RESIDUAL ANALYSIS OF COVID-19'S IMPACT ON RETURNS OF RESTAURANT INDUSTRY STOCKS WHEN GROUPED BY MARKET CAPITALIZATION

Aaron Yuan, Ben He, Chujie Zhang, Jun Wei Foo, & Yujia Wang

CAPSTONE PROJECT – TEAM 20

Abstract

Since its discovery in 1992 by Fama and French, size in the form of market capitalization has been widely accepted as a factor in determining stock performance.[i] However, the disruptive and unprecedented nature of the COVID-19 pandemic led us to question whether stock performance in the restaurant industry varies significantly due to market capitalization. If size influenced the magnitude of the initial market shock and the speed of recovery in this study, we then sought to explain these differences with a model. To determine the impact of size, we estimated a Fama-French 3-Factor Model (FF3FM) for each restaurant stock before analyzing the residuals during the COVID-19 landscape. If the residuals varied based on market capitalization, we would perform an auxiliary regression on the residuals, using fundamental ratios as potential explanatory variables. Through our residual analysis, we found that size does play a significant role due to increased variance in residuals for small-cap firms in comparison to their mid-cap and large-cap counterparts. Furthermore, we built a statistically significant regression model to explain the residuals using certain fundamental ratios, which were return on assets (ROA), debt to assets (DTA), receivables turnover, quick ratio, and cash ratio.

Introduction

The COVID-19 pandemic led to one of the sharpest declines in U.S. stock market history, which was a 34% reduction in market value within a month span.[ii] While the virus was first announced in Wuhan province of China in early January of 2020, the markets did not react to the news of the COVID-19 until late February of 2020. The negative market reaction as well as the unknown dangers of the virus led to many government policies with widespread economic impact, such as shutdowns, limited occupancy rates, and mask mandates. These adverse effects ultimately triggered a recession within the U.S. economy with job loss and GDP shrinkage.[iii]

For the purpose of our study, we wanted to examine an industry that dealt with the ramifications of both the economic downturn and the government response to COVID-19. Given that we treated the COVID-19 environment as a new economic landscape, it would not make sense to consider marginally affected industries. Of the sectors most affected by the virus, we believe that the restaurants industry would provide us with the best insights.[iv] First of all, the abrupt nature of public policy forced restaurant specific firms and holdings to quickly adapt while hindering revenue streams due to limited occupancy or mandatory business shutdowns. Likewise, public sentiment around COVID-19 greatly impacted demand for restaurant firms due to the possibility of consumers contracting the virus from interacting with other individuals. More importantly, from publicly available information, restaurant industry stocks had the most diverse balance of companies within each size classification. This, in contrast, was not the case for other industries, such as the airline industry.

In addition, we wanted to examine more than stock return data when assessing the impact of COVID-19 across different market capitalizations. Given the severe economic complications caused by the virus, we included fundamental ratios from financial balance sheet data in our analysis as another benchmark to relative performance during this economic environment. More importantly, we believe such information may have explanatory capability if disparities exist between different size classifications and stock return performance. For example, we hypothesize that large-cap firms can take on more debt without risk than small-cap firms, allowing these large-cap firms more flexibility in raising capital without over-levering the company. Similarly, we imagine that large-cap firms have more cash on hand as established businesses, unlike their small-cap counterparts. Again, this would allow large-cap firms navigate through the COVID-19 economic environment more successfully than small-cap firms.

Data

Given the scope of our question, we extracted our data from two sources: the Wharton Research Database Services (WRDS) and the Dr. Kenneth French's website. The data from WRDS comprised of stock return data from the Center of Research on Stock Prices (CRSP) and quarterly financial balance sheet and income statement ratios from S&P Global Intelligence (Compustat).

The CRSP data had monthly price, volume, and returns between January 2015 and December 2020. It included 51 restaurant stocks that publicly traded on U.S. stock exchanges as of April 2021, referenced by ticker and PERMNO (a unique identifier for CRSP).[v] Furthermore, we classified each ticker by size relative to its total market capitalization at the end of December 2019, before the COVID-19 pandemic. We classified them based on the finance industry standard which classifies companies with capitalization over $10 billion as large cap, companies with capitalization between $2 and $10 billion as mid cap, and companies with less than $2 billion capitalization as small cap. Based on this classification, 30 firms were small-cap, 12 firms were mid-cap, and the remaining 9 firms were large-cap. Given that not some companies began trading on the market after January 2015, this dataset had 3,235 observations.

During the data-cleaning process, there was an occasional observation removed due to unavailable price, returns, and volume data for a specific month, usually the month of inception for that traded ticker. In addition, we found two small-cap companies that started to trade publicly during the COVID-19 pandemic. Given that we needed observations before January 2020 to generate residual values from a fitted FF3FM, these two firms were excluded from our analysis. Finally, we primarily focused on the ticker, date, and returns columns provided in the dataset for our calculations and analysis.

Similar to the CRSP data, the Compustat data had fundamental ratios between January 2015 and December 2020, albeit on a quarterly schedule.[vi] Again, it included the same 51 publicly traded restaurant stocks as of April 2021, referenced this time by ticker, cusip, gvkey (a unique identifier for Compustat), and PERMNO. Likewise, each ticker was classified by size, leading to the same 30/12/9 split between small-cap, mid-cap, and large-cap firms. As for the fundamental ratios, there were a total of 69 different metrics, based off information on each firm's 10-Q (quarterly financial statement document). Finally, this dataset provided three dates: the date of the 10-K (annual financial statement document) that coincided with the 10-Q information, the date of the 10-Q, and the date of publication from Compustat. In total, there were 1,562 observations given that not every ticker publicly traded before January 2015.

