# A white logo with a blue background Description automatically generatedPerformance Analysis of US-listed Securities

* During the COVID-19 Market Crash

Author: Chufan He

## **Executive Summary**

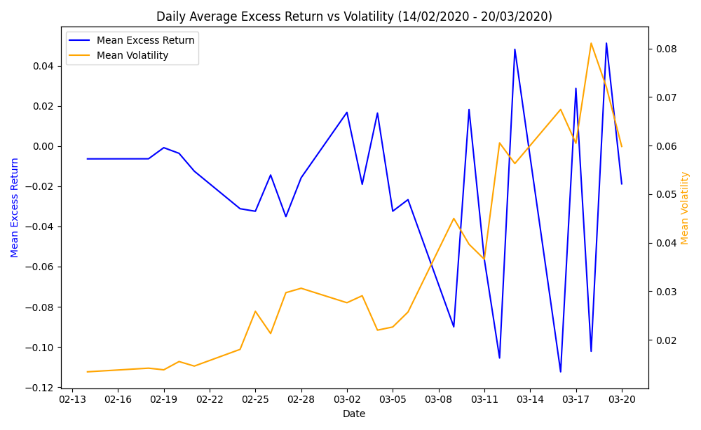
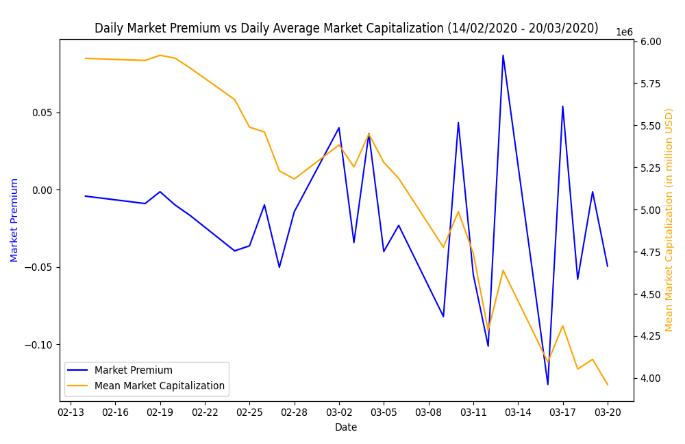
This report provides a comprehensive analysis of the performance of US-listed securities during the COVID-19 market crash, employing data analytics including exploratory data analysis, visualizations, descriptive statistics, supervised machine learning (OLS and Logit Regressions), unsupervised machine learning (K-means clustering and principal component analysis), analyzing experimental evidence using Difference-in-Difference regressions. The analysis purpose is to offer insights to AQR Asset Management on the behavior of securities during market turmoil and provide actionable recommendations for investment and trading strategies.

## Key Findings[[1]](#footnote-1):

* Without regard to any factors including market premium, volatility, dollar volume, and Robust Minus Weak factor, a stock or ETF decrease in price during the COVID-19 market crash[[2]](#footnote-2).
* The market premium factor has the strongest positive correlation with monthly excess returns. Additionally, higher profitability, as captured by the profitability factor, is associated with increased monthly returns. However, the size factor shows a low negative impact on monthly excess returns, suggesting that larger companies may not necessarily offer higher returns.
* Quoting in $0.05 increments but continuing to trade at current price increment, quoting and trading in $0.05 increments and with certain exemptions allowed would increase relative spread so that reduce market liquidity; nevertheless, quoting and trading in $0.05 increments and with certain exemptions allowed, but also subject to a "trade-at" requirement and block size order exemptions, would decrease on relative spread, so that increase market liquidity.

**Task 1**

To characterize on the effect of the COVID-19 pandemic on US securities’ performance between February 14, 2020, and March 20, 2020, these variables were chosen: Excess Return (ex\_ret), Volatility, Market Premium, Market Capitalization (MktCap), Sharpe Ratio, and Relative Spread (RelSpread). The following time series and summary statistics of these variables present the market performance and behaviour during the period from different aspects.

