Kantonsschule Hottingen

Maturitätsarbeit

Value Investing with Big Data

The short-term performance of streamlined fundamental security analysis based on Benjamin Graham.

Lion Six

G4b-IM

Supervised by Thomas Kotur

13.12.2021

Content

1 Preamble ... 1

2 Introduction ... 2

2.1 Goal of This Paper ... 3

2.2 The Decreased Time Frame Explained ... 4

2.3 What This Paper Does Not Focus on ... 4

3 Procedure of the Experiment ... 5

3.1 Initial Graham Method ... 5

3.1.1 Defensive ... 5

3.1.2 Graham Number ... 6

3.1.3 Enterprise ... 6

3.1.4 NCAV ... 6

3.2 Development of the Adjusted Method ... 7

3.2.1 Coding an Algorithm for Data Filtering ... 8

3.2.2 Optimization and Splitting of the Algorithms ... 12

3.3 Storing the Data ... 12

3.4 Maintaining the Data ... 13

3.5 Overview and File Structure ... 14

4 Resources Used for the Experiment ... 15

4.1 Hardware ... 15

4.2 Software ... 15

5 Results of the Experiment ... 16

5.1 Plain Results ... 16

5.1.1 Exclusion of Certain Values and the Meaning of Fundamental Indicators ... 16

5.1.2 The Bubble Chart Explained in Detail ... 17

5.1.3 Eight-week Performance ... 18

5.2 Optimized Results ... 20

5.2.1 Cutting Winners, Letting Looser Run ... 20

5.2.2 Only Investing in Newcomers ... 20

5.2.3 Filtering for Graham 21
5.2.4 Filtering for NCAV 22
5.3 Comparing Key Metrics 23
5.3.1 The Adjusted Method and Major Indices as Benchmarks 24
5.3.2 Volatility and Sharpe Ratios 25
6 Conclusion 27
6.1 The Hypothesis and its Validity 27
6.2 Can the Adjusted Method be Profitable? 27
6.3 Challenges Faced During Development 28
6.3.1 Problems With Web Scraping 28
6.3.2 Problems With Data Maintenance 29
6.4 Outlook 29
6.5 Closing Statement 30
7 Bibliography 31
8 Index of Equations, Illustrations, and Tables 35
8.1 Index of Equations 35
8.2 Index of Illustrations 35
8.3 Index of Tables 35
9 Erklärung 36
10 Arbeitsjournal 37
11 Glossary 40
12 Appendix 43
12.1 Code of the Main Algorithm in Python 3.9 44

1 Preamble

Imagine you are driving a car; this is your beautiful car. But there is a catch: there is fog in front of your car. The fog is so thick that you can barely see the end of the hood. This fog will never go away, and you still need to get from point A to point B. Now you can modify this car however you like with almost no restrictions: You can add all the maps you want into your satnav system, consult all the experts on road navigation you may find, and you can drive at any speed you like. The only thing you cannot do is look ahead; you cannot see what is in front of you. This is what trading comes down to at its core. No matter who you are, how powerful or wealthy, you will never be able to see ahead. The future is unknown.

But what you can do is try to make a guess and make that guess as well informed as possible. In the analogy of driving a car through the fog, this would be using a map, retrieving past knowledge of the road, or taking other cars around you as an indicator for where the road might lead you. In the stock market, there are similar tools to aid you. There are the industry and market indices, economic data and surveys, the monetary policy decisions of central banks, the earnings and fundamental figures of individual companies, and the movement in other financial markets like bonds, futures, or in recent times even cryptocurrencies, as well as truckloads of historical data. Additionally, like driving slowly on a foggy road, one can take measures to minimize the risk in the market further. These measures include but are not limited to diversification, position-sizing, stop-losses, and *technical analysis*. I believe that if one uses a wide range of these tools given to them by the market, the demanding task of achieving above-average returns gets substantially easier.

In the early days of the modern stock market, respectable investors could be seen going through newspapers and calculating economic figures with pen and paper. Nowadays, the aids from the paragraph above are available to everyone with an internet connection, for free, at any time and place. Pen and paper have been replaced by spreadsheets and the newspaper by a Bloomberg Terminal or Yahoo Finance. Institutional traders, one storage above the trading floor in New York or Chicago, once had an information advantage that now is mostly gone.

It was this change in conditions that motivated me to write this paper. Being interested in the world of financial markets and computer science, I wanted to connect these two fields by creating an algorithm that utilises some of the tools listed above and weaponizes them through the power of possible infinite data mining. The result of this algorithm would be a strategy that should be a worthy opponent for any professional in the stock market.

I want to thank my family, my classmates, Mrs Schmoelzl and my supervisor Mr Kotur for their extensive support in the conception and the writing of this paper.

2 Introduction

Predicting the stock market is a conundrum. It is not possible. However, one can try to identify occasions where the market is ‘wrong’, with assets being either valued more or less than they are worth in the eyes of the individual investor. Expecting a correction of these mispriced assets, one can then try to profit from these movements. Although, according to the Efficient Market Hypothesis developed by Samuelson and Fama[1], this should not be possible. The Efficient Market Hypothesis states that firstly, all assets are valued correctly, and secondly, that there is no such thing as a free lunch as it is further argued below:

> *There is no free lunch on wall street. A free lunch in investing cannot exist because of the trade-off investors make between risk and reward. The greater the inherent risk in an investment, the greater the reward. This is a fundamental truism. Conversely, securities with less risk generally have commensurately lower returns. So, the notion of riskless reward is, for the most part, a theoretical concept that provides fodder for academic discussions. On the rare occasions when this does occur, it will quickly be snuffed out by arbitrageurs who, by their actions, eliminate the inefficiencies that gave rise to the free lunch.* (Ganti, 2021)

Learning about the Efficient Market Hypothesis may be frustrating because this means there is no edge to be gained in the market. If an investor or trader now still attempts to be successful in their endeavours at the stock market, he or she might want to start by looking at people that have already succeeded in this field. While copying someone else’s work gets people thrown out of college, it does not get anyone thrown out of the stock market. Looking at working strategies that others have developed might help guide an aspiring investor or trader in the right direction[2].

One notable strategy that has been copied and reapplied numerous times is value investing based on *fundamental analysis*. Benjamin Graham is credited with creating this strategy in the 1920s[3]. Warren Buffet, a renowned investor and the first student of Graham, rose to fame due to his success in the stock market[4]. Coincidently, both were also strong opponents of the Efficient Market Hypothesis[5], believing that there were, in fact, stocks that were trading below or above their fair value.

Benjamin Graham wrote multiple books, one of which lays out a complete plan to employ his value investing strategy and identify undervalued stocks. Valuation of stocks based on the points laid out in his book is referred to as the Graham Method. The different Graham Methods

[1] (Vamvakaris, Pantelous, & Zuev, 2017, pp. 401-417)

[2] (Yahoo Finance, 2020)

[3] (The Heilbrunn Center for Graham & Dodd Investing, 2021)

[4] (Sommer, 2020)

[5] (Ekweruo, 2011)

are further explained in chapter 3.1. Graham came up with his methods in 1949, a time in which the process of investing was more time consuming and expensive than it is now[6]. Among other factors, stock information was not readily available, and the markets were illiquid. However, times have changed, and the financial world is more democratized than ever[7].

2.1 Goal of This Paper

The main hypothesis of this paper is:

To generate an above-average return by utilizing the power of big data in combination with a value investing method deployed as part of a short-term strategy.

This paper tries to utilize the dramatic changes in the financial world over the last decades and reapplies Graham's mantra of value investing. Therefore, this paper aims to adjust, tune, and combine the already known Graham Methods to create a new method. This new method, henceforth referred to as the Adjusted Method, tries to preserve most of the already existing benefits and risk mitigation characteristics of the original Graham Methods while delivering a high average yearly return. As part of the experiment of this paper, stocks deemed undervalued by the Adjusted Method are collected. Then, these stocks are implemented into strategies that focus on short holding periods of individual positions, among other characteristics, as explained in chapter 2.2.

Further, this experiment is greatly supported by the technology of the 21st century, allowing a great majority of tradable stocks to be combed through with an algorithm carrying most of the workload. This task will be completed with no funds allocated for fees or any other costs that might occur. This step should demonstrate that in today's day and age, stock market information of any kind is publicly available at no cost to the average trader and, by that, an almost level playing field between institutional and retail traders, at least on the information level, is in theory possible.

Essentially, this paper will reiterate the known Graham Methods and describe the development of an Adjusted Method derived from them. The paper will then describe how an algorithm was coded and set up to collect data. Finally, the paper will describe the findings, identify subgroups that had favourable performance, and discuss the experiment results.

[6] (Investopedia, 2021)

[7] (Insana, 2021)

2.2 The Decreased Time Frame Explained

The shorter time frame compared to the original strategy by Graham was motivated by the fact that compounding returns are more significant when the frequency of compounding increases[8]. For example, as laid out in the table (Table 1), the yearly return differs depending on the compounding frequency when investing 1000 U.S. dollars of capital with an interest rate of 20%.

Compounding Periods	Annual	Quarterly	Monthly	Daily
Excess Return in USD	200	216	219	221

Table 1: One year return on 1000 USD with a 20% interest rate and different compounding frequencies.

So, if a trader can achieve a constant return of 1% per week and these returns compound, the trader would be way better off than an investor who averages a return rate of 52% per year. Even though, at first sight, both receive the same interest rate. The difference can be shown again with the 1000 USD initial capital example. While the investor receives 520 dollars in interest at the end of the year, the trader gets 678 dollars. Again, this is due to the compounding of the interest payments.

2.3 What This Paper Does Not Focus on

This paper tries to be as objective as possible and focuses on the quantifiable characteristics of stock price moves. The experiment involved zero investment decisions made by humans. The experiment does not test the Graham strategy as it is known for long term investing. This has already been done in other papers successfully and showed that Graham's Methods generally outperformed[9]. Furthermore, the countless value investors that followed Graham's rules and became profitable can further be seen as proof of his findings. This paper focuses on stocks that fit the base requirement, meaning being listed on the NASDAQ or NYSE. This decision was made because the US markets are the most liquid and active and offer retail customers the most accessible and reliable data[10]. However, this criterion excludes most non-US companies, penny stocks, and other micro-caps almost exclusively listed on over-the-counter[11] markets.

[8] (Chen, Boyle, & Courage, 2021)

[9] (Net Net Hunter, 2021)

[10] (Largest stock exchange operators worldwide as of September 2021, by market capitalization of listed companies, 2021)

[11] OTC equity markets are subject to fewer regluations. Major equity OTC markets are the OTCQX, OTCQB and OTC pink.

3 Procedure of the Experiment

3.1 Initial Graham Method

Benjamin Graham laid out core principles concerning stock valuation in his book, “The Intelligent Investor”[12]. These principles can be looked at as a checklist ensuring what Graham called a margin of safety when purchasing a stock. This margin of safety materialises in the stock being traded for less than the investor believes the company and the stock itself are worth. Graham further stated that the standards set by these principles would ensure two things: First of all, a company had a minimum level of quality in its past performance and current financial position. Second of all, a company had a minimum level of quantity in terms of earnings and asset per dollar of price. Graham split stocks that he analysed into three categories: defensive, enterprise and NCAV.

A stock that conformed to the criteria and was sufficiently undervalued, allowing for the margin of safety, was added to the portfolio. The stock was then held for two years or sold after the stock was not undervalued anymore, which means that a positive price change, an adverse change in fundamentals or the simple passing of time could lead to a sell.

3.1.1 Defensive

As laid out by Graham in chapter 14 of his book, the defensive checklist contains seven points[13] that a stock needs to fulfil. The defensive criteria are very restrictive, which according to Graham ensured the highest quality of stocks[14]. The seven points are:[15]

1) More than one billion dollars in annual sales[16].
2) *Current assets* should be at least twice the size of *current liabilities* and long term debt should be lower than net current assets[17].
3) Earnings Stability: Some earnings for the common stock in each of the past ten years.
4) Uninterrupted dividend payments for at least the past 20 years.
5) At least a 33% increase in per-share earnings in the past ten years using a three-year average.

[12] (Graham, 1949, p. 347)

[13] (Graham, 1949, pp. 348-349)

[14] (Graham, 1949, pp. 347-350)

[15] The checklist is recited by the author but equivallent to (Graham, 1949, pp. 348-349).

[16] Adjusted for inflation, 50 million in 1949.

[17] An irrelevant exception for public utility companies has been left out.

6) Stock price should be less than 15 times average *earnings per share* of the past three years.

7) Stock price should be less than one and a half times the *book value* per share.

3.1.2 Graham Number

The last two principles of the defensive checklist can be summarized together to create what is called the Graham Number[18]. Which is calculated as follows:

$$Graham\ Number = \sqrt{22.5 * EPS * BVPS}$$

Equation 1: Graham Number Calculation

22.5 in Equation 1 is the product of the values 15.0 from point six and 1.50 from point seven in chapter 3.1.1. Coincidentally this number is close to the current expected *P/E ratio* of the *S&P 500 index*[19].

