Are changing margins factored into stock prices?

By

Alfred Lin Kaishuo

B.E. Computer Engineering, National University of Singapore, 2014
M.B.A. Tsinghua University, 2021

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2021

Signature of Author: _____
MIT Sloan School of Management
May 7, 2021

Certified by: _____
Christopher F. Noe
Senior Lecturer
Thesis Supervisor

Accepted by: _____
Jacob Cohen
Senior Associate Dean for Undergraduate & Master's Program
MIT Sloan School of Management

Are changing margins factored into stock prices?

By

Alfred Lin Kaishuo

Submitted to MIT Sloan School of Management
on May 7, 2021 in Partial Fulfillment of the
requirements for the Degree of Master of Science in
Management Studies.

ABSTRACT

Stocks returns are often associated with the value of the company. There are many ways in which research has found viable variables to predict the future stock return, including autocorrelation of the stock prices, using the P/E ratio or more recently using the GP/A ratio. Having clear evidence that the GP/A ratio is useful to predict future stock return, this paper asks whether margin, which is GP/net sales, if changed, is factored into stock prices. Qualitatively, a common hypothesis is that when a company improves its profit margin, the company becomes more profitable and increases its returns. In this paper, I verified the hypothesis quantitatively.

In this paper, I explored three types of studies to verify the hypothesis. The three studies are 1) a correlational study, 2) a portfolio study, and 3) a regression study. From the above studies, I found out that there is a significant linear relationship between changing a margin and stock returns, the margin change is directly proportional to both current price difference and forward price difference, and finally the margin change variable is significant irrespective of the industry and the specific year the company is in.

Thesis Supervisor: Christopher F. Noe
Title: Senior Lecturer

Acknowledgements

Firstly, I would like to express my deepest thanks for my thesis supervisor, Dr. Noe. He has been a very influential figure since I was a student for his accounting class to understand and tell a story behind the financial data. Throughout the journey, he has helped me to refine my methodology and to provide insights whenever I met with huge roadblocks. His constant patience and availability have encouraged me along the way to complete this endeavor.

I would also like to thank Telis and members of the MIT investment management club for creating a platform for investing in equities. From the weekly meetings of the club, I grew to be more interested in understanding different company stocks and developed new research ideas with my friends.

I would like to thank fellow classmates of MSMS class, especially Amelia and Matt, for setting up weekly online writing sessions to motivate me to divide and conquer my thesis throughout the semester.

Last but not least, I would like to thank my girlfriend for supporting me throughout this journey. It has been a difficult time for me to write a thesis during the pandemic. Her encouragement has made this journey much easier for me.

Contents

1 Introduction 6
1.1 Aim of the research 7
1.2 Research methodology 7
2 Literature review 8
2.1 Efficient market hypothesis 8
2.2 Common variables affecting stock prices 9
2.3 Using gross profits-to-assets to understand the changes in stock prices 10
3 Hypothesis description 10
3.1 Fundamentals 11
3.2 Correlational study 12
3.3 Portfolio study 13
3.4 Regression study 14
4 Data description 15
4.1 Type of data 15
4.2 Data collection 15
4.3 Data cleaning 17
4.3.1 Spurious data 17
4.3.2 Outliers and invalid data 18
4.4 Final data 19
5 Results 21
5.1 Correlational study ..... **Error! Bookmark not defined.**
5.2 Portfolio study 21
5.3 Regression study 25
5.3.1 Regression of margin difference vs current returns 25
5.3.2 Regression of margin difference vs future returns 26
5.3.3 Summary of results ..... **Error! Bookmark not defined.**
6 Conclusion and suggestions for future work 27

1 Introduction

Stock prices have always been one of the focal points of discussion among both working professionals and common people on the street. Stock prices reflect the level of confidence that people have in the company. For example, if more people are confident in a hot technology company to do well, buyers will be willing to buy shares of the company at a higher price, increasing the demand and thus the price of the company stock. On the other hand, if more people are pessimistic about the performance of a technology company, there will be more investors selling than buying the shares of the company, reducing the demand and thus the price of the company stock.