*** Figure 1 Figure 2***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **mean** | **std** | **min** | **1%** | **25%** | **50%** | **75%** | **99%** | **max** | **var** |
| **ex\_ret** | -0.02 | 0.08 | -0.82 | -0.24 | -0.05 | -0.02 | 0.00 | 0.19 | 5.84 | 0.01 |
| **Volatility** | 0.04 | 0.04 | 0.00 | 0.00 | 0.01 | 0.02 | 0.05 | 0.19 | 0.75 | 0.00 |
| **MktCap** | 5,078,008.61 | 30,904,935.86 | 328.18 | 1,998.00 | 66,003.04 | 323,874.00 | 1,791,247.92 | 83,974,852.26 | 1,424,460,482.16 | 955,115,060,571,953.62 |
| **Market\_Premium** | -0.02 | 0.05 | -0.13 | -0.13 | -0.05 | -0.02 | -0.00 | 0.09 | 0.09 | 0.00 |

***Table 1: Summary statistics for ex\_ret, Volatility, MktCap, and Market\_Premium (February 14th, 2020 to March 20th, 2020)***

**Time Series 1**

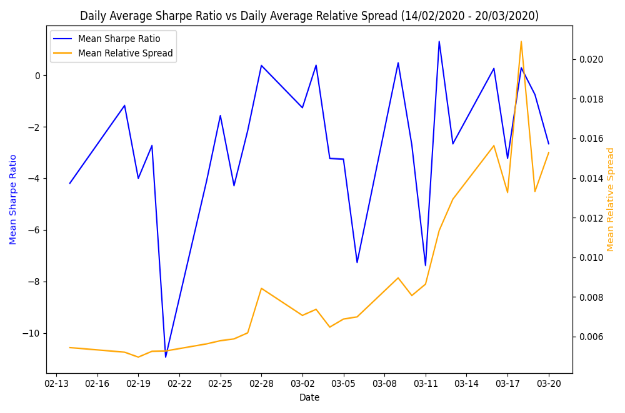
The increasing trend of volatility correlates to massive fluctuations of excess return.

* **Excess Return (ex\_ret):** This variable refers to the return on a security or portfolio over and above the risk-free rate. It is a key measure of performance and is crucial for evaluating the impact of the COVID-19 pandemic on securities' returns during the specified period. As Figure 1 indicates above, the average excess return mostly remains negative throughout the period, reflecting the overall downturn in the market. In addition, Table 1 shows the mean of -0.02 with a standard deviation of 0.08 indicates a general decline in returns with substantial variability, provide statistical evidence to the excess return trend displaying on Figure 1.
* **Volatility:** This measures the degree of variation in a security's returns. During the selected periods, volatility is sensitive, thus it is selected to present the changing market reaction. It can be found that, volatility increases noticeably as the market crash progresses, peaking around mid-March; moreover, min of 0.04, a standard deviation of 0.04, and maximum value of 0.75 (Table 1), indicates increased risk and uncertainty in the market during this period.

**Time Series 2**

The falling trend of market capitalization was sticky to the high volatility of market premium.

* **Market Premium:** refers to the excess return of the market portfolio over the risk-free rate. It is a measure of the overall market performance. Analysing the market premium helps assess the broader market conditions and how securities' returns compare to the overall market during the specified period. Figure 2 presents that, the market premium shows a declining (demonstrated by the mean of -0.13) and highly volatile trend over the selected period, which aligns with the negative sentiment and declining stock prices during the crash.
* **Market Capitalization (MktCap):** This is a company's total market value, and can be used to indicate a company's size. The high standard deviation indicates obvious differences in the market capitalization of different securities. Examining market capitalization provides insights into how different categories of securities were affected by the COVID-19 pandemic during the specified period. During the selected period, the average market capitalization shows a significant downward trend, reflecting the loss in value of securities over the period; and these sharp declines correspond with the periods of increased volatility.

 ***Figure 3***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **mean** | **std** | **min** | **1%** | **25%** | **50%** | **75%** | **99%** | **max** | **var** |
| **Sharpe\_ratio** | -2.66 | 54.76 | -11,289.30 | -23.62 | -2.79 | -1.11 | 0.14 | 4.43 | 6,833.91 | 2,998.28 |
| **RelSpread** | 0.01 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.11 | 1.98 | 0.00 |

***Table 2: Summary statistics for Sharpe Ratio and Relative Spread (February 14th, 2020 to March 20th, 2020)***

**Time Series 3**

The extremely volatile Sharpe ratio might have correlation with the reduced liquidity caused by relative spread.

