Historical Analysis of the S&P 500 During the COVID Period Using PCA

Carlos Gafa'

1 Introduction

This analysis seeks to identify stock and sector weightings that would have outperformed the S&P500 during the COVID-19 recession through the application of Principal Component Analysis (PCA).

The dataset comprises stocks that were included in the S&P500 prior to 2019 and remained constituents throughout the analysis period. Stock price data for these 402 securities were downloaded from Yahoo Finance spanning January 2019 through December 2021. The timeframe from January 2019 to March 2020 is designated as the pre-COVID period, while the subsequent period encompasses the COVID-19 market disruption and recovery phases. The normalized price trajectories of these securities are shown in Fig. 1.

2 First Principal Component

Standardized log returns were calculated for the period from January 2019 to December 2021, normalized to exhibit mean zero and unit variance [1]. A PCA model was subsequently fitted to these returns utilizing the *sklearn* library. The explained variance ratio of the first 30 components is presented in Fig. 2.

The first principal component, which explains the majority of variance among stock returns during this period, was found to exhibit loadings exclusively in one direction,

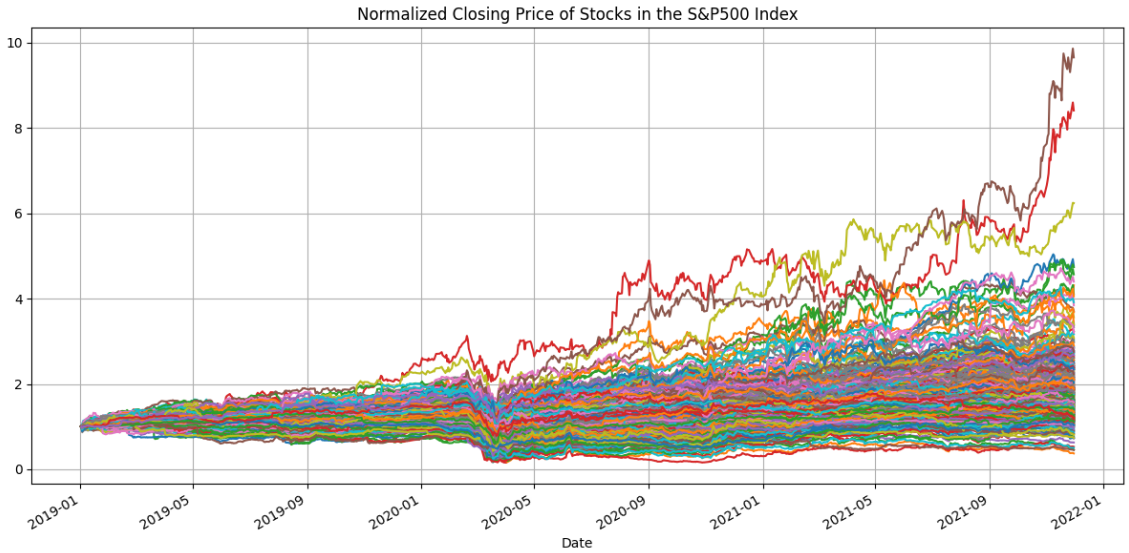

Figure 1: Normalized closing prices of stocks in the S&P500 index.

