FIN 557 Project Report

Stock Valuation Across Industries

4/25/23

By: Tzu-Mien Liu, Tung-Han Lee, Pratik Relekar, Amin Azimi, Hinduja Hanumanthappa, Gurpratap Singh Thind, Xiying Zhao

**Introduction**

The project involves examining how stocks perform in diverse industries and using this analysis to devise an investment strategy. As a part of this initiative, we selected 50 companies from five industries and evaluated whether each stock is undervalued, fairly valued, or overvalued in its respective industry. The primary goal is to identify the most undervalued stocks across all industries and create an investment portfolio using these stocks.

**Data source**

WRDS to collect annual and quarterly data for 10 companies in each of the 5 sectors spanning from January 2010 to April 2023.

**Links -**

<https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/compustat/north-america-daily/fundamentals-quarterly/>

<https://finance.yahoo.com/screener/predefined/undervalued_growth_stocks/>

**Research question**

The primary goal of this project is to analyze 50 sample stocks from 5 different industries and develop an investment portfolio that is optimal.

**Approach**

​​Firstly, we categorized the 50 stocks into three groups according to their average Price-to-Book ratio (PB ratio) for the period spanning from 2015 to 2017. The PB ratio is calculated by dividing a share’s market price to its book value.

Secondly, we determined the average returns of all companies in each of the three groups over the subsequent five-year period from 2018 to 2022.

Finally, we calculated the number of stocks in each group and their respective average returns within each industry for the 2018-2022 period.

**Data Manipulation and Analysis**

**Analysis 1: Based on P/B ratio**

1. Variables constructed

* Outstanding: Calculate the number of shares outstanding by dividing the total market value by stock price.
* PB\_ratio: Calculate the P/B ratio by dividing the market Capitalization by Book value of Equity.
* Year: Extract the year from DATADATE.
* Quarter: Extract the quarter from DATADATE
* Stock\_group: Define whether the stock is value stock or growth stock based on its P/B ratio.

1. Data manipulation and techniques applied

* Filter to only include raw data in 2015, 2016 and 2017.
* Calculate the average PB ratio over a three year period.
* We assume that the stocks in the bottom 25% of the PB ratio are "value stocks". We assume that the stocks in the top 25% of the PB ratio are "growth stocks". Others are defined as "general stocks".
* Filter to only include raw data in 2018 to 2022.
* Merge stock\_group and stock price(after 2017) by company name.
* Calculate the mean return of all companies in each group over the 5-year period.

/\*Create new colums: Outstanding, PB\_ratio, Year, QUARTER \*/

/\*Only select positive PB\_ratio\*/

**Code -**

data stock1;

set project.stock;

Outstanding = MKVALTQ/PRCCQ;

PB\_ratio = PRCCQ\*Outstanding/CEQQ;

Year = year(DATADATE);

QUARTER = qtr(DATADATE);

if PB\_ratio > 0;

run;

/\*Filter three years' data to calculate average PB\_ratio\*/

**Code -**

data stock2;

set stock1;

keep GVKEY DATADATE YEAR QUARTER SIC CONM PRCCQ PB\_ratio;

where YEAR in (2015,2016,2017);

run;

/\*Calculate three years average PB ratio of different companies\*/

**Code -**

proc means data=stock2;

var PB\_ratio;

class CONM;

ways 1;

output out=PBmean mean=AvgPB;

run;



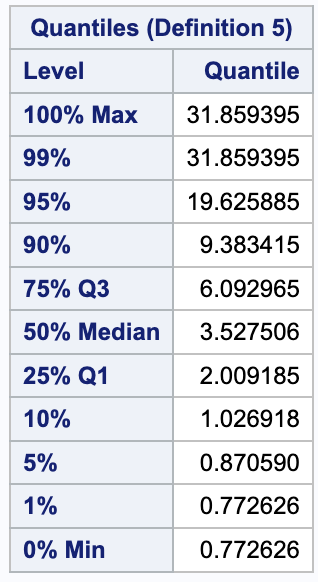
/\*Find p25 and p75 of companies' average PB\_ratio\*/

**Code -**

proc univariate data= PBmean;

var AvgPB;

run;



/\*Average PB\_ratio p25 =2.009185 and p75 = 6.092965\*/

/\*We assume that the bottom 25% of the stocks in the PB ratio are "value stocks" \*/

/\*We assume that the top 25% of the stocks in the PB ratio are "growth stocks" \*/

/\*Others are "general stocks"\*/

/\*Divide all stocks into three groups(based on p25 and p75)\*/

**Code -**

data stock4;

set PBmean;

length stock\_group $14;

keep CONM AvgPB stock\_group;

where AvgPB is not missing;

if AvgPB=<2.009185 then do;

stock\_group="value stock";

output;

end;

else if AvgPB>6.092965 then do;

stock\_group="growth stock";

output;

end;

else do;

stock\_group="general stock";

output;

end;

run;



/\*Filter five years' data\*/

**Code -**

data stock\_after\_2017;

set stock1;

keep GVKEY DATADATE YEAR QUARTER CONM PRCCQ;

where YEAR in (2018,2019,2020,2021,2022);

run;