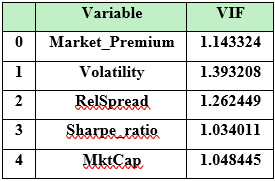
**Sharpe Ratio**: The Sharpe ratio measures the risk-adjusted return of a security or portfolio. It is calculated as the excess return divided by the volatility. A high Sharpe ratio means good risk-adjusted performance. Analyzing the Sharpe ratio helps assess how effectively securities managed risk and generated returns during the specified period. From Figure 3, the sharpe ratio shows significant volatility with values fluctuating drastically; furthermore, mean of -2.66 with an extremely high standard deviation of 54.76 (Table 2), reflects the extreme variability in risk-adjusted returns. These indicate high variability in risk-adjusted returns during the period, reflecting increased market uncertainty and instability due to the COVID-19 pandemic.

**Relative Spread**: The relative spread is a measure of liquidity, representing the difference between the bid and ask prices of a security relative to its price. A higher relative spread indicates lower liquidity. Studying the relative spread provides insights into the liquidity conditions of securities during the specified period, which can be influenced by market uncertainties like the COVID-19 pandemic. In aligning with the general increasing trend of Relative Spread shown on the plot above, reflecting a widening bid-ask spreads, and suggesting increased trading costs and lower liquidity, and this is common in periods of market distress; furthermore, mean of 0.01 with a standard deviation of 0.03 suggests increased trading costs and reduced liquidity.

**Task 2**

This part aims to explore the factors influencing security returns between February 14, 2020, and March 20, 2020. During this period, significant market volatility occurred due to the outbreak of the COVID-19 pandemic. Therefore, OLS regression is conducted for predicting the future excess returns (ex\_RET\_t+1) for each security, and various variables are selected to explain these returns[[3]](#footnote-3); additionally, by splitting the dataset into training and testing samples (8:2), the following outputs evaluated the accuracy of the regression models: model evaluation metrics (MSE, RMSE, MAE), and an assessment of which factors are most important for explaining the price changes.

**OLS Regression Model 1[[4]](#footnote-4)**

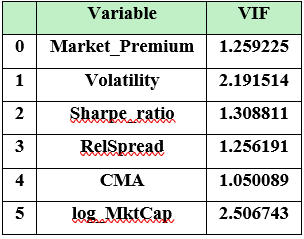
***A table with numbers and letters

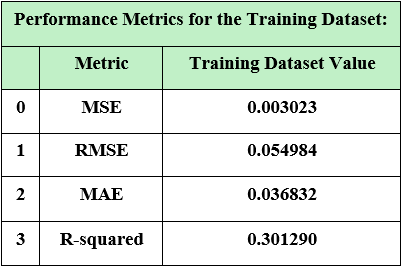
Description automatically generatedA table with numbers and letters

Description automatically generatedFigure 4: Summary Output for OLS Regression Model 1***

This model includes Market Premium, Volatility, Relative Spread (RelSpread), Sharpe Ratio, and Market Capitalization (MktCap) as independent variables; according to R-squared shown on Figure 4 left regression output table, these variables explain 29.1% of variance of future excess return. By comparing the model evaluation metrics (MSE, RMSE, MAE) between the testing and training datasets, it can be concluded that this model is reliable for prediction. This conclusion is supported by the fact that these metrics, including the R-squared values, are similar across the model's performance on both datasets. Furthermore, VIF table on the right side of Figure 4 indicates multicollinearity is not a major issue in the regression model, as each value is less than 10. However, the coefficient of RelSpread is not significant since its p-value is less than 5%, plus the coefficient of MktCap is very small, therefore, this model is not perfect enough, the new model is needed.