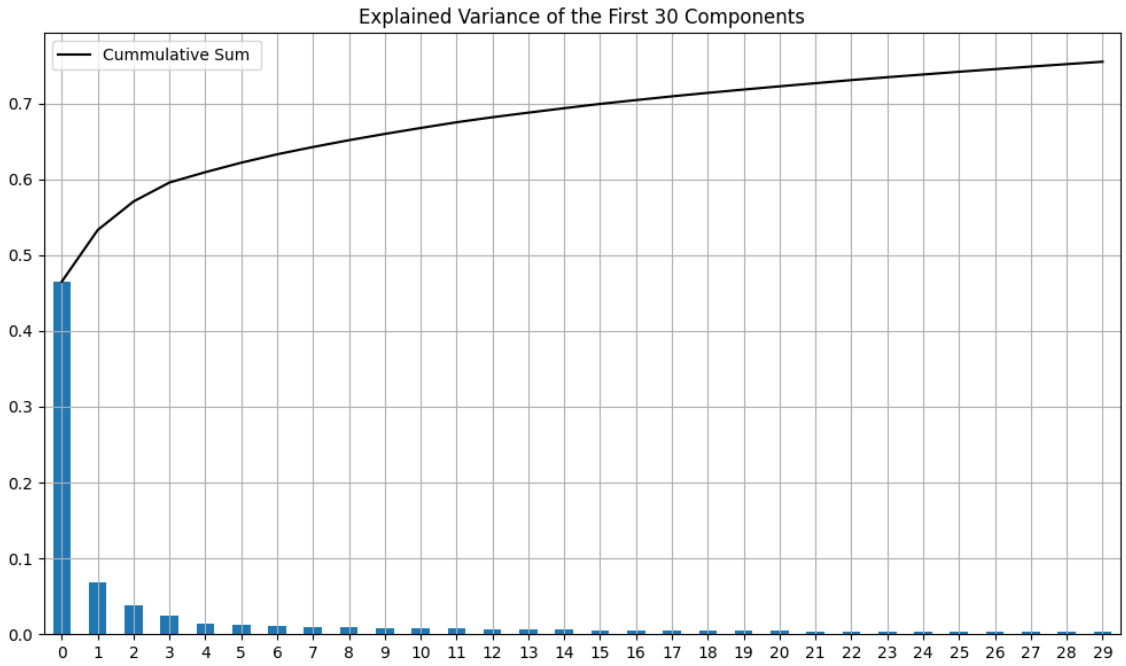

Figure 2: Explained variance ratio of the first 30 principal components.

thereby indicating its representation of the market factor wherein all securities in the market move together [2]. The normalized stock weights (scaled such that their sum equals unity) for the top securities in the first principal component are presented in Fig. 3.

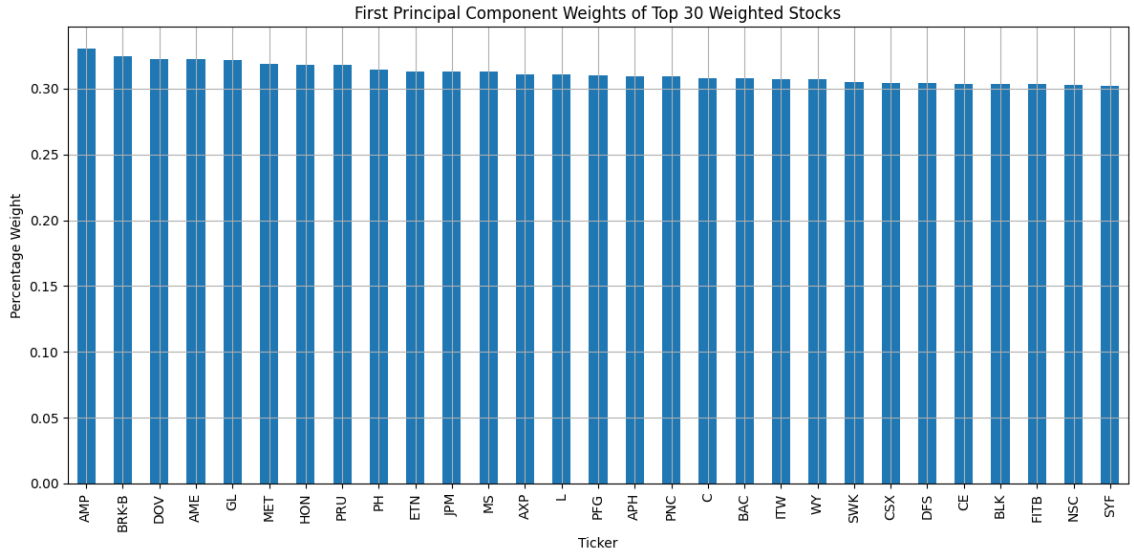

Figure 3: Stock weights in the first principal component.

A comparison of the cumulative returns of the S&P500 index and the portfolio constructed from the first principal component, as shown in Fig. 4, reveals that this portfolio closely tracks the index, further confirming its characterization as the market portfolio.