Unlike the CRSP data, the Compustat data required more data cleansing. While many columns had unavailable data for certain metrics, we only removed observations if no data existed for all 69 different metrics. As mentioned before, we removed the two companies that did not have observations before January 2020. However, much of our work revolved around determining which fundamental ratios to consider for our analysis. Given the redundancy in some of the fundamental ratios, we prioritized the following five distinct categories of ratios: liquidity, solvency, coverage, efficiency, and profitability.

Liquidity ratios determine a company's ability to pay off current debt. In the dataset, we focused on the current ratio, quick ratio, and cash ratio as our liquidity metrics. Solvency ratios, on the other hand, determine a company's ability to pay off long-term debt. We used the interest coverage ratio, DTA, and debt to equity ratio (D/E) as our solvency metrics. Coverage ratios are like liquidity and solvency ratios, and we looked at after-tax interest coverage as the standard. Efficiency ratios demonstrate a company's ability to turn assets into income. Ratios of this nature in the dataset included inventory turnover, payables turnover, receivables turnover, and asset turnover. Finally, profitability ratios look at a company's profit compared to other balance sheet or income statement items. For this category, we chose gross profit margin (GPM), net profit margin (NPM), ROA, return on equity (ROE), and return on capital employed (ROCE). This effectively helped reduce our scope of focus into 16 key metrics out of the original 69.

Despite our reduced focus into these following metrics, we still noticed some observations with unavailable data in some of these companies. To combat this issue, we chose to keep the dataset as is and run two different models. One model would try to explain the future residuals based off both the available and unavailable fundamental ratios. The other model would only regress with the necessary complete data for all its key variables. In addition to the 16 previously mentioned financial statement metrics, we also needed the ticker and quarter date variables as unique identifiers for the dataset.

Finally, the Kenneth French data had factor estimates for Rm-Rf (market risk premium), SMB (size factor), and HML (value factor) on a monthly basis between 1926 and 2021.[vii] In addition, Dr. French provided the monthly estimate for the risk-free rate at the time. In addition, the original dataset provided its numbers in percentage form rather than decimal form. Although the data starts 1926, we reduced the data by extracting monthly observations between January 2015 and December 2020. This led to a total of 72 observations, which will be used for our FF3FM estimation.

Methodology

Initial Exploratory Data Analysis

Our analysis first began with original exploratory data analysis (EDA) of the fundamental ratios relative to its size classification. To accomplish this, we would use the Compustat dataset and visualize the different fundamental metrics through Tableau. Information would be conveyed on a year-to-year basis while being aggregated within each market capitalization category. The purpose of this exercise was to identify whether different trends existed by market capitalization class in a thematic way. If small-cap, mid-cap, and large-cap companies had similar movements, on average, in both direction and magnitude, this would create doubt that size impacts the behavior of restaurant stocks in a COVID-19 economic environment.

Fama-French Three-Factor Model

After these visualizations, we would estimate a FF3FM for each restaurant stock. This would require us to break down CRSP dataset into 49 separate data frames, each representing a separate restaurant industry stock. Once complete, these individual stock files would be merged with the French dataset. During this merge, we would keep all elements of the French dataset but only the returns, ticker, and size columns, the latter two being unique identifiers for later exploration. This would leave us with a dataset of monthly observations between January 2015 and December 2020. Given that we want a pre-fitted model to estimate the returns and residuals during COVID-19, we chose to estimate the FF3FM based on data from January 2015 and December 2019. This segmentation of data would allow us to see effects of the COVID-19 economic environment in the residuals when running the fitted model on 2020 monthly observations. This would not be possible if we used the whole dataset to fit the model as COVID-19 would be already accounted for in such model.

Given our new dataset between January 2015 and December 2019, we needed to reorder our variables so that the specific beta coefficients for each stock could be easily calculated in a linear regression. The following formula represents the FF3FM:

$$r_i = r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \varepsilon$$

Because the risk-free rate is a constant that never changes, we calculated the excess monthly returns for each restaurant stock by subtracting the returns by the risk-free rate. By doing so, we could run a fitted regression model to solve for the unknown beta coefficients in R. This calculation led to this formula:

$$r_i - r_f = \alpha + \beta_1\left(r_m - r_f\right) + \beta_2(SMB) + \beta_3(HML) + \varepsilon$$

With this slight alteration, now we can run a linear regression where the outputs match coefficients. For example, the intercept would represent alpha, the regressor coefficients would represent the betas relative to its order, and the residuals would represent the error term. Once we calculate a fitted regression for each stock, we would then take the given alpha and beta coefficients and predict each monthly stock return in 2020. This would require us retrieving the merged dataset that only included 2020 monthly observations for each stock. After calculating the predicted results in R, we would compare these results to the actual excess returns. The difference of two would equal the residuals, or error term, as seen by the simplification of the FF3FM below:

$$Actual\ Excess\ Returns = Predicted\ Returns + \varepsilon$$

Residual Exploratory Data Analysis

Continuing from our FF3FM estimation, we would perform more EDA on the residuals of the fitted regressions to see the potential impact of size. The key areas of focus would be comparing the fitted residuals by the predicted residuals, the residuals by size over time, and the residuals by size based on fundamental ratios. To achieve this, we would need to retrieve both sets of residual values from the FF3FM estimation. In addition, we would need to aggregate each set of residuals based on market capitalization categorization. Finally, we would convert the monthly predicted residual values into a quarterly basis so that we could merge this information with the Compustat dataset, allowing for comparison with residuals and fundamental ratios. The data manipulation would occur in R while the data visualization would require Tableau.