**OLS Regression Model 2[[5]](#footnote-5)**

***Figure 5: Summary Output for OLS Regression Model 2***

A table with numbers and letters

Description automatically generated

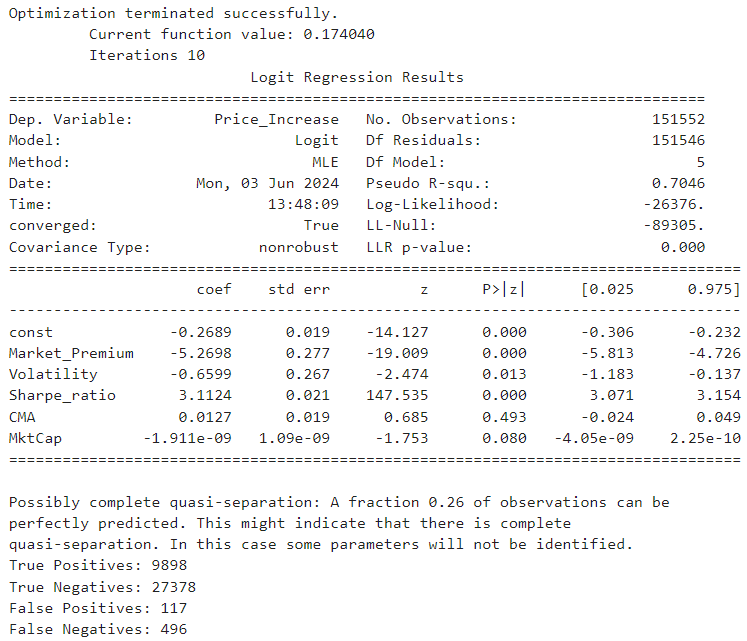
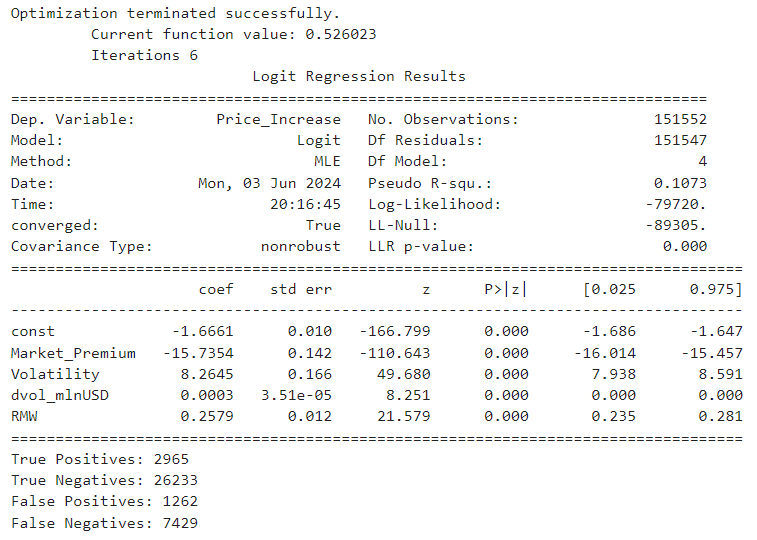
In comparing with OLS Regression Model 1, this model added Conservative Minus Aggressive (CMA)[[6]](#footnote-6) and transformed Market Capitalization into log form (log\_MktCap). According to the output shown on Figure 5, R-squared has increased to 0.301, and the coefficient of every variable is significant at 5% confidence level; also, multicollinearity is not a major issue in this model. Therefore, to interpret this outcome, holding everything constant:

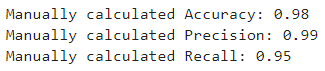
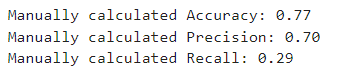
* Market\_Premium: one unit increase in market premium would decrease the future excess return by 0.4690 on average, indicating the negative effect of market condition during COVID-19 market crash.
* Volatility: one unit increase in volatility would decrease the future excess return by 0.2890 on average
* Sharpe\_ratio: one unit increase in Sharpe ratio would increase the future excess return by 0.0071 on average
* RelSpread: one unit increase in relative spread would decrease the future excess return by 0.1044 on average
* CMA: one unit increase in CMA would decrease the future excess return by 0.0083 on average. In the context of the Fama-French three-factor model, a negative coefficient for CMA indicates that, on average, conservative stocks (with lower investment-to-assets ratios) tend to outperform aggressive stocks (with higher investment-to-assets ratios) in terms of future excess returns. This relationship is consistent with the idea that conservative stocks are less risky and may offer more stable returns compared to aggressive stocks.
* log\_MktCap: one percent increase in market capitalization would decrease the future excess return by 0.0023 on average

**Task 3**

This analysis employs a logistic regression (logit) model to predict the probability of a security increase in price between February 14, 2020, and March 20, 2020. The dataset includes various variables to explain this price change (Price change is defined by whether future excess return (ex\_RET\_t+1) is > 0 or < 0). By splitting the dataset into training and testing samples (8:2), we evaluate the model's accuracy and recall. The key outputs include regression coefficients and model evaluation metrics. This analysis aims to identify the factors that characterize securities that increased in price during the specified period.

**The First Logit Model The Second Logit Model**





Both logit models applied the same split ratio of 8:2 for training and testing dataset, and the threshold is 0.5. Through comparing, the first logit model has issues for CMA is not significant at 5% confidence level; moreover, the model evaluation metrics look not normal, while the regression outcome report other issues. Therefore, the second logit model swapped Sharpe ratio, CMA, MktCap by dvol\_minUSD and RMW.