2.1 PCA during COVID period

The analysis was then restricted to the COVID period by limiting the return series to March 2020 through December 2021. The first component was again found to correspond to the market factor, with all stock loadings exhibiting the same directional bias. The top five securities with the highest loadings in this factor were identified as belonging to the following industries: Communication Services, Consumer Defensive, Financial Services, and Industrials. Sector classifications were obtained using the yfinance library's sector identification methodology.

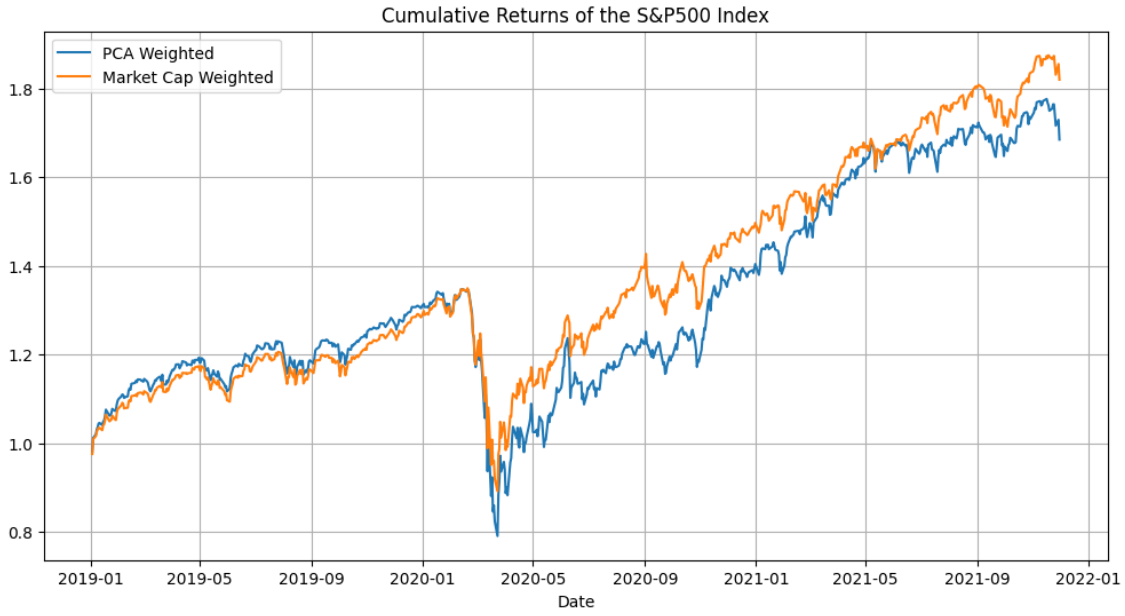

Figure 4: Cumulative returns of the portfolio obtained from the first principal component and the S&P500 Index.

2.1.1 Second Principal Component

The second principal component, accounting for 7.6% of the variance, comprises 224 securities with positive weights and 178 securities with negative weights. Accordingly, the securities were partitioned into two distinct portfolios based on the sign of their respective loadings. While the absolute magnitude of the loadings is the primary consideration rather than their sign, this binary classification convention is maintained for analytical clarity.

The performance of these portfolios during the COVID period is illustrated in Fig. 5. The analysis reveals that the positive-weight portfolio experienced a more modest drawdown during the COVID crisis and demonstrated faster recovery relative to the broader market, though it did not achieve superior performance over the entire period under examination.

These performance characteristics are further accentuated when considering an investment initiated in January 2019, as demonstrated in Fig.7. The composition of the positive-weight portfolio is detailed in Fig.6, which highlights the predominance of wholesale and

