Project

Quantitative Equity Portfolio Creation

By: Group 5
Ankit Rawat
Jay Gupta
Rajarajan Mohan
Shruti Maindola
Shuxin Zhan

Table of Contents

1 Executive Summary

As part of this project, we worked towards creation of equity portfolio of growth stocks using Quantitative Equity Portfolio technique and tracking portfolio return against the selected Benchmark (S&P 100) index. By using multiple unique factors in our model, we aim to build a more concentrated portfolio of stocks with the highest scoring factor profiles.

2 Research Methodology

2.1 Factors and Factors Choice

Factors are explanatory variables that represents different types of risk. A factor model shows that the average stock return is proportional to the stock exposure to the risk that the factor represents and to the payoff for each unit of exposure to the risk. As part of this project, we picked different fundamental and technical factors to understand their correlation and impact stocks movement.

We grouped Fundamental factors into mainly 4 Sub-Categories:

- **Valuation factors:** Measures whether the stocks are relatively cheap or expensive
- **Solvency factors:** Measures a company's ability to meet future short-term obligations
- **Operating Efficiency/Profitability factors:** Explains how well management is running the company
- **Financial Risk factors:** Measures the financial health of the company

2.2 Data Decision

The data set of the model is determined with a type of factor and a set of factors in mind. We considered cross sectional data of S&P 100 stocks of approximately 20 years to perform analysis of different Fundamental and technical factors. As our focus on fundamental factors hence we mainly captured Quarterly data for our analysis.

Date	AAPL	ABBV	ABT	ACN	ADBE	AIG	ALL	AMGN	AMT	AMZN	...	UNH	UNP	UPS
2000Q1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2000Q2	-0.228716	NaN	0.272783	NaN	0.168118	0.073531	-0.059655	0.144603	-0.155696	-0.458022	...	0.438156	-0.044967	-0.062074
2000Q3	-0.508353	NaN	0.072058	NaN	0.194433	0.222067	0.571027	-0.006005	-0.095952	0.058520	...	0.151603	0.052509	-0.040044
2000Q4	-0.422330	NaN	0.022356	NaN	-0.250403	0.030453	0.259345	-0.084359	0.004975	-0.595122	...	0.243038	0.308978	0.045074
2001Q1	0.483698	NaN	-0.021481	NaN	-0.398638	-0.182902	-0.032622	-0.058651	-0.511551	-0.342651	...	-0.033944	0.112289	-0.028355
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2019Q4	0.315049	0.203853	0.042288	0.099427	0.193882	-0.072646	0.039303	0.254053	0.043941	0.064479	...	0.357953	0.122238	-0.015417
2020Q1	-0.131982	-0.127859	-0.087668	-0.221678	-0.035081	-0.521603	-0.180232	-0.153071	-0.052519	0.055135	...	-0.148037	-0.215276	-0.194040
2020Q2	0.450243	0.307839	0.163541	0.321137	0.367867	0.298923	0.063217	0.171294	0.197274	0.414983	...	0.187827	0.205464	0.202591
2020Q3	0.272136	-0.096974	0.194931	0.056343	0.126622	-0.107187	-0.023716	0.084807	-0.060484	0.141332	...	0.061382	0.170267	0.508289
2020Q4	-0.060012	-0.015197	-0.030963	-0.036420	-0.088351	0.143843	-0.057255	-0.146443	-0.049973	-0.035754	...	-0.021266	-0.099964	-0.057133

2.3 Factor Selection

We followed multiple approaches to do factor selection:

- **Univariate Regression Tests**: We simplify the process of searching for relevant factors by performing a series of simple univariate regression of factors following "Univariate Regression Tests" on the group of factors that theoretically justify explaining stock returns

$$r_{i,t} = \alpha + f\beta_{i,t} + \epsilon_{i,t}$$

- β is the factor exposure of stock (i) at time t, and the estimate of 'f' from this panel regression will show the relationship between the factor and stock return. If factor has significant value of f then factor is useful in explaining stock returns
- It helped us to give early idea on set of key cut of key factors but it didn't help to give details about co-relation between them.