According to the output of the second logit model, it can be defined that the coefficients of all variables are statistically significant, and each coefficient can be interpreted as follows:

* Market Premium has a strong negative correlation to the future price increase
* Volatility has a strong positive correlation to the future price increase
* Dollar volume (dvol\_mlnUSD) has an extremely low positive correlation to the future price increase
* RMW has a slight positive correlation to the future price increase

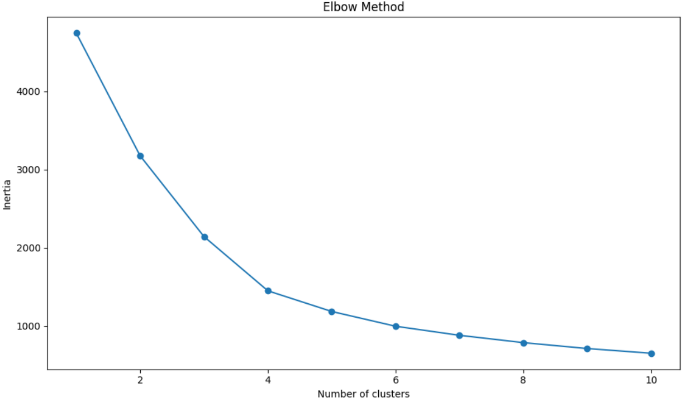
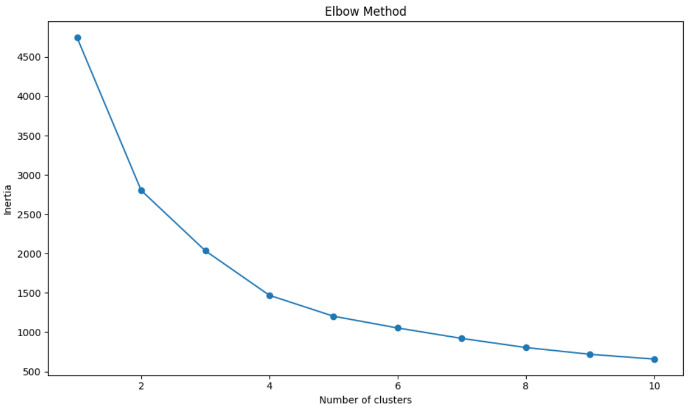
On the other hand:

* Accuracy of 0.77 means that 77% of predictions were correct, indicating a relatively good overall performance.
* Precision of 0.70 means that out of all instances predicted as positive (price increase), 70% were actually positive. This suggests a moderate level of false positives.
* Recall of 0.29 means that out of all actual positive instances, the model correctly identified 29% of them. This indicates a relatively low sensitivity or high rate of false negatives, suggesting that the model may be missing many actual positive cases.

**Task 4**

This analysis will combine valuation metrics such as the Enterprise Value Multiple (evm) and Net Profit Margin (npm) to provide a comprehensive K-means clustering analysis. The goal is to discern patterns indicating which securities were potentially over- or under-valued as of January 31, 2020, and to assess the temporal stability of these clusters by extending the analysis to June 30, 2020.

In this study, the optimal number of clusters is determined by using the elbow method, and also detailed cluster characteristics are provided, including average values of evm and npm in the form of standardization within each cluster. Additionally, the visualization the distribution of these metrics across clusters, offered insights into market valuation trends and investment implications for the specified dates. Finally, this analysis aims to support informed investment decisions by identifying valuation trends and their stability over time.

 ***Figure 6 (January 31st, 2020) Figure 7 (June 30th, 2020)***

In the study, by using the elbow method, both elbow plots above indicate that the optimal number of clusters is 4 for the clustering of each period.

A graph of blue and orange bars

Description automatically generatedA graph with blue and orange squares

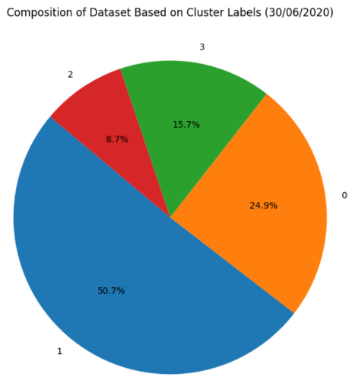
Description automatically generated ***Figure 8 (January 31st, 2020) Figure 9 (June 30th, 2020)***

*Notes: The y-axis values of both mean evm and mean npm shown on Figures 8 & 9 are standardized values*

The characteristics of Clusters are as follows:

* Cluster 0: securities with low evm and high npm
* Cluster 1: securities with both evm and npm are low
* Cluster 2: securities with both evm and npm are extremely low
* Cluster 3: securities with high evm and low npm

By comparing both Cluster 0, 1, 2 and 3 on Figure 8 and Figure 9, it can be found that the changes on clusters' characteristics on evm and npm are generally not obvious, but Cluster 2’s mean evm has improved a bit.