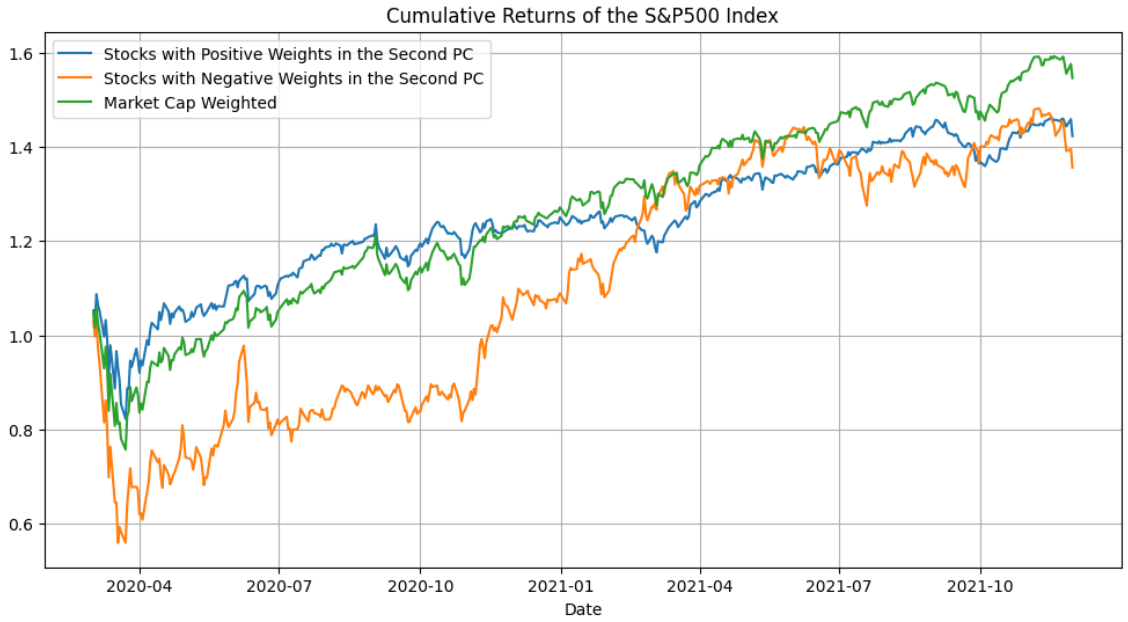

Figure 5: Cumulative returns of the portfolios obtained from the second component during the COVID period.

Table 1: Portfolio Metrics for the positive and negative second principal component portfolios

	Annualized Mean %	Annualized Volatility %	Annualized Sharpe Ratio	Maximum drawdown %
Positive Weights	21.55	19.91	1.08	26.81
Negative Weights	10.32	37.16	0.28	57.08
S&P 500	20.55	22.75	0.90	33.92

consumer goods companies, including Costco and Procter&Gamble Co. Notably, the top 10 securities in this portfolio are exclusively from the Consumer Defensive sector. Conversely, the top 10 securities comprising the negative-weight portfolio span the following industries:** Consumer Cyclical, Energy, Financial Services, Industrials, and Real Estate.

Over the combined pre-COVID and COVID periods, the positive-weight portfolio achieved an annualized Sharpe ratio of 1.08 (calculated assuming a zero risk-free rate), compared to 0.90 for the S&P500 index. Additional performance metrics for these portfolios are presented in Table 1.