3.1.3 Enterprise

As laid out by Graham in chapter 15 of his book[20], the enterprise checklist contains six points, which a stock needs to fulfil. The enterprise checklist is suited for more aggressive investors, allowing for a broader selection of companies while still upholding the mantra of discovering undervalued stocks. The six points are:[21]

1) P/E ratio of less than 10.

2) Current assets more than one and a half current liabilities and debt no more than 110% of current assets.

3) Earnings stability: No deficit in the last five years.

4) Some current dividends.

5) Earnings to be higher than those four years ago.

6) Stock price less than 120% net tangible assets.

3.1.4 NCAV

NCAV or Net current asset value is another critical number on which Graham based a check-list. It is calculated on the next page as follows:

[18] (Serenity, 2016)

[19] (Wall Street Journal, 2021)

[20] (Graham, 1949, pp. 385-386)

[21] The checklist is recited by the author but equivallent to (Graham, 1949, pp. 385-386).

$$NCAV = \frac{Current\ Assets - Total\ Liabilities}{Ordinary\ Shares\ outstanding}$$

Equation 2: Net Current Asset Value Calculation

This checklist was simple and only contained two points, laid out by Graham at the end of chapter 15 of his book[22]. This checklist, while rather elementary, is quite popular among value investors. "Net Current Asset Value is Graham's most well-known category of stocks […]"[23]. The two points are as follows:[24]

1) A diversified group of common stocks at a price less than the NCAV.
2) Eliminating those that reported a net loss in the last 12-month period.

As mentioned in the first point, for a diversified group, Graham recommended 30 stocks or no more than 3.3% of the Portfolio allocated to a single stock of the NCAV category[25].

3.2 Development of the Adjusted Method

Graham focused on value stocks, and his checklists represent that. While the Adjusted Method shall be close to Graham's mantra, it needed to be adjusted for two things. Firstly, the planned extreme downsizing on the time axis from two years to a couple of weeks and secondly the current market conditions. Even though value and growth switch back and forth regularly depending on the business cycle, growth stocks have dominated for the last ten years[26]. So, the Adjusted Method needs to shift its focus away from value stocks and start broadening the conditions to allow, not necessarily typical growth stocks but instead value stocks with growth potential and vice versa. The Adjusted Method was created from the defensive checklist in chapter 3.1.1. However, all points except the last two were removed. Not all of them were removed because they necessarily hindered the inclusion of growth stocks but rather because they were either unnecessary or too restrictive. The first point in the defensive checklist is a perfect example. While only including stocks with a relatively solid market cap is a great idea most of the time, in this case, it may exclude valuable stock information from being gathered during the data collection, and the more information, the better. The Adjusted Method now consisted of what essentially was the Graham Number from chapter 3.1.2. This alone was not a strict enough requirement to satisfy at least some connection with Graham's mantra, so the Adjusted Method was then perpetuated by adding the first point of the NCAV checklist of chapter 3.1.4 to it. It needs to be further noted that the EPS from point six in the defensive checklist

[22] (Graham, 1949, pp. 390-391)

[23] (Serenity, 2016)

[24] The checklist is recited by the author but equivallent to (Graham, 1949, pp. 390-391).

[25] (Graham, 1949, p. 390)

[26] (Hardin & Bischof, 2021)

of chapter 3.1.4 was not calculated with the values of the last three years, but rather with the EPS over the last year[27] like described in point two of the NCAV checklist in chapter 3.1.4. This allowed for younger companies and even IPOs to be included as well. Eventually, the Adjusted Method, or rather the checklist that was going to be used in the preliminary data collection, consisted out of the following three points:

1) The Graham Number as a percentage of the stock price shall be at least 150%.
2) The NCAV value as a percentage of the stock price shall be at least 35%.
3) Only ordinary shares are allowed.

The cutoff values for the Graham Number were heightened compared to the defensive checklist, while the NCAV one was significantly lowered. This allowed more stocks to be picked up and included by the data collection, even if they had fewer assets than what Graham would have called ideal. The higher Graham Number countered this surplus of stocks, created by the less restrictive condition in point two. In addition, the higher Graham Number ensured that stocks with weak or no growth potential would not be included. A Graham Number as a percentage of the stock price being 100% would have been acceptable in the original defensive checklist. Because even if the stock was not technically undervalued, it was at least not overvalued and was still expected to outperform due to continued revenue growth and dividends. Nevertheless, for the fact that both conditions are missing in the Adjusted Method, a rise to 150% for the first condition was vital to preserve the spirit of the original Graham Method. This condition also ensured that companies with no earnings were excluded as demanded by Graham.

3.2.1 Coding an Algorithm for Data Filtering

Due to the sheer amount of data to be collected, the data collection could not be done manually, and an automated solution had to be created. Therefore, an algorithm was coded in *Python* 3.9. For interested readers, the entire algorithm can be viewed in the Appendix on page 44. The algorithm was split into five main steps, shown below (Figure 1) and are explained further in the following paragraphs.

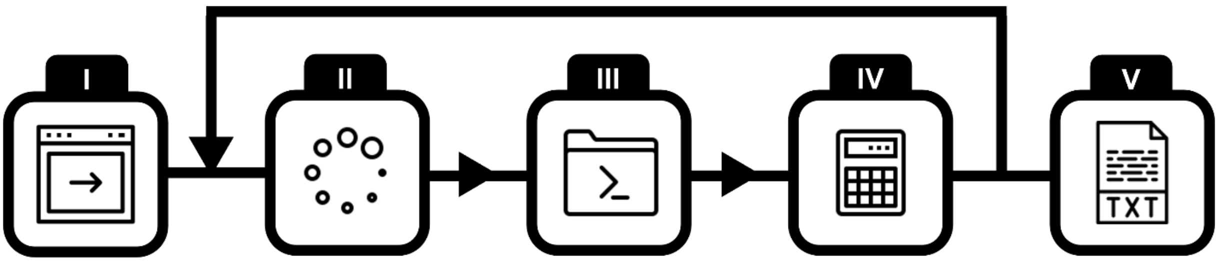


Figure 1: Process of the algorithm.

[27] Trailing twelve months, abbreviated as TTM

Step I

In the first step (I), the algorithm was set up. This included opening, reading out and converting a text (.txt) file into an array that contained a list of all tickers that were to be checked through. The tickers, which were to be checked against the conditions set in chapter 3.2, had been gathered from a website[28] run by Nasdaq, which offers certain additional services to traders. One of these services is a list of all traded tickers on the Nasdaq. After the list was imported into an excel sheet, it had to be filtered for any ETFs, ETPs, warrants or other tickers that traded like a stock but did not have the same characteristics as one. This had to be done because some non-stock tickers, like ETFs, would not show values like “Current Assets” on Yahoo Finance. Therefore, calculating the key figures would not even be possible for these assets. Subsequently, the excel sheet was exported as a text file.The *Selnium* extension and the artificially created Google Chrome window were initiated and opened. Furthermore, the algorithm declared and assigned global variables and opened the output text file. If there was no existing output file, he created one to store the positive results. The input in this step contained around 3000 individual stocks.

Step II

In step two, the Yahoo Finance site of the first ticker in the array was loaded. This was by far the most time-consuming step. After loading the site, it was read out (“scraped”) by the extension *Beautiful Soup* and assigned to a variable of the type “soup”. After that, the website was saved and could be used for further reference in the following steps without reading out the whole website again. This was primarily done in the code lines 47 to 59.

Step III

In step three, the web scraping continued by combing through the website, now saved as the soup variable locally. The locations of the fields of interest were trying to be located. This was rather difficult because Yahoo Finance lacks usability when their website is accessed by scraping. This is due to the fact that Yahoo Finance is providing financial information for free in return for the user’s toleration of advertisements. When accessing the website with an algorithm through the “backdoor”, Yahoo Finance does not get to display advertisements to a human user, and by that takes a slight loss due to the hosting costs. These circumstances lead to Yahoo Finance actively and passively combating web-scraping, a deeper explanation of which can be found in chapter 6.3.1. The algorithm first located the table which displayed all the data, highlighted in the picture (Figure 2) with a dashed black border. After that, the algorithm identified all table rows by their shared class name and saved them in an array called “rows” (Figure 3, 079).

[28] ftp.nasdaqtrader.com

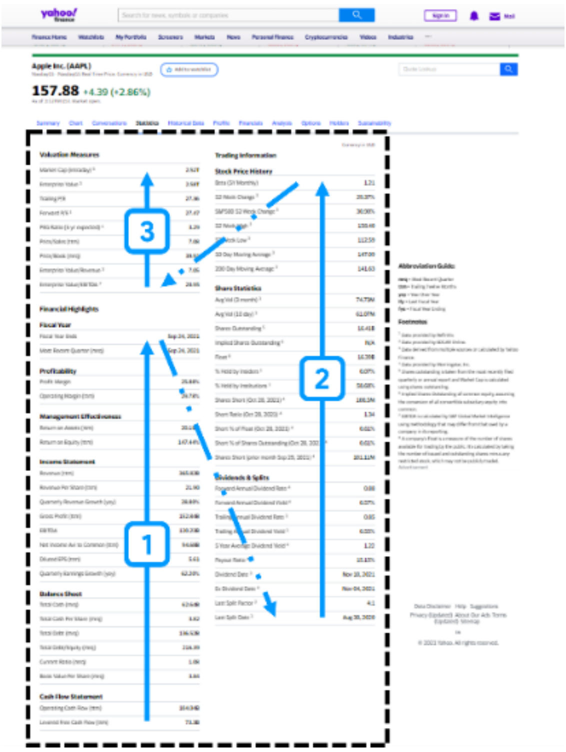


Figure 2: Screenshot with sections of interest highlighted (Yahoo Finance, 2021).

This allowed referencing only specific rows just by their index position. However, finding the position of these rows that contained values of interest was tricky because the table was arranged somewhat counterintuitively. In the website's hierarchy, the rows had been sorted and arranged differently and were also saved in this order in the array. So, as depicted in the illustration (Figure 2) by the light blue arrows, the table had to be read out from the bottom left column, jumping to the bottom right column after reaching a bit more than the midway point, to then lastly, jump back to the left column again.

This arrangement obviously was not identifiable to the user's eye and had to be discovered through tedious trial and error. Eventually, the rows of interest were identified and could be referenced by their now discovered index position (Figure 3, 106).

To read out a row of interest, the row value was assigned to the soup variable (Figure 3, 107). The value, in this case EPS, could then be read out and formatted easily as a string, made possible by the previous conversion to the soup variable (Figure 3, 112). After all values of interest were read out and assigned to their variables, step III and the web scraping were concluded.

```
078    #get all rows of the stockdata table
079    rows = list(StockData.find_all(class_="Fw(500) […] Miw(60px)"))
[…]
106    if i == 51:
107        soup = rows[i - 1]
[…]
112        EPS = float(soup.text.replace(',',''))
```

Figure 3: Selected lines of the code (78, 79, 106, 107, 112).

Step IV

In step four, all values collected were used in their respective formula to calculate the Graham and NCAV percentage. This step slightly varied, depending on if the algorithm was in the first or second cycle, as explained in chapter 3.2.2. Generally, the code looked similar as illustrated

(Figure 4), with the code snippet simplified to be more legible. First, the Graham Number is calculated as laid out in the formula in 3.1.2 (Figure 4, 138 & 142). The Graham Number is then transformed into a percentage to ensure it is proportional to stock price (Figure 4, 147). In lines 149 to 151, it is noticeable that the algorithm is checking two things. Firstly, it checks if the Graham percentage is sufficient to be added to the second cycle, and secondly, it examines if this is the first cycle. If both conditions are true, the stock is shown to the user in the console log. The user can then copy and paste all the results from the first cycle into the file for the input data in the second run cycle. If the Graham percentage was below 120%, the cycle ended here, and the next stock request was made. If the Graham percentage was sufficient, but the algorithm was no longer in the first run-through, the NCAV percentage was calculated. The calculations of the NCAV percentage involved loading a different site and gathering more data but boiled down to eventually employing the formula from chapter 3.1.4 (Figure 4, 195-197). If the NCAV percentage was sufficient, the stock and some other info, as explained in chapter 3.3, were written to a text output file.

```
138    Graham = 22.5 * BVPS * EPS
[…]
142    Graham = math.sqrt(Graham)
[…]
146    #Graham percentage
147    GrahamPer = (100/float(StockPrice.replace(',','')))*Graham
[…]
149    #first run trough: check if this value meets the treshhold
150    if GrahamPer > 120:
151        if firstRunTrough:
152            print(tickers[y])
[…]
154        #second run trough: calculate NCAV
155        else:
[…]
195            Denominator = (CurrentAssets - CurrentLiabilities) * 1000
196            ncav = Denominator / SharesOutstanding
197            NCAVpercent = (100.0 / float(StockPrice) ) * ncav
```

Figure 4: Selected lines of the code (Graham and NCAV calculations).

Step V

In the last step, the driver and the output file were closed, as they were no longer needed. Furthermore, the timestamp noted at the start of the algorithm was compared to the current time, and the difference was shown to the user, letting them know how long the algorithm had taken. With this, the algorithm of the Adjusted Method was concluded.

