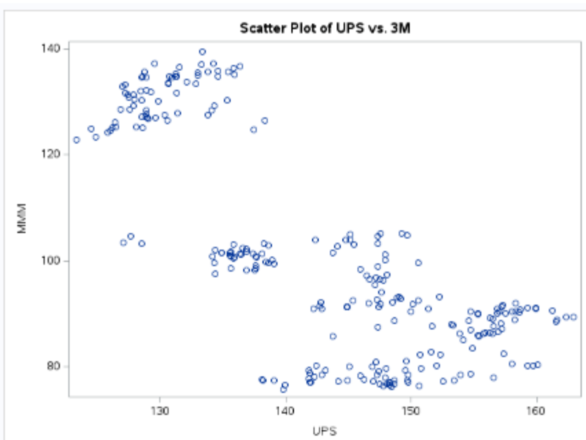
direction, as well as a linear trend. The relationship appears to be quite strong as the points follow a clear upward trend. There may be the presence of one outlier but it is not very extreme, so it is difficult to say for sure. The scatter plot for VLTO (Veralto) shows a negative direction. The



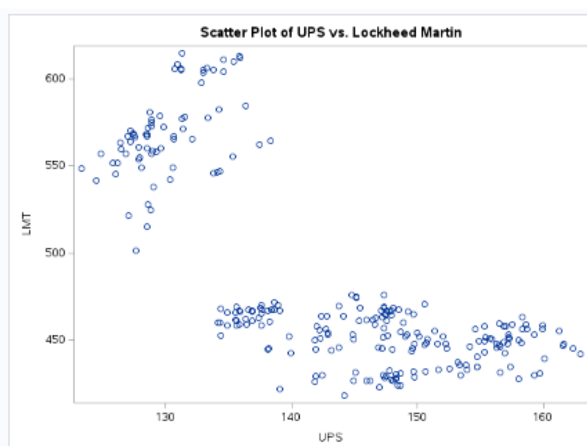
plot displays a linear trend and moderate strength, with some points straying from a clear path. Because of the distribution of points, it is difficult to identify any distinct outliers. The scatter plot for CTAS (Cintas) is very similar to the scatter plot for VLTO



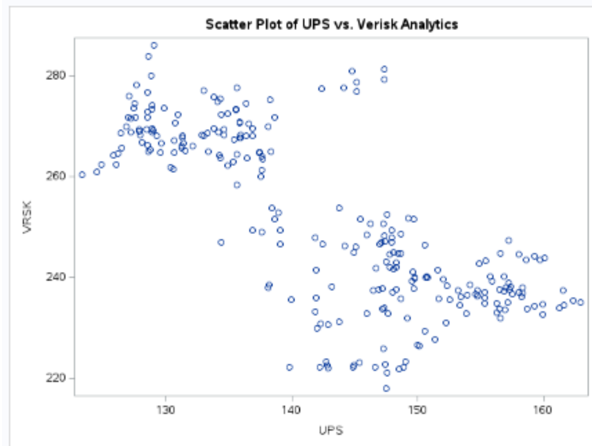
(Veralto). With a negative direction, a linear trend and moderate strength, with some slight deviations, as well as no extreme outliers. The scatter plot for MMM (3M) displays a weak negative direction, as well as a weak



linear trend with several deviations. The strength of the relationship is weak due to lack of points along a clear path. There does appear to be some outliers as some points deviate significantly to the path. The



scatter plot for LMT (Lockheed Martin) is clearly displaying a weak negative direction. Additionally, there is not a clear trend, although slightly linear with more scatter. The strength of the plot displays a weak relationship with points widely spread. There are possible outliers towards the higher end but it is hard to clearly identify them due to the scattered distribution of points. Lastly the scatter plot for VRSK (Verisk Analytics) shows a negative direction and a linear trend. The distribution of points displays a moderate relationship, as the points somewhat follow a path, although there is still scatter. There may appear to be a few outliers but again, it is difficult to identify them due to the scattered distribution of points. With this essential first step in understanding the relationships between variables it was revealed that the majority of these scatter plots display clear trends. These visualizations help to provide a solid foundation before diving into the more in depth regression analysis.



Multiple Regression

Jumping into the Multiple Regression analysis, using UPS as the dependent variable and the eight stocks as the independent variables, the relationship between fluctuations of stock pricing was conducted. This multi-step process incorporates all combinations of the eight variables within several multiple regression models with the goal of identifying which combinations of variables most effectively predict the daily stock prices for UPS. This requires several models for the initial selection of variables, followed by two more complex models and a final model for testing highly correlated variables.

For the initial selection of variables, Stepwise Regression was implemented. This is a technique that iteratively adds or removes variables based on their statistical significance, focusing on P-values and F-statistics. The result will contain the best model based on the Stepwise Regression process that includes a set of potentially important independent variables. Four additional selection criteria methods were employed including: R-square Criterion, Adjusted R-square Criterion, CP Criterion and Press Criterion. These models are employed to help support the conclusions made from Stepwise. From this, a model was selected to move forward in the regression process. Following this, all models were analyzed based on the following processes.

Each model was analyzed and evaluated based on key statistical criteria within the outputs of each model. These metrics included the Adjusted R-Square, F Test P-Value, Standard Deviation (Root MSE), and Coefficient Variation, as well as the interpretation of relevant residual plots and Variance Inflation (VIF). It is crucial to consider these metrics when examining multiple regression models because they provide a comprehensive view of the model's overall performance and accuracy.

Why are each of these statistical metrics important? The importance of the adjusted r square lies in its ability to display how well the independent variables explain the variability in the dependent variable. A higher adjusted r square indicates a better fit, therefore the model that displays the highest adjusted r square has potential to be the best. Next, it is important to look at the F Test P-Value because it reveals the overall significance of the model. Having p-values less than 0.05, or alpha, indicates that the variables are statistically significant. The standard deviation, or root mse, indicates the predictive accuracy of a model. Having a lower root mse means that the model's predictions are closer to the actual values, which indicates better predictive accuracy. Next it is important to look at the coefficient of variation because it provides insights to how much variability there is relative to the mean of the dependent variable. Additionally, the interpretation of relevant residual plots is important to validate the final model of choice. Important plots analyzed are Residuals vs. Predicted Values plot, Studentized Residual plot, QQ plot, Cook's D plot, and the Histogram of Residuals. Each of these are analyzed based on the specific distribution of points. The plots collectively assess model assumptions including linearity, homoscedasticity, and the presence of outliers or other key trends. The Residuals vs. Predicted Values plot, checks for homoscedasticity, or equal variances. For this plot there should be a complete scatter of points and zero patterns visible. The Studentized Residual plot helps to identify potential outliers or influential points in the data. It is important to identify and be aware of any points beyond the boundaries within the graph. The QQ plot assesses the normality of residuals. When interpreting, points should hug the line present in the graph, this represents normally distributed data. The Cook's D plot is used to identify any significant outliers or other influential observations. Any spike in the graph indicates potential outliers that could significantly affect the data. The Histogram of Residuals plot is essential for visualizing the distribution of residuals. Residuals following the bell curve, indicate normally distributed data. Any spikes or values that go beyond the curve suggest a lack of symmetry and violations of normality. Lastly, Variance Inflation (VIF) was tested and analyzed in order to identify multicollinearity between variables. Any variables with a VIF greater than 10 indicates that two or more of the independent variables used in the model contribute redundant information, meaning that they will be correlated with each other causing multicollinearity. The presence of multicollinearity is to be identified and fixed by removing variables iteratively until all remaining variables had acceptable VIFs less than 10. The importance of this is to ensure that independent variables are not excessively correlated with one another. Through this analysis, all of these key metrics are considered and used in determining the best overall multiple regression model that effectively predicts stock prices for UPS.

Results

Beginning with the initial selection of variables, a Stepwise Regression model was run. As well as the four additional selection criteria methods (R-square Criterion, Adjusted R-square Criterion, CP Criterion and Press Criterion) to help support the conclusions made from Stepwise. The models were analyzed based on the specific key metrics stated above. In order to better visualize the results of each model and its key statistics, the following table was generated.

Variable Selection Key Statistic Table

Selection Methods	*Selected Model*	*Selection Method*
Stepwise Regression	*6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO)*	*Highest Adjusted R-Square of 0.8902*
R-Square Criterion	*7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM)*	*Highest R-Square of 0.8944 (Second highest of 0.8891)*
Adjusted R-Square Criterion	*7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM)*	*Highest Adjusted R-Square of 0.8913 (Second highest of 0.8902)*
Cp Criterion	*7-variable model (RTX, ROK,*	*Lowest Cp value of 8.000*

	VRSK, HWM, LMT, VLTO, MMM)	*(Second lowest of 9.5966)*
Press Criterion	*6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO)*	*Lowest Press value of 3042.83734*

This table displays the key statistics gathered from all of the selection method models employed, these statistics help to determine the best model. Each model selects the best model based on its specific criterion. Beginning with Stepwise, the 6 variable model was chosen (RTX, ROK, VRSK, HWM, LMT, VLTO). The model successfully identified the most statistically significant combination of variables while maintaining model simplicity. Through Stepwise, MMM and CTAS were removed as they did not improve the model significantly enough to outweigh the cost of increased complexity. From this, the four criterion methods were implemented in order to support the decision made from Stepwise. Starting with the R-Square criterion, the 7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM) came out on top with the highest R-Square of 0.8944. Although the 6 variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) was not too far behind with an R-Square of 0.8891. Adjusted R-Square also displayed that the 7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM) had the highest Adjusted R-Square of 0.8913, although again the 6-variable model does not differ too much with an Adjusted R-Square of 0.8902. For Cp criterion, the model with the lowest Cp value of 8.0007 was the 7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM). Lastly, the model chosen for Press criterion is the 6 variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) with the lowest Press value of 3042.83734. While both the 7-variable and 6-variable models displayed strong statistical performance, the 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) was ultimately chosen for several reasons. It is clear to see that the differences in R-Squared and Adjusted R-Square between the two models are minimal (less than 0.005). This suggests that including the variable MMM, does not significantly improve the model and displays that choosing the 6-variable model balances performance and simplicity. The 6-variable model also resulted in a lower Press value, indicating it has better predictive accuracy when applied to new data. Finally, this result supports the decision made from Stepwise, that the simpler 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) is the best at predicting daily stock prices of UPS.