Example :
We picked P/E ratio and P/B ratio explain stock returns but as both explains the same idea hence we need to find corelation between them. Idea is higher the factors we include in our model higher will be our information ratio, provided average contribution of each factor does not decrease.

- **Corelation Study:** We further worked to understand the corelation of different factors filtered out after univariate test to understand the corelation between different factors.
  We found below result based on our study :
  - Low correlation between the returns and fundamental factors
  - Ranked the highest positive and lowest negative correlation with the objective of these factors meeting the criteria of minimum correlation.

Correlation Study

quarterly_return	1.000000	int_debt	-0.007081
bm	0.054341	staff_sale	-0.008132
roe	0.052856	inv_turn	-0.013081
cash_ratio	0.051710	lt_ppent	-0.013934
ptb	0.049653	cash_conversion	-0.014020
pcf	0.046081	short_debt	-0.014295
curr_ratio	0.044234	invt_act	-0.016483
quick_ratio	0.043232	rect_turn	-0.017463
cash_lt	0.038368	intcov_ratio	-0.017630
adv_sale	0.037935	sale_equity	-0.019097
ps	0.030053	intcov	-0.020269
fcf_ocf	0.028890	totdebt_invcap	-0.020629
sale_invcap	0.027699	de_ratio	-0.021659
at_turn	0.025159	dltt_be	-0.029017
pay_turn	0.024111	rect_act	-0.054903
GProf	0.021760	divyield	-0.106951
CAPEI	0.020839		

- **PCA Analysis. :** Performed PCA analysis to find the statical significance between the different set of factors and observation as follows :
  - No statistical significance and low absolute coefficient values
  - Unable to achieve the dimensionality reduction
  - No concrete relationship between fundamental factors and returns

```
                            OLS Regression Results
===============================================================================
Dep. Variable:                return_m   R-squared (uncentered):          0.159
Model:                             OLS   Adj. R-squared (uncentered):     0.098
Method:                  Least Squares   F-statistic:                     2.598
Date:                 Sun, 15 Nov 2020   Prob (F-statistic):             0.0461
Time:                         17:41:45   Log-Likelihood:                 47.323
No. Observations:                   59   AIC:                            -86.65
Df Residuals:                       55   BIC:                            -78.34
Df Model:                            4
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
PC1           -0.0039      0.003     -1.483      0.144      -0.009       0.001
PC2            0.0020      0.004      0.529      0.599      -0.006       0.009
PC3           -0.0139      0.007     -1.980      0.053      -0.028       0.000
PC4            0.0174      0.009      1.998      0.051   -5.28e-05       0.035
==============================================================================
Omnibus:                       12.128   Durbin-Watson:                   2.087
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               35.470
Skew:                          -0.271   Prob(JB):                     1.99e-08
Kurtosis:                       6.760   Cond. No.                         3.28
==============================================================================
```

- We further filtered our factors based on Secondary research and narrowed it down to 6 factors to perform OLS regression test . Insignificant relationship observed with the six factors against the returns. However, we proceeded with these six factors based on secondary research and the final correlation outcome.

<u>List of Factors:</u>

- Book/ Market
- Return on Equity
- Cash_Ratio
- Price / Book Ratio
- Current_Ratio
- Dividend Yield

OLS Regression Results

Dep. Variable:	quarterly_return	**R-squared (uncentered):**	0.087
Model:	OLS	**Adj. R-squared (uncentered):**	0.085
Method:	Least Squares	**F-statistic:**	57.42
Date:	Sun, 15 Nov 2020	**Prob (F-statistic):**	5.85e-79
Time:	22:39:00	**Log-Likelihood:**	3028.4
No. Observations:	4235	**AIC:**	-6043.
Df Residuals:	4228	**BIC:**	-5998.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
bm	0.0597	0.007	8.247	0.000	0.045	0.074
roe	-0.0044	0.008	-0.527	0.598	-0.021	0.012
cash_ratio	0.0013	0.005	0.251	0.802	-0.009	0.011
ptb	0.0017	0.000	3.876	0.000	0.001	0.003
pcf	0.0006	0.000	2.644	0.008	0.000	0.001
curr_ratio	0.0057	0.004	1.370	0.171	-0.002	0.014
divyield	-0.0057	0.001	-4.665	0.000	-0.008	-0.003