Thus, some believe that stock price is a proxy for the health of the company (Murphy C, 2021). An entire collection of stocks, which represent the stock market, is therefore agreed by many to be an important indicator of a country's economy. Others are simply interested in knowing how stock prices move in order to profit from the momentum of the price movement (Wiley, 2021). In spring of 2021, a band of Redditors – users who gather to trade ideas on an online forum called Reddit – wanted to bet against two hedge fund firms which shorted GameStop, driving prices of GameStop to astronomical levels. This unusual case emphasizes the strong role of investor perception in raising the price of a stock.

The first known stock market started in Amsterdam in 1611, and the first stock market in the United States started in the 1700s (Sofi, 2008). Even though we have over 300-400 years of analysis and data, we are still unable to predict for certain the stock price movements of different companies. Nevertheless, the demands and the monetary incentives for understanding how stock prices move have persuaded many researchers and analysts to spend most of their lives studying what drives stock prices.

In order to gain some sense of stock price movements, investors pay attention to several key indicators. Annual financial statements are seen as one of the most reliable sources of financial data. The three most popular predictors in annual financial statements are increasing revenue in the balance sheet, increasing net income in the income statement, and positive cash flow in the cash flow statement.

Changing margins, which are equivalent to changing profitability, indicate the profitability of the company. When there is a margin expansion within a firm, the firm is producing goods and services more efficiently than before. Thus, even if the firm has constant revenue, margin expansion results in a larger net income, keeping other factors constant.

Running a correlational study, portfolio study and different types of regression studies, I found sufficient evidence of a relationship between margin differences of the company and its stock returns. The correlational study and regression studies showed strong evidence that changing margins are a significant variable for stock returns. The portfolio study shows a positive relationship between stock returns and margin changes.

From the three studies, the paper found some instances to suggest that margin differences are factored into stock prices. However, more research is suggested to confirm this phenomenon.

1.1 Aim of the research

The aim of the research is to understand the impact of changing gross profit margins on stock prices. Fundamentally, when the gross profit margin of a company increases, keeping other factors constant, the company is more efficient at providing goods and services than before. Thus, an increase in gross profit at the top-line consequently translates to an increase in net income at the bottom line, which is often associated with a well-performing company. An increase in net income is usually accompanied by an increased amount of cash and returns to both the shareholders and the company, and thus looks appealing to the investors, increasing the stock price (Adkins W., 2018).

However, a changing gross profit margin is not an indicator that is well used by institutional or retail investors to predict the stock price's performance. Furthermore, there is an issue in which an increase in gross profit margin does not always translate to an immediate boost in net income, a metric that investors watch closely. This is because an investment in improving its operations or adopting new technologies, which improves the company's ability to increase gross profit margin, takes some time to take effect. In some cases, the company does not observe an increase in net income for a prolonged period of time. In other cases, the company might even experience a reduction of net income in the short term. However, after the investments to increase the gross profit margin take effect, the company enjoys an eventual increase in net income. Thus, net income is a lagging indicator of the company's value. It will be interesting to observe if the changing gross profit margin helps to increase the overall value of the company in the long run.

As such, in this study, this paper aims to explore whether it is meaningful to associate a changing gross profit margin to an increase in performance of the company. A common metric that is widely available to measure the performance of a company is stock price.

1.2 Research methodology

The scope of the research in this paper will be limited to listed companies in the U.S. from 1950 to 2019.

The advantage of looking at publicly listed companies instead of private companies or start-up companies is the availability of financial information (CFI, 2021). Though it would be interesting to look at private companies or start-up companies to observe the financial performance of these companies, it will be difficult to obtain a continuous set of financial data or any financial data on the company. As such, analyzing private companies or start-up companies would not give us a definitive answer on the question. On the other hand, publicly listed companies are mandated by the Securities and Exchange Commission (SEC) to adhere to a fixed frequency and timing of publishing financial reports (Nguyen T., 2021). As such, the timing between each data point of the company's performance will be more or less equal in duration, for example, on a yearly basis and a comparison of how a company performs in one duration as compared to another duration would be much more meaningful.