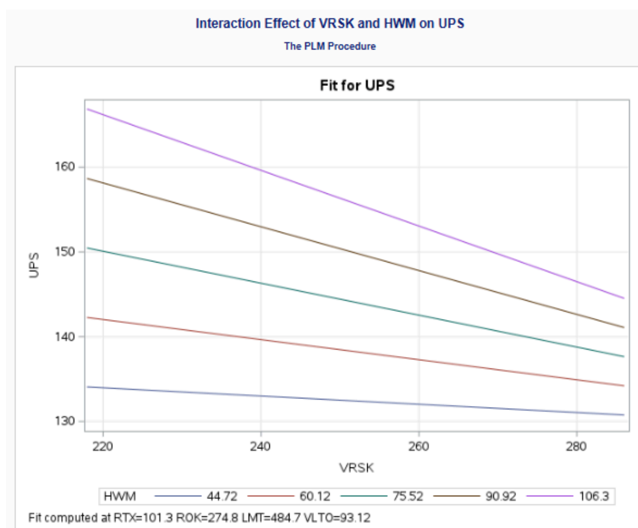
Now that the Each model 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) has been revealed as the best model at predicting daily stock prices of UPS we can proceed with the analysis. The 6-variable model was analyzed based on the specific key metrics stated above. In order to better visualize each model and its key statistics, the following table was generated.

Stepwise Regression Model Key Statistic Table

Stepwise Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 1: *UPS = 143.24744 - 0.36729(RTX) + 0.34846(ROK) - 0.18196(VRSK) + 0.42064(HWM) - 0.04488(LMT) - 0.24603(VLTO)*	*0.8902*	*<.0001*	*3.42645*	*2.39991*

This table displays the key statistics gathered from the Stepwise model, these statistics help to determine the best model. Again each model is evaluated based on the following key statistical criteria: Adjusted R-Square, F Test P-Value, Standard Deviation (Root MSE), and Coefficient Variation. The stepwise model has an adjusted r square of 0.8902, which indicates a strong fit. The model has a p value <.0001, which is <0.05, and indicates that the model is highly statistically significant. The standard deviation lies at 3.42645, which indicates good predictive accuracy. Lastly, the model appears to have a coefficient variation of 2.39991,indicates low relative variability and suggests more consistency in predictions. Overall, this model performs well with a high adjusted r square, a significant p value, a low standard deviation, and low coefficient of variation. Given these factors, it is clear that the 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) is a strong candidate for predicting UPS daily stock prices.

To further enhance the analysis, an interaction term model was developed based on the best overall model chosen from Stepwise. This is done to ensure that other possible relationships were explored in the exploration of finding the best overall model for predicting stock prices for UPS. The first step of this process was deciding which interaction term to include. Initial thoughts when choosing an interaction term was to look for variables that have strong individual effects in the model and or are logically related in industries. The most effective process in determining whether an interaction and/or second order term might be needed is the creation of an interaction plot. The plot generated displays the slopes between variables tested, if the slopes appear to be different, an interaction term could be useful in improving the model, if slopes are parallel, an interaction term is not necessary. Multiple interaction terms were tested



based on these factors and finally the interaction term VRSK & HWM was chosen. Because the plot reveals different slopes for the relationship between variables, it suggests that an interaction term could enhance the model. Based on the output of the plot, an interaction term was added and tested in the regression analysis with the objective being to examine how this interaction of variables is correlated with UPS.The model was analyzed based on the same specific key metrics stated above. In order to better visualize each model and its key statistics, the following table was generated.

Interaction Model Key Statistic Table

Interaction Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 2: *UPS = 52.91801 - 0.35333(RTX) + 0.35033(ROK) + 0.15470(VRSK) + 1.52330(HWM) - 0.02440(LMT) - 0.26897(VLTO) - 0.00455(VRSK_HWM)*	*0.8914*	*<.0001*	*3.40817*	*2.38710*

This table displays the key statistics gathered from the interaction model, these statistics help to determine the best model. Again each model is evaluated based on the following key statistical criteria: Adjusted R-Square, F Test P-Value, Standard Deviation (Root MSE), and Coefficient Variation. From the table, it can be seen that the interaction model slightly outperforms the original 6-variable model.

The interaction model has an adjusted r square of 0.8914, which is a bit higher than previous results (adjusted r square of previous model: 0.8902). The model has a p value of <.0001, <0.05, which indicates that the model is statistically significant (the same as the previous model). The standard deviation lies at 3.40817, which is slightly lower than the previous model (standard deviation of previous model: 3.42645). This indicates that the interaction model has better predictive accuracy. Lastly, this model appears to have a coefficient variation of 2.38710, which again is lower than the last model (coefficient variation of previous model: 2.39991). This suggests that the interaction model is more consistent in its predictions with interaction terms. Overall, this model improves slightly on all key metrics with a higher adjusted r square, a p value <0.05, a lower standard deviation, and lower coefficient of variation, across the previously evaluated model, making it a slightly better model model. This result suggests that including an interaction term to the original 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) increased the model's predictive accuracy of the daily stock prices for UPS. Therefore, of the models we’ve analyzed so far, the Interaction Model, Model 2: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & VRSK_HWM, appears to be the best at predicting daily stock prices of UPS.

In addition to the interaction term, a second-order term model was also generated based on the best model, from Stepwise, to explore the possibility of non-linear relationships between the other stocks and UPS. The second order term was chosen based on the individual correlation with UPS. Because RTX displayed the strongest correlation to the dependent variable, UPS, it was selected as the second order term. This second order term was added and tested in the regression analysis with the objective being to examine how the chosen stock is correlated with UPS. The model was analyzed based on the same specific key metrics stated above. In order to better visualize the model and its key statistics, the following table was generated.

Second Order Model Key Statistic Table

Second Order Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 3: ***UPS = 201.99778 - 1.41583(RTX) + 0.34489(ROK) - 0.17753(VRSK) + 0.35565(HWM) - 0.07547(LMT) - 0.15356(VLTO) + 0.00558(rtx_2)***	***0.8915***	***<.0001***	***3.40631***	***2.38580***

This table displays the key statistics gathered from the second order model, these statistics help to determine the best model. Again each model is evaluated based on the following key statistical criteria: Adjusted R-Square, F Test P-Value, Standard Deviation (Root MSE), and Coefficient Variation. From the table, it can be seen that this model displays an r square adjusted of 0.8915, even higher than the adjusted r square for the interaction model (0.8914). This suggests that adding a second order term increased the model's predictive ability. This model displays a p value of <.0001, again <0.05, indicating that the model is highly statistically significant. This model also appears to have the lowest standard deviation, or root mse, of 3.40631, again even lower than the interaction model (3.40817). This means that this model's predictions were more accurate on average compared to the other models, including the original 6-variable model. Lastly, this model appears to have the lowest coefficient variation coming in at 2.38580 (previously 2.38710). This means the model has less variability relative to the mean than both of the other models. This suggests that this second order model is more consistent in its predictions. Overall, this model presented the highest adjusted r square, a p value <0.05, the lowest standard deviation, and lowest coefficient of variation, across both previously evaluated models, making it the most reliable model. This model's results suggest that adding the chosen second order term, RTX^2, significantly impacted the model. Given these factors, the Second Order Model, Model 3: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2, is revealed to be the best performing model for predicting the daily stock prices for UPS.

Finally, a test for multicollinearity was conducted on the 6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO) in order to test highly correlated variables. The model was analyzed based on the specific key metrics stated above, VIF values (<10). In order to better visualize the results of the model, the following table was generated.

VIF Model 1 Key Statistic Table

Variable	*VIF*
RTX	***26.29748***
ROK	*1.59078*
VRSK	*2.94067*
HWM	***37.23414***
LMT	*4.92445*
VLTO	***23.58768***

This table displays the key statistics gathered from the regression model testing for multicollinearity, these statistics help to determine the best model. Here it is clear to see that there are three variables with VIF values greater than 10, HWM, RTX, and VLTO. This is an issue and suggests that independent variables contribute redundant information, are highly correlated and therefore cause multicollinearity. As stated above, in order to fix this issue, variables must be removed iteratively, starting with the highest VIF value, until all remaining variables had acceptable VIFs less than 10. Following this process, HWM is removed from the model and a new 5-variable model is tested.

Again, the test for multicollinearity continues now on the 5-variable model (RTX, ROK, VRSK, LMT, VLTO) in order to test highly correlated variables. The model is still analyzed based on the specific key metrics stated above, VIF values (<10). In order to better visualize the results of the model, the following table was generated.

VIF Model 2 Key Statistic Table

Variable	*VIF*
RTX	***20.78550***
ROK	*1.56398*
VRSK	*2.51149*
LMT	*4.82157*
VLTO	***15.44196***

This table displays the key statistics gathered from the second regression model testing for multicollinearity, these statistics help to determine the best model. It is clear to see that there still remain two variables with VIF values greater than 10, RTX, and VLTO. This is an issue and suggests that independent variables contribute redundant information, are highly correlated and therefore cause multicollinearity. As stated above, in order to fix this issue, variables must be removed iteratively, again starting with the highest VIF value, until all remaining variables had acceptable VIFs less than 10. Following this process, RTX is now removed from the model and a new 4-variable model is tested.

Once more, the test for multicollinearity continues now on the 4-variable model (ROK, VRSK, LMT, VLTO) in order to test highly correlated variables. The model is still analyzed based on the specific key metrics stated above, VIF values (<10). In order to better visualize the results of the model, the following table was generated.

VIF Model 3 Key Statistic Table

Variable	*VIF*
ROK	*1.40291*
VRSK	*2.40900*
LMT	*3.21413*
VLTO	*4.10776*

This table displays the key statistics gathered from the third regression model testing for multicollinearity, these statistics help to determine the best model. Here it is clear to see that all variables present VIF values less than 10, meaning there is no longer the presence of multicollinearity. The issue involving highly correlated independent variables contributing redundant information, has been resolved. Variables were removed iteratively, starting with the highest VIF value, and now the variables within this model all present acceptable VIFs less than 10. This process has successfully fixed the issue of multicollinearity, therefore we can proceed by analyzing the new 4-variable model inorder to determine if it is impactful in predicting the daily stock prices for UPS.

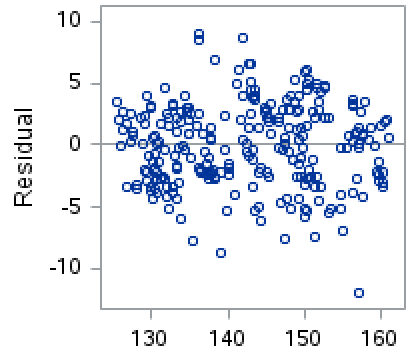
In addition to the interaction and second-order term model, a 4-variable model (ROK, VRSK, LMT, VLTO) was generated through the process of testing for multicollinearity. This was done by iteratively removing variables with high VIF values in order to fix the issue of multicollinearity. The model was analyzed based on the same specific key metrics stated above. In order to better visualize the model and its key statistics, the following table was generated.