/\*Sort by company name\*/

**Code -**

proc sort data=stock4;

by CONM;

run;

proc sort data=stock\_after\_2017;

by CONM;

run;

/\*Merge stock group and stock price(after 2017) by company name\*/

**Code -**

data stock\_price;

merge stock4 stock\_after\_2017;

by CONM;

run;

proc sort data=stock\_price;

by CONM;

run;

/\*Get stock price in 2018 and 2022 to calculate stock return\*/

**Code -**

data stock\_return;

set stock\_price;

by CONM;

if first.CONM = 1 then output;

else if last.CONM = 1 then output;

run;

/\*Calculate stock return\*/

**Code -**

data stock\_return2;

set stock\_return;

by CONM Year;

PRCCQ2=lag(PRCCQ);

if first.CONM then PRCCQ2=.;

ret=(PRCCQ-PRCCQ2)/PRCCQ2;

run;

/\*Keep non-missing return\*/

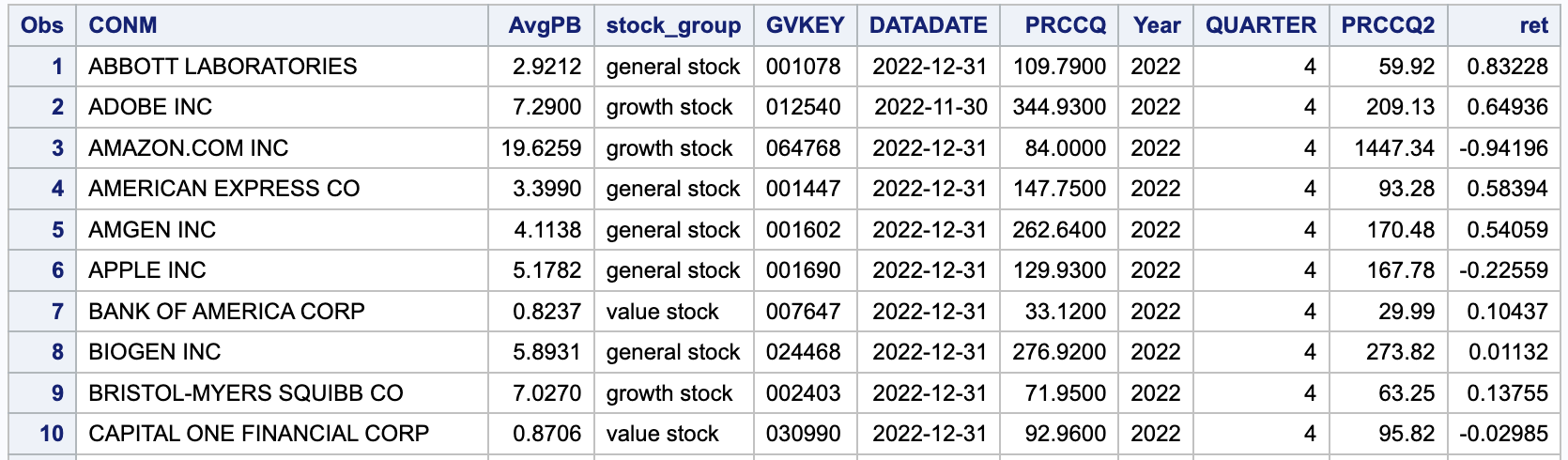
**Code -**

data stock\_return3;

set stock\_return2;

where ret~=.;

run;



/\*Get average returns of all companies in each group\*/

**Code -**

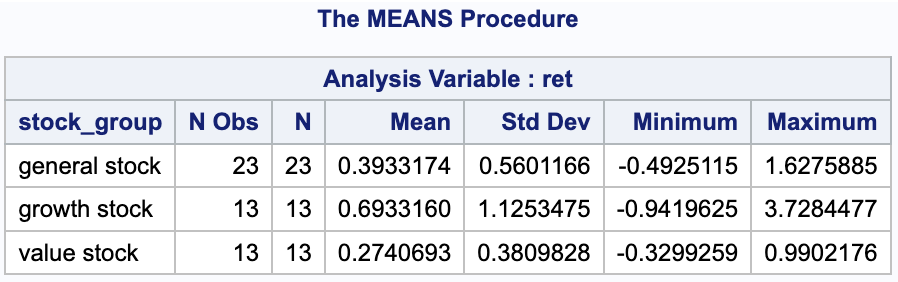
proc means data=stock\_return3;

var ret;

class stock\_group;

ways 1;

run;



Key observations: Growth stocks had the highest return over the past five years, while value stocks had the lowest return of the three groups. This result does not meet our expectations. We speculate that the opposite trend was caused by the pandemic.

**Analysis 2: Based on industry**

In the previous analysis, we examined the average PB ratios of 50 stocks between 2015-2017 and their average returns over 5 years (2018-2022). Moving forward, we aim to incorporate the industry of the companies into our analysis and assess the stock performance across various industries.