3.2.2 Optimization and Splitting of the Algorithms

The amount of data, which needed to be checked through, was significant. Because of that, every split second saved during a single data request would save a substantial amount of time. Therefore, it was vital for the usability of the algorithm to run as fast as possible. Now, this obviously needs to be put in perspective when combing through most of the Western stock market because there are a lot of different stocks. Therefore, the term “fast” in the context of this algorithm meant anything less than five hours. With this, the algorithm could be started at the beginning of a day and would be done around noon, leaving enough time to rerun it in case of problems. It was nowhere near the magic five-hour limit when the algorithm was first created. To make it more efficient, one would naturally look at the steps of the process, which take much time. It came to light that these were not “internal” factors like calculating or processing data but rather an “external” one, with the limiting factor being the speed of the Chrome driver loading the next website.

Possible improvements could have been using a faster workstation, switching to a system that was a server-side run or getting a faster broadband connection. The workstation and the internet connection were already pretty good, though, and setting up a server to run the algorithm would have cost a good amount of time and money. A much easier way to fix this was simply to reduce the occurrences where a page had to be loaded. For example, to calculate the NCAV, one would have to look at the “Current Assets” value, which could only be accessed from the “Financials” subsection of the stocks Yahoo Finance site. While the “Summary” subsection contained all values needed for the Graham calculation, it did not so for the NCAV calculations. Additionally, the spam protection, further explained in chapter 6.3.1, was way more restrictive on the “Financials” subsection than on the “Summary” one. While one could make about 100 web requests to read out individual stocks before hitting a cooldown for five minutes on the summary page, the same five-minute cooldown started after only a bit more than ten requests on the financials page. So, while calculating the NCAV needed two pages loaded, one of which with a very restrictive cooldown mechanism, calculating the Graham Number only needed one page.

The solution was to split up the two steps into two cycles. In the first cycle, all tickers had their Graham Number calculated, which left over about 4% of the original stocks and filtered out the others which were not eligible. Only now, the second cycle, which included calculating NCAV, was concluded with the remaining 4%. This reduced the time the algorithm needed dramatically to under five hours.

3.3 Storing the Data

The data was stored in an excel file. Excel was used for simplicity and efficiency because it is relatively easy to transfer the output data files into the excel sheet. The excel sheet contained all the information of the algorithm's output file. The output file contained six data points further explained in the table below (Table 2).

Attribute	Description	Relevance
Ticker symbol	The symbol through which the stock is traded.	The ticker was the key identifier to every stock and could be used for future algorithms or data requests.
Website	The Yahoo Finance website corresponding to the ticker symbol.	For first time entries, the data had to be re-checked. Especially values like the EPS or if the stock was even actively trading were vital.
Graham %	Chapter 3.1.2	Key data point that was the focus of the scrap-ing.
NCAV %	Chapter 3.1.4.	Key data point that was the focus of the scrap-ing.
Spot price	The price at which the stock last traded.	Used for the hypothetical P/L calculations when sourcing future prices (chapter 3.4).
Date of analy-sis	The date on which the data was sourced.	Firstly, to match up the index moves with the weekly data collection, secondly just general data collection "housekeeping".

Table 2: Attributes of the data collected.

3.4 Maintaining the Data

The data was maintained with another algorithm, a second python script. Henceforth referred to as the Refresh algorithm. The Refresh algorithm featured some of the functions of the main algorithm but was slimmer and, with that, incredibly more efficient. First, a list of tickers, the ones added to the excel file up to this point, was fed to it. It then scraped the websites of these stocks in the same fashion as explained for the main algorithm from chapter 3.2.1.

In contrast to the primary algorithm, only one value, "price", had to be collected. This meant no loading of the financials subsection – which saved a lot of time. Additionally, no calculations whatsoever had to be done to decide if a stock was eligible or not, which also cut down on a little bit of time. The Refresh algorithm was a lot faster than the main one for several reasons. For one, instead of around 3000 individual tickers, the Refresh algorithm only had to deal with the ones already collected in previous weeks, which at the peak were about 600. By this, the time needed for webpage loading was reduced by more than 80%. A simple caching mechanism further expedited this process. Every stock price, which was fetched by the Refresh algorithm, was added to a local list. Before loading the webpage of a ticker, the Refresh algorithm would first check the list. Therefore, if the stock had already been scraped this session, the algorithm could take the value from the list, skipping the web scraping. All the improvements above led to the Refresh algorithm taking only five minutes to collect all current price data.

3.5 Overview and File Structure

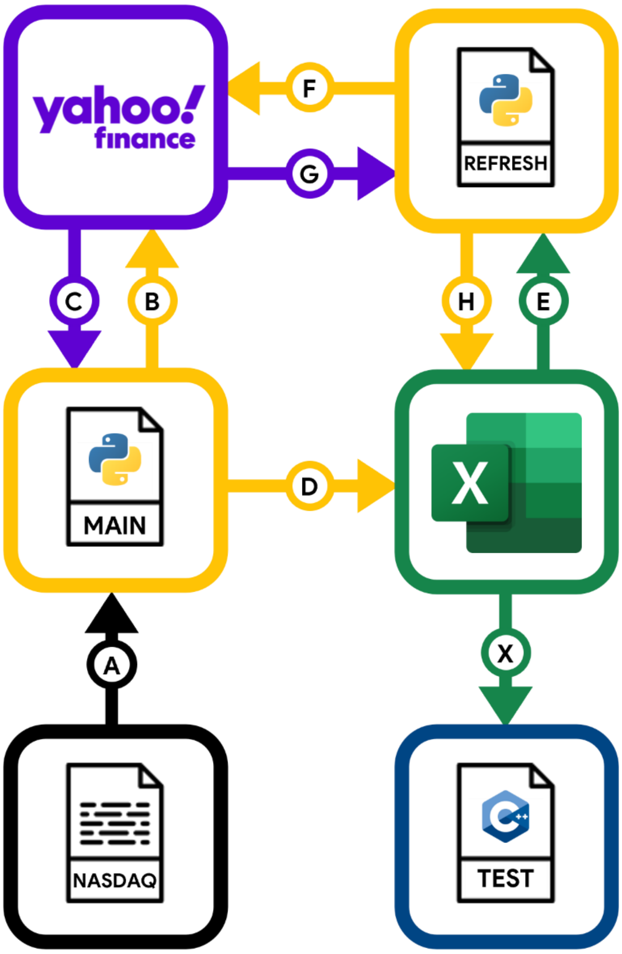


Figure 5: Overview and file structure of the experiment.

A: Input file containing the tickers to be checked, excluded ETFs, ETPs, Warrants, Units or Preferred Shares.

B: Main algorithm described in chapter 3.2.1 gathers information from Yahoo Finance.

C: Main algorithm processes the data from the web scraping and calculates Graham and NCAV percentages. In an average week, the first cycle (Graham) would shrink the original input data set of 3000 stocks down to 120. After the second cycle (NCAV), the 120 stocks from the first cycle would come down to 27 final outputs.

D: The output text file is imported into excel as described in chapter 3.3.

E: The collected tickers from the excel file are also saved in a separate text file. This text file is sorted chronologically and separated by weeks. This text file is then fed to another python script.

F & G: The Refresh algorithm then fetches the newest price from Yahoo Finance every week and saves it into a text file, as explained in chapter 3.4.

H: The text file is then implemented into the excel file. This had to be done manually because excel did not support the required import behaviour. Through the separation of step E in the input file, the output file is also separated. The sections are copy-pasted one by one, each being shifted accordingly to their original collection date to preserve the integrity of the data collected.

X: Certain data windows could be exported from excel into text files, e.g., all stock with an eight-week price history. These text files can then be read in by a C++ script, which can recreate past and simulate future market environments. The C++ script boosted efficiency when analysing the results, one of its example applications was in chapter 5.2.1, where it helped speed up the process exponentially.

4 Resources Used for the Experiment

4.1 Hardware

The algorithm was coded on a Windows desktop machine from 2019. The hardware specifications were an Intel CPU type i7-7700 with four cores, a graphics card from Nvidia type GeForce GTX 1050 and 8 GB of RAM. This is a somewhat low-level setup compared to today's standards, receiving only a score of 38% on a benchmark website[29]. This fact was not very hindering during the coding phase because the code was tested only on a small scale. The data collection was also run on the same machine, at which point performance problems did begin to show occasionally.

4.2 Software

The main code for the data collection was written in coding language Python version 3.9. Python was chosen for multiple reasons, the main ones being ease of use and effectiveness. Additionally, the framework Beautiful Soup was used. "Beautiful Soup is a Python library for pulling data out of HTML and XML files"[30]. Beautiful Soup controlled and navigated the Chrome browser for Python. Both Python and Beautiful Soup are open source and free of charge. The Python code was written in an IDE short for "Integrated Development Environment", called PyCharm. PyCharm is a product by JetBrains, which requires a yearly subscription at a current cost of around 200 USD[31]. Luckily, the company also offers a free version, the PyCharm Community version, which is Apache 2 licensed, meaning it is free, even for commercial use, and open source[32]. The collected data was stored in an Excel file. Excel is part of the Microsoft 365 Personal yearly subscription, costing around 70 USD[33]. A second algorithm was created in the coding language C++ for the detailed analysis of the excel data and the custom simulation of past and future market moves. C++ was used because it is one of the fastest programming languages, which comes in handy when playing around with massive amounts of data[34]. The windows text editor was used for the data interchange between the data collection algorithm, the excel sheet, and the data analysis algorithm. This was an easy solution to transfer data points between different programmes that were not already internally connected.

[29] (Rate my PC, 2021)

[30] (crummy.com, 2021)

[31] (JetBrains, 2021)

[32] (Haagsman, 2017)

[33] (Microsoft, 2021)

[34] (@suryadbdeveloper, 2021)

5 Results of the Experiment

The experiment was executed over 27 weeks, with the first data collected on April 4th, 2021, and the final data collected on September 25th, 2021. The data set was maintained until October 2nd, 2021. The data was always collected on a Saturday, during which all stock markets and the futures market are closed. This ensured that no data was falsified during the rather lengthy cycle by major market or price movements. There were 717 results collected, which averages around 27 results per week. The 717 results consisted of 61 unique stocks, which averaged to around two unique results per week. Some of these stocks were collected in the weekly data collections only once, while others showed up nearly all the time. After adding a stock to the excel sheet, the price-performance data was collected through the Refresh algorithm as laid out in chapter 3.4. Therefore, data points collected in the first couple of weeks had substantially more performance data than those collected in the last weeks. For example, results collected in the first seven weeks of the experiment can be evaluated and compared to each other with performance-price data of the following 20 weeks after collection. On the other hand, results collected in the last week can only be evaluated based on the very next week. Because results decrease in significance the smaller the sample size, the data evaluation focused on the one to eight-week time frame. This ensured that there was a fair amount of data available, making the results more accurate. Subsequently, these two framework conditions apply in the following pages:

Data evaluated during a one-week time frame includes all 717 results. Data evaluated during an eight-week time frame includes only results which have price data for at least these eight weeks (around 510 results).

5.1 Plain Results

The 717 results, meaning 717 stocks with a Graham percentage greater than 150% and a NCAV percentage over 35%, had an average weekly performance of -0.13%. This weekly performance would net a return of -6.54% if scaled up to one year. The one-week performance is displayed in the bubble chart below (Figure 6, page 20). The bubbles represent individual results; their positions represent the Graham and NCAV percentage of these results.

5.1.1 Exclusion of Certain Values and the Meaning of Fundamental Indicators

Extreme Graham and NCAV percentage values were excluded in the bubble chart. This had two reasons. For one, Graham values that were not in the normal frame of around 150% to 700% are not always a good sign. In the same way, a dividend yield above 10% should be a warning signal in the current market, a Graham or NCAV percentage in the thousands should also raise some red flags. One example in the dataset for this would be the stock QIWI. The QIWI stock, a Russian fintech company, was collected on 26.06.2021. It had a Graham percentage of 1400% and a NCAV percentage of 2000%. The stock was trading at 11 USD on that date. While writing this paper (November 2021), the QIWI stock trades at 8 USD, a fall in price of 27%. How is this possible? The stock was deemed severely undervalued according to the Adjusted Method and projected to rise in price twenty-fold. However, the answer to this question and the explanation for this continued downtrend of QIWI was not displayed in the Graham or NCAV percentage. It was due to current events. The Russian Central Bank had fined the company earlier this year after an audit showed their books were not in order. Additionally, the Russian Central Bank restricted QIWI's core business activities[35]. These factors made QIWI highly unappealing for investors, which caused a massive price drop.

This example also helps to show a more general point about fundamental indicators. Namely, that certain Graham and NCAV values or even other more widely used indicators like the P/E ratio do not necessarily indicate that a stock is being undervalued or overvalued by the market, but rather that almost all stocks are priced correctly as according to the Efficient Market Hypothesis from chapter 2. Fundamental key figures are merely that and nothing more, and a high or low value does not have to indicate anything beyond the number itself. One well-known example of this is the company Tesla (TSLA), which has been called overvalued for years while still rising in price.

The second reason to exclude specific data points in the chart was design related. If the bubble chart displayed the few (around 20) breakout values, with similarly ridiculous high Graham or NCAV percentages, it would have made the chart a lot less readable because of the extreme scaling on the axes that would have been necessary.