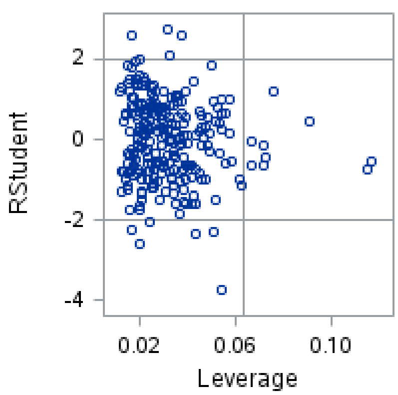
Multicollinearity Model Key Statistic Table

Multicollinearity Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 4: *UPS = 113.81604 + 0.34959(ROK) - 0.12604(VRSK) - 0.05003(LMT) - 0.12046(VLTO)*	*0.8747*	*<.0001*	*3.65962*	*2.56322*

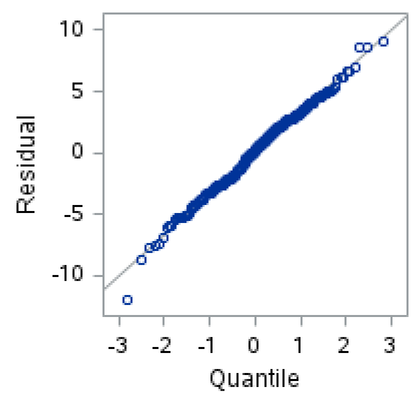
This table displays the key statistics gathered from the 4-variable (Multicollinearity) model, these statistics help to determine the best model. Again each model is evaluated based on the following key statistical criteria: Adjusted R-Square, F Test P-Value, Standard Deviation (Root MSE), and Coefficient Variation. From the table, it can be seen that this model displays an r square adjusted of 0.8747, which is much lower that the previous model tested, the Second Order term model (0.8915). Similar to all of the previous models, this model displays a p value of <.0001, again <0.05, indicating that the model is highly statistically significant. The standard deviation, or root mse, of this model also appears to be much higher than previous models, coming in at 3.65962 (previously 3.40631). Lastly, this model appears to have a higher coefficient variation of 2.56322, than the previous model (2.38580). While still significant, this model is not better than previously tested models. Despite resolving the issue of multicollinearity, the second order model, , Model 3: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2, remains the best performing multiple regression model for predicting daily stock prices for UPS.

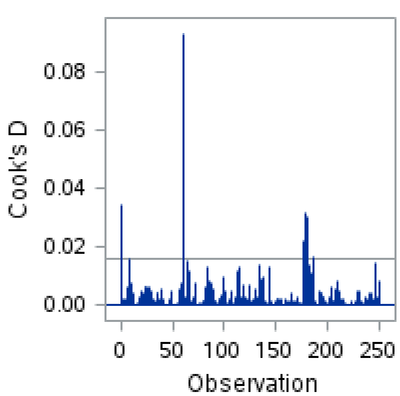
Now that the best model for predicting the daily UPS stock prices has been revealed, the Second Order model, Model 3: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2, we can proceed with further analysis of the model's relevant residual plots to validate its effectiveness. The 5 plots to be interpreted are the Residuals vs. Predicted Values plot, Studentized Residual plot, QQ plot, Cook's D plot, and the Histogram of Residuals. Starting with the Residuals vs. Predicted Values plot, checking for homoscedasticity, a complete scatter can be seen. This graph suggests that the residuals are randomly distributed and that the model is correctly specified. The Studentized Residual plot displays that most of the



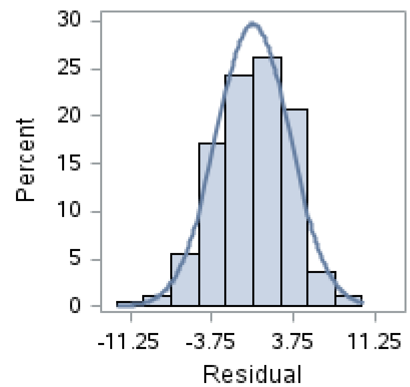


residuals lie within the -2 to +2 range, although there appears to be some potential outliers. Although few, it is important to note the residuals that reside outside of the specified range and monitor their influence on the model. Overall, this graph represents a good distribution of the majority of residuals. The QQ plot shows residuals strongly hugging the line, which suggests that they are normally distributed. There appears to be slight deviations at each end which could indicate slight skewness. The Cook's D plot displays major spikes which represent potential outliers within the residuals





around 0, a much higher spike at around 50 and a bit of a lower spike at around 190. Although few, it is important to note these spikes or possible outliers as they could significantly affect the model. Lastly, the Histogram of Residuals displays a normal distribution of residuals. The majority of residuals follow the bell curve, which indicates the normally distributed data. There may be a slight spike in values around 3.75 that seems to go beyond the curve but not significant. The interpretations of these relevant residual plots further validate the effectiveness of the Second Order model, which was previously determined as the best model at predicting daily stock prices for UPS. These plots provide insights into the model's adequacy and performance and solidify the models efficiency.



To take the analysis one step further, the model is to be displayed in its equation form and each coefficient within the model is interpreted in the context of the data. These insights provide additional support for the selection of this model as the best fit for the data. The equation of the best model for predicting daily stock prices of UPS is displayed below:

UPS = 201.99778 - 1.41583(RTX) + 0.34489(ROK) - 0.17753(VRSK) + 0.35565(HWM) - 0.07547(LMT) - 0.15356(VLTO) + 0.00558(rtx_2).

In order to better visualize and interpret each coefficient with the model, the following table was generated:

Coefficient Interpretation Key Statistic Table

Coefficient	*Interpretation*
Intercept(201.99778)	*The intercept represents the predicted UPS stock price when all other independent variables are equal to zero.*
RTX(-1.41583)	*For every 1 unit increase in RTX stock price, the UPS stock price decreases by 1.41583 units, holding all other variables constant.*

ROK(0.34489)	*For every 1 unit increase in ROK stock price, the UPS stock price increases by 0.34489 units, holding all other variables constant.*
VRSK(-0.17753)	*For every 1 unit increase in VRSK stock price, the UPS stock price decreases by 0.17753 units, holding all other variables constant.*
HWM(0.35565)	*For every 1 unit increase in HWM stock price, the UPS stock price increases by 0.35565 units, holding all other variables constant.*
LMT(-0.07547)	*For every 1 unit increase in LMT stock price, the UPS stock price decreases by 0.07547 units, holding all other variables constant.*
VLTO(-0.15356)	*For every 1 unit increase in VLTO stock price, the UPS stock price decreases by 0.15356 units, holding all other variables constant.*
RTX_2(0.00558)	*The second order term suggests a non linear relationship between RTX and UPS stock prices. For every 1 unit increase in RTX^2 stock price, the UPS stock price increases by 0.00558 units, holding all other variables constant.*

The interpretations of each of these variables help to explain how fluctuations in the other six companies' stock prices impact UPS. The interpretation of the second order term captures non-linear trends. These insights further support the selection of this model as the best fit for the data.

Through this multi-step analysis, the most effective model for predicting daily stock prices based on the other eight relevant stocks of companies within the S&P 500 was successfully identified. The best overall model for predicting daily stock prices for UPS is the Second Order model, Model 3: UPS = RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2. While each model was revealed to be statistically significant and all close in performance, the Second Order model remained on top. This model was chosen as the best overall model due to its strong statistical performance and interpretability. In the end, the analysis identified Model 3: UPS = RTX & ROK & VRSK

& HWM & LMT & VLTO & RTX^2 as the most effective model for predicting the daily stock prices of United Parcel Service (UPS).

Methodology

For the methodology of this analysis, several statistical techniques were implemented within SAS. SAS, also known as the Statistical Analysis System, is a software suite that allows users to perform data management and analysis, advanced analytics, predictive analytics and much more. To begin, a ‘proc print’ statement was used in SAS to gather important basic information regarding the dataset including the number of observations. For the purpose of the exploratory data analysis portion of this study, histograms were implemented within SAS. In SAS, histograms are generated with the ‘proc sgplot’ statement. Each variable was displayed with this technique in order to gather a better understanding before jumping into the bulk of the analysis. Histograms help to visualize distribution of variables in order to identify potential skewness, outliers, and variability. The purpose of using histograms in regards to this study is to better visualize each batting variable. Additionally, in SAS, a correlation analysis was done to find the most relevant variables in regards to the dependent variable. This is done with a ‘proc corr’ statement and includes all 77 variables that were correlated with UPS. After discovering the most relevant variables in regards to UPS, the ‘sgplot’ was utilized once more for the creation of scatterplots. Scatter plots are an impactful tool that help provide visual insights of the relationships between variables. Each of the variables was displayed with this technique alongside the dependent variable in order to gather a better understanding of the distributions of data before jumping into the bulk of the analysis. Following this, our multiple regression models were built with the ‘proc reg’ statement. This was utilized in the testing of all combinations of variables to find the best model with a Stepwise Regression model. A stepwise regression model was used for the primary variable selection. This is an iterative method for selecting variables in a multiple regression model. The goal is to find the best subset of independent variables. Following this, in order to support the conclusions made from stepwise, four major selection criteria methods are used. These include R-square Criterion, Adjusted R-square Criterion, CP Criterion and Press Criterion. These four selection criteria methods are also used within multiple regression models and unique perspectives on model performance. The combination of Stepwise and these four selection criteria models ensures that the chosen model is efficient. The ‘proc reg’ statement was also utilized in the creation of the interaction term model and the second order term model. Although before that step, a ‘proc glm’ and ‘proc plm’ statement was used to generate an interaction plot for the justification of adding an interaction term. These techniques create a plot or a graph in which the slopes of the variables can be seen. This is helpful in determining if an interaction term is needed. Additionally, for the creation of both the interaction and second order models, two new data sets were created in order to introduce the interaction terms and second order terms. Once generated, the new data sets including these new terms were utilized in each corresponding model. Finally, a test for multicollinearity was executed, again within a multiple regression model, by simply adding a ‘VIF’. This is to ensure that no two variables are highly correlated with each other. If two or more of the independent variables used in the model contribute redundant information, they will be correlated with each other causing multicollinearity or a VIF value greater than 10. The presence of multicollinearity was tested and fixed by removing variables iteratively until all remaining variables had acceptable VIFs less than 10. In summary, several statistical techniques were used within SAS for the purpose of this analysis in order to achieve the overall goal of identifying the best model for predicting the daily stock prices of UPS.

Conclusion

In conclusion, this analysis revealed that Model 3: RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2 (the Second Order model) is the best overall model for predicting the daily stock prices of UPS based on several other daily stocks listed within the S&P 500. While each model additionally revealed to perform very well in predicting the daily stock prices of UPS, it was not enough to outshine the Second Order model. Model 3: RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2 presented the highest overall adjusted r square, a p value <0.05, the lowest overall standard deviation, and lowest overall coefficient of variation, across all evaluated models, making it the most reliable and effective model. Despite testing an interaction model and multicollinearity, the Second Order model presented the best results. This suggests that when combined, the companies that are the most impactful predictors of the daily stock prices of UPS are RTX Corporation (RTX), Rockwell Automation (ROK), Verisk Analytics (VRSK), Howmet Aerospace (HWM), Lockheed Martin (LMT), Veralto (VLTO), as well as the second order term of RTX Corporation (RTX). Given these outcomes, Model 3: RTX & ROK & VRSK & HWM & LMT & VLTO & RTX^2 is the best performing model for predicting the daily stock prices of UPS based on several other daily stocks listed within the S&P 500 from November 1st, 2023 to October 31st, 2024.