Omnibus:	294.549	**Durbin-Watson:**	2.070
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1481.179
Skew:	0.045	**Prob(JB):**	0.00
Kurtosis:	5.896	**Cond. No.**	82.2

2.4 Stock screening and ranking based on aggregate z-score:

In context of stock screening and ranking, we used simultaneous screening and the aggregate Z-Score approach.

We worked to find factor exposure of every stock. Then, for each stock, we subtract the universe's mean factor exposure from the stock's individual factor exposure and divide the difference by the standard deviation of factor exposures for the universe

$$z_{i,k} = \frac{\beta_{i,k} - \overline{\beta}_k}{S(\beta_k)}$$

- After computing Z-score for each stock across the selected factors, we aggregated them into one final score.

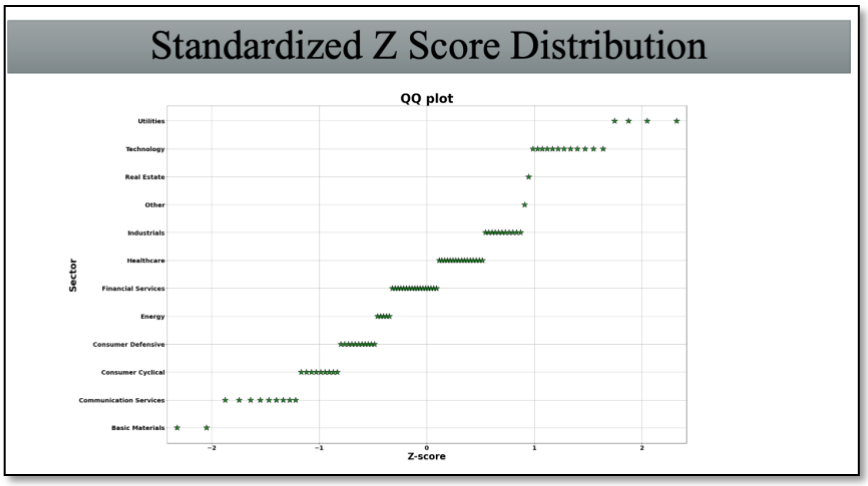
- We used equal weighting of the Z-Score factors however for dividend yield we reversed the sign of z-score considering its negative correlation with the returns.

TICKER	Book to Market	ROE	Cash Ratio	Price to Book	Price to Cashflow	Current Ratio	Dividend Yield	Returns(Q)	Z_score	Sectors	Z_score_normalize
FB	-0.926656	-0.058489	5.844937	0.081713	1.126092	5.843921	0.000000	1.861885	13.773401	Communication Services	4.393716
GOOG	-0.793941	0.114578	3.952892	0.076460	0.978791	3.930755	0.000000	1.047348	9.306883	Communication Services	2.968896
NVDA	-0.477809	0.115645	1.527002	0.335230	2.144890	1.653791	-1.130176	2.647727	9.076652	Technology	2.895453
MA	-1.142563	0.887063	0.131109	2.223538	1.384835	-0.038239	-1.730139	2.229412	7.405294	Financial Services	2.362289
ADBE	-0.815909	0.103820	0.972096	0.583069	1.047579	0.681541	-2.005389	0.507609	5.085194	Technology	1.622177
NFLX	-1.055876	0.153806	0.175607	2.287506	-0.758493	-0.096422	0.000000	4.073831	4.779959	Communication Services	1.524807
QCOM	-0.666536	-0.057228	1.830293	0.266203	1.523230	1.778976	-0.308641	-0.579209	4.404371	Technology	1.404995
CRM	-1.139918	-0.435406	-0.252274	1.490769	3.254273	-0.564876	0.000000	1.621020	3.973587	Technology	1.267575
GILD	-1.024570	0.441850	1.568710	1.076655	-0.546939	1.876098	0.362581	0.866259	3.895482	Healthcare	1.242659
CL	-1.307629	2.387695	-0.528050	3.505146	0.505980	-0.491025	-0.285149	-0.625894	3.731372	Consumer Defensive	1.190308