Furthermore, since there is evidence that different accounting standards affect earnings management (Zhang et al., 2013), it will be difficult to compare companies with different accounting standards, which might occur across different countries. Thus, the scope of the research will be limited to listed companies in the U.S.

In this paper, I will first explore the current consensus on the reliability of using publicly available data to predict stock prices. After which, I will look into what the current research has done to use gross profit margin to measure returns in stocks. Since using gross profit margin to predict stock prices is currently not well documented, I will list several assumptions, make a hypothesis, and describe how the study will be conducted. Lastly, a discussion of the results will be given.

2 Literature review

2.1 Efficient market hypothesis

The stock market is a place in which market participants buy, sell, and issue stocks (J. Chen, 2021). Since the performance of the stock market relies on the performance of the stocks, which in turn relies on the expectations of the retail and institutional traders, the behavior of the market, and thus the stock prices of each company listed on the market, is largely dependent on the behavior of the market participants.

The Efficient Market Hypothesis (EMH), coined by Eugene F. Fama (Fama, 1970), states that all available information and expectations are reflected in the current stock prices. It also states that the current prices of publicly traded stocks reflect the intrinsic value of the stocks. If this hypothesis is true, it implies that there is no way to take advantage of buying undervalued stocks or selling overvalued stocks. Thus, people who believe in EMH believe that it is futile to trade in the market to profit in the long run (Downey, 2021).

On the other hand, there are a number of researchers who suggest that the stock prices are not market efficient and do not reflect economic reality. Summers suggested that the rational expectations of present value of the cash flows of stocks and bonds can differ widely from the current market valuations (Summers, 2010). Brainard, Shoven, and Weiss had also discovered that there is a large discrepancy between market value of U.S. corporate capital stock and its present value of after-tax cash flows (Brainard et al., 1980). In addition, Shiller had concluded that the volatility of stock prices are significantly higher than expected volatility due to uncertainties and adjustments to new information in the market (Shiller, 1981). In a more recent case, many analysts agreed that the stock price of GameStop is different from the fundamental value of the company. A group of Redditors drove the market price of GameStop to an astronomical price by going against the bets of hedge fund firm, Melvin Capital (McDermott, 2021). These findings suggest that the market does not fully reflect the information in the market, and thus to some extent is not efficient.

Furthermore, research has also shown that investors can use the autocorrelation of stock prices to reliably predict the future stock prices (Anderson et al., 2008), especially in the short term. This finding illustrates that the market is indeed not efficient since it is constantly adjusting, which allows investors to profit in the short run.

In addition, adjustments may take a long while to complete. Short-term or long-term gains can be achieved in the market during the time that the market adjusts to the correct price. With a significant number of studies that go against Efficient Market Hypothesis, it is safe to assume that opportunities exist in the market, while these adjustments are taking place.

In this paper, I will use changing margins, which is a measurement of company fundamentals, to conduct a study to understand its effect on stock returns. As there is no definitive way to determine how long markets take to adjust to new information, for simplicity, the paper will assume the duration of one year for market adjustments to occur.

2.2 Common variables affecting stock prices

Before exploring margins, it is important to discuss the common variables that authors have taken notice of when predicting stock returns. Earnings per share is one of the most popular ways of understanding stock prices. In Ohlson's paper, strong earnings translated to improved long term cash flow of the company, increasing the value of the company (Ohlson, 1995).

While experimenting in the Athens market, Glezakos had also found that the price-to-earnings ratio is one of the most significant variables in predicting stock prices (Glezakos et al., 2012).

Dividend provision is also shown to be a significant independent variable to predict changes in stock price. Lack of dividend provision when there is a sustainable strong cash flow affected stock prices (Omokhudu & Ibadin, 2015).

Many of the above research suggests that the income statement contains some of the most important variables to predict stock prices. However, Kwon had reported that the book value from the balance sheet and cash flow from the cash flow statement were much more relevant in affecting security prices than earnings upon experimentation in the South Korean market (Gee-Jung, 2009).