If I were to complete this analysis again, I would consider incorporating a wider range of variables, including additional highly correlated stocks to expand the scope of the analysis. Likewise, I would be interested in expanding data sources to include factors that may provide more valuable context for fluctuations in UPS stock prices, like interest rates or inflation. Investigating more potential highly correlated variables could ultimately enhance the models overall efficiency and further improve the prediction accuracy of the stock. Additionally, I would consider extending the timeframe of the dataset to include additional years. This would offer valuable insights into how the daily stock prices of UPS have performed and evolved overtime, since joining the S&P 500. Furthermore, I would consider exploring the use of different software applications for better visualizations and analysis. Software I would be interested in using is R or Python as they offer powerful libraries such as ggplot2 (R) and Matplotlib (Python). These libraries are extremely useful for creating detailed, interactive plots that could provide better visualizations of the relationships between variables. I would keep these ideas in mind if I were to conduct a similar analysis, with the goal of refining the model by integrating more variables and testing different combinations of highly correlated stocks and incorporating the use of different softwares. With this I could improve predictions and provide deeper insights into the daily stock prices of UPS.

I gained a substantial amount of knowledge through my analysis of building a multiple regression model that can accurately predict the daily stock prices of UPS based on several other daily stocks listed within the S&P 500. Through this, I was able to successfully determine the best performing model and therefore I am happy with the results. I am looking forward to implementing what I have learned from this analysis on future projects involving more advanced predictive modeling and statistical techniques. I hope to build upon this foundation, and continue to explore more complex datasets and methodologies in both my future academic and professional career.

IMPORTANT TABLES

Variable Selection Key Statistic Table

Selection Methods	*Selected Model*	*Selection Method*
Stepwise Regression	*6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO)*	*Highest Adjusted R-Square of 0.8902*
R-Square Criterion	*7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM)*	*Highest R-Square of 0.8944 (Second highest of 0.8891)*
Adjusted R-Square Criterion	*7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM)*	*Highest Adjusted R-Square of 0.8913 (Second highest of 0.8902)*
Cp Criterion	*7-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO, MMM)*	*Lowest Cp value of 8.000 (Second lowest of 9.5966)*
Press Criterion	*6-variable model (RTX, ROK, VRSK, HWM, LMT, VLTO)*	*Lowest Press value of 3042.83734*

Stepwise Regression Model Key Statistic Table

Stepwise Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 1: *UPS = 143.24744 - 0.36729(RTX) + 0.34846(ROK) - 0.18196(VRSK) + 0.42064(HWM) - 0.04488(LMT) - 0.24603(VLTO)*	*0.8902*	*<.0001*	*3.42645*	*2.39991*

Interaction Model Key Statistic Table

Interaction Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 2: *UPS = 52.91801 - 0.35333(RTX) + 0.35033(ROK) + 0.15470(VRSK) + 1.52330(HWM) - 0.02440(LMT) - 0.26897(VLTO) - 0.00455(VRSK_HWM)*	*0.8914*	*<.0001*	*3.40817*	*2.38710*

Second Order Model Key Statistic Table

Second Order Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 3: ***UPS = 201.99778 - 1.41583(RTX) + 0.34489(ROK) - 0.17753(VRSK) + 0.35565(HWM) - 0.07547(LMT) - 0.15356(VLTO) + 0.00558(rtx_2)***	***0.8915***	***<.0001***	***3.40631***	***2.38580***

VIF Model 1 Key Statistic Table

Variable	*VIF*
RTX	***26.29748***
ROK	*1.59078*
VRSK	*2.94067*
HWM	***37.23414***
LMT	*4.92445*
VLTO	***23.58768***

VIF Model 2 Key Statistic Table

Variable	*VIF*
RTX	***20.78550***
ROK	*1.56398*
VRSK	*2.51149*
LMT	*4.82157*
VLTO	***15.44196***

VIF Model 3 Key Statistic Table

Variable	*VIF*
ROK	*1.40291*
VRSK	*2.40900*
LMT	*3.21413*
VLTO	*4.10776*

Multicollinearity Model Key Statistic Table

Multicollinearity Model	*R Squared Adjusted*	*F Test P-Value*	*Standard Deviation*	*Coefficient Variation*
Model 4: *UPS = 113.81604 + 0.34959(ROK) - 0.12604(VRSK) - 0.05003(LMT) - 0.12046(VLTO)*	*0.8747*	*<.0001*	*3.65962*	*2.56322*

Coefficient Interpretation Key Statistic Table

Coefficient	*Interpretation*
Intercept(201.99778)	*The intercept represents the predicted UPS stock price when all other independent variables are equal to zero.*
RTX(-1.41583)	*For every 1 unit increase in RTX stock price, the UPS stock price decreases by 1.41583 units, holding all other variables constant.*
ROK(0.34489)	*For every 1 unit increase in ROK stock price, the UPS stock price increases by 0.34489 units, holding all other variables constant.*
VRSK(-0.17753)	*For every 1 unit increase in VRSK stock price, the UPS stock price decreases by 0.17753 units, holding all other variables constant.*
HWM(0.35565)	*For every 1 unit increase in HWM stock price, the UPS stock price increases by 0.35565 units, holding all other variables constant.*
LMT(-0.07547)	*For every 1 unit increase in LMT stock price, the UPS stock price decreases by 0.07547 units, holding all other variables constant.*
VLTO(-0.15356)	*For every 1 unit increase in VLTO stock price, the UPS stock price decreases by 0.15356 units, holding all other variables constant.*
RTX_2(0.00558)	*The second order term suggests a non linear relationship between RTX and UPS stock prices.* *For every 1 unit increase in RTX^2 stock price, the UPS stock price increases by 0.00558 units, holding all other variables constant.*

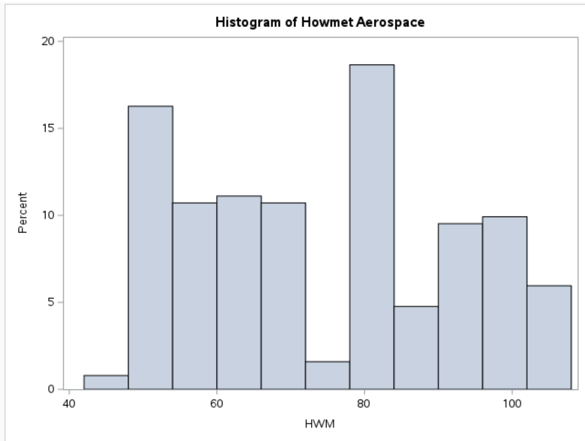
SAS CODE + OUTPUT

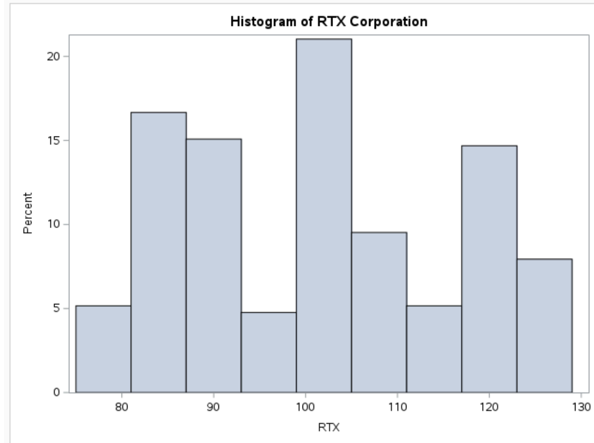
```
/* Generated Code (IMPORT) */
/* Source File: Fall 2024 Lab #3 Closing Prices.xlsx */
/* Source Path: /home/u63987812/sasuser.v94/Lab 3 */
/* Code generated on: 11/21/24, 9:18 PM */

%web_drop_table(WORK.IMPORT);


FILENAME REFFILE '/home/u63987812/sasuser.v94/Lab 3/Fall 2024 Lab #3 Closing Prices.xlsx';

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

PROC CONTENTS DATA=WORK.ups; RUN;


%web_open_table(WORK.IMPORT);

proc print data=ups; run;
```

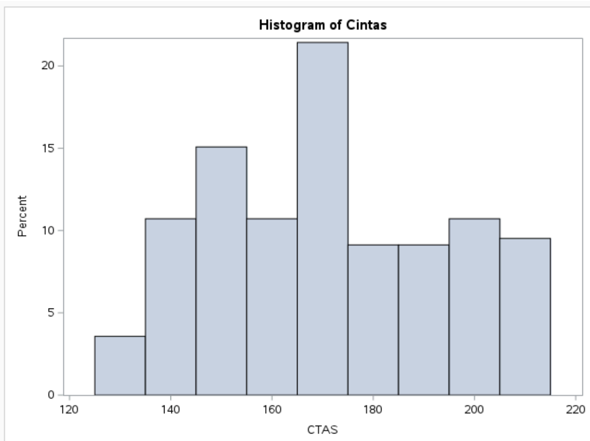
Alphabetic List of Variables and Attributes