5.1.2 The Bubble Chart Explained in Detail

When explaining the chart in more detail, it is helpful to use an example. At the bottom centre of the chart, the company trading under the ticker FAMI can be spotted with multiple bubbles. They represent different data collection times and, more importantly, different fundamental values. In the case of FAMI, the NCAV percentage stayed relatively constant around the 35% level, but the Graham percentage can be seen ranging from 300 to 500% depending on the bubble in focus. While the NCAV percentage and the stock price remained almost constant, the Graham value, i.e., either the company's EPS or BVPS, changed drastically. Moreover, while NCAV and book value are not precisely the same, they share many similar characteristics, so it was more likely that EPS was the essential factor in this case. However, eventually, the stock also changes on the Y-axis, as can be seen by the small FAMI bubble at position (280/220), probably caused by a price decrease. A move on only the X- or Y-axis is a rather unusual sight on this chart as most datapoints move in a diagonal. Because NCAV and

[35] (Smith, 2021)

Graham percentages are proportional to price, a diagonal move typically represents a move in the spot price without a change in the fundamental values.

Some bubble moves might seem unexplainable by any of these three concepts; thus, it needs to be noted that this chart does not account for any gaps in data collection. Meaning, a stock could be collected because it fulfilled the conditions set in chapter 3.2 and added to the bubble chart. It could then trade outside these boundaries for an extended time without being collected. When collected later, an earnings report or price action could have moved the bubble's position on the chart in ways that are not directly comprehensible. However, most bubbles on the chart follow a rather logical trend, moving primarily in a diagonal line. Considering consistent indicators, upwards / right-hand side moves indicate a fall in price, while downwards / left-hand side moves indicate a rise in price.

When looking at the bubble chart, one can see how certain areas are more clustered than others. One example is the lower-left corner of the graph between the origin, the Graham level 360 (X-axis), and the NCAV level 100 (Y-axis). This can be explained because generally, there are more companies, the lower the Graham or NCAV value. This trend can not only be seen in the bubble chart but can also be proven by the simple fact that while over 3000 tickers were scanned each week, only around 27 made it through both Graham and NCAV checks. This implies that if the Graham and NCAV cut off would have been set at zero, there would have been substantially more results every week and by that on the final bubble chart as well. Additionally, it can be assumed that a bubble chart including all NCAV and Graham values would have shown a distribution like Figure 6, with the bottom-left being the most densely populated region and the top-right being the most thinly populated one.

5.1.3 Eight-week Performance

The plain results were still not satisfying when stretching out the time frame from one week to eight. The table below shows the average change of the price between the time of the data collection and the referenced week in percentage (Table 3).

TIME	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
CHANGE	-0.11%	-0.39%	-0.31%	-0.14%	0.14%	-0.10%	-0.82%	-1.38%

Table 3: Performance over eight weeks in percentage.

One might notice that the first-week performance in the table is different from the first-week performance stated in chapter 5.1. The difference stems from the fact that only 511 stocks had price data up to eight weeks compared to the 717 that had price data for at least one week, so the database was not exactly the same. In this table, week 5 was the only profitable holding period. Nevertheless, with 0.14% over five weeks, this performance is annualized at a measly 1.4%.

Figure 6: Bubble chart of the one-week performance for every result collected, excluding certain Graham and NCAV % values.

5.2 Optimized Results

The total results are not reflecting an above-average return but rather a severe underperformance of the general market. No matter this fact, the data set can still be sorted and evaluated in numerous ways, identifying certain subgroups in the set that net a better return.

Multiple strategies, therefore, were explored based on the Adjusted Method. Some of these are listed and further explained in the following paragraphs.

5.2.1 Cutting Winners, Letting Looser Run

Because it is known that the initial Graham Method returns a positive Alpha over the long run, one could assume that the initial and Adjusted Method share enough characteristics to produce similar results over the long-term time horizon. Consequently, when managing a portfolio based on the Adjusted Method, one could cut the winners and let the losing position run. Because based on the initial method, even most losing positions will eventually turn “green”. So, a considerable strategy might be to open positions in all collected stocks of the week and close all winning positions at the end of the week. Losing positions are kept on until they turn positive or four trading weeks since the position's opening has passed. The returned capital is then directly invested into the newly collected stocks of the respective weeks. This strategy, tested on a small scale, showed some promising results while also causing problems. The strategy returned a positive 0.2% weekly average over the test period of three months. This is equivalent to a 10.4% yearly return. However, the strategy was very capital intensive because the first four weeks had to be funded upfront to ensure liquidity. This caused a small 100 USD individual position size to amount to a roughly 8500 USD capital commitment.

5.2.2 Only Investing in Newcomers

In this strategy, the data was sorted after the indicators collected in columns eight and nine in the main data table (chapter 11). In theory, a stock that had never been collected before and just now has reached favourable Graham and NCAV levels would perform better due to high volatility. Similarly, a stock that had once been collected then traded outside the range and has recently been recollected, with at least one week of no collection between the two instances, might also outperform. Furthermore, stocks that had not moved out of their “undervalued” position were less likely to do so at all. When looking at the chart (Figure 7), one can see that the stocks, which were collected for the first time (“First time collected”), outperformed every other group collected, including the S&P 500 that was collected as a reference index.

On the other hand, the “Recollected” stocks not only performed worse than the “First time collected” stocks but worse than all the collected stocks on average (“All collected stocks”). This was rather surprising, especially because when collecting data for an infinite amount of time, eventually, all stocks would have been collected at least once. Even though only slowly, the “First time collected” category would shrink away and mutate fully into the “Recollected” category. Consequently, however, it would be wise to invest only in stocks collected for the first time instead of all stocks collected and to stay away from “Recollected” stocks at all cost.

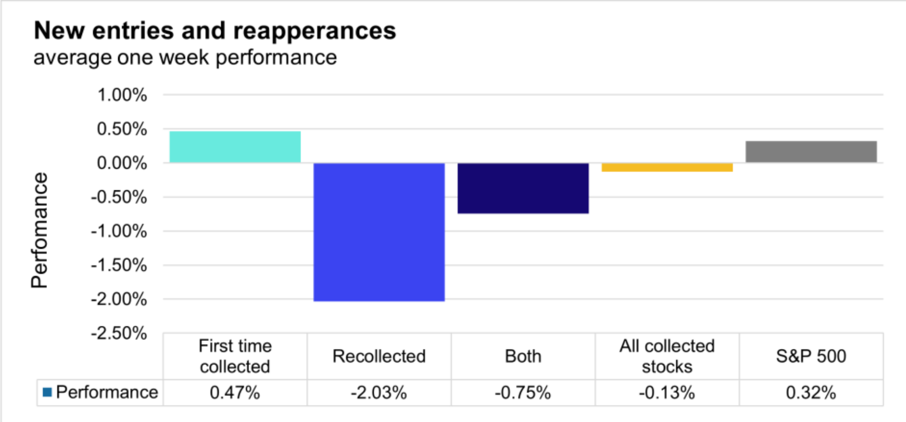


Figure 7: The weekly performance of new entries and recollected data points.

5.2.3 Filtering for Graham

Looking at all the stocks collected, it was clear that not all behaved indifferently. One attempt to sort the stocks and possibly identify the ones bound to perform better than others was done through the Graham Number itself. To accommodate this idea, the data set of stocks with at least eight weeks of price action data were sorted after their Graham percentage and grouped into meaningful categories, ensuring that each subgroup had about the same size. The groups were then ranked according to performance. Of the 18 subgroups created by the split, most were not yielding interesting results. They were performing not incredibly well and showed little deviation from the average by being relatively constant. Because these subgroups did not show interesting patterns and were either of negative value or insignificantly sized positive, they were not included in the optimized results. Six of the 18 groups, however, stood out. These are depicted in the graph (Figure 8). The lines represent different groups that are further specified in the graph's legend. The label of the lines stands for the group's starting and ending values. The black line, for example, represents the group of stocks with a Graham percentage between 160 and 170. For the analysis, one can note multiple fascinating things. First of all, even though stocks with a Graham value between 210 and 220 underperformed considerably in the first week, losing 1%, they vastly outperformed after that. They even set the maximum seen in the graph in week 3 with almost 4% returns on the initial investment. Similar characteristics are shared by the teal line representing the group with a Graham percentage between 270 and 280, even though they performed even worse initially and took considerably longer to return a positive result than the gold one.

In contrast to these two outlier lines stand the four other lines that all follow a similar trend of peeking in the first week and then falling sharply down to or below zero over the remaining period. The dark grey line group is most notable, with a Graham percentage between 320 and 350 returning 2.5% per week on average. Another captivating behaviour is seen in the blue line concerning Graham values between 220 and 230. This line nets a positive return in the first week of 1% and then drops in the red for the second week to come back to the 1% mark in the third week[36]. It needs to be added that duplicates were not excluded. Hence, it is very well possible that the behaviour seen in the blue line is caused by good performing stocks being collected in one week of the experiment while simultaneously having been collected two weeks before already.

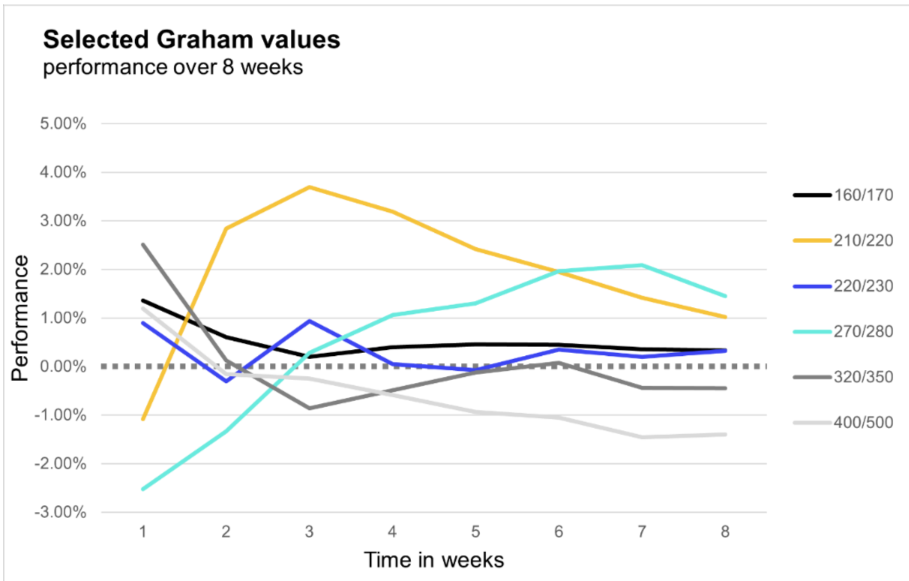


Figure 8: Line Graph of the performance of stocks with selected Graham values.

5.2.4 Filtering for NCAV

Similarly to how the data could be sorted after selected Graham values, the same can be done for the second key metric used in the experiment, the NCAV value. The sorting for the NCAV value was done in the same fashion as for the Graham values. Consequentially, individual data points were once again split, this time by their NCAV value, into groups. Of the 13 groups created by the division, only four showed an interesting pattern. For the same reasons as stated in chapter 5.2.3, the other subgroups were left out of the optimized NCAV results. However, these four groups of interest are displayed in the graph below (Figure 9). The patterns of the Graham graph (Figure 8) share certain characteristics to this one. The teal line representing the lowest NCAV percentages collected, between 35 and 40, shows a very high return in

[36] Assumed that not precisely the same positions cause the one per cent rise in the first week as the one in the third week, it could be viable to pursue a similar strategy as laid out in 5.2.1 but with a time horizon of three weeks.

the first week with 4%. The line then descends over the remaining seven weeks and even crosses the zero thresholds to become negative in the end.

Nevertheless, a 4% return per week is an impressive result. However, it would be advisable not to hold stocks with these NCAV values for more than a week. The blue and gold lines show a pattern also already seen in the Graham graph with peaks in weeks two and three. A possible strategy here would be to wait the first week before adding a stock with an NCAV value in the range between 70 and 100 to a portfolio. Finally, the grey line shows a similar trend as the teal one, starting at a high of 2% to slowly flattening out towards the end. The NCAV filtering showed that the stocks with the lowest NCAV value performed best in the short term. This is contrary to the initial Graham Method, in which a NCAV value of at least 100% in relation to the stock price was the requirement. Furthermore, this shows that having a high NCAV value may actually not be a sign for a company that will deliver short term trading gains. This is opposite to the filtering from chapter 5.2.3, in which the selected Graham categories with high cut off values performed equally or better than lower ones.