***A pie chart with numbers and a few different colored circles

Description automatically generated with medium confidence Figure 10 Figure 11***

|  |  |
| --- | --- |
| **Cluster** | **Count** |
| 0 | 605 |
| 1 | 1249 |
| 2 | 138 |
| 3 | 382 |

|  |  |
| --- | --- |
| **Cluster** | **Count** |
| 0 | 591 |
| 1 | 1202 |
| 2 | 207 |
| 3 | 373 |

Both Figure 5 and Figure 6 report the sizes of each cluster in different periods, in terms of percentage and count. Notably, Cluster 1 comprises more than 50% of the securities in both periods, indicating that around half of the securities maintained low EVM and NPM from January 31, 2020, to June 30, 2020.

A diagram of different colored circles

Description automatically generated

***A diagram of different colored dots

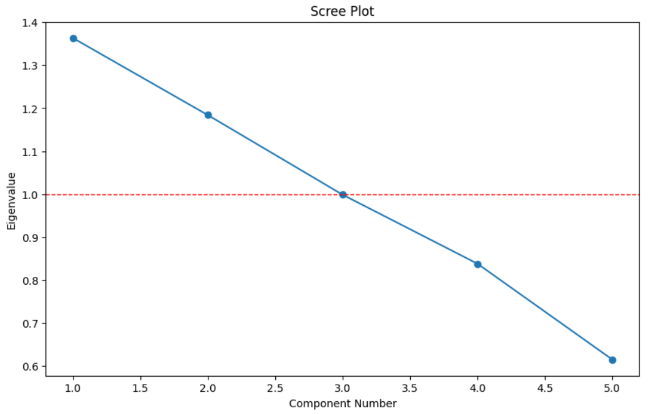
Description automatically generated Figure 12 Figure 13***

**Interpretation and Analysis**

* **Valuation Insights:** According to the previous information noticed, and by looking at both Figure 8 and Figure 9 displayed above, securities in Cluster 0 are considered undervalued, because low evm and high npm, and suggests that these securities that are priced lower relative to their earnings and have high profitability; securities in Cluster 3 are considered overvalued, because high evm and low or negative npm, and suggests that these securities are priced higher relative to their earnings and have low profitability.
* **Temporal Stability:** Based on the information extracted by comparing Figure 10 and Figure 11, it can be concluded that Cluster 0, the group of undervalued securities, are relatively stable in comparing with securities in other clusters. This suggests consistent market recognition of these securities' values. On the other hand, notably, the size of Cluster 2, has increased significantly, representing more securities suffered a downward shift on evm and npm, that might be caused by market fluctuations between January 31, 2020, to June 30, 2020. Moreover, the reduced size of Cluster 3 might prove the previous valuation insights, as these securities are overvalued, the market adjusted the mismatch in value, thus some of them moved to Cluster 2 in June 30, 2020; also, same reason can be applied to the size reduction of Cluster 1.
* **Investment Implications:** According to the previous analysis, the optimal suggestion is to invest securities in Cluster 0 on both January 31, 2020 and June 30, 2020, this is because the characteristic of Cluster 0 remained unchanged.

**Task 5**

The object of this analysis derived a comprehensive set of market-wide and company-specific factors into core principal components using Principal Component Analysis (PCA). This technique helped to identify the most significant factors influencing monthly excess returns. By using variables like Market Capitalization (MktCap), Shillers Cyclically Adjusted P/E Ratio (CAPEI), Net Profit Margin (npm), Return on Asset (roa), Enterprise Value Multiple (evm) the principal components are extracted and interpreted; in addition, Market Premium (Mkt-RF), as a factor extracted from Fama French Factors, added into the final principal component regression model[[7]](#footnote-7). The following sections present a detailed summary of the principal components, their factor loadings, interpretations, and implications for monthly excess returns.