1. Variables Constructed:

* Sector: define the industry of company based on the company list(see Appendix)

1. Data Manipulation with Techniques Applied:

* Add a new column named “Sector” to define industry using case…when… sentence in sql
* Combine the data using merge sentence in SAS data step

*# Create a new column: "Sector" to define industry*

*# In our case, we have five industries and each industry contains 10 companies*

*# Define the sectors and assign TIC to them*

**Code -**

proc sql;

create table sector\_data as

select distinct TIC,CONM,

case

when TIC in ('AAPL','MSFT','IBM','INTC','CSCO','ORCL','HPE','DELL','ADBE','TXN') then 'Technology'

when TIC in ('XOM', 'CVX', 'COP', 'HAL','OXY','MPC','VLO','EOG','SLB','WMB') then 'Energy'

when TIC in ('WMT','AMZN','HD','COST','TGT','LOW','CVS','WBA','DLTR','TJX') then 'Retail'

when TIC in ('JNJ','PFE','MRK','BMY','LLY','ABT','AMGN','UNH','GILD','BIIB') then 'Healthcare'

when TIC in ('JPM','BAC','WFC','C','GS','AXP','MS','V','MA','COF') then 'Finance'

end as Sector

from stock1;

quit;

*# Check if all assigned well*

**Code -**

proc sql;

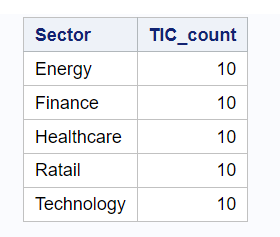
select Sector, count(distinct TIC) as TIC\_count

from sector\_data

group by Sector;

quit;

**Output -**



*# Sort stock1 and sector\_data by CONM*

**Code -**

proc sort data=stock1;

by CONM;

run;

proc sort data=sector\_data;

by CONM;

run;

*# Merge the sector data with the main data*

**Code -**

data main\_data;

merge stock1 sector\_data;

by CONM;

run;

1. Data Analysis:

* The number of stocks of different sectors in each stock group (2015-2017)

# *Stock average performance in different industries*

**Code -**

proc sql;

select stock\_group, sector, count(\*) as stock\_number

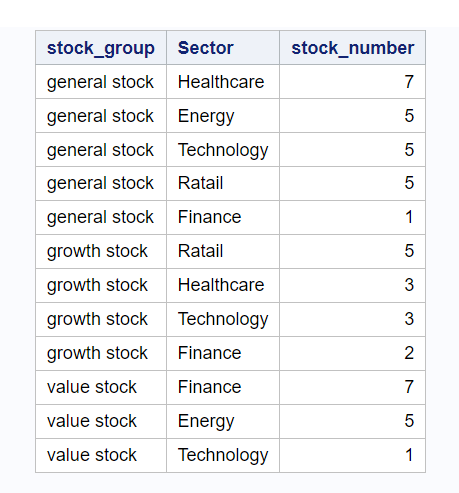
from stock\_group\_sector

group by stock\_group, sector

order by stock\_group, stock\_number desc;

quit;

**Output -**



Key observations: The retail sector boasts the highest number of growth stocks (5), while the finance sector has the highest number of value stocks (7). Based on this, we can predict that the finance sector is likely to outperform other sectors in the next five years since it has the largest concentration of value stocks.

* The average return ratio of each industry (2018-2022)

*#Get average returns of each industry*

**Code -**

proc means data=stock\_industry;

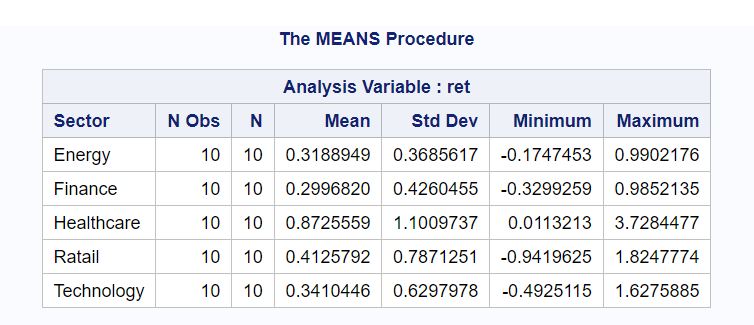
var ret;

class Sector;

ways 1;

run;

**Output -**



Key observations: The healthcare industry has the highest return ratio, almost twice as much as the second-highest industry, which contradicts our previous prediction based on the PB ratio. The healthcare industry has been significantly impacted by the pandemic, which could have contributed to its impressive performance. This reinforces our assumption that the COVID-19 pandemic has had a significant impact on the stock market.

**Conclusion and Findings**

In general, the stock price of value stocks should be the one with the most potential for growth. However, this trend has reversed in recent years. Over the past five years, growth stocks have outperformed value stocks, possibly due to the pandemic.

It is worth noting that the retail sector has a larger proportion of growth stocks, while the finance sector has more value stocks. Nonetheless, the healthcare industry has demonstrated the highest return ratio, providing evidence that COVID-19 has had a significant impact on the stock market.

**Appendix** - Company List