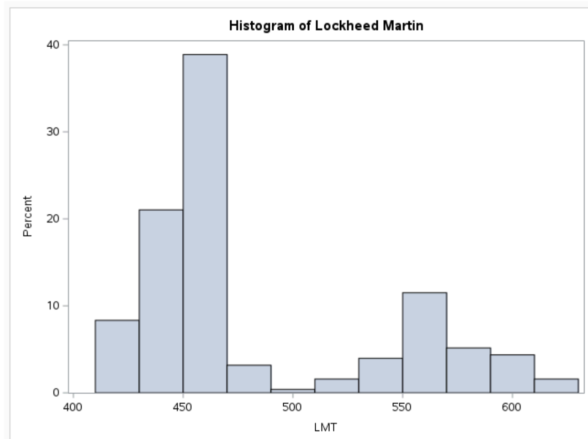
#	Variable	Type	Len	Format	Label
2	ADP	Num	8	BEST.	ADP
3	ALLE	Num	8	BEST.	ALLE
4	AME	Num	8	BEST.	AME
5	AOS	Num	8	BEST.	AOS
6	AXON	Num	8	BEST.	AXON
7	BA	Num	8	BEST.	BA
8	BLDR	Num	8	BEST.	BLDR
9	BR	Num	8	BEST.	BR
10	CARR	Num	8	BEST.	CARR
11	CAT	Num	8	BEST.	CAT
12	CHRW	Num	8	BEST.	CHRW
13	CMI	Num	8	BEST.	CMI
14	CPRT	Num	8	BEST.	CPRT
15	CSX	Num	8	BEST.	CSX
16	CTAS	Num	8	BEST.	CTAS
17	DAL	Num	8	BEST.	DAL
18	DAY	Num	8	BEST.	DAY
19	DE	Num	8	BEST.	DE
20	DOV	Num	8	BEST.	DOV
1	Date	Num	8	MMDDYY10.	Date
21	EFX	Num	8	BEST.	EFX
22	EMR	Num	8	BEST.	EMR
23	ETN	Num	8	BEST.	ETN
24	EXPD	Num	8	BEST.	EXPD
25	FAST	Num	8	BEST.	FAST
26	FDX	Num	8	BEST.	FDX
27	FTV	Num	8	BEST.	FTV
28	GD	Num	8	BEST.	GD
29	GE	Num	8	BEST.	GE
30	GNRC	Num	8	BEST.	GNRC
31	GWW	Num	8	BEST.	GWW
32	HII	Num	8	BEST.	HII
33	HON	Num	8	BEST.	HON
34	HUBB	Num	8	BEST.	HUBB
35	HWM	Num	8	BEST.	HWM
36	IEX	Num	8	BEST.	IEX
37	IR	Num	8	BEST.	IR
38	ITW	Num	8	BEST.	ITW
39	J	Num	8	BEST.	J
40	JBHT	Num	8	BEST.	JBHT
41	JCI	Num	8	BEST.	JCI
42	LDOS	Num	8	BEST.	LDOS
43	LHX	Num	8	BEST.	LHX
44	LMT	Num	8	BEST.	LMT
45	LUV	Num	8	BEST.	LUV
46	MAS	Num	8	BEST.	MAS
47	MMM	Num	8	BEST.	MMM
48	NDSN	Num	8	BEST.	NDSN
49	NOC	Num	8	BEST.	NOC
50	NSC	Num	8	BEST.	NSC
51	ODFL	Num	8	BEST.	ODFL
52	OTIS	Num	8	BEST.	OTIS
53	PAYC	Num	8	BEST.	PAYC
54	PAYX	Num	8	BEST.	PAYX
55	PCAR	Num	8	BEST.	PCAR
56	PH	Num	8	BEST.	PH
57	PNR	Num	8	BEST.	PNR
58	PWR	Num	8	BEST.	PWR
59	ROK	Num	8	BEST.	ROK
60	ROL	Num	8	BEST.	ROL
61	RSG	Num	8	BEST.	RSG
62	RTX	Num	8	BEST.	RTX
63	SNA	Num	8	BEST.	SNA
64	SWK	Num	8	BEST.	SWK
65	TDG	Num	8	BEST.	TDG
66	TT	Num	8	BEST.	TT
67	TXT	Num	8	BEST.	TXT
68	UAL	Num	8	BEST.	UAL
69	UBER	Num	8	BEST.	UBER
70	UNP	Num	8	BEST.	UNP
71	UPS	Num	8	BEST.	UPS
72	URI	Num	8	BEST.	URI
73	VLTO	Num	8	BEST.	VLTO
74	VRSK	Num	8	BEST.	VRSK
75	WAB	Num	8	BEST.	WAB
76	WM	Num	8	BEST.	WM
77	XYL	Num	8	BEST.	XYL

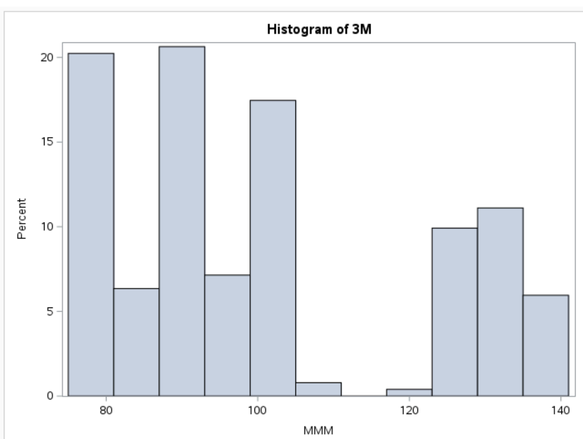
```
/* EDA */
/* HISTOGRAMS */
/* 8 variables */
/* (Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) ,
3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK)) */
/* histogram Howmet Aerospace */
proc sgplot data=ups;
    histogram HWM;
    TITLE "Histogram of Howmet Aerospace";
    run;

/* histogram RTX Corporation */
proc sgplot data=ups;
    histogram RTX;
    TITLE "Histogram of RTX Corporation";
    run;

/* histogram Cintas */
proc sgplot data=ups;
    histogram CTAS;
    TITLE "Histogram of Cintas";
    run;

/* histogram Lockheed Martin */
proc sgplot data=ups;
    histogram LMT;
    TITLE "Histogram of Lockheed Martin";
    run;

/* histogram 3M */
proc sgplot data=ups;
    histogram MMM;
    TITLE "Histogram of 3M";
    run;

/* histogram Rockwell Automation */
proc sgplot data=ups;
    histogram ROK;
    TITLE "Histogram of Rockwell Automation";
    run;

/* histogram Veralto */
proc sgplot data=ups;
    histogram VLTO;
    TITLE "Histogram of Veralto";
    run;


/* histogram Verisk Analytics */
proc sgplot data=ups;
    histogram VRSK;
    TITLE "Histogram of Verisk Analytics";
    run;

/* histogram United Parcel Service */
proc sgplot data=ups;
    histogram UPS;
    TITLE "Histogram of United Parcel Service";
    run;
```

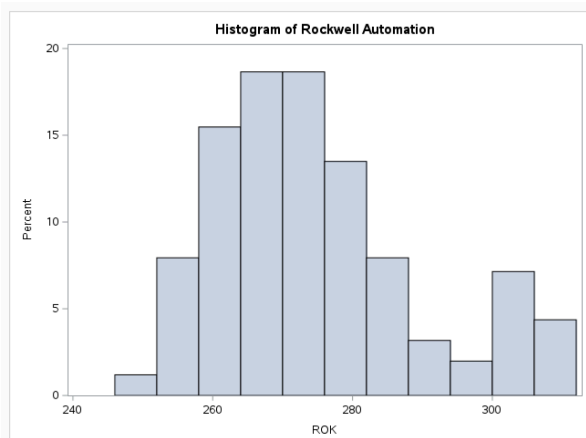

Histogram of Howmet Aerospace
Percent
HWM

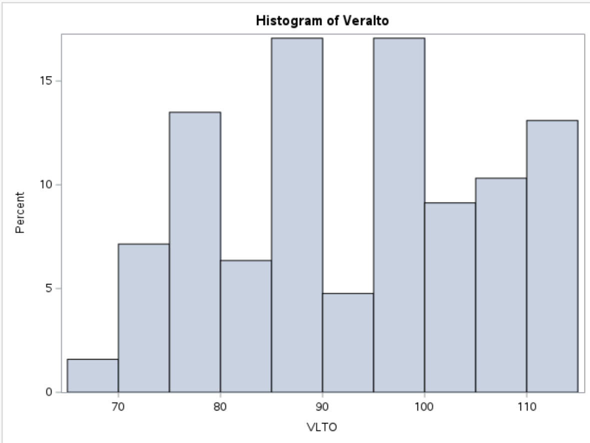

Histogram of RTX Corporation
Percent
RTX

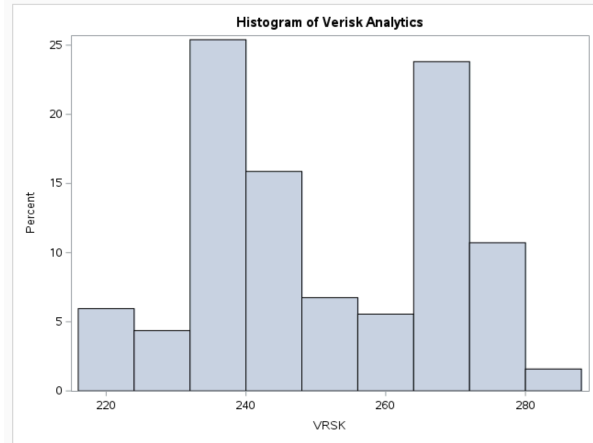

Histogram of Cintas
Percent
CTAS

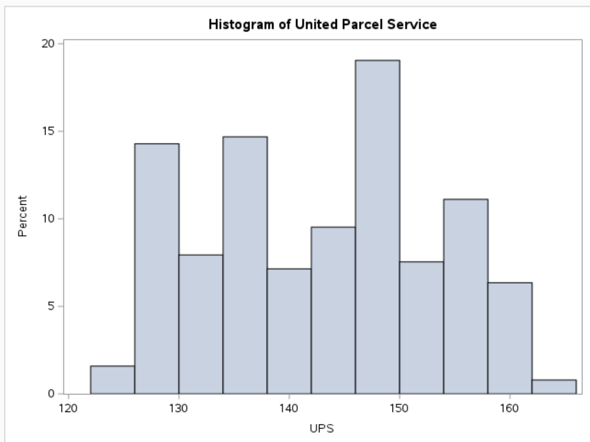

Histogram of Lockheed Martin
Percent
LMT


Histogram of 3M
Percent
MMM


Histogram of Rockwell Automation
Percent
ROK








```
/* PROC CORR */
proc corr data=ups;
var _numeric_;
with ups;
    run;

/* variables most strongly associated ups based on correlation coefficients:
highest proc corr outputs:
Howmet Aerospace (HWM): -0.80339
RTX Corporation (RTX): -0.81707
Cintas (CTAS): -0.77796
Lockheed Martin (LMT): -0.75456
3M (MMM): -0.77309
Rockwell Automation (ROK): 0.79696
Veralto (VLTO): -0.78342
Verisk Analytics (VRSK): -0.75028
*/
```

	Date	ADP	ALLE	AME	AOS	AXON	BA	BLDR	BR	CARR	CAT	CHRW	CMI	CPRT	CSX	CTAS	DAL	DAY	DE	DOV	EFX	EMR
UPS UPS	-0.79916 <.0001	-0.65186 <.0001	-0.40164 <.0001	0.07197 0.2550	0.03939 0.5336	-0.62673 <.0001	0.74755 <.0001	0.07803 0.2170	-0.45857 <.0001	-0.68319 <.0001	-0.44325 <.0001	-0.63742 <.0001	-0.61461 <.0001	-0.13477 0.0325	0.33245 <.0001	-0.77796 <.0001	-0.25570 <.0001	0.64807 <.0001	0.11352 0.0720	-0.59587 <.0001	-0.50108 <.0001	-0.31638 <.0001