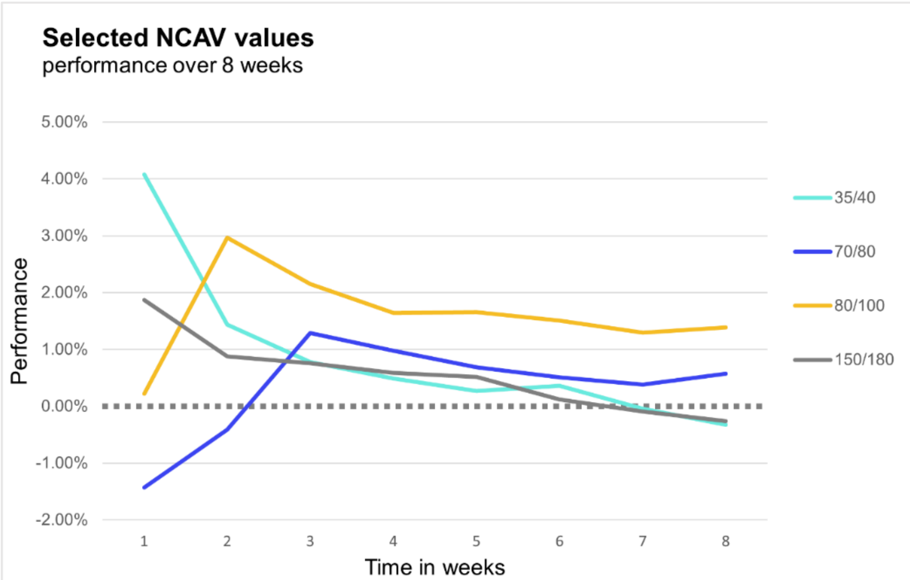


Figure 9: Line Graph of the performance of stocks with selected NCAV values.

5.3 Comparing Key Metrics

The values collected throughout almost six months of data collection can be compared to benchmarks. One of these benchmarks could be the risk-free rate of return, which is often measured by the return on a three-month Treasury bill[37]. This return rate would currently, October the 13th, amount to 0.05% for the US[38] and -0.83% for Switzerland[39]. With the risk-free

[37] (Hayes, Risk-Free Rate of Return, 2021)

[38] (Board of Governors of the Federal Reserve System, 2021)

[39] (Investing.com, 2021)

rate of return being close to zero, or even negative depending on which T-bill is used to measure, simply returning a profit will be counted as beating this benchmark.

5.3.1 The Adjusted Method and Major Indices as Benchmarks

More interesting benchmarks are the larger stock market indices: the S&P 500 index, the Russel 2000 index, and the DOW JONES Industrial index. Looking at the chart (Figure 10), the grey lines represent these three indices, while the coloured lines show the performance of the collected stocks split into subcategories. In the chart, all lines started at a value of 100, representing 100% of an original investment. The chart was created by taking the weekly average return of the categories displayed and multiplying them to the current value consecutively for the 27 weeks of the experiment. The gold line in the graph represents all collected stocks of the experiment, performing worst with a loss of 2.88%, not even beating the risk-free rate of return. Slightly better but still in the red was the Russel 2000 with a net loss of 0.54%. The S&P 500 was the top performer of the three indices, with an 8.39% return on capital during the measured time frame.

Outperforming all benchmarks were the “Graham fulfilled” and “NCAV fulfilled” subcategories. The blue line represented the return of the subcategory of stocks collected, which fulfilled the top Graham values. However, the Graham category did not represent the performance of all stocks with a sufficient Graham percentage over 150 because every stock collected had to have at least this value already. Due to this, the two categories of “Graham” and “all collected stocks” would have coincided. Instead, the “Graham fulfilled” category consisted of the top values of interest as shown in Figure 8 and included the following percentage ranges: 160-170, 210-230, 270-280, 320-350, and 400-500. The stocks that fulfilled this requirement achieved a total return of 21.62% over the experiment period.

The teal line in Figure 10 represented the return of the stocks collected, which fulfilled the top NCAV values. The “NCAV fulfilled” category was similarly structured to the “Graham fulfilled” category and consisted of the top values of interest as shown in Figure 9 and included the

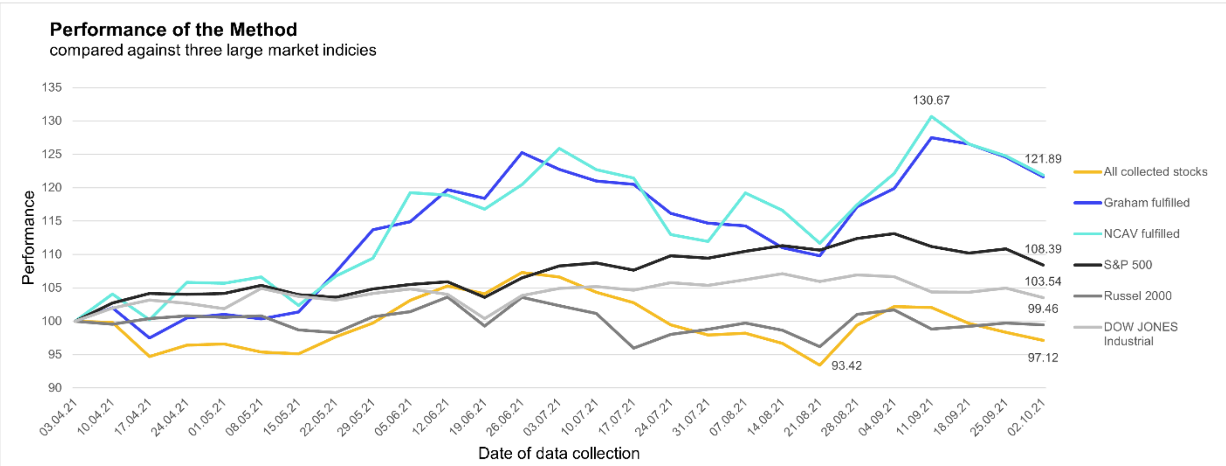


Figure 10: Performance of the Adjusted Method with benchmark indices.

following percentage ranges: 35-40, 70-100, and 150-180. The stocks which fulfilled this requirement achieved a total return of 21.89% over the 27 weeks, performing slightly better than the Graham subcategory. The "NCAV fulfilled" category not only outperformed all other displayed results but also achieved the highest return on the chart, being up 30.67% at one point. Additionally, it is interesting to note that the NCAV and Graham fulfilled subcategories show a fair correlation on the chart, clearly distinguishing themselves from the average performance of all collected stocks.

Extending the optimization of these results, one last category was created. This category was called "Graham x NCAV" and consisted of the "crème de la crème" of collected stocks. The top three performing subgroups from the Graham filtering (5.2.3) and the top two performing subgroups from the NCAV filtering (5.2.4) were hybridized. This meant that a stock had to be in the top three performing Graham groups and the top two NCAV groups. This created a slim intersection group of just 30 results out of the 717. These 30 results, however, produced a staggering 192.43% return throughout the experiment. Because of the magnitude of the return and the questionable scientific significance, it was not included in the chart (Figure 10).

In conclusion of this subsection, concerning the S&P 500 in particular and the stock market in general, it must be noted that the market was acting relatively bullish during the experiment. This resulted in a weekly return of 0.32% for the S&P 500 over the course of the experiment and a one-year performance of more than 25% as of this writing (November 2021). These values are considerably higher than the historical average for this index of around 7%[40]. With an annual return of 7%, the weekly performance of a given strategy would have to be around 0.13% to produce similar returns over the long run.

5.3.2 Volatility and Sharpe Ratios

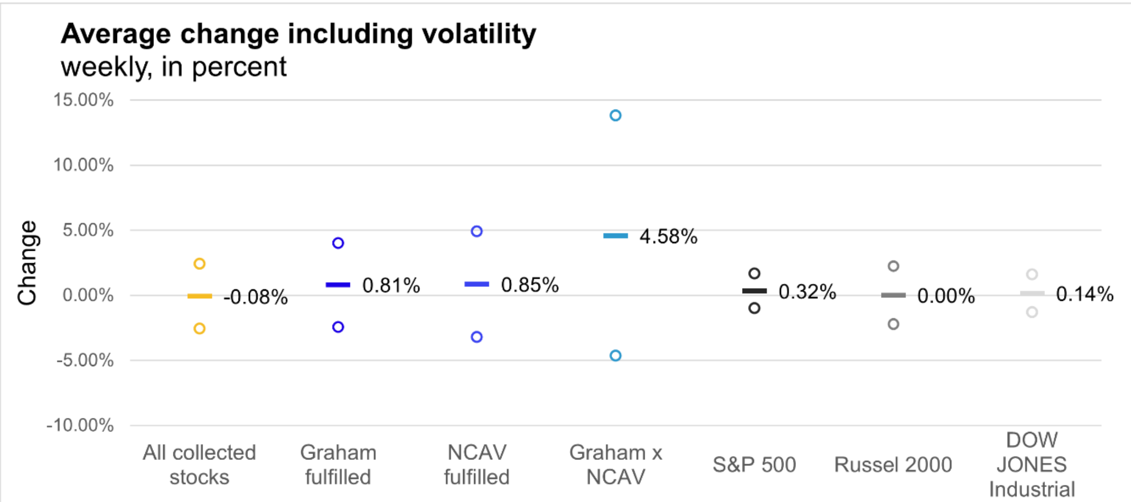
Another way of measuring the different methods is by using volatility. One way of measuring volatility is by calculating the standard deviation of the asset[41]. In the chart (Figure 11), the average change of the different strategies is displayed as well as the standard deviation or volatility from these moves. When adding and subtracting the standard deviation to the mean return, a range between an upper and lower value is created. Assuming a normal distribution, 68% of values will be in this calculated range. In the case of the subcategory "all collected stocks", this means that a trader employing the Adjusted Method, as laid out in 3.2, can assume to lose 0.08% per week, with 68% of all weekly moves falling between a -2.57% and +2.40% return, indicating a volatility value of 2.48%. The most obvious observation is the "Graham x NCAV" subcategory. While the return of this method, as described in chapter 5.3.1, is rather exceptional, so is its volatility. Based on this observation, it is to be noted that there also exists

[40] (Maverick & Mansa, 2021)

[41] (Hayes, Potters, & Rathburn, Volatility, 2021)

a disproportional relationship between the number of stocks one strategy included and its volatility. One might wonder why the average weekly return of the Russel 2000 is zero when in Figure 10, the performance was -0.54%. This comes down to the fact that in Figure 10, compounding returns were accounted for, while in Figure 11, the weekly returns alone were used to calculate the true average. Another notable occurrence in the chart is the rather high volatility of the benchmark indices compared to their returns. This shows once more that, in the stock market especially, average values are to be taken with a grain of salt and that while averages over the long run can be a trusted indicator, they are almost useless when the time frame is scaled down.

The Sharpe ratio of a portfolio is its expected return minus the risk-free rate, divided by volatility[42]. Considering a risk free rate of 0%, the formula can be applied to the subcategories. A Sharpe ratio of 2.37 and 2.62 for the Graham and NCAV subcategories, respectively, resulted from the calculations. The S&P 500 had a ratio of 2.65 and the “Graham x NCAV” one of 3.27. The Sharpe ratio of the subcategories with negative returns were not calculated due to having no useful meaning[43]. According to these ratios, the S&P 500, while being closely followed by Graham and NCAV subcategories, offered a better risk to reward ratio than all other categories except the outlier “Graham x NCAV”. However, this outlier plays into the Sharpe ratio's limitations because a normal distribution is assumed when calculating the ratio. This does not pose a problem when calculating with the S&P 500, a prime example of a diversified stock portfolio, but can be fatal when using a data set of only 30 data points that the “Graham x NCAV” category consisted of.



	All collected stocks	Graham fulfilled	NCAV fulfilled	Graham x NCAV	S&P 500	Russel 2000	DOW JONES Industrial
Average	-0.08%	0.81%	0.85%	4.58%	0.32%	0.00%	0.14%
upper	2.40%	4.03%	4.91%	13.80%	1.65%	2.24%	1.59%
lower	-2.57%	-2.41%	-3.22%	-4.65%	-1.01%	-2.23%	-1.31%

Figure 11: Average change per week and volatility, including benchmark indices.

[42] (Fernando, James, & Munichiello, Sharpe Ratio, 2021)

[43] (Fernando, James, & Munichiello, Sharpe Ratio, 2021)

6 Conclusion

6.1 The Hypothesis and its Validity

The hypothesis of this paper was to generate an above-average return with the Adjusted Method as part of a short-term strategy.

When trading a stock over short-term time frames, such as between one to eight weeks, the results were not profitable. The exception was the five-week time frame, but trading was only slightly profitable even in that case. The Adjusted Method did not outperform the broader market. Furthermore, when applied like in Figure 10, the Adjusted Method did underperform not only in comparison to the market, but also to the risk-free rate, and generated a loss overall. It is therefore not advisable to trade stocks during short time frames with this Adjusted Method based on the original method of Benjamin Graham. The experiment results further reinforced the prevailing dogma that fundamental factors are virtually worthless in short-term trading.

Looking at the results, the main hypothesis has to be rejected.

However, the question of whether competition with institutional trading is possible can be answered in the affirmative. Even though the plain results were negative, data collection and evaluation were possible and had little-to-no barrier of entry. The total cost for this experiment was close to zero, with the only expense being the initial purchase price of the hardware and the time invested. Even though the applied strategy might have been flawed, in a more general sense, this experiment has proven that data collection and calculation on scale is possible for anyone with a working internet connection and enough spare time. Therefore competition with institutional investors and traders is possible, at least on the information level. It can be argued that retail money even has an advantage over institutional money when investing in stocks filtered out by this method.