3 Portfolio construction

3.1 Picking stocks based on normalized Z factors:

Based normalizing all data and comparing it over Z score we found that Utilities and Technology has Z score compare to other factors



3.2 List Of Selected Stocks:

As we mainly want to keep focus of our portfolio on growth stocks hence picked top 10 stocks from the constructed Z scale. And it's gave us good mix of cross-sectional stocks.

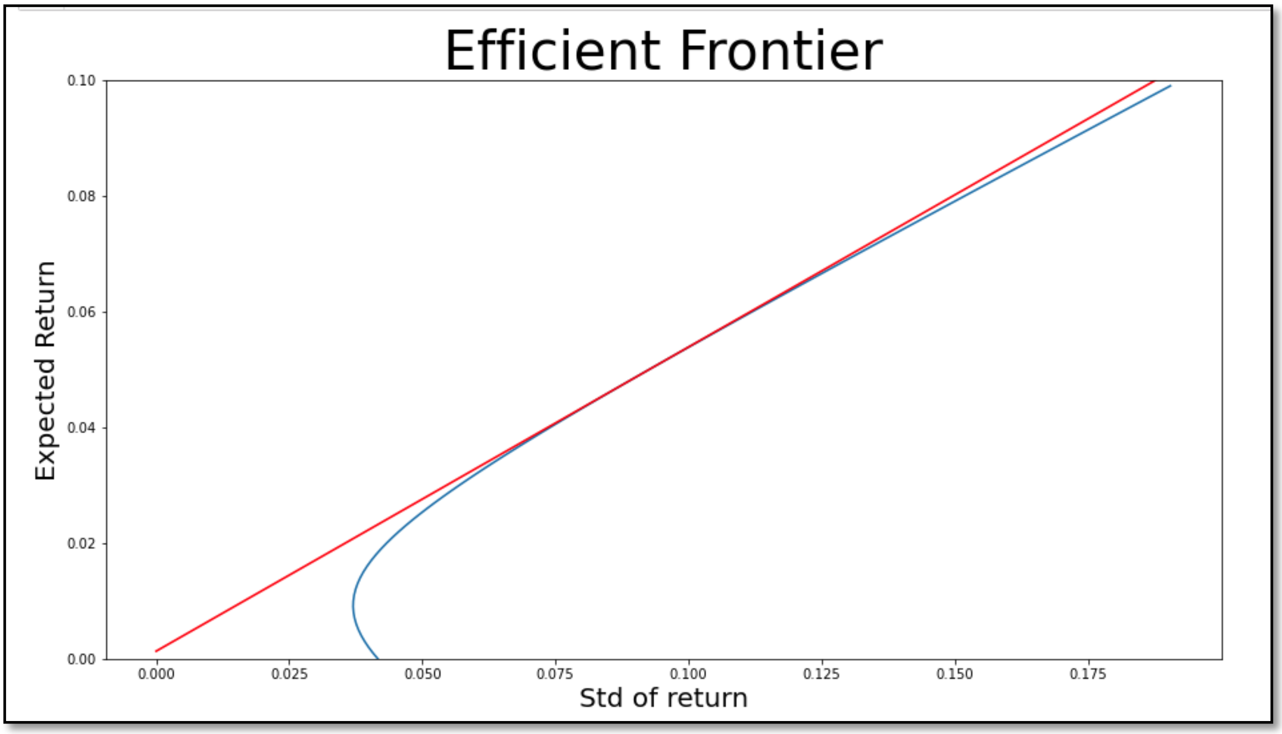
Stock
Facebook
Google
Nvidia
Mastercard
Adobe
Netflix
Qualcomm
Salesforce
Gilead
Colgate - Palmolive