The reason behind this EDA is that if residuals vastly differ between the fitted residuals and the predicted residuals, this justifies that COVID-19 had tangibly measured economic impact on the restaurant industry as a whole. Furthermore, if residuals are not consistent between size, there are variables not present in the FF3FM that cause significant differences between size classification. Finally, if the behavior of the residuals within different market capitalization categorizations differs both in direction and magnitude over time, this would confirm that the size of a company in the restaurant industry has disproportional impact.

Auxiliary Regression of Predicted Residuals

Finally, after our analysis of the residuals from the FF3FM in a COVID-19 economic environment, we would build an auxiliary regression to explain the residuals using fundamental ratios. The linear regression would ultimately be formed like the following:

$$\varepsilon = Intercept + \beta_1(Fundamental\ Ratio_1) + \beta_2(Fundamental\ Ratio_2) + \cdots$$

This would necessitate us to retrieve the previously merged dataset with quarterly residuals and fundamental ratios of each stock during the COVID-19 economic environment. To perform the regression, we would select a stepwise regression, which iteratively adds and removes potential predictors that generates the lowest prediction error and highest R squared.[viii] Two models would be generated using the stepwise regression in R. One would feed uncleaned data with some unavailable information in the aforementioned ratios of importance. Meanwhile, the other would feed cleaned data that contains available information in the necessary ratios. Subsequently, a check for robustness would be performed on both models with a Principal Components Analysis (PCA).[ix] The purpose behind this instrument is to see whether redundancy exists within the variables of the model given the amount of variability explained by each component. Ultimately, one of the two models would be chosen as the proper auxiliary regression based on the PCA analysis and consistency with previous EDA.

Results and Findings

Initial Exploratory Data Analysis

When performing the original EDA on fundamental ratios, interesting trends appeared when comparing liquidity metrics by market capitalization classification. As seen in Figure 1 below, large-cap restaurant stocks have the largest average current ratio when compared to their mid-cap and small-cap counterparts. However, the four-year movement of the current ratio differed by size categorization. Large-cap restaurants stocks have experienced a consistent decrease in average current ratio while small-cap restaurant stocks have seen the exact opposite trend. Meanwhile, mid-cap restaurant stocks have stagnated with their average current ratio. When focusing specifically on cash equivalents in Figure 2, we noticed that the rate of decrease is less pronounced in the cash ratio than the current ratio for large-cap restaurant firms. On the contrary, both mid-cap and small-cap restaurant firms have seen an increasing trend for their respective cash ratios. Small-cap firms, though, have the sharpest slope for average cash ratio.

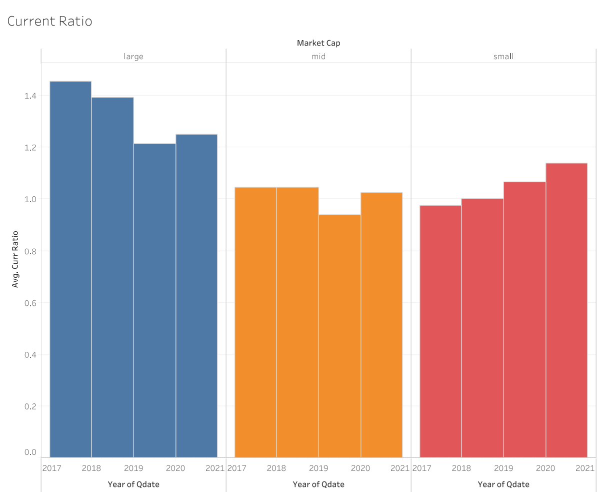

Figure 1. Current Ratio By Market Cap

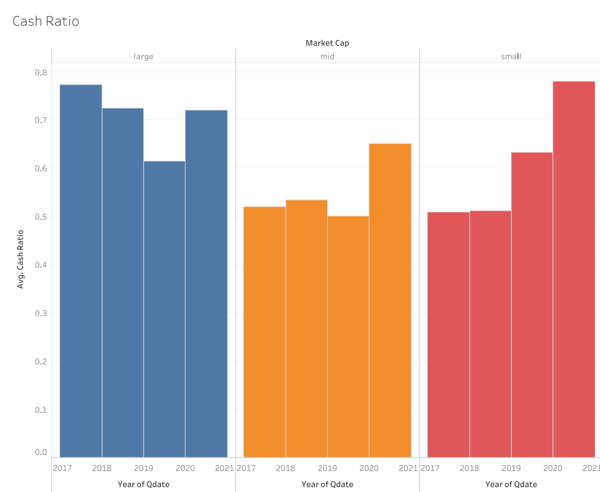

Figure 2. Cash Ratio By Market Cap

At first thought, this seemed counterintuitive as a decrease in the current ratio and cash ratio usually represents a reduced ability to generate cash and potential financial difficulty for the firm. In addition, an increase in the current ratio and cash ratio usually represents more financial stability and higher liquidity. That being said, these trends are understandable when juxtaposed with Figure 3 below, which compares average DTA by market capitalization.