Because the stocks mainly were lower market cap, high beta companies, institutional money in these cases need to worry about price impact and liquidity while retail money with their limited amount of capital does not. Institutional investors and traders are theoretically beatable during any time frame, like any other market participant. But it must be noted that, on average, the performance of any given trader gets worse the shorter the holding period and that there is no evidence of persistent skill in short-term trading[44].

6.2 Can the Adjusted Method be Profitable?

Even though the Adjusted Method was not profitable, this does not mean there can be no money made from it at all. The data displayed in chapter 5.2.2 suggest that investing in the

[44] (Charkrabarty, Moulton, & Trzcinka, 2017)

subcategories “newcomers” can be profitable and generate an above-average return. The selected subcategories in chapters 5.2.3 and 5.2.4 outperformed all benchmarks, and the subcategory “Graham x NCAV” even outperformed the S&P 500 twenty-fold. However, for all these subcategories, it must be noted that the scientific significance diminishes with decreasing sample size. Furthermore, it is generally accepted that handpicked subcategories of a strategy only perform in this fashion during the selected time frame and are likely to perform vastly different under ever-changing market conditions.

While the experiment was done with great care to preserve the validity of the data set, it is also possible that certain data points are skewed. This is not problematic when looking at a subcategory with many data points but should be considered when looking at categories with smaller-sized data pools. For example, these errors could include the difference between Bid and Ask due to Yahoo Finance basing price data on the last trade. Moreover, timing errors due to the weekend data collection, including Friday night and Monday morning price fluctuations, which are notoriously hard to catch for retail traders, can not be ruled out.

Additionally, the two subcategories in chapters 5.2.3 and 5.2.4 had a similar Sharpe ratio to the S&P 500, indicating that a long-term investment in a broad market index fund would yield a similar risk to reward balance.

In conclusion, the experiment has shown that it is possible to be profitable with the Adjusted Method if the correct subcategories are picked, as seen in the Optimized Results. However, due to the many challenges, questionable scientific significance, and skewed risk to reward ratio, the recommendation has to be made that this strategy is not applied to short-term trading.

6.3 Challenges Faced During Development

6.3.1 Problems With Web Scraping

Letting an algorithm run through 3000 web requests creates two primary sources of problems. For one, a problem can occur when one of these 3000 requests yields an exception, and the output of that request, therefore, is flawed. Secondly, the environment of the algorithm must work without error for the time of the cycle, which was around five hours.

The first source of problems had to be tackled continuously. The algorithm would stop functioning if it had to do a calculation that was unsolvable. This meant that if Yahoo Finance could not find a value for EPS, and the algorithm then received a value like “0” or “N/A” or “ “ from the web-scraping, that the algorithm would still try to calculate the Graham Number, which would of course fail and cause a crash. So, every single value that was read out had to be checked against two primary fail-safes: Is the value available? (e.g., line 097). And, if only a certain range of values were excepted, is the value in this range? (e.g., only positive values, line 102). Another cause for crashes was the formatting of Yahoo Finance. Balance sheet values were separated by a comma every three digits, which confused the algorithm because it took those commas for the decimal point, and it crashed. Additionally, the field for "Shares Outstanding" was formatted in a way whereby the thousands and millions were abbreviated instead of written out. Hence, the algorithm had to take the value as a string, cut off the last character, convert the rest of the string to an integer and multiple with a factor depending on the cut-off character (line 117-134).

Problems with the environment were a lot easier to detect. For one, the Yahoo Finance website would deploy a spam protection after a certain number of requests. The workaround was to count the number of failed calculations, which could only occur because of a bad input or the deployed spam protection. If the counter reached five, that meant that the last five calculations had failed. Because failed inputs only occurred in about 1% of calculations, with an extremely high degree of certainty, this indicated the activated spam protection. The algorithm was then paused for the duration of the spam protection (around 5 minutes) and then picked up where it had first detected a failed calculation (lines 217-227). Other problems were the computer shutting down during the algorithm cycle and software updates for the Chrome browser, which had to be done manually every two months.

6.3.2 Problems With Data Maintenance

The problem with the data maintenance was mainly that stocks could get delisted, change their ticker, or get halted. Hence, sometimes a stock would be collected and no more data for that stock would be found later on. Sometimes the problem could be fixed by manually changing the tickers if that was the cause. However, in many cases, the company had been delisted. Stocks that were collected but did not show any price change over the following weeks had likely been halted, went bankrupt or were bought out without proper delisting procedures followed. Delisted stocks posed a challenge for the data collection because the calculation of key figures could be skewed. For example, a stock that had declined by -60% previously and was now delisted would show up as 0.00% in the data table, which had to be taken into account when evaluating the results.

6.4 Outlook

While the Adjusted Method did not yield promising results during most time frames analysed, it would be interesting to look at other time frames in future studies. While the data collection indicated that the performance got worse the longer the asset was held, no conclusion could be made about what would happen during even shorter time frames. Because it was not possible in this experiment to run the algorithm daily, the smallest time frame investigated was one week. Therefore, more work is needed to provide results on the use of this method during an intraday-based time frame.

The performance of the “Graham x NCAV” subcategory was so unusual that more work is needed to provide accurate results. This would be possible if the method was applied in the future, during new time frames and market conditions.

Another possibility to test the application of the Adjusted Method would be an experiment with different input data groups. As explained in chapter 2.3, some stocks, especially micro-caps and non-US companies, were left out. Hence, while a transformation to a different exchange region such as Asia would require adjustments to the methodology, it would certainly be interesting to see whether differences in the results would be found for other regions.

I enjoyed writing this paper, exploring the hypothesis, and I look forward to expanding my knowledge in this field in the future.

6.5 Closing Statement

This paper took the methods created by Benjamin Graham and applied them, modified, on a short-term time frame. The paper showed that some excess returns during certain time frames were possible, although the main hypothesis of this paper had to be rejected. In accordance with the car analogy stated in the Preamble, this paper has tried to provide tools to retail investors, aiding them through the fog of financial markets. However, driving a car in heavy fog is driving blindly, without any idea of the road ahead. And when driving without foresight, events, good or bad, will always carry a certain element of surprise.

7 Bibliography

@suryadbdeveloper. (2021, September 19). Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/top-10-fastest-programming-languages/

Adinarayan, T. (2021, June 30). *Retail traders account for 10% of U.S. stock trading volume - Morgan Stanley*. Retrieved from REUTERS: https://www.reuters.com/business/retail-traders-account-10-us-stock-trading-volume-morgan-stanley-2021-06-30/

Amadeo, K., Mansa, J., & Ernsberger, E. (2021, October 31). *The S&P 500 and How It Works*. Retrieved from the balance: https://www.thebalance.com/what-is-the-sandp-500-3305888

Bankrate. (2021, December 03). *Earnings per share*. Retrieved from Bankrate: https://www.bankrate.com/glossary/e/earnings-per-share/

Board of Governors of the Federal Reserve System. (2021, October 5). *3-Month Treasury Bill Secondary Market Rate*. Retrieved from Federal Reserve Economic Data: https://fred.stlouisfed.org/series/DTB3

Charkrabarty, B., Moulton, P. C., & Trzcinka, C. (2017, August). The Performance of Short-Term Institutional Trades. *Issue 4*(Volume 52), pp. 1403-1428. doi:https://doi.org/10.1017/S0022109017000400

Chen, J., Boyle, M. J., & Courage, A. (2021, February 23). *Compounding*. Retrieved from Investopedia: https://www.investopedia.com/terms/c/compounding.asp

crummy.com. (2021, September 19). Retrieved from Beautiful Soup Documentation: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

Ekweruo, C. (2011). *Efficient Market Hypothesis.* Retrieved from Grin: https://www.grin.com/document/183394

Fernando, J., Brock, T., & Li, T. (2021, November 28). *Price-to-Earnings (P/E) Ratio*. Retrieved from Investopedia: https://www.investopedia.com/terms/p/price-earningsratio.asp

Fernando, J., James, M., & Munichiello, K. (2021, October 04). *Sharpe Ratio*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/sharperatio.asp

Folger, J., & Schmitt, K. (2021, July 3). *Investing vs. Trading: What's the Difference?* Retrieved from Investopedia: https://www.investopedia.com/ask/answers/12/difference-investing-trading.asp

Ganti, A. (2021, July 22). *Free Lunch*. Retrieved from Investopedia: https://www.investopedia.com/terms/f/free-lunch.asp

Graham, B. (1949). *The Inteligent Investor.*

Haagsman, E. (2017, September 1). Retrieved from The PyCharm Blog: https://blog.jetbrains.com/pycharm/2017/09/pycharm-community-edition-and-professional-edition-explained-licenses-and-more/

Hardin, T., & Bischof, B. (2021, May 6). Retrieved from NASDAQ: https://www.nasdaq.com/articles/value-vs-growth%3A-a-brief-historical-view-2021-05-06

Hargrave, M., Anderson, S., & Bellucco-chatham, A. (2021, December 8). *Standard Deviation*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/standarddeviation.asp

Hayes, A. (2021, August 30). *Risk-Free Rate of Return*. Retrieved from Investopedia: https://www.investopedia.com/terms/r/risk-freerate.asp

Hayes, A., & Anderson, S. (2021, September 14). *Graham Number.* Retrieved from Investopedia: https://www.investopedia.com/terms/g/graham-number.asp

Hayes, A., & Scott, G. (2021, February 11). *Book Value*. Retrieved from Investopedia: https://www.investopedia.com/terms/b/bookvalue.asp

Hayes, A., & Scott, G. (2021, February 17). *Retail Investor*. Retrieved from Investopedia: https://www.investopedia.com/terms/r/retailinvestor.asp

Hayes, A., Battle, A., & Jackson, A. (2021, October 28). *Technical Analysis.* Retrieved from Investopedia: https://www.investopedia.com/terms/t/technicalanalysis.asp

Hayes, A., James, M., & Kvilhaug, S. (2021, October 20). *Current Assets*. Retrieved from Investopedia: https://www.investopedia.com/terms/c/currentassets.asp

Hayes, A., Potters, C., & Rathburn, P. (2021, October 30). *Volatility*. Retrieved from Investopedia: https://www.investopedia.com/terms/v/volatility.asp

Insana, R. (2021, July 31). *Insana: This is why Robinhood's democratization of investing is nothing new.* Retrieved from CNBC: https://www.cnbc.com/2021/07/31/insana-robinhoods-democratization-of-investing-is-nothing-new-.html

Investing.com. (2021, October 13). *Switzerland 3-Month Bond Yield*. Retrieved from Investing.com: https://www.investing.com/rates-bonds/switzerland-3-month-bond-yield

Investopedia. (2021, January 26). *Stocks Then And Now: The 1950s And 1970s*. Retrieved from Investopedia: https://www.investopedia.com/articles/stocks/09/stocks-1950s-1970s.asp

JetBrains. (2021, October 17). Retrieved from Subscription options & Pricing: https://www.jetbrains.com/pycharm/buy/#commercial?billing=yearly

Largest stock exchange operators worldwide as of September 2021, by market capitalization of listed companies. (2021, November 15). Retrieved from Statista:

https://www.statista.com/statistics/270126/largest-stock-exchange-operators-by-market-capitalization-of-listed-companies/

Maverick, J., & Mansa, J. (2021, May 31). *What Is the Average Annual Return for the S&P 500?* Retrieved from Investopedia: https://www.investopedia.com/ask/answers/042415/what-average-annual-return-sp-500.asp

Microsoft. (2021, October 4). Retrieved from Microsoft 365: https://www.microsoft.com/en-us/microsoft-365/buy/compare-all-microsoft-365-products

Net Net Hunter. (2021, December 01). *Is Benjamin Graham Still Relevant Today?* Retrieved from Net Net Hunter: https://www.netnethunter.com/benjamin-graham-still-relevant-or-a-complete-waste-of-time/#Just_What_Kind_of_Returns_Can_Investors_Expect

Palmer, B., & Schmitt, K. (2021, August 5). *Institutional vs. Retail Investors: What's the Difference?* Retrieved from Investopedia: https://www.investopedia.com/ask/answers/06/institutionalinvestor.asp

python. (2021, December 3). *What is Python? Executive Summary*. Retrieved from python: https://www.python.org/doc/essays/blurb/

Rate my PC. (2021, October 19). Retrieved from PC Game Benchmark: https://www.pcgamebenchmark.com/ratemypc?platform=windows&cpu=intel-core-i7-7700&memory=8gb&gpu=nvidia-geforce-gtx-1050

Segal, T., Boyle, M. J., & Munichiello, K. (2021, September 9). *Fundamental Analysis.* Retrieved from Investopedia: https://www.investopedia.com/terms/f/fundamentalanalysis.asp

Serenity. (2016, March 19). *Quick Reference: Graham and Serenity*. Retrieved from Serenity Stocks: https://www.serenitystocks.com/quick-reference#FootNotes

Smith, R. (2021, December 10). *Why Qiwi Stock Just Crashed 20%*. Retrieved from The Motley Fool: https://www.fool.com/investing/2020/12/10/why-qiwi-stock-just-crashed-20/

Sommer, U. (15. May 2020). *Warren Buffetts großer Lehrer: Wie Benjamin Graham Anlegern zu Rendite verhilft*. Von Handelsblatt: https://www.handelsblatt.com/finanzen/anlagestrategie/trends/serie-die-strategien-der-boersengurus-warren-buffetts-grosser-lehrer-wie-benjamin-graham-anlegern-zu-rendite-verhilft/26239024.html?ticket=ST-3415246-EOq4Gr4UserpTHJsER9V-cas01.example.org abgerufen

The Heilbrunn Center for Graham & Dodd Investing. (2021, November 12). *Benjamin Graham Value Investing History*. Retrieved from Columbia Buisness School: https://www8.gsb.columbia.edu/valueinvesting/about/history

Tuovila, A., Mansa, J., & Munichiello, K. (2021, October 8). *Current Liabilities*. Retrieved from Investopedia: https://www.investopedia.com/terms/c/currentliabilities.asp

Vamvakaris, M. D., Pantelous, A. A., & Zuev, K. (2017). Investors' Behavior on S&P 500 Index During Periods of Market Crashes: A Visibility Graph Approach. In K. G. Fotini Economou, *Handbook of Investors' Behavior During Financial Crises* (p. 401). Academic Press.