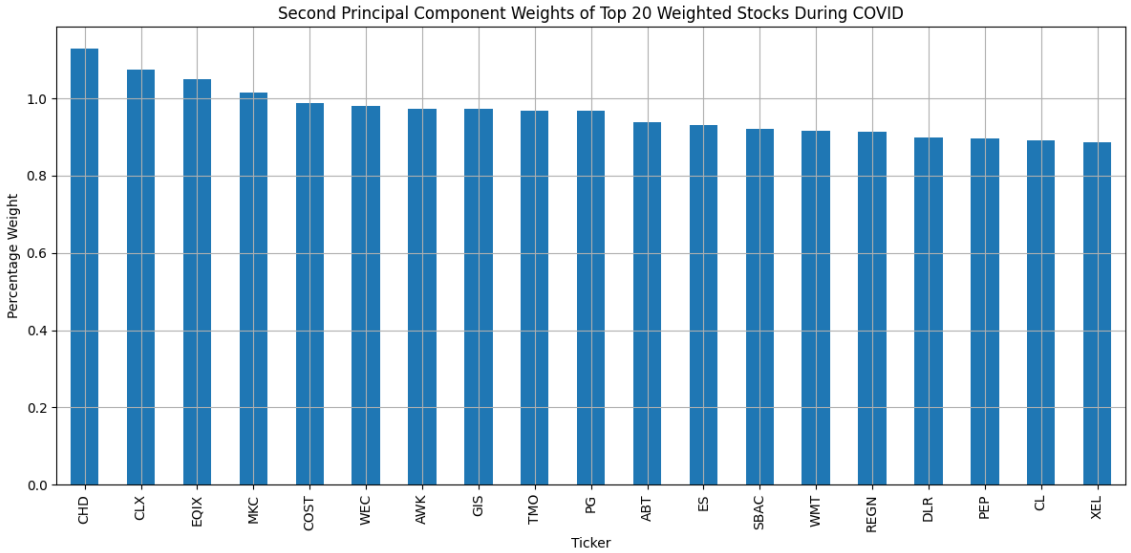

Figure 6: Top 20 stocks in the portfolio of stocks with positive weights in the second PC.

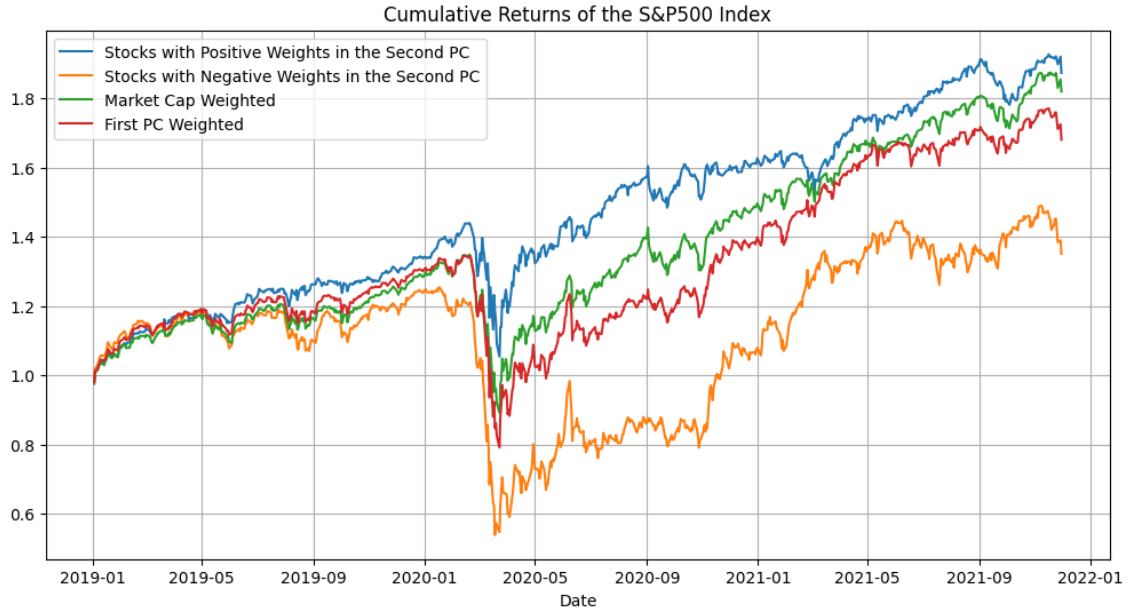

Figure 7: Cumulative returns of the portfolios obtained from the second component.

2.1.2 Third Principal Component

The third principal component, accounting for 3.6% of the variance, encompasses 196 securities with positive loadings and 206 with negative loadings. Consistent with the previous analysis and acknowledging that the sign magnitude is less critical than the opposing nature of the loadings, two portfolios were constructed: those with negative weights and those with positive weights in the third component.