****** ***Figure 14 Table 3: Eigenvalue Table***

Usually, a factor having an eigenvalue higher than 1, can be considered as a principal component; however, in this case, the outcomes in the scree plot and the Eigenvalues Table shown above, suggest that 4 factors should be selected as principal components, for they collectively explain around 90% of the data variance.

***Table 4: Factor Loading Table***

*Notes: The factor pattern table presents the loadings of each variable on every single factor. Both CAPEI and evm have a relatively high loading on Factor 1, so this Factor might represents overall market valuation; and Factor 2 might be the combination of size and profitability factors, for market capitalization and npm has high loading respectively. Factor 3 is considered as profitability factor, for roa having a strong positive impact and npm having a strong negative impact.*

Mainly about Factor 4, which is similar to Factor 2, it suggests another dimension of size and profitability, but This represents a more complex relationship. Although large companies with high market capitalization increase the factor's value, high profitability and high CAPEI decrease the factor's value. This may indicate that in certain situations, while large companies have significant size, if they have high profitability or high valuations, the factor's value decreases instead. This reflects a negative correlation between company size and profitability or the suppressive effect of valuation on the factor.

**Principal Component Regression**

***Table 5: Principal Component Regression Output[[8]](#footnote-8)***

With the factors identified, the regression model is ran based on the equation above. Each PC here in the variable column of Table 5 refers to principal component shown on the equation in order, and market premium is added into the model. Taking PC3 as an example, this is Factor 3, or profitability factor, the coefficient is 0.0067, which means that, holding everything constant, one unit increase in profitability factor, would increase the monthly excess return by 0.0067, on average. All variables are significant, except for PC4. In addition, the coefficient of market premium is highest, this indicates that it might have a more effective and positive correlation with monthly excess return, in comparing with other variables. On the other hand, PC1 and PC2 (size & profitability), have a low negative impact on monthly excess return, but when it comes to PC3, the profitability factor, it has positive impact. This information suggests that size and profitability may not jointly contribute to the increase in monthly excess return, and reveals that the size factor along may have negative impact.

**Task 6**

The Securities & Exchange Commission (SEC) launched the Tick Size Pilot Program on October 3, 2016, after approving the National Market System (NMS) Plan. The program's objective was to assess the effects of widening tick sizes on securities of smaller capitalization companies. This pilot program include a control group and three test groups, for securities in the control group were quoted and traded at their current tick size increment, whereas those in other groups are distinguished by treatment settings.

Therefore, this analysis will mainly employ Difference-in-Differences regressions to investigate how the securities in test groups were affected in terms of **relative spread**, in order to provide an insight on adjusting AQR trading strategy to minimize the transaction costs.

Furthermore, relative spread is calculated as the bid-ask spread divided by the mid-price, offers a normalized measure of spread that adjusts for price levels. This metric facilitates comparisons across securities, particularly useful for assessing liquidity and trading costs, as it identifies securities with relatively tighter or wider spreads independent of their price levels. Additionally, relative spread is less affected by absolute price changes, ensuring a more stable measure for spread analysis.

**The First Test Group**

H0: Quoting in $0.05 increments but continuing to trade at current price increment has no impact on relative spread.

H1: Quoting in $0.05 increments but continuing to trade at current price increment has an impact on relative spread.

***Table 6: Regression Model 1****[[9]](#footnote-9)*

By mainly looking at the highlighted part of the table above, we found that the coefficient for the intersection term d\_after:d\_G1 is 0.0019, and it is significant at 1% confidence level; thus, we have to reject the null hypothesis (H0). This information suggests that, holding everything constant, after October 3rd, 2016, securities quoted in $0.05 increments would increase their relative spread by 0.0019, on average.

**The Second Test Group**

H0: Quoting and trading in $0.05 increments and with certain exemptions allowed[[10]](#footnote-10) has no impact on relative spread.

H1: Quoting and trading in $0.05 increments and with certain exemptions allowed has an impact on relative spread.

***Table 7: Regression Model 2****[[11]](#footnote-11)*

By mainly looking at the highlighted part of the table above, we found that the coefficient for the intersection term d\_after:d\_G2 is 0.0433, and it is significant at 1% confidence level; thus, we have to reject the null hypothesis (H0). This information suggests that, holding everything constant, after October 3rd, 2016, securities quoted and traded in $0.05 increments and with certain exemptions allowed, would increase their relative spread by 0.0433, on average.