Wall Street Journal. (2021, October 15). *P/Es & Yields on Major Indexes*. Retrieved from WSJ MARKETS: https://www.wsj.com/market-data/stocks/peyields

Yahoo Finance. (2020, January 26). *What Is Copy Trading and Should You Do It?* Retrieved from Yahoo Finance: https://finance.yahoo.com/news/copy-trading-004534084.html

Yahoo Finance. (2021, November 18). Apple Inc. (AAPL). Retrieved from Yahoo Finance: https://finance.yahoo.com/quote/AAPL/key-statistics?p=AAPL

Zucchi, K., & Scott, G. (2021, August 18). *Institutional Traders vs. Retail Traders: What's the difference?* Retrieved from InvestopediaI: https://www.investopedia.com/articles/active-trading/030515/what-difference-between-institutional-traders-and-retail-traders.asp

8 Index of Equations, Illustrations, and Tables

Where not explicitly mentioned otherwise, illustrations and tables have been created solely by the author. The equations are retrieved from Serenity Stocks[45].

8.1 Index of Equations

Equation 1: Graham Number Calculation 6
Equation 2: Net Current Asset Value Calculation.. 7

8.2 Index of Illustrations

Figure 1: Process of the algorithm. 8
Figure 2: Screenshot with sections of interest highlighted (Yahoo Finance, 2021)................10
Figure 3: Selected lines of the code (78, 79, 106, 107, 112). 10
Figure 4: Selected lines of the code (Graham and NCAV calculations). 11
Figure 5: Overview and file structure of the experiment. 14
Figure 6: Bubble chart of the one-week performance for every result collected, excluding certain Graham and NCAV % values...19
Figure 7: The weekly performance of new entries and recollected data points. 21
Figure 8: Line Graph of the performance of stocks with selected Graham values. 22
Figure 9: Line Graph of the performance of stocks with selected NCAV values. 23
Figure 10: Performance of the Adjusted Method with benchmark indices. 24
Figure 11: Average change per week and volatility, including benchmark indices. 26

8.3 Index of Tables

Table 1: One year return on 1000 USD with a 20% interest rate and different compounding frequencies.. 4
Table 2: Attributes of the data collected...13
Table 3: Performance over eight weeks in percentage. 18

[45] (Serenity, 2016)

9 Erklärung

Erklärung

„Ich erkläre, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Alle wörtlichen und sinngemässen Übernahmen aus andern Werken habe ich als solche kenntlich gemacht.

Ich nehme ausserdem zur Kenntnis, dass meine Arbeit zur Überprüfung der korrekten und vollständigen Angabe der Quellen mit Hilfe einer Software (Plagiaterkennungstool) geprüft wird."

Datum Unterschrift

10 Arbeitsjournal

Das Arbeitsjournal selbst wurde als Datei 04 nummeriert und ist unter Berücksichtigung der Wegleitung in der Arbeit (Datei 11) eingebunden. Alle Dateien, auf welche im Arbeitsjournal verwiesenen wird, können beim Autor angefragt werden.

DATUM	Zeit (h)	RUBRIK	INHALT	DOKUMENT
KW 4-11	5	Organisation	Themenfindung, Suche des Betreuers, Vorbereitung erstes Gespräch.	00_Maturathemen
25.03.21	0.5	Vorbereitung Arbeit	Vorgespräch mit Herr Kotur, Festlegung Thema, Kein Protokoll.	00_Maturathemen
KW 13-20	4	Vorbereitung Arbeit	Erstellen der Projektskizze.	02_PROJEKTSKIZZE
21.05.21	-	Organisation	Abgabe der Projektskizze via E-Mail.	02_PROJEKTSKIZZE
18.06.21	0.5	Gespräch	Erstes Gespräch mit Herr Kotur, Besprechung der Projektskizze, Vertrag und Arbeitsumfang.	02_PROJEKTSKIZZE, 201_Protkoll_19.06
19.06.21	2	Organisation	Ausfüllen des Arbeitsvertrags und Abgabe via E-Mail.	01_Vertrag
04.07.21	2	Daten Auswertung	Sortierung der Daten, Vergleich mit Schlüsselwerten und Indexen.	1000_Data
10.07.21	1	Vorbereitung Arbeit	Wegleitung lesen und die Layout Richtlinien im Word Dokument umsetzen.	401_Maturitaetsarbeit_Wegleitung_2020
27.07.2021	3	Daten Auswertung	Sichtung, Sortierung und Subjektive Auswertung der Daten.	1000_Data
28.07.2021	2	Schreiben	Inhaltsverzeichnis erstellen.	06_Inhaltsverzeichnis_Entwurf
29.07.2021	3	Schreiben	Einleitung und Vorwort beginnen.	07_TestArbeitText
08.08.2021	1	Organisation	Aktualisierung des Codes.	11_Reinschrift
12.08.2021	1	Recherche	Gespräch mit einem Mathematiker.	08_Referenz_M_Seyfried
15.08.2021	1	Organisation	Aktualisierung des Codes.	11_Reinschrift
21.08.2021	4	Daten Auswertung	Sortierung der Daten, Vergleich mit Schlüsselwerten und Indexen.	1000_Datatable
29.08.2021	1	Schreiben	Inhaltsverzeichnis erstellen.	06_Inhaltsverzeichnis_Entwurf
29.08.2021	1	Vorbereitung Arbeit	Vorbereitung auf zweites Gespräch.	Persönliche Notizen

31.08.2021	0.5	Gespräch	Zweites Gespräch mit Herr Kotur, Besprechung Bewertungsraster.	202_Protkoll_31.07
03.09.2021	1	Organisation	Bearbeitung und Abgabe des Protokolls zum zweiten Gespräch.	202_Protkoll_31.07
04.09.2021	3	Daten Auswertung	Sortierung der Daten, Vergleich mit Schlüsselwerten und Indexen.	1000_Data
05.09.2021	2	Daten Auswertung	Simulator Entwicklung zum Testen einer möglichen Strategie.	1000_Data
10.09.2021	1	Schreiben	Reinschrift Dokument erstellt, Text übertragen.	11_Reinschrift
12.09.2021	1	Daten Auswertung	Auswertung der Daten, Überprüfung auf Unregelmässigkeiten, vergleich mit Indexen.	1000_Data
19.09.2021	1	Daten Auswertung	Auswertung der Daten, Überprüfung auf Unregelmässigkeiten, vergleich mit Indexen.	1000_Data
26.09.2021	1	Daten Auswertung	Auswertung der Daten, Überprüfung auf Unregelmässigkeiten, vergleich mit Indexen.	1000_Data
11.10.2021	3	Schreiben	Einleitung, Begriffserklärungen, Layout.	11_Reinschrift
14.10.2021	3	Schreiben	Begriffserklärung.	11_Reinschrift
14.10.2021	1	Recherche	Quellenrecherche, Lesen «The Intelligent Investor».	402_The_Intelligent_Investor
15.10.2021	2	Schreiben	Kapitel 2.	11_Reinschrift
17.10.2021	1	Recherche	Quellenrecherche, nachlesen «The Intelligent Investor».	402_The_Intelligent_Investor
17.10.2021	3	Schreiben	Kapitel 2.	11_Reinschrift
18.10.2021	2	Schreiben	Kapitel 3.	11_Reinschrift
24.10.2021	1	Schreiben	Kapitel 2.	11_Reinschrift
31.10.2021	1	Schreiben	Erstellung eigener Grafik und schreiben Kapitel 2.	101_Algorithmusprozess
04.11.2021	2	Vorbereitung Arbeit	Formatieren und erstellen des Zwischenabgabe Dokuments, Vorbereiten auf drittes Gespräch (Fragen und Anmerkungen).	13_Zwischenabgabe_Reinschrift
05.11.2021	1	Gespräch	Drittes Gespräch mit Herr Kotur via Teams.	203_Protkoll_05.11
05.11.2021	1	Organisation	Bearbeitung und Abgabe des Protokolls zum dritten Gespräch.	203_Protkoll_05.11
05.11.2021	1	Schreiben	Erstellung eigener Grafik.	105_Ablauf_Prozess
06.11.2021	5	Schreiben	Kapitel 4.	11_Reinschrift

06.11.2021	3	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	403_bubble_over-view, 106_bubble_color
07.11.2021	1	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	107_FT_RE_1w_gra ph
07.11.2021	2	Schreiben	Kapitel 4.	11_Reinschrift
08.11.2021	2	Schreiben	Kapitel 4, Verbesserungen andere Kapitel.	11_Reinschrift
11.11.2021	1	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	108_Gra-ham_8w_graph
11.11.2021	1	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	109_NCAV_8w_gra ph
13.11.2021	3	Schreiben	Kapitel 4.	11_Reinschrift
14.11.2021	1	Schreiben	Kapitel 4.	11_Reinschrift
14.11.2021	2	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	110_total_perfor-mance
18.11.2021	1	Schreiben	Kapitel 4.	11_Reinschrift
19.11.2021	1	Daten Auswer-tung	Visualisierung der Erho-benen Daten.	111_change_volatil-ity
21.11.2021	3	Schreiben	Kapitel 5.	11_Reinschrift
24.11.2021	2	Schreiben	Einfügen und formatieren des Computer Codes.	11_Reinschrift
25.11.2021	2	Schreiben	Kapitel 5.	11_Reinschrift
27.11.2021	3	Schreiben	Formatierung und Korrektur.	11_Reinschrift
04.12.2021	2	Schreiben	Korrektur.	11_Reinschrift
05.12.2021	3	Schreiben	Gestaltung Titelblatt und Layout Änderungen.	11_Reinschrift
06.12.2021	4	Organisation	Organisieren der File-Strukturen und Datenbank bereitstellen.	1000_Data
07.12.2021	1	Vorbereitung Arbeit	Vorbereitung auf viertes und letztes Gespräch.	Persönliche Notizen
08.12.2021	1	Gespräch	Letztes Gespräch mit Herr Kontur.	204_Protkoll_08.12
08.12.2021	1	Organisation	Erstellen Versuchsanlei-tung	14_Versuchsanlei-tung
08.12.2021	3	Schreiben	Korrektur und Layout	11_Reinschrift
09.12.2021	1	Organisation	Bearbeitung und Abgabe des Protokolls zum letzten Gespräch.	204_Protkoll_08.12
10.12.2021	7	Schreiben	Korrektur.	11_Reinschrift
11.12.2021	2	Organisation	Ausdrucken.	11_Reinschrift

11 Glossary

Array: *(computer science) An ordered collection of different values that are all the same variable type. In contrast to a list, an array cannot change its length and is therefore non-dynamic.*

Beautiful Soup: *A python library allowing the conversion of web pages to python variables like floats and integers.*

Book value: *The total value of a company's assets that shareholders would theoretically receive if a company was liquidated[46]. Divide by the number of shares to receive Book Value per share or BVPS.*

Current assets: *"Current assets are all the assets of a company that are expected to be sold or used as a result of standard business operations over the next year"[47].*

Current liabilities: *"Current liabilities are a company's short-term financial obligations that are due within one year or within a normal operating cycle"[48].*

Earnings per share (EPS): *A company's quarterly or annual net income divided by the number of its shares of stock outstanding[49].*

Fundamental analysis: *Fundamental analysts study anything that can affect the value of a stock, from macroeconomic factors such as the state of the economy and industry conditions to microeconomic factors like the effectiveness of the company's management[50].*

Index: *(computer science) The position of a value inside an array or a list, starting at the top of the list with 0 and counting in whole numbers. The integer array "int[] example_array = {0, 5, 10, 15, 20}" would have a value of 0 at index position zero and a value of 15 at index position three.*

Institutional money: *Institutional investors and traders use other people's money on their behalf instead of their own[51]. Hedge funds, investment banks or money managers all represent institutional money. They are the big guys on the block and account for more than 85% of the volume of trades[52]. Institutional investors and especially traders often can negotiate better fees because of the large size of their transactions. They have access to more exotic*

[46] (Hayes & Scott, Book Value, 2021)

[47] (Hayes, James, & Kvilhaug, Current Assets, 2021)

[48] (Tuovila, Mansa, & Munichiello, 2021)

[49] (Bankrate, 2021)

[50] (Segal, Boyle, & Munichiello, 2021)

[51] (Zucchi & Scott, 2021)

[52] (Adinarayan, 2021)

securities like forwards and swaps. Additionally, institutional money is solicited for investments in IPOs.