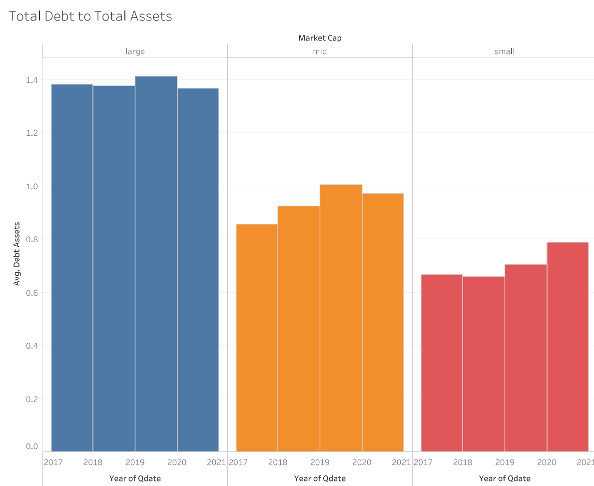

Figure 3. Debt to Assets By Market Cap

When looking at solvency ratios, we observed large-cap restaurant firms have the largest average DTA on average, followed by mid-cap restaurant firms then small-cap restaurant firms. In addition, we noticed that average DTA has remained relative stationary for large-cap restaurant stocks. On the other hand, average DTA for mid-cap and small-cap restaurant stocks have increased over time. What is important in the lenses of the COVID-19 economic environment is the fact that large-cap and mid-cap restaurants witnessed a decrease in their average DTA from 2019 to 2020. Meanwhile, small-cap stocks witnessed an increase in their average DTA during the same time frame. Given our previous EDA on the liquidity ratios, this suggests that the increases seen in the current ratio and cash ratio for small-cap restaurant stocks may come from raising debt capital as this increases current assets. Furthermore, this could imply that large-cap and mid-cap restaurant stocks were in a better financial without needing to bring on significant leverage during the COVID-19 pandemic.

In regards to the profitability of restaurant stocks, each market capitalization saw a deacrease in both GPM in Figure 4 and ROA in Figure 5. In both figures, large-cap restaurant companies have the highest GPM and ROA, followed by their mid-cap counterparts and finished by their small-cap counterparts. However, the drop in GPM due to the COVID-19 economic environemnt seem more severe for smaller restaurant firms. This is due to the fact that GPM stayed relatively the same for large-cap restaurant stocks while falling for mid-cap and small-cap restaurant stocks. Similarly, small-cap restaurant firms faced the worst impact in ROA, effectively equaling 0% in 2021, in spite of matching the same trajectory as their large-cap and mid-cap counterparts. Again, this suggests that the COVID-19 pandemic had significant and disparate impact on the restaurant industry when segregated by size classification.

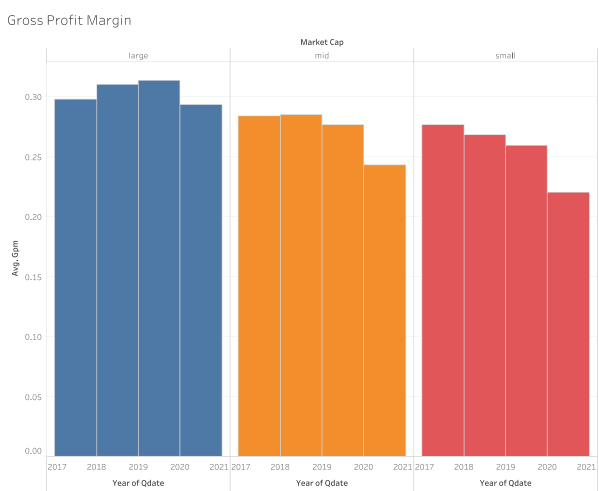

Figure 4. Gross Profit Margin By Size

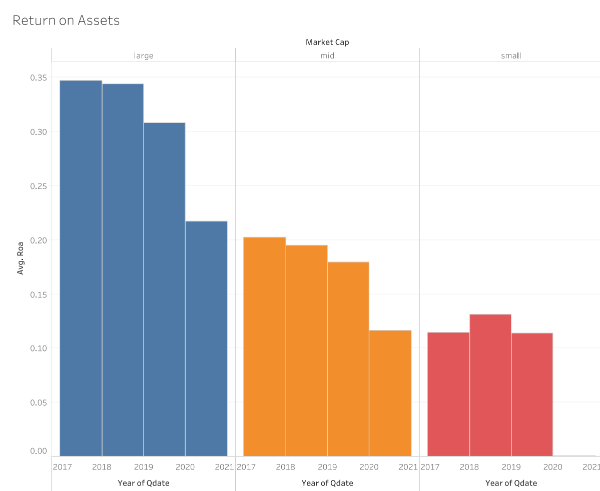

Figure 5. Return on Assets By Size

However, the most stark contrast between size classification within the restaurant industry demonstrates itself in Figure 6, which focuses on average inventory turnover. From the beginning, large-cap restaurant firms tend to have lower inventory turnover numbers than their mid-cap and small-cap counterparts. However, the differences in trends exemplify the economic impact of the

COVID-19 pandemic. While large-cap restaurant firms have a slightly decreasing trend in inventory turnover, small-cap restaurant firms have an exacerbated increasing trend in inventory turnover. Meanwhile, mid-cap restaurant firms seem to maintain their current inventory turnover. A giant inventory turnover either represents increased sales or a drop in potential inventory. Given the dramatic drops in GPM and ROA for small-cap restaurants, it is fair to assume that small-cap firms drastically cut down on their inventory, whether that be due to decreased demand or supply chain disruptions.