**The Third Test Group**

H0: Quoting and trading in $0.05 increments and with certain exemptions allowed, but also subject to a “trade-at” requirement and block size order exemptions has no impact on relative spread.

H1: Quoting and trading in $0.05 increments and with certain exemptions allowed, but also subject to a “trade-at” requirement and block size order exemptions has an impact on relative spread.

***Table 8: Regression Model 3****[[12]](#footnote-12)*

By mainly looking at the highlighted part of the table above, we found that the coefficient for the intersection term d\_after:d\_G3 is -0.0041, and it is significant at 1% confidence level; thus, we have to reject the null hypothesis (H0). This information suggests that, holding everything constant, after October 3rd, 2016, securities quoted and traded in $0.05 increments and with certain exemptions allowed, but also subject to a "trade-at" requirement and block size order exemptions, would decrease their relative spread by 0.0041, on average.

**Conclusion**

In conclusion, based on the results, AQR should adjust their trading strategies to minimize transaction costs. For control group stocks unaffected by new quoting or trading rules, AQR should continue trading as before. However, for the first test group AQR may consider reducing bid-ask spreads to mitigate the potential increase in relative spread and reduce costs. With the second test group, AQR should adjust strategies to utilize exemptions, reducing spreads to minimize costs. For the third test group, AQR could consider widening bid-ask spreads to potentially reduce relative spreads and lower transaction costs.

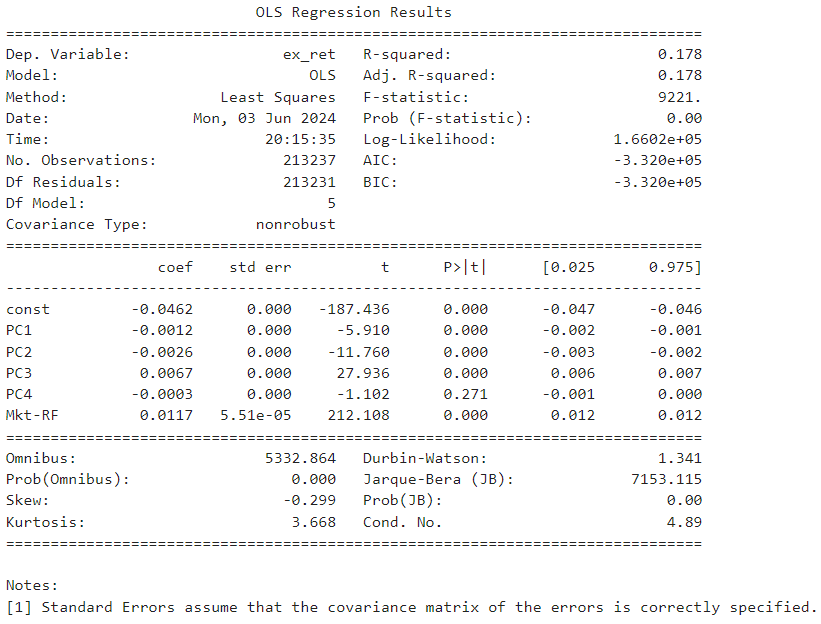
# Appendix

### Note 1

### A screenshot of a computer Description automatically generatedNote 2

### Note 3

### Note 4



A white background with black text

Description automatically generated

### Note 5



### Note 6



### Note 7



1. Answers to Task 7 [↑](#footnote-ref-1)
2. According to the output from the logit regression analysis, [↑](#footnote-ref-2)
3. Excess Return is lagged by 1 day, for predicting future values. [↑](#footnote-ref-3)
4. For detailed information, see [Note 1](#_Note_1) in [Appendix](#_Appendix) [↑](#footnote-ref-4)
5. [Note 2](#_Note_2) [↑](#footnote-ref-5)
6. From Fama French Factors [↑](#footnote-ref-6)
7. The previous PCA found that market premium is dominating this component for it has an extremely high loading on a factor thus it should be excluded from PCA, and it can be included directly in the regression model. For detailed information, please see [Note 3](#_Note_3). [↑](#footnote-ref-7)
8. [Note 4](#_Note_4) [↑](#footnote-ref-8)
9. [Note 5](#_Note_5). [↑](#footnote-ref-9)
10. e.g., midpoint executions, retail investor executions, and negotiated trades. [↑](#footnote-ref-10)
11. [Note 6](#_Note_6). [↑](#footnote-ref-11)
12. [Note 7](#_Note_7). [↑](#footnote-ref-12)