Investor: *An investor looks at the stock market and the returns he wishes to generate with a long-time horizon in mind. His goal is to gradually create wealth through buying and holding positions in stocks, bonds, or other investment instruments. These positions are often held for longer than a year and up to multiple decades while making use of interest and dividends along the way. Investors do not give too much attention to market fluctuations or extreme volatility but rather focus on the underlying fundamentals of their position, like price-to-earnings ratio or management forecasts. Furthermore, they wait out any market downturn with the expectation that prices will eventually rebound.*[53] *Institutional money is also considered more sophisticated and knowledgeable than retail and therefore is less likely to make uneducated investments or trades. Because of this fact, it is subject to fewer protective regulations from the Securities and Exchange Commission*[54]*.*

P/E ratio: *The P/E ratio helps investors determine the market value of a stock as compared to the company's earnings. In short, the P/E ratio shows what the market is willing to pay today for a stock based on its past or future earnings*[55]*. It is calculated by dividing the current value of a share by EPS.*

Python: *"Python is an interpreted, object-oriented, high-level programming language with dynamic semantics"*[56]*.*

Retail money: *Retail investors and traders execute their trades with their own money and through brokerage firms. They often trade in dramatically smaller amounts as compared to institutional money. Some see retail money as undisciplined and without much understanding of the market. However, the retail market is rather big, including retirement accounts, brokerage firms, and robo-advisors. Retail investors and traders can be used to gauge the market sentiment through survey data or first day IPO performance.*[57]

S&P 500: *"The S&P 500 is a stock market index that tracks the stocks of 500 large-cap U.S. companies. It represents the stock market's performance by reporting the risks and returns of the biggest companies. Investors use it as the benchmark of the overall market, to which all other investments are compared"*[58]*.*

---

[53] (Folger & Schmitt, 2021)

[54] (Palmer & Schmitt, 2021)

[55] (Fernando, Brock, & Li, Price-to-Earnings (P/E) Ratio, 2021)

[56] (python, 2021)

[57] (Hayes & Scott, Retail Investor, 2021)

[58] (Amadeo, Mansa, & Ernsberger, 2021)

Selnium: A python library allowing the control of a web browser through python without human input.

Standard deviation: *"The standard deviation is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The standard deviation is calculated as the square root of variance by determining each data point's deviation relative to the mean. […] A volatile stock has a high standard deviation, while the deviation of a stable blue-chip stock is usually rather low."*[59]

Technical analysis: *Method employed to evaluate investments and identify trading opportunities by analyzing statistical trends from trading activity, such as price movement and volume.*[60]

Total liabilities: *All liabilities that a company owes.*

Trader: *A trader represents the counterpart to an investor. A trader usually acts more frequently in the market and holds their positions for a shorter time than an investor. In addition to stocks, traders may also buy or sell commodities, currencies, other instruments, or derivatives, like options or futures, on one of these instruments. The general goal is to outperform a buy and hold strategy. While investors may "ride out" a less profitable position, traders use stop-losses to close out losing positions automatically. Traders are concerned with the current market momentum and trend. They often use technical analysis tools to further help them find trading opportunities. Traders have multiple subcategories, differentiated by their hold time of positions like a day trader or a swing trader.*[61]

[59] (Hargrave, Anderson, & Bellucco-chatham, 2021)

[60] (Hayes, Battle, & Jackson, Technical Analysis, 2021)

[61] (Folger & Schmitt, 2021)

12 Appendix

A file “data.xlsx” containing the data set from this paper and other files like an interactive version of Figure 6 can be requested from the author. The code of the main algorithm can be found in chapter 12.1.

12.1 Code of the Main Algorithm in Python 3.9

```python
#Copyright 2021 Lion Six

#import libraries
import time
from bs4 import BeautifulSoup
from selenium import webdriver
import math

#setup

#Save current time as start_time, so we can track how long the algorithm took
start_time = time.time()

#Open and read in the file containing a list of all tickers that should be searched trough, convert the file in to a string array and close the file
inFileLocation = "C:\\Users\\...\\source_stocks.txt"
ticker_file = open(inFileLocation, "r")
tickers = ticker_file.read().splitlines()
ticker_file.close()

#declare and assing running variable
y = len(tickers)
StartingValue = len(tickers)

#Open the output file
outFileLocation = "C:\\Users\\...\\ticker_output.txt"
f = open(outFileLocation,"w+")

#declare and assing the artificial chromedriver
driver = webdriver.Chrome(executable_path = r'C:\Users\...\chromedriver.exe')

#varaible responsible to counter anti-spam protection
spamDetectionCounter = 0
ncavDetectionCounter = 0

#Todays date and if this is the first or second run trough
firstRunTrough = True
```

```
    TodaysDate = "02.10.2021"

    #main loop, go trough every stock in the string array
    while y > 0:
        y -= 1

        #calculate progress of the loop in percent for debug purposes
        progress = ((y / StartingValue)-1)*-100
        progress = "{:.2f}".format(progress)

        #open Yahoo Finance website
        url = "https://finance.yahoo.com/quote/"+tickers[y]+"/key-statistics?p="+tickers[y]
        driver.get(url)
        #after opening the Yahoo Finance website, scroll down all the way to make sure the whole page is loaded and extend all minimzed fields
        try:
            iframe = driver.find_element_by_id("nvpush_popup_background_iframe")
            cross = driver.find_element_by_id("nvpush_cross")
            cross.click()
        except:
            pass

        #grab the "soup" of the website and save it as a variable for further reference
        soup = BeautifulSoup(driver.page_source,'html.parser')

        #locate the table which contains the values of interest in the soup
        StockData = soup.find(class_="Mstart(a) Mend(a)")

        #check if there acrtually is any data about this stock, otherwise the stocker either does not exist or the anti-spam protection activated
        if StockData:

            #locate the field were the stockprice is displayed and save it to a variable
            StockPriceSoup = soup.find(class_="Trsdu(0.3s) Fw(b) Fz(36px) Mb(-4px) D(ib)")

            #check if there even is a stockprice displayed, if there is save it, if there is not, or it is an empty field, save the price as 1
            if StockPriceSoup:
                StockPrice = StockPriceSoup.text
                if StockPrice == "":
```

```
                StockPrice = "1"
        else:
            StockPrice = "1"

        #get all rows of the stockdata table
        rows = list(StockData.find_all(class_="Fw(500) Ta(end) Pstart(10px) Miw(60px)"))
        #assing a int with the value of the length of the array of the rows
        i = len(rows)

        #declare calculation metrics and assign default values
        BVPS = 0
        EPS = 0
        #sahres outstanding gets two variables because it needs to be formatted first before being saved as a float
        SOS = "0"
        SharesOutstanding = 1

        #go trough the rows of the stock data table
        while i > 0:
            i -= 1

            #BVPS is displayed in this section, check if it does exist and if it is above zero, if not set it to zero
            if i == 58:
                soup = rows[i-1]
                if soup.text == "N/A":
                    BVPS = 0
                else:
                    #replace all , with empty spaces. Yahoo Finance places , every 3 digits while Python converts this to a float with decimal points
                    BVPS = float(soup.text.replace(',',''))
                    if BVPS < 0:
                        BVPS = 0

            # EPS is displayed in this section, check if it does exist and if it is above zero, if not set it to zero
            if i == 51:
                soup = rows[i - 1]
                if soup.text == "N/A":
                    EPS = 0
```

```
                else:
                    # replace all , with empty spaces. Yahoo Finance places , every 3 digits while Python converts this to a float with decimal points
                    EPS = float(soup.text.replace(',',''))
                    if EPS < 0:
                        EPS = 0

            # Shares outstanding is displayed in this section, check if it does exist and if it is above zero, if not set it to zero
            if i == 19:
                soup = rows[i - 1]
                if soup.text == "N/A":
                    SOS = "Unknown"
                else:
                    #Yahoo Finance abbreviates the shares outstanding value with the letters T, B, M and k. Format respectively.
                    SOS = soup.text
                    appendix = SOS[-1]
                    SOS = SOS[:-1]
                    if appendix == 'T' or appendix == 't':
                        SharesOutstanding = float(SOS) * 1000000000000
                    if appendix == 'B':
                        SharesOutstanding = float(SOS) * 1000000000
                    if appendix == 'M':
                        SharesOutstanding = float(SOS) * 1000000
                    if appendix == 'k':
                        SharesOutstanding = float(SOS) * 1000



        #All values have been read out, start calculating the results
        Graham = 22.5 * BVPS * EPS

        #if Graham is positive, take the square root
        if Graham > 0:
            Graham = math.sqrt(Graham)
        else:
            Graham = 0

        #Graham percentage
```

```
            GrahamPer = (100/float(StockPrice.replace(',','')))*Graham

            #If we are in the first run trough, check if this value meets the threshold
            if GrahamPer > 120:
                if firstRunTrough:
                    print(tickers[y])

                #if we are in the second run trough, the stock has already passed and now we load the second page containing more detailed fiancial info
                else:
                    ncav = 0.00
                    NCAVpercent = 0.00

                    #load balance sheet
                    url = "https://finance.yahoo.com/quote/" + tickers[y] + "/balance-sheet?p=" + tickers[y]

                    #the driver is set up again to make sure everything runs properly
                    driver.get(url)
                    try:
                        iframe = driver.find_element_by_id("nvpush_popup_background_iframe")
                        cross = driver.find_element_by_id("nvpush_cross")
                        cross.click()
                    except:
                        pass

                    soup = BeautifulSoup(driver.page_source, 'html.parser')

                    #if there is a button for "expand all", find it and press it
                    if soup.find(class_="D(tbr) fi-row Bgc($hoverBgColor):h"):
                        driver.find_element_by_xpath('//*[@id="Col1-1-Financials-Proxy"]/section/div[2]/button').click()

                        #find the field containing current Assets by its css selector and save it
                        Row1 = driver.find_element_by_css_selector(
                            "#Col1-1-Financials-Proxy > section > div.Pos\(r\) > div.W\(100\%\).Whs\(nw\).Ovx\(a\).BdT.Bdtc\(\$seperatorColor\) > div > div.D\(tbrg\) > div:nth-child(1) > div:nth-child(2) > div:nth-child(1) > div.D\(tbr\).fi-row.Bgc\(\$hoverBgColor\)\:h > div:nth-child(2)")
                        soup = Row1
                        row1string = soup.text

```

```
                        # find the field containing Total liabilities by its css selector and save it
                        Row2 = driver.find_element_by_css_selector(
                            "#Col1-1-Financials-Proxy > section > div.Pos\(r\) > div.W\(100\%\).Whs\(nw\).Ovx\(a\).BdT.Bdtc\(\$seperatorColor\) > div > div.D\(tbrg\) > div:nth-child(2) > div.D\(tbr\).fi-row.Bgc\(\$hoverBgColor\)\:h > div:nth-child(2)")
                        soup = Row2
                        row2string = soup.text

                        #only keep calculating if all of the values are above zero and exist
                        if row1string != '-' and row2string != '-' and SharesOutstanding != 0:
                            CurrentAssets = float(row1string.replace(',',''))
                            CurrentLiabilities = float(row2string.replace(',',''))

                            #calculate ncav
                            Denominator = (CurrentAssets - CurrentLiabilities) * 1000
                            ncav = Denominator / SharesOutstanding
                            NCAVpercent = (100.0 / float(StockPrice.replace(',','')) ) * ncav

                            #second parameter check, if the stock also passes this if test, it is added to the output file
                            if NCAVpercent > 30:
                                #write the stock to file, do so in a way were it can be read out easily by excel with ; as seperators
                                f.write(tickers[y] +";"+ url +";"+str(GrahamPer)+";"+str(NCAVpercent)+";"+str(StockPrice.replace(',',''))+";"+TodaysDate+"\n")

                    #if ncav could not be calculated print an error message and increase the spam counter
                    else:
                        print(tickers[y] +";could not calculate NCAV")
                        ncavDetectionCounter += 1
                        #if the ncav calculation fails for 4 times in a row, the anti spam protection is with high certainty active
                        #stop the algorithm for 5 minutes and try again then
                        if ncavDetectionCounter > 3:
                            time.sleep(300)
                            y = y + 4
                            ncavDetectionCounter = 0


            #if all calculations went trough well reset the anti spam detection
```

```
            spamDetectionCounter = 0

    #calculations of Graham failed
    else:
        spamDetectionCounter += 1
        # if the Graham calculation fails for 5 times in a row, the anti spam protection is with high certainty active
        # stop the algorithm for 5 minutes and try again then
        if spamDetectionCounter > 4:
            time.sleep(300)
            y = y + 5
            spamDetectionCounter = 0

#the loop is done, close the artificial google chrome window and the output file
driver.close()
f.close()

#inform the user that the algorithm has finished
print("PROGRESS: 100%")
#calculate and output the time it took for this run through
print("time elapsed: {:.2f}s".format(time.time() - start_time))


```