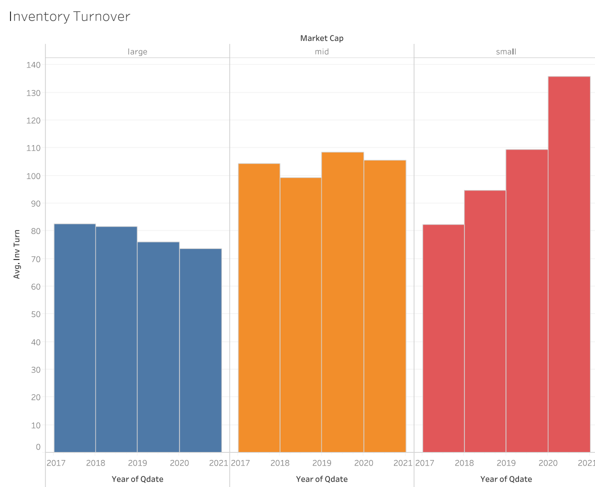

Figure 6. Inventory Turnover By Market Cap

From this basic EDA, we can confidently conclude that restaurant firms perform differently in fundamental ratios when aggregated by market capitalization classification, both intrinsically and trend-dependent. Furthermore, the visualizations of the liquidity, solvency, profitability, and efficiency ratios from Tableau confirm that COVID-19 had a significant and disparate impact on the restaurant industry, especially when grouped by size. Given these results, we can confidently move into our FF3FM estimation knowing that our average estimated model and residual metrics should differ between size classifications. Moreover, we can also trust choosing fundamental ratios as potential explanatory regressors for the auxillary regression on residuals due to inherently different qualities and characteristics among size groupings.

Fama-French Three-Factor Model and Residual Exploratory Data Analysis

After performing the steps outlined in the methodology in regards to FF3FM estimation, we found the average intercept and coefficient values by market capitalization, seen in Figure 7.

Market Cap	Intercept	Market Beta	SMB Beta	HML Beta
Small Cap	-0.005	0.530	0.545	-0.092
Mid Cap	0.009	0.618	0.324	-0.353
Large Cap	0.007	0.639	-0.110	-0.328

Figure 7. Fama French Three Factor Estimation Average Coefficients

As expected, each model projected to have an alpha close to 0, given the miniscule values for the intercept. In addition, large-cap restaurant stocks had the highest beta coefficient for market risk premium, while small-cap restaurant stocks had the highest beta coefficient for the size factor. Furthermore, all beta coefficients for the value factor were negative. However, these insights only demonstrate that stock returns act differently depending on size. When comparing the performance of fitted models residuals between 2015-2019 and during COVID-19 pandemic, we noticed a great disparity in root mean square error (RMSE).[x]

Market Cap	Fitted RMSE	Predicted RMSE
Small Cap	0.034	0.168
Mid Cap	0.043	0.076
Large Cap	0.026	0.073

Figure 8. Fama French Three Factor Fitted and Predicted RMSE

As displayed in Figure 8 above, the average RMSE for each model, aggregated by size classification, increased during the COVID-19 economic environment. However, the difference in RMSE is most prevalent between the fitted and predicted values for the FF3FM of small-cap restaurants. This is statistically important given that the RMSE is greater that the mid-cap and large-cap predicted RMSE combined even though the dataset has 30 small-cap firms compared to 9 large-cap firms and 12 mid-cap firms, respectively.

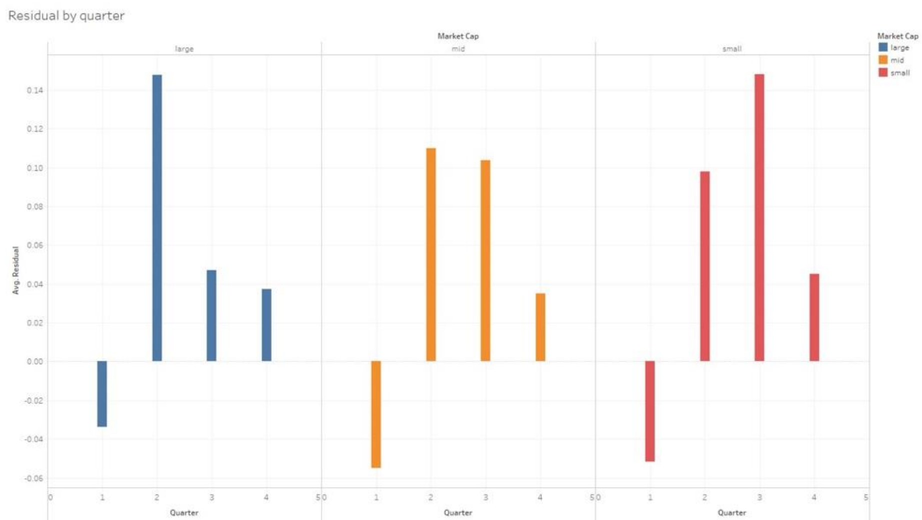

Figure 9. Predicted Quarterly Residuals By Market Cap During COVID-19

When analyzing the differences in residuals by quarter, as portrayed in Figure 9 above, we noticed significantly different trends among each size designations. The only shared characteristics between all restaurant stocks was that Q1 of 2020 overestimated returns due to negative residuals on average while Q2, Q3, and Q4 of 2020 underestimated returns due to positive residuals. For large-cap restaurant stocks, Q2 of 2020 had the largest residuals in the COVID-19 economic environment before residuals minimized significantly in Q3 and Q4 of 2020. Mid-cap restaurant stocks, on the other hand, had both Q2 and Q3 of 2020 with the greatest residuals before they minimized drastically in Q4 of 2020. Finally, small-cap restaurant stocks had very high residuals in Q2 of 2020 and peaked with the magnitude of residuals in Q3 of 2020 before dipping in Q4 of 2020. Such interpretation of these residuals could be that large-cap restaurant stocks adjusted the quickest from the initial shock caused by COVID-19, while mid-cap and small-caps struggled to return to normality until Q4 of 2020.