Pearson Correlation Coefficients, N = 252
Prob > |r| under H0: Rho=0

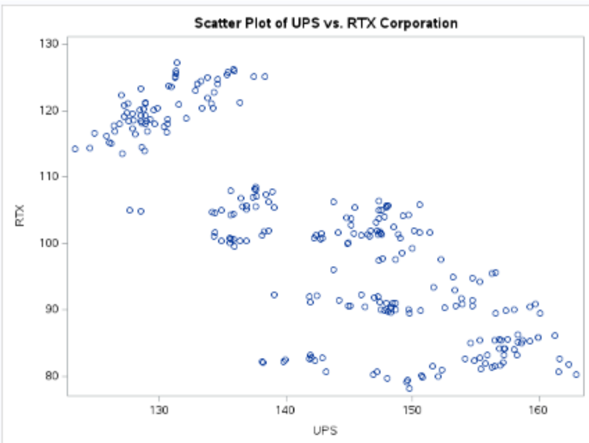
ETN	EXPD	FAST	FDX	FTV	GD	GE	GNRC	GWW	HII	HON	HUBB	HWM	IEX	IR	ITW	J	JBHT	JCI	LDOS	LHX	LMT	LUV
-0.49459	0.04288	-0.06659	-0.46534	0.24500	-0.68052	-0.73172	-0.63793	-0.47314	0.06460	-0.26220	-0.43588	-0.80339	0.48899	-0.50419	0.41024	-0.48439	0.65247	-0.68756	-0.74258	-0.63526	-0.75456	0.20105
<.0001	0.4980	0.2924	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.3070	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0013

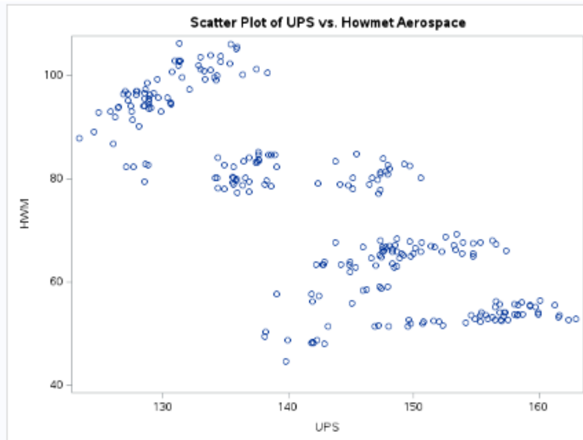
MAS	MMM	NDSN	NOC	NSC	ODFL	OTIS	PAYC	PAYX	PCAR	PH	PNR	PWR	ROK	ROL	RSG	RTX	SNA	SWK	TDG	TT	TXT	UAL
-0.42960	-0.77309	0.34108	-0.44070	-0.07741	0.30411	-0.36757	0.74289	-0.46669	0.22519	-0.58937	-0.58128	-0.59287	0.79696	-0.65773	-0.71493	-0.81707	0.09848	-0.20165	-0.67647	-0.73541	-0.25234	-0.34569
<.0001	<.0001	<.0001	<.0001	0.2207	<.0001	<.0001	<.0001	<.0001	0.0003	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1189	0.0013	<.0001	<.0001	<.0001	<.0001

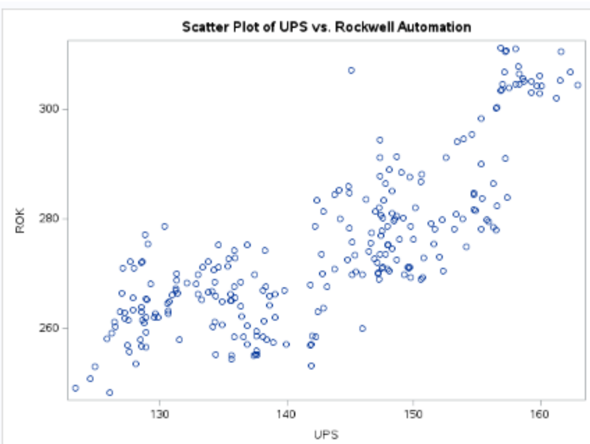
UBER	UNP	UPS	URI	VLTO	VRSK	WAB	WM	XYL
-0.23669	0.03029	1.00000	-0.53172	-0.78342	-0.75028	-0.65998	-0.48334	-0.51336
0.0001	0.6323		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

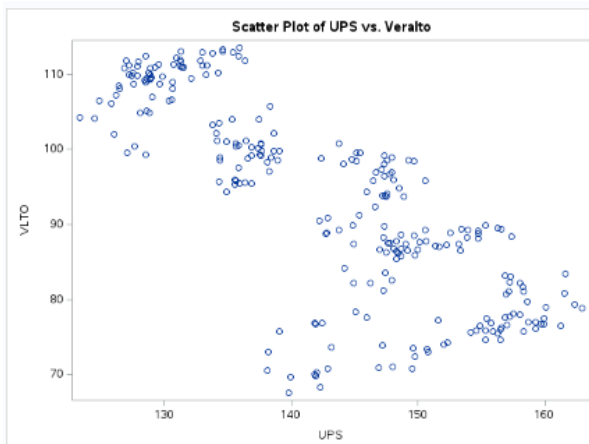
```
/* PROC CORR */
/* scatterplots */
/* (Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) ,
3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK)) */

/* ups vs Howmet Aerospace scatterplot */
proc sgplot data=ups;
    scatter x=ups y=hwm;
    title "Scatter Plot of UPS vs. Howmet Aerospace";
    xaxis label="UPS";
    yaxis label="HWM";
    run;

/* ups vs RTX Corporation scatterplot */
proc sgplot data=ups;
    scatter x=ups y=rtx;
    title "Scatter Plot of UPS vs. RTX Corporation";
    xaxis label="UPS";
    yaxis label="RTX";
    run;

/* ups vs Cintas scatterplot */
proc sgplot data=ups;
    scatter x=ups y=ctas;
    title "Scatter Plot of UPS vs. Cintas";
    xaxis label="UPS";
    yaxis label="CTAS";
    run;

/* ups vs Lockheed Martin scatterplot */
proc sgplot data=ups;
    scatter x=ups y=lmt;
    title "Scatter Plot of UPS vs. Lockheed Martin";
    xaxis label="UPS";
    yaxis label="LMT";
    run;

/* ups vs 3M scatterplot */
proc sgplot data=ups;
    scatter x=ups y=mmm;
    title "Scatter Plot of UPS vs. 3M";
    xaxis label="UPS";
    yaxis label="MMM";
    run;

/* ups vs Rockwell Automation scatterplot */
proc sgplot data=ups;
    scatter x=ups y=rok;
    title "Scatter Plot of UPS vs. Rockwell Automation";
    xaxis label="UPS";
    yaxis label="ROK";
    run;
```

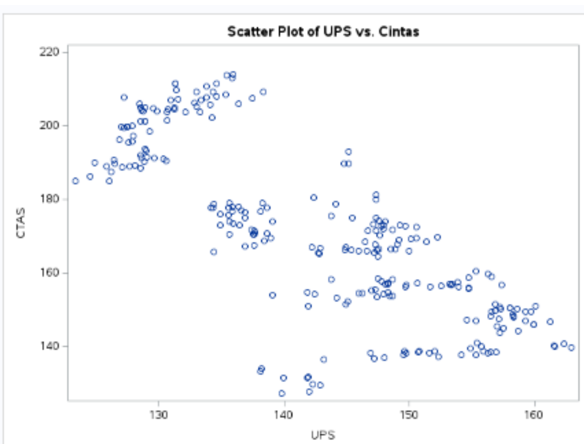
```
/* ups vs Veralto scatterplot */
proc sgplot data=ups;
    scatter x=ups y=vlto;
    title "Scatter Plot of UPS vs. Veralto";
    xaxis label="UPS";
    yaxis label="VLTO";
    run;

/* ups vs Verisk Analytics scatterplot */
proc sgplot data=ups;
    scatter x=ups y=vrsk;
    title "Scatter Plot of UPS vs. Verisk Analytics";
    xaxis label="UPS";
    yaxis label="VRSK";
    run;
```