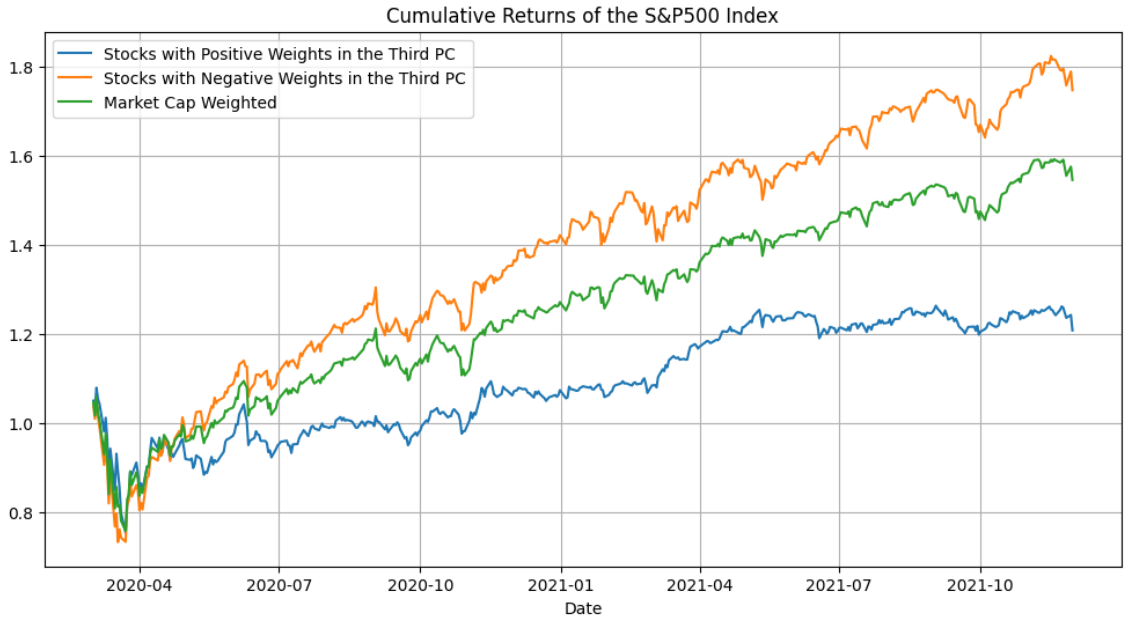

Figure 8: Cumulative returns of the portfolios obtained from the third component during COVID.

As demonstrated in Fig.8, the negative-weight portfolio significantly outperformed the S&P500 index during the COVID period, though this superior performance was accompanied by larger drawdowns and elevated volatility. This outperformance becomes even more pronounced when the pre-COVID period is included in the analysis, as illustrated in Fig.9.

The composition of the negative-weight portfolio, as presented in Fig. 10, reveals a concentration in large technology companies such as Amazon, Microsoft, Google, and Meta. The top 10 securities in this portfolio span the following industries: Communication Ser-