However, there is limited research in using profitability as a measure of the company to predict the stock prices. Based on the popularity of the indicators for analyzing stock performance, I observed that investors are much more interested to know if the stocks they acquire are at a good price as opposed to if the stocks they acquire have good fundamental value. For instance, investors are interested to know price-to-earnings (P/E) in order to figure out how much premium they are willing to pay for the stock. If the P/E ratio is high, it would indicate that they are paying too much. On the other hand, if the P/E ratio is low, the ratio would communicate that the stock is too cheap. However, in both cases, the ratio does not reflect the fundamental value of the stock.

To illustrate this, the common P/E ratio in the market is around 25 (Murphy, 2021). However, the P/E ratio of Apple is around 36 in 2021[1] . Even though Apple is trading at a significantly higher P/E ratio than its peers outside of the technology industry, Apple's fundamentals are considered sound. On the other hand, a low P/E ratio does not necessarily mean that the company is performing well. As Warren Buffet puts it, it is much more important to buy a great business at a fair price than a fair business at a great price (Velasco, 2013). Thus, this paper explores changing margin to understand the fundamentals of the company, and test if changing margin is related to the stock prices.

[1] Apple PE Ratio | AAPL (ycharts.com)

2.3 Using gross profits-to-assets to understand the changes in stock prices

In Novy-Marx's paper, the author identified that gross profits-to-assets can be a standalone indicator to predict future stock returns (Robert, 2013). Using stock price data on U.S. companies between the period of 1962-2009, the author arrived at several conclusions: 1) Gross profits-to-assets is more superior than earnings and free cash flow in predicting future returns of individual stocks, 2) Using gross profits-to-assets as an indicator to predict future returns is much more powerful when used within the same industry as compared to across industries, and 3) A four-factor model based on market, momentum, industry-adjusted value and gross profits-to-assets is useful to explain a wide range of anomalies in the company's financial data such as free cash flow growth, earnings yield, and return of equity.

Thus, I can observe that gross profitability, also known as gross profits-to-assets, is powerful in not only understanding the state of the company, but also by extension strong in predicting stock returns. With the trend moving towards understanding the fundamentals of the company rather than understanding the premium or discount of purchasing a stock, this paper examines a subset of gross profits-to-assets, which is gross margin, in order to understand if it predicts stock returns. As the "DuPont Model" goes, gross profits-to-assets is comprised of the product of asset turnover and gross margins. Asset turnover is the sales-to-asset ratio, which measures efficiency of the company. On the other hand, gross margin is the gross profits-to-sales ratio, which is a measure of market power of the company.

However, the value of asset turnover ratio of a company is unique to a particular industry. For example, companies in a labor-intensive industry are going to have a higher asset turnover ratio than companies in a capital-intensive industry. This is because a labor-intensive company does not require as much physical assets as a capital-intensive company in order to generate revenue. As such, companies in the retail sector are known to have a high asset turnover ratio, having an asset turnover ratio of 2.05 in 2014[2]. On the other hand, companies in other industries such as aviation can have a low asset turnover ratio. For example, Boeing, known to have a lot of assets, has an asset turnover ratio of 0.1 in 2020[3]. Since asset turnover ratio varies widely across industry, gross profits-to-assets is not powerful in comparing returns between companies across different industries if I do not exclude the industry effect.

Thus, this paper focuses on the other aspect, the market power of the company, gross margin, to understand if it is also a suitable indicator to predict stock returns, after removing the efficiency of the company and the industry effect. Fundamentally, it makes sense that as a company grows market power, regardless of industry, it is expected to be able to produce superior returns and experience a higher stock return.

3 Hypothesis description

In order to study the relationship between gross margin and stock prices, the paper has developed several experiments: a correlational study, portfolio study, and regression study.

[2] What is Asset Turnover Ratio? Definition of Asset Turnover Ratio, Asset Turnover Ratio Meaning - The Economic Times (indiatimes.com)
[3] Boeing Asset Turnover 2006-2020 | BA | MacroTrends

The correlational study provides a starting point to investigate specific phenomena, if there is a correlation found between margin difference and stock returns.

However, the original structure of data might contain too much variability or noise to find a relationship. As such, a portfolio study will be also used to divide the original sample into multiple sections, so as to observe how the variables interact at the sectional level. In this way, the relationships at the sectional level, which may not be detected in the correlational study, can be captured and examined.