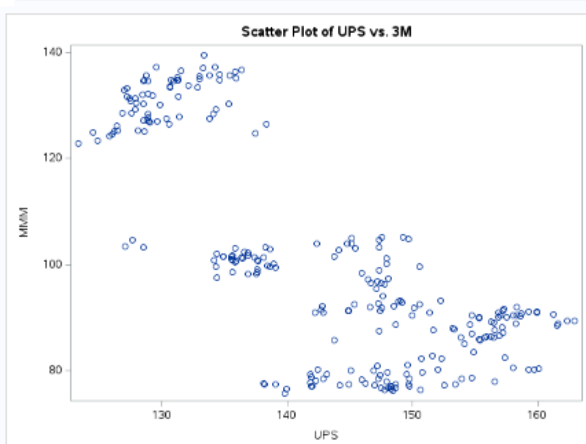


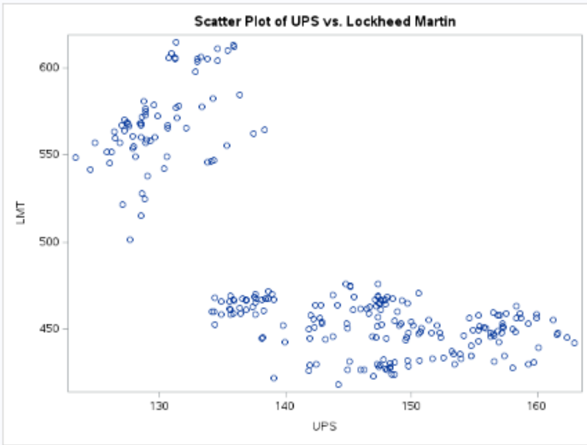


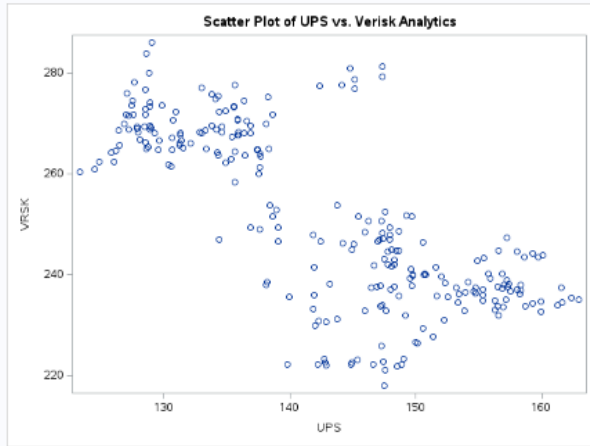













```
/* (Howmet Aerospace (HWM) , RTX Corporation (RTX) , Cintas (CTAS) , Lockheed Martin (LMT) ,
3M (MMM) , Rockwell Automation (ROK) , Veralto (VLTO) , Verisk Analytics (VRSK)) */

/* stepwise */
proc reg data=ups;
    model ups = hwm rtx ctas lmt mmm rok vlto vrsk/selection = stepwise slentry = .05;
    run;

/* variables remaining after stepwise:
 1 RTX Corporation (RTX)
 2 Rockwell Automation (ROK)
 3 Verisk Analytics (VRSK)
 4 Howmet Aerospace (HWM)
 5 Lockheed Martin (LMT)
 6 Veralto (VLTO)
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Stepwise Selection: Step 1

Variable RTX Entered: R-Square = 0.6676 and C(p) = 517.9150

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	17917	17917	502.12	<.0001
Error	250	8920.92216	35.68369		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	202.84039	2.70685	200378	5615.39	<.0001
RTX	-0.59277	0.02645	17917	502.12	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable ROK Entered: R-Square = 0.8413 and C(p) = 119.7564

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	22578	11289	659.84	<.0001
Error	249	4260.11264	17.10889		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	86.59044	7.28837	2414.90840	141.15	<.0001
RTX	-0.39412	0.02192	5532.04843	323.34	<.0001
ROK	0.34973	0.02119	4660.80953	272.42	<.0001

Bounds on condition number: 1.4317, 5.727

Stepwise Selection: Step 3

Variable VRSK Entered: R-Square = 0.8699 and C(p) = 55.8602

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	23346	7781.88610	552.57	<.0001
Error	248	3492.59272	14.08304		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	119.43468	7.96989	3162.65822	224.57	<.0001
RTX	-0.28221	0.02500	1794.04597	127.39	<.0001
ROK	0.32427	0.01953	3881.80744	275.64	<.0001
VRSK	-0.14806	0.02006	767.51991	54.50	<.0001

Bounds on condition number: 2.2637, 17.504

Stepwise Selection: Step 4

Variable HWM Entered: R-Square = 0.8795 and C(p) = 35.6150

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	23605	5901.18927	450.78	<.0001
Error	247	3233.49393	13.09107		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	141.31408	9.12316	3140.91126	239.93	<.0001
HWM	0.28837	0.06482	259.09879	19.79	<.0001
RTX	-0.57952	0.07104	871.09608	66.54	<.0001
ROK	0.32393	0.01883	3873.71423	295.91	<.0001
VRSK	-0.20040	0.02263	1026.19775	78.39	<.0001

Bounds on condition number: 24.372, 193.51

Stepwise Selection: Step 5

Variable LMT Entered: R-Square = 0.8888 and C(p) = 16.3353

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23853	4770.52200	393.06	<.0001
Error	246	2985.64103	12.13675		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	140.01055	8.78907	3079.91175	253.77	<.0001
HWM	0.28466	0.06242	252.43032	20.80	<.0001
RTX	-0.44970	0.07419	445.88649	36.74	<.0001
LMT	-0.03799	0.00841	247.85290	20.42	<.0001
ROK	0.33556	0.01831	4074.71114	335.73	<.0001
VRSK	-0.18588	0.02203	864.06134	71.19	<.0001

Bounds on condition number: 24.376, 282.62

Stepwise Selection: Step 6

Variable VLTO Entered: R-Square = 0.8928 and C(p) = 8.9601

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	23962	3993.63443	340.16	<.0001
Error	245	2876.44445	11.74059		
Corrected Total	251	26838			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	143.24744	8.70935	3176.08606	270.52	<.0001
HWM	0.42064	0.07587	360.84951	30.74	<.0001
RTX	-0.36729	0.07781	261.58601	22.28	<.0001
LMT	-0.04488	0.00857	321.84670	27.41	<.0001
ROK	0.34846	0.01850	4164.34304	354.70	<.0001
VLTO	-0.24603	0.08067	109.19658	9.30	0.0025
VRSK	-0.18196	0.02171	825.07958	70.28	<.0001

Bounds on condition number: 37.234, 579.45

Bounds on condition number: 37.234, 579.45

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.0500 significance level for entry into the model.

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	RTX		RTX	1	0.6676	0.6676	517.915	502.12	<.0001
2	ROK		ROK	2	0.1737	0.8413	119.756	272.42	<.0001
3	VRSK		VRSK	3	0.0286	0.8699	55.8602	54.50	<.0001
4	HWM		HWM	4	0.0097	0.8795	35.6150	19.79	<.0001
5	LMT		LMT	5	0.0092	0.8888	16.3353	20.42	<.0001
6	VLTO		VLTO	6	0.0041	0.8928	8.9601	9.30	0.0025

```
/*other 4 selection criteria methods:
r-square,adjusted r-square,cp,press*/

/* r-square criterion */
proc reg data=ups;
model ups = rtx rok vrsk hwm lmt vlto mmm/ selection=rsquare start=3 best=2;
run;
/* starts looking at 3-var models; */
/* best gives best # var models; */

/* adjusted r-square criterion; */
proc reg data = ups;
model ups = rtx rok vrsk hwm lmt vlto mmm/ selection=adjrsq;
run;

/* cp criterion (lowest is best) ; */
proc reg data = ups;
model ups = rtx rok vrsk hwm lmt vlto mmm/ selection=cp;
run;

/* press criterion, also gives press statistici */
proc glmselect data=ups;
model ups = rtx rok vrsk hwm lmt vlto mmm/ selection=stepwise (choose=press) ;
run;

/* gives graphs, easiest way to compare models */
proc glmselect data=ups plots=criterionpanel;
model ups = rtx rok vrsk hwm lmt vlto mmm/ selection=stepwise stats=all;
run;

/* final model:
6 variables: RTX, ROK, VRSK, HWM, LMT, VLTO */
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS

R-Square Selection Method

Number of Observations Read	252
Number of Observations Used	252

Number in Model	R-Square	Variables in Model
3	0.8711	ROK VRSK LMT
3	0.8699	RTX ROK VRSK
4	0.8795	RTX ROK VRSK HWM
4	0.8793	RTX ROK VRSK LMT
5	0.8888	RTX ROK VRSK HWM LMT
5	0.8844	RTX ROK VRSK HWM MMM
6	0.8928	RTX ROK VRSK HWM LMT VLTO
6	0.8891	RTX ROK VRSK HWM LMT MMM
7	0.8944	RTX ROK VRSK HWM LMT VLTO MMM

The REG Procedure
Model: MODEL1
Dependent Variable: UPS

Adjusted R-Square Selection Method

Number of Observations Read	252
Number of Observations Used	252

Number in Model	Adjusted R-Square	R-Square	Variables in Model
7	0.8913	0.8944	RTX ROK VRSK HWM LMT VLTO MMM
6	0.8902	0.8928	RTX ROK VRSK HWM LMT VLTO
5	0.8865	0.8888	RTX ROK VRSK HWM LMT
6	0.8863	0.8891	RTX ROK VRSK HWM LMT MMM
6	0.8832	0.8860	RTX ROK VRSK HWM VLTO MMM
5	0.8820	0.8844	RTX ROK VRSK HWM MMM
6	0.8808	0.8837	ROK VRSK HWM LMT VLTO MMM
5	0.8807	0.8831	ROK VRSK HWM LMT VLTO
5	0.8784	0.8808	RTX ROK VRSK HWM VLTO
5	0.8777	0.8801	RTX ROK VRSK LMT MMM
4	0.8776	0.8795	RTX ROK VRSK HWM
4	0.8774	0.8793	RTX ROK VRSK LMT

The GLMSELECT Procedure

Data Set	WORK.UPS
Dependent Variable	UPS
Selection Method	Stepwise
Select Criterion	SBC
Stop Criterion	SBC
Choose Criterion	PRESS
Effect Hierarchy Enforced	None

Number of Observations Read	252
Number of Observations Used	252

Dimensions	
Number of Effects	8
Number of Parameters	8

The REG Procedure
Model: MODEL1
Dependent Variable: UPS

C(p) Selection Method

Number of Observations Read	252
Number of Observations Used	252

Number in Model	C(p)	R-Square	Variables in Model
7	8.0000	0.8944	RTX ROK VRSK HWM LMT VLTO MMM
6	9.5966	0.8928	RTX ROK VRSK HWM LMT VLTO
5	16.9960	0.8888	RTX ROK VRSK HWM LMT
6	18.2760	0.8891	RTX ROK VRSK HWM LMT MMM
6	25.4186	0.8860	RTX ROK VRSK HWM VLTO MMM
5	27.0773	0.8844	RTX ROK VRSK HWM MMM
5	30.1133	0.8831	ROK VRSK HWM LMT VLTO
6	30.7456	0.8837	ROK VRSK HWM LMT VLTO MMM
5	35.3003	0.8808	RTX ROK VRSK HWM VLTO
4	36.3305	0.8795	RTX ROK VRSK HWM
4	36.7245	0.8793	RTX ROK VRSK LMT
5	36.9359	0.8801	RTX ROK VRSK LMT MMM

The GLMSELECT Procedure

Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	SBC	PRESS
0	Intercept		1	1181.9043	27052.5276
1	RTX		2	909.8737	9067.0606
2	ROK		3	729.1490	4348.0815
3	VRSK		4	684.6183	3606.2747
4	HWM		5	670.7233	3369.4993
5	LMT		6	656.1560	3135.4484
6	VLTO		7	652.2961*	3042.8373*

* Optimal Value of Criterion

Selection stopped at a local minimum of the SBC criterion.

Stop Details

Candidate For	Effect	Candidate SBC		Compare SBC
Entry	MMM	654.1380	>	652.2961
Removal	VLTO	656.1560	>	652.2961

The GLMSELECT Procedure
Selected Model

The selected model, based on PRESS, is the model at Step 6.

Effects:	Intercept RTX ROK VRSK HWM LMT VLTO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	23962	3993.63443	340.16
Error	245	2876.44445	11.74059	
Corrected Total	251	26838		

Root MSE	3.42645
Dependent Mean	142.77441
R-Square	0.8928
Adj R-Sq	0.8902
AIC	881.59005
AICC	882.18264
PRESS	3042.83734
SBC	652.29606

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	143.247436	8.709349	16.45
RTX	1	-0.367294	0.077813	-4.72
ROK	1	0.348458	0.018502	18.83
VRSK	1	-0.181959	0.021706	-8.38
HWM	1	0.420639	0.075874	5.54
LMT	1	-0.044880	0.008572	-5.24
VLTO	1	-0.246032	0.080674	-3.05

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 6).