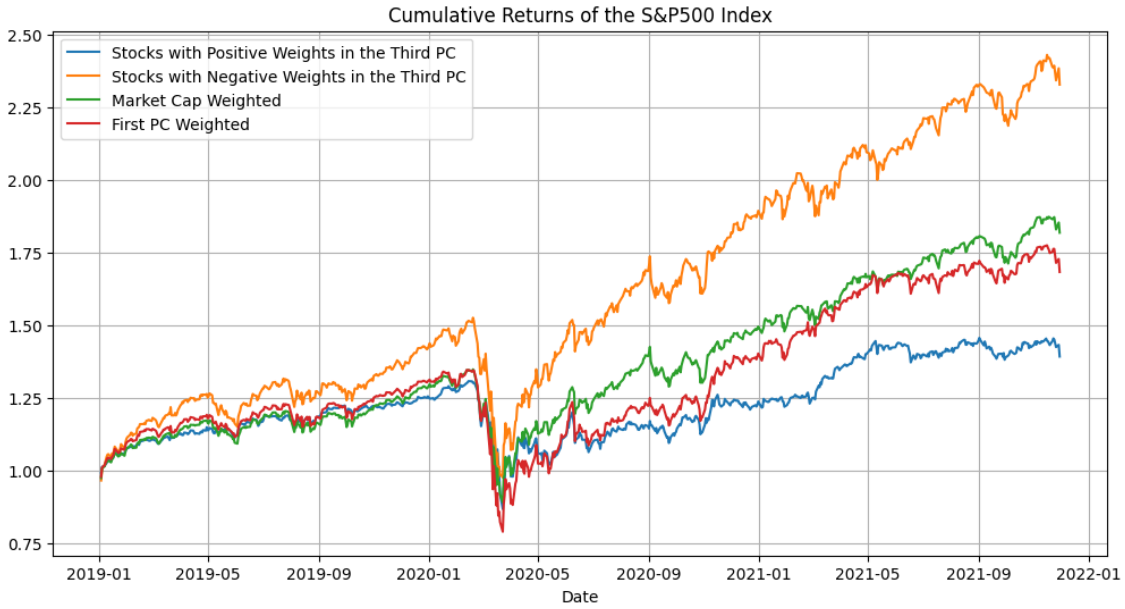

Figure 9: Cumulative returns of the portfolios obtained from the third component.

Table 2: Portfolio Metrics for the positive and negative third principal component portfolios

	Annualized Mean %	Annualized Volatility %	Annualized Sharpe Ratio	Maximum drawdown %
Positive Weights	11.33	21.61	0.52	33.47
Negative Weights	29.00	25.45	1.14	36.17
S&P 500	20.55	22.75	0.90	33.92

vices, Consumer Cyclical, Financial Services, Healthcare, and Technology.

Despite exhibiting higher volatility than the S&P500, this portfolio delivered a superior Sharpe ratio during the combined pre-COVID and COVID periods, as detailed in Table 2.

2.1.3 Conclusion

Through the application of PCA analysis, two portfolios were identified that outperformed the S&P500 index during the COVID period from a risk-adjusted return perspective, where one of the portfolios offered investors at the time a lower risk with a slightly lower return, and the other portfolio provided investors at the time with a higher return at higher a volatility.

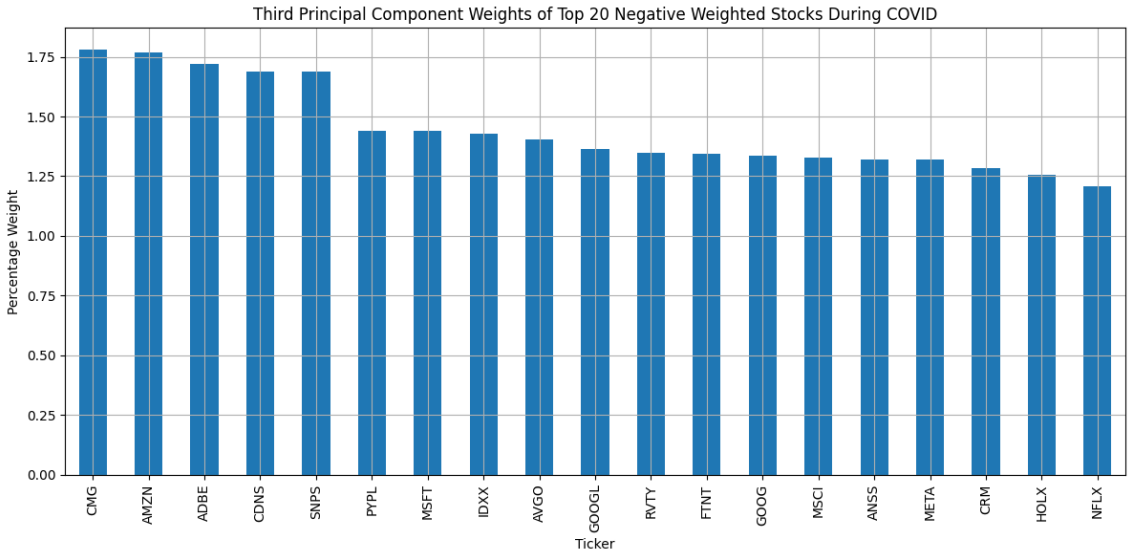

Figure 10: Weights of the top 20 stocks in the portfolio corresponding to those stocks with a negative weighing in the third component.

References

[1] M. Avellaneda and J.-H. Lee. Statistical arbitrage in the u.s. equities market. *Quantitative Finance*, 10:761–782, 07 2008.

[2] M. Avellaneda and J. A. Serur. Hierarchical pca and modeling asset correlations, 2020.