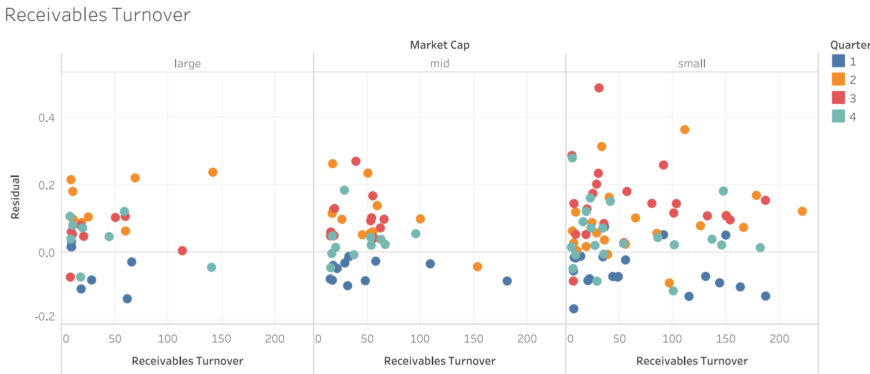

Figure 10. Quarterly Residuals By Recievables Turnover Divided By Market Cap

In addition to generic residual analysis and time-series residual analysis, we compared residuals based on the key fundamental ratios to potentially identify significant trends within the residuals. As displayed in Figure 10 above, one can see that there was greater variability among the distribution of residuals and for receivable turnover for small-cap stocks. Even when accounting for quarters, small-cap stocks show greater volatility than mid-cap and large-cap restaurant stocks. This supports the argument that the COVID-19 pandemic had more disparate impact towards small-cap restaurant stocks than their mid-cap and large-cap counterparts. Similarly to Figure 10, Figure 11 demonstrated the same trends in volatility and variability for small-cap restaurant stocks compared to its mid-cap and large-cap counterparts. All residuals for large-cap restaurant firms had a positive ROA while most mid-cap restaurant firms had a positive ROA. Small-cap restaurant firms, on the other hand, tend to lean closer to zero or negative ROA.

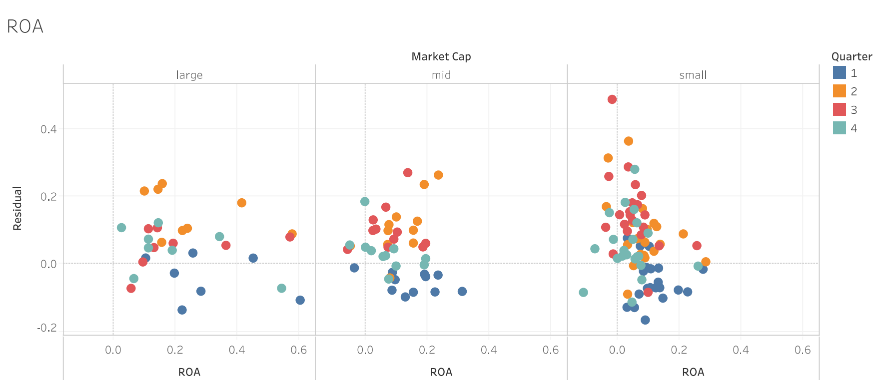

Figure 11. Quarterly Residuals By ROA Divided By Market Cap

Based on the following visualizations, we can arrive to these following conclusions. First of all, there was a significant difference in residuals between the fitted regression and the predicted regression of the FF3FM estimation, confirming the economic impact of COVID-19. Next, the interaction between the quarter of 2020 and size classification exemplified underlying aspects that contribute to a quick bounceback from large-cap restaurant firms and a slower recovery for mid-cap and small-cap restaurant firms. Finally, the volatility of fundamental ratios when grouped by market capitalization demonstrated fundamental differences between restaurant firms of distinctive sizes. Given these observations, we have confidence in building an auxiliary regression that explains the difference in residuals using fundamental ratios.

Auxiliary Regression of Predicted Residuals

Finally, we tried to model an auxiliary regression that would explain the variance of the residuals. When using slightly uncleaned data for the stepwise regression, its iterative process returns the following equation:

$$\varepsilon = 0.261 + 0.259(DTA) - 1.411(ROA) - 0.252(Quick\ Ratio) +$$
$$0.249(Cash\ Ratio) - 0.001(Receivables\ Turnover)$$

Of these five variables, the intercept, DTA, and ROA are very statistically significant while quick ratio has some statistical significance. However, the model itself has a low adjusted R squared, as only 10.57% of the variance in residuals can be explained by this model. Furthermore, the inclusion of the five variables makes intuitive sense, but the directional movement for each would need more inquiry. For example, it seems that the quick ratio and cash ratio variables cancel each other out while the receivables turnover has a negligible effect. Meanwhile, ROA is significantly negative, yet DTA is significantly positive. A potential reason why ROA is negative may be since the inverse nature of the market and ROA during the COVID-19 pandemic. The original market shock occured while ROA was high, but once the market recovered as a whole, ROA metrics deteriorated for all restaurant companies. The biggest revelation of this model is the lack of size as a relevant factor. This potentially hints at the fact that these fundamental ratios act as a proxy for size and the inherently distinct characteristics seen among restaurant stocks of different market capitalizations.