Effects:	Intercept RTX ROK VRSK HWM LMT VLTO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	23962	3993.63443	340.16
Error	245	2876.44445	11.74059	
Corrected Total	251	26838		

Root MSE	3.42645
Dependent Mean	142.77441
R-Square	0.8928
Adj R-Sq	0.8902
AIC	881.59005
AICC	882.18264
BIC	629.83840
C(p)	9.59663
PRESS	3042.83734
SBC	652.29606
ASE	11.41446

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	143.247436	8.709349	16.45
RTX	1	-0.367294	0.077813	-4.72
ROK	1	0.348458	0.018502	18.83
VRSK	1	-0.181959	0.021706	-8.38
HWM	1	0.420639	0.075874	5.54
LMT	1	-0.044880	0.008572	-5.24
VLTO	1	-0.246032	0.080674	-3.05

The GLMSELECT Procedure

Data Set	WORK.UPS
Dependent Variable	UPS
Selection Method	Stepwise
Select Criterion	SBC
Stop Criterion	SBC
Effect Hierarchy Enforced	None

Number of Observations Read	252
Number of Observations Used	252

Dimensions

Number of Effects	8
Number of Parameters	8

The GLMSELECT Procedure

Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Model R-Square	Adjusted R-Square	AIC	AICC	BIC	CP	SBC	PRESS	ASE	F Value	Pr > F
0	Intercept		1	0.0000	0.0000	1432.3749	1432.4231	1177.0056	2060.1647	1181.9043	27052.5276	106.5010	0.00	1.0000
1	RTX		2	0.6676	0.6663	1156.8148	1156.9116	901.2248	519.8891	909.8737	9067.0606	35.4005	502.12	<.0001
2	ROK		3	0.8413	0.8400	972.5607	972.7226	718.4883	120.6991	729.1490	4348.0815	16.9052	272.42	<.0001
3	VRSK		4	0.8699	0.8683	924.5006	924.7445	671.1541	56.6330	684.6183	3606.2747	13.8595	54.50	<.0001
4	HWM		5	0.8795	0.8776	907.0761	907.4190	654.1122	36.3305	670.7233	3369.4993	12.8313	19.79	<.0001
5	LMT		6	0.8888	0.8865	888.9794	889.4385	636.7454	16.9960	656.1560	3135.4484	11.8478	20.42	<.0001
6	VLTO		7	0.8928	0.8902*	881.5901*	882.1826*	629.8384*	9.5966*	652.2961*	3042.8373*	11.4145	9.30	0.0025

* Optimal Value of Criterion

Selection stopped at a local minimum of the SBC criterion.

Stop Details

Candidate For	Effect	Candidate SBC		Compare SBC
Entry	MMM	654.1380	>	652.2961
Removal	VLTO	656.1560	>	652.2961

```
/* final model from stepwise:
6 variables: RTX, ROK, VRSK, HWM, LMT, VLTO */
proc reg data = ups;
    model ups = rtx rok vrsk hwm lmt vlto;
    run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	23962	3993.63443	340.16	<.0001
Error	245	2876.44445	11.74059		
Corrected Total	251	26838			

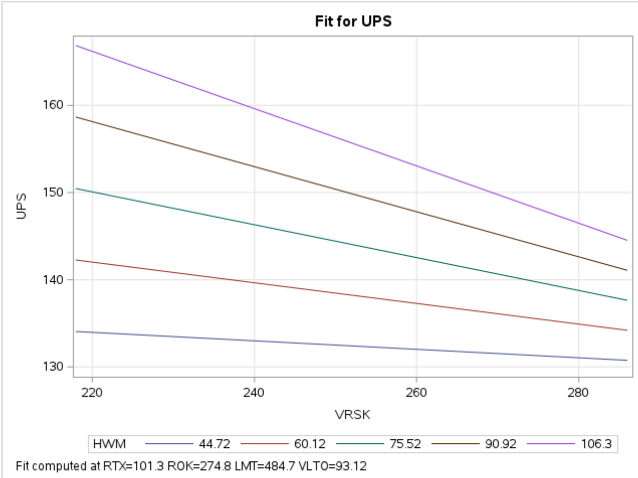
Root MSE	3.42645	R-Square	0.8928
Dependent Mean	142.77441	Adj R-Sq	0.8902
Coeff Var	2.39991		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	143.24744	8.70935	16.45	<.0001
RTX	RTX	1	-0.36729	0.07781	-4.72	<.0001
ROK	ROK	1	0.34846	0.01850	18.83	<.0001
VRSK	VRSK	1	-0.18196	0.02171	-8.38	<.0001
HWM	HWM	1	0.42064	0.07587	5.54	<.0001
LMT	LMT	1	-0.04488	0.00857	-5.24	<.0001
VLTO	VLTO	1	-0.24603	0.08067	-3.05	0.0025

```
/* Test one interaction term and one quadratic term to determine if its significant in the model  */
/* interaction term model */
proc glm data=ups;
    model ups = rtx rok vrsk hwm lmt vlto vrsk*hwm / solution;
    store GLMMODEL;
run;

proc plm restore=GLMMODEL noinfo;
    effectplot slicefit(x=vrsk sliceby=hwm);
    title "Interaction Effect of VRSK and HWM on UPS";
run;

data ups2;
    set ups;
    vrsk_hwm=vrsk*hwm;

proc reg data=ups2;
    model ups = rtx rok vrsk hwm lmt vlto vrsk_hwm;
    run;


/* quadratic term model */
data ups3;
    set ups;
    rtx_2=rtx*rtx;

proc reg data=ups3;
    model ups = rtx rok vrsk hwm lmt vlto rtx_2;
    run;
```

Interaction Effect of VRSK and HWM on UPS

The PLM Procedure



The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	24004	3429.14813	295.22	<.0001
Error	244	2834.21409	11.61563		
Corrected Total	251	26838			

Root MSE	3.40817	R-Square	0.8944
Dependent Mean	142.77441	Adj R-Sq	0.8914
Coeff Var	2.38710		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	52.91801	48.15934	1.10	0.2729
RTX	RTX	1	-0.35333	0.07774	-4.54	<.0001
ROK	ROK	1	0.35033	0.01843	19.01	<.0001
VRSK	VRSK	1	0.15470	0.17788	0.87	0.3853
HWM	HWM	1	1.52330	0.58320	2.61	0.0096
LMT	LMT	1	-0.02440	0.01371	-1.78	0.0764
VLTO	VLTO	1	-0.26897	0.08114	-3.31	0.0011
vrsk_hwm		1	-0.00455	0.00238	-1.91	0.0577

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	24007	3429.58972	295.58	<.0001
Error	244	2831.12301	11.60296		
Corrected Total	251	26838			

Root MSE	3.40631	R-Square	0.8945
Dependent Mean	142.77441	Adj R-Sq	0.8915
Coeff Var	2.38580		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	201.99778	30.96168	6.52	<.0001
RTX	RTX	1	-1.41583	0.53615	-2.64	0.0088
ROK	ROK	1	0.34489	0.01848	18.66	<.0001
VRSK	VRSK	1	-0.17753	0.02169	-8.18	<.0001
HWM	HWM	1	0.35565	0.08228	4.32	<.0001
LMT	LMT	1	-0.07547	0.01767	-4.27	<.0001
VLTO	VLTO	1	-0.15356	0.09285	-1.65	0.0994
rtx_2		1	0.00558	0.00282	1.98	0.0492

```
/* Test for multicollinearity */
proc reg data = ups;
    model ups = rtx rok vrsk hwm lmt vlto / VIF;
    run;
```

```
/* Test for multicollinearity */
proc reg data = ups;
    model ups = rtx rok vrsk hwm lmt vlto / VIF;
    run;
/* Remove highest vif, HWM 37.23414.  */
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	23962	3993.63443	340.16	<.0001
Error	245	2876.44445	11.74059		
Corrected Total	251	26838			

Root MSE	3.42645	R-Square	0.8928
Dependent Mean	142.77441	Adj R-Sq	0.8902
Coeff Var	2.39991		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	143.24744	8.70935	16.45	<.0001	0
RTX	RTX	1	-0.36729	0.07781	-4.72	<.0001	26.29748
ROK	ROK	1	0.34846	0.01850	18.83	<.0001	1.59078
VRSK	VRSK	1	-0.18196	0.02171	-8.38	<.0001	2.94067
HWM	HWM	1	0.42064	0.07587	5.54	<.0001	37.23414
LMT	LMT	1	-0.04488	0.00857	-5.24	<.0001	4.92445
VLTO	VLTO	1	-0.24603	0.08067	-3.05	0.0025	23.58768

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23601	4720.19141	358.68	<.0001
Error	246	3237.29396	13.15973		
Corrected Total	251	26838			

Root MSE	3.62763	R-Square	0.8794
Dependent Mean	142.77441	Adj R-Sq	0.8769
Coeff Var	2.54082		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	118.88291	7.96068	14.93	<.0001	0
RTX	RTX	1	-0.16979	0.07324	-2.32	0.0213	20.78550
ROK	ROK	1	0.33514	0.01942	17.26	<.0001	1.56398
VRSK	VRSK	1	-0.13599	0.02124	-6.40	<.0001	2.51149
LMT	LMT	1	-0.03801	0.00898	-4.23	<.0001	4.82157
VLTO	VLTO	1	0.01680	0.06911	0.24	0.8082	15.44196

```
/* Test for multicollinearity */
proc reg data = ups;
    model ups = rok vrsk lmt vlto / VIF;
    run;
/* Remove highest vif, HWM 37.23414.  */
/* Remove next highest, RTX 20.78550. */
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	23530	5882.55744	439.23	<.0001
Error	247	3308.02128	13.39280		
Corrected Total	251	26838			

Root MSE	3.65962	R-Square	0.8767
Dependent Mean	142.77441	Adj R-Sq	0.8747
Coeff Var	2.56322		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	113.81604	7.72227	14.74	<.0001	0
ROK	ROK	1	0.34959	0.01856	18.84	<.0001	1.40291
VRSK	VRSK	1	-0.12604	0.02098	-6.01	<.0001	2.40900
LMT	LMT	1	-0.05003	0.00740	-6.76	<.0001	3.21413
VLTO	VLTO	1	-0.12046	0.03596	-3.35	0.0009	4.10776

```
/* The final independent variables should be regressed on the dependent variable */
/* best model: quadratic term model */
/* final proc reg */
proc reg data=ups3;
    model ups = rtx rok vrsk hwm lmt vlto rtx_2;
    run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS

Number of Observations Read	252
Number of Observations Used	252

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	24007	3429.58972	295.58	<.0001
Error	244	2831.12301	11.60296		
Corrected Total	251	26838			

Root MSE	3.40631	R-Square	0.8945
Dependent Mean	142.77441	Adj R-Sq	0.8915
Coeff Var	2.38580		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	201.99778	30.96168	6.52	<.0001
RTX	RTX	1	-1.41583	0.53615	-2.64	0.0088
ROK	ROK	1	0.34489	0.01848	18.66	<.0001
VRSK	VRSK	1	-0.17753	0.02169	-8.18	<.0001
HWM	HWM	1	0.35565	0.08228	4.32	<.0001
LMT	LMT	1	-0.07547	0.01767	-4.27	<.0001
VLTO	VLTO	1	-0.15356	0.09285	-1.65	0.0994
rtx_2		1	0.00558	0.00282	1.98	0.0492

The REG Procedure
Model: MODEL1
Dependent Variable: UPS UPS