3.3 Efficient Frontier Creation and Weight allocation

Using the given Lagrange multipliers estimators, we created efficient frontier and later used the same estimators to derive tangency portfolio on efficient frontier. As part of project, we consider Quarterly Risk-Free rate as 0.13 %

Portfolio Weights

Stock	Weight
Facebook	0.09
Google	-0.10
Nvidia	0.46
Mastercard	0.31
Adobe	0.89
Netflix	0.20
Qualcomm	-0.13
Salesforce	-0.53
Gilead	-0.40
Colgate - Palmolive	0.21



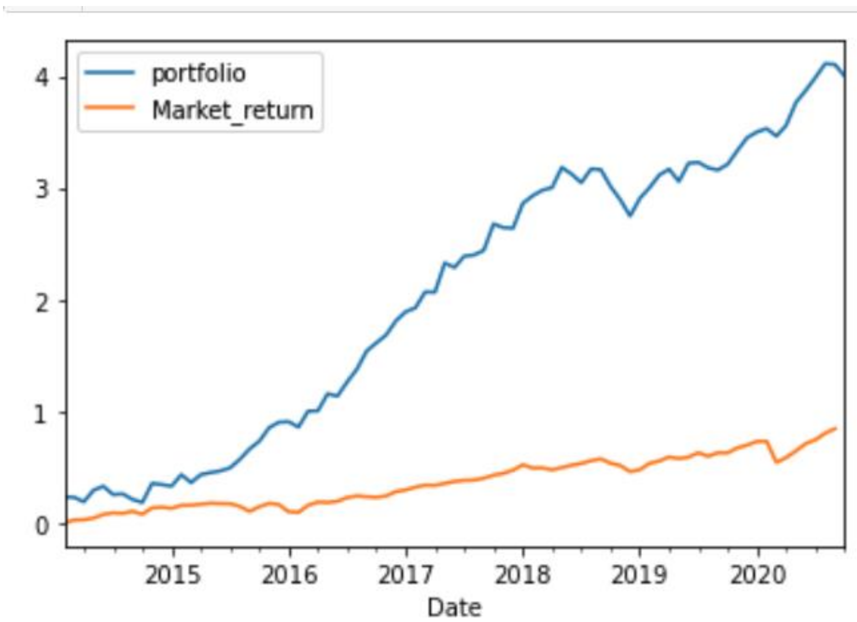
4 Overview of Key Findings

4.1 Sharpe Ratio

Using tangency portfolio we got Sharpe ratio around 0.5258

4.2 Comparison: Constructed portfolio Vs S&P 100 Index

Based On the analysis, we found that performance of portfolio constructed using growth stocks is higher than S&P 100 in the last 20 years.



5 Improvements

There are a number of potential improvements that can be made to the project. First and foremost, the data we acquired was quarterly for the last 20 years. As a result, we obtained a large dataset with around 60 fundamental factors for each company. It is possible for us to use a correlation matrix to find the correlations between the different fundamental factors for all the companies. This would enable us to weed out a number of fundamental factors that potentially allows us to reduce the size of our dataset and make the modelling more manageable.

Another potential improvement to reduce the number of factors would be to use certain tests such as calculating the Cronbach Alpha to check how closely related a group of factors were with each other. For example, in our dataset, under the profitability factor, we had five other factors making up the profitability portion. Using this test would enable to us cut down the number of factors that make up the profitability heading. In addition to a correlation matrix and PCA analysis, we could have also used common factor analysis. The drawback of PCA is that it assumes there is no unique variance, however in our case, this is extremely unlikely. Common factor analysis deals with this assumption and is ideal to use considering there is some underlying connection between the various fundamental factors used.

For our regression model, instead of just using the standard ordinary least squares (OLS) regression, we could've been a bit more specific and considered other types of models, such as ridge or lasso regression. This would potentially give us a more accurate reading and results on the exploratory and predictive power of our model. Furthermore, when building our portfolio, we assigned equal weights to the factors that would be used. Instead, with the results of the final model, we can just use the coefficients outputs of the factors as the weights to choose the stocks that would fit into our portfolio.