When using cleaned data for the stepwise regression, though, it generates this model:

$$\varepsilon = 0.208 - 3.882\ (Net\ Profit) + 4.682\ (Operating\ Profit\ Margin\ AD) - 6.880\ (Operating\ Profit\ Margin\ BD) + 3.989\ (Cash\ Flow\ Margin) - 1.042\ (Gross\ Profit) + 1.303\ (DTA) + 0.000\ (Debt\ to\ EBITDA) + 1.028\ (Current\ Debt) + 0.547\ (Profit\ to\ Current\ Liabilities) - 0.349\ (OCF\ to\ Current\ Liabilities) - 0.626\ (Debt\ to\ Capital) - 0.007\ (Debt\ to\ Equity) - 0.091\ (Quick\ Ratio) - 0.666\ (Asset\ Turnover) + 0.320\ (Sales\ to\ Inventory) + 1.061\ (Accruals)$$

Of these 17 variables, two are very statistically significant in gross profit and asset turnover while operating profit margin before depreciation, DTA, current debt, profit to current liabilities, OCF to current liabilities, debt to capital, quick ratio, sales to inventory, and accruals are statistically significant. Finally, operating profit margin after depreciation, cash flow margin, and debt to equity have some statistical significance. While the adjusted R squared improves when compared to the previous stepwise regression, it is still low at 16.45%. This means only 16.45% of all variances can be explained by this regression.

Furthermore, there seems to be a lot of redundancy among the 17 variables, leading to potential overfitting issues during the iterative process. Of the very statistically significant variables, the negative direction of gross profit and asset turnover probably have the same reasons for the negative direction of the ROA in the previous stepwise regression. That is, the COVID-19 market shock occurred at higher levels of gross profit and asset turnover. When the market recovered from the initial shock, gross profit and asset turnover had greatly diminished across all companies. Like the previous model, size itself was not a relevant factor. Again, this suggests that the chosen fundamental ratios act as a good proxy for the size effect among stocks in the restaurant industry.

To test the robustness of both stepwise regressions, we chose to run a PCA on each model to determine the number of components of key significance within each model. The purpose of PCA is to reduce dimensionality as minimizations of errors in regressions can lead to overfitting issues. Furthermore, PCA serves as a tool to identify potentially redundant aspects that add little value to the model. For the sake of simplicity, we believe that eigenvalues that are at least 0.05 are key components. Ideally, we would run a simulated distribution to determine the statistical significance of each eigenvalue. However, the nature of a large eigenvalue implies a low p-value and statistical importance. Given this threshold, we ran our first PCA on the stepwise function of five variables, leading to this result:

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	0.4096	0.2597	0.1686	0.1321	0.0283	0.0017

Figure 12. Principal Component Analysis Table of Stepwise Regression 1

From the breakdown of the proportion of variance, it seemed that four of the five variables within the model are needed as PC1 through PC4 have eigenvalues greater than 0.05. Intuitively, this makes sense as 4 of the 5 variables represent different characteristics. The only variables that mimic one another in this model are quick ratio and cash ratio. In addition, this helps confirm the robustness of the regression as it seems that the existing variables are necessary in determining the variance of the residuals.

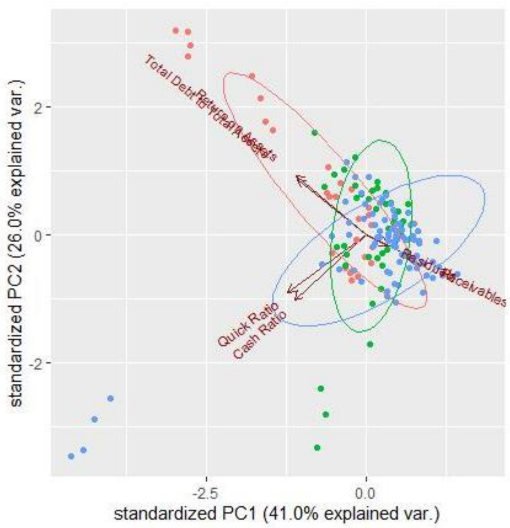

Figure 13. PC1 vs PC2 Biplot

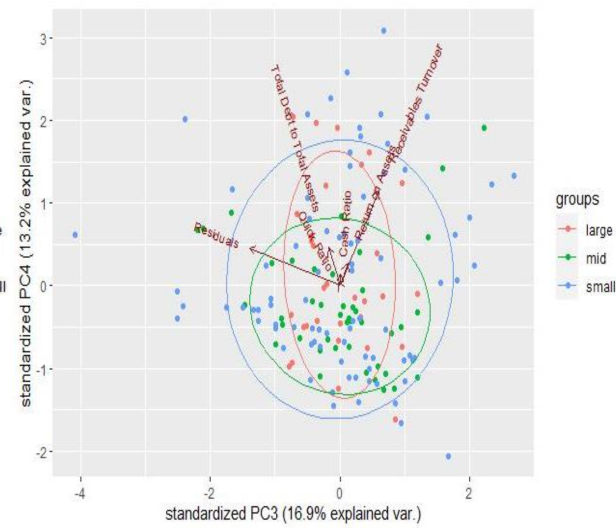

Figure 14. PC3 vs PC4 BiPlot

However, the PCA biplots in Figure 13 and Figure 14 above gave us more insight on the tendencies of this stepwise regression. In Figure 13, when we focused of the first 2 components of the PCA, we notice that most of residuals cluster in an eliptical shape, regardless of size classification. However, large-cap restaurant stocks seemed to have a wider elliptical to better encompass the centrality of its cluster. As for explanatory effect, it seemed that the same four variables for both PC1 and PC2, which are DTA, ROA, quick ratio and cash ratio. We could tell this due to the length and magnitude of the arrows. As for correlation to the residuals, only recievables turnovers went in the same direction as the residuals, meaning this relationship was positively correlated. Given that quick ratio and cash ratio were perpendicular to the residuals in this plot, no relationship exists for PC1 and PC2. Finally, DTA and ROA were negatively correlated to residuals in the PC1 vs PC2 biplot as they moved in the opposite direction of the residuals.

As for Figure 14, we noticed that the dispersion of residuals widened for each size classification when reaching PC3 and PC4. In fact, mid-cap restaurants started to mimic a circular centrality distribution. As for explanatory effect, recievables turnover seemed to be the most important for PC3 and PC4, given the length of its arrow. This, however, was suprising given the fact that recievables turnover had no correlation with the residuals, along with ROA and cash ratio. If anything, the PC3 vs. PC4 biplot showed that only quick ratio and DTA have a slightly positive correlation with the given dimensions. Given the different plots in Figure 13 and 14 encompassed every variable within the model with some explanatory effect, this confirms our faith in the robustness of the first stepwise regression.

As for the stepwise regression of 17 variables, the PCA generated the following result:

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	0.3528	0.1889	0.1374	0.0761	0.0612	0.0493
	PC7	**PC8**	**PC9**	**PC10**	**PC11**	**PC12**
Eigenvalue	0.0479	0.0451	0.0315	0.0130	0.0098	0.0068
	PC13	**PC14**	**PC15**	**PC16**	**PC17**	
Eigenvalue	0.0047	0.0015	0.0007	0.0002	0.0000	

Figure 15. Principal Component Analysis Table of Stepwise Regression 2

From this breakdown of the proportion of variance, we noticed that roughly six of the 17 variables within the model are needed as PC1 through PC6 have eigenvalues either greater than or almost equal to 0.05. This is problematic due to the suggestions of overfitting. However, this is also intuitive due to the significant use of redundant variables within this specific stepwise regression. While we could run a PCA biplot on this stepwise regression for further details on interactions of variables with the residuals and principal components, this seems unneccesary given the amount of eigenvalues under 0.05. In spite of this stepwise regression having a higher adjusted R squared than the previous stepwise regression, it is hard to look past overfitting issues. Therefore, we believe the best model for explaining COVID-19 residuals is the stepwise regression with ROA, DTA, receivables turnover, quick ratio, and cash ratio. This is due to its simplistic nature of the model as well as being validated by the PCA analysis and PCA biplots.

Conclusions

To summarize, we analyzed restaurant stocks within the U.S. stock exchange to study whether the effects of COVID-19 were disproportionate based on size classification. Our initial exploratory data analysis demonstrated to us that not only restaurant firms of different market capitalization reacted to the COVID-19 economic environment differently, but they have inherently distinct characteristics. After using data between 2015 and 2019 to properly fit the FF3FM and estimate the beta coefficients for each restaurant stock, we noticed that the RMSE worsened for each size grouping when trying to predict returns during 2020. However, further exploratory data analysis of the residuals proved that COVID-19 had a disparate impact on small-cap restaurant stocks due to a wider dispersion and volatility of its residuals. This could not be said for their mid-cap and large-cap counterparts. Later, we built two stepwise regression models to act as auxiliary regressions that helped explain the variance present in the residuals during a COVID-19 economic environment not explained by the FF3FM. After performing a robustness check with PCA, we found the stepwise regression of ROA, DTA, receivables turnover, quick ratio, and cash ratio to be our most accurate model. However, the variance explained by this model was only 10.57%.

While pursuing this analysis, we realized that there were many limitations to our data. First, we only had one year of data within the COVID-19 economic environment. This hurts our ability to confidently trust our model due to the lack of degrees of freedom and potential worries of over-extrapolation of data. Given that the pandemic has been a recent development, though, our lack of data makes sense. A future expansion of this research would be to wait for more data to be released to improve the existing model and potentially model for long-term lag effects of COVID-19 on the restaurant industry as a whole.

Next, we could only use quarterly fundamental ratios as these numbers aren't updated on a more frequent basis. Again, this hurts our degrees of freedoms and forced us to convert monthly observations into quarterly observations. This standardization of data could potentially have smoothed out significant trends that can only be seen on a shorter time frame. Unfortunately, the lack of frequency from fundamental ratios can't necessarily be fixed unless we chose a different set of company-specific fundamental data with higher posting frequency. A future expansion of this research would be trying to find company-specific KPIs within the restaurant industry and using this alternative data to develop our auxiliary regression on the residuals of a COVID-19 economic environment.

Finally, we had limitations in the functionality of the stepwise regression. While the stepwise regression is a powerful tool that helps reduce dimensionality, it does not take into consideration interaction effects. Given that small-cap restaurant stocks had wider dispersion in residuals than mid-cap and large-cap counterparts, it is possible that certain ratios have different magnitudes of effect depending on size. For example, a large-cap restaurant with high leverage could be less effected and less volatile than a small-cap restaurant with high leverage. A future expansion of this research would be to experiment with interaction effects, whether that be with trial-and-error testing or the use of data science tools such as LASSO. However, LASSO would require significantly more data to properly train the model via k-fold validation.

---

[i] https://rady.ucsd.edu/faculty/directory/valkanov/pub/classes/mfe/docs/fama_french_jfe_1993.pdf
[ii] https://nextlevel.finance/2020-stock-market-crash/
[iii] https://www.brookings.edu/research/ten-facts-about-covid-19-and-the-u-s-economy/
[iv] https://www.spglobal.com/marketintelligence/en/news-insights/blog/industries-most-and-least-impacted-by-covid19-from-a-probability-of-default-perspective-september-2020-update
[v] https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/
[vi] https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/
[vii] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[viii] https://www.investopedia.com/terms/s/stepwise-regression.asp
[ix] https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202
[x] http://statweb.stanford.edu/~susan/courses/s60/split/node60.html