Lastly, a simple regression model will be used to generate a model using gross margins to predict changes in stock prices. By setting up and running a regression model, I can find out if the variables used in the model are statistically significant, and if the margin difference variable is still significant when other variables are added into the model.

To summarize, these three methods will be helpful to demonstrate the relationship between gross margin and stock prices, and to answer the question on whether changing margin is factored into stock prices.

3.1 Fundamentals

Before elaborating on the three types of studies, I will discuss fundamentals, using formulas, to generate important data inputs for the analysis. The fundamentals are gross profit, change in price, and change in margin.

Gross Profit Margin

Gross profit margin is known as (*Net Sales – COGS) / Net Sales* (Bloomenthal A., 2021). Net sales is equivalent to net sales minus discounts, cost of sales returns and allowances. For cyclical industries such as airlines, hotels, retails, restaurants, and retails, discounts play a big part of their business (DiLallo M., 2021). There might be some implications to take note of if I compare between cyclical and non-cyclical industries. In addition, there are companies that expect a large amount of cost of sales returns and allowances as part of their business, especially in e-commerce companies such as Amazon (Connolly B., 2021), Taobao and Rakuten.

The Gross Profit Margin formula used in this paper is as follows:

$$Gross\ Profit\ Margin = \frac{Gross\ Profit}{Net\ Sales} \tag{3.1}$$

A benefit of looking at gross profit margin, in terms of a ratio, as compared to gross profit in levels is that it allows me to normalize margins across companies. Since different companies are different sizes, normalization allows me to compare different companies of different sizes across different industries with a common base value.

Margin Difference

New information causes stock price to change, and I am interested to know the change in price. As margin difference is equivalent to new information on margin, I used margin difference to explore the relationship between margins and prices.

Thus, the margin difference formula is seen in 3.2 below:

$$Margin\ Difference = \frac{Gross\ Profit_{t+1}}{Net\ Sales_{t+1}} - \frac{Gross\ Profit_t}{Net\ Sales_t} \quad (3.2)$$

In the formula, I am taking the difference of Gross Profit Margin of a single company between two time periods, $t+1$ and t, where t is any arbitrary time period in terms of the year, and $t+1$ is exactly one year after t.

Because comparisons across different companies and different time periods will create inconsistencies in comparisons, leading to incorrect deductions, this study will only include data points that satisfy the following two requirements:

- Gross Profit Margin across the same company
- Time period between $t+1$ and t, where t represents a single year

Price Difference

I utilize price difference to understand the relationship between gross profit margin and stock prices. The formula for price difference is similar to the investment percentage gain formula[4].

This paper focuses on comparing the price change and margin difference across a year. By comparing the independent variable and dependent variable on the same scale, this prevents erroneous deductions due to measurements on a different scale.

Thus, the formula to calculate price difference is seen below:

$$Price\ Difference = \frac{Price_{t+1} - Price_t}{Price_t} \quad (3.3)$$

Similar to the formula in margin difference, the t subscript corresponds to an arbitrary time period of the stock price of the company whereas $t+1$ corresponds to 1 year from the arbitrary time period. Likewise, a negative value in the price difference indicates that the price decreases across a year whereas a positive value indicates that the price increases across the year.

The price difference is bounded between -1 and infinity. This mimics the actual behavior of the stock whereby the maximum loss of a particular stock for a stockholder is -100% in which the stock had become worthless, and the maximum gain of a particular stock is infinite in which the stock had become very valuable. This phenomenon cautions me that a few overperforming stocks can result in outliers, which may skew the sample. Any outliers present in the dataset would need to be handled appropriately.

3.2 Correlational study

Correlation analysis has been commonly used in previous studies to compare stock prices across industries (Sharma & Banerjee, 2015) and across countries (Buda, 2010). In my paper, I will be using correlation analysis to test if margin difference is correlated with price difference. The formula is as follows:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3.4)$$

[4] How to Calculate the Percentage Gain or Loss on an Investment (investopedia.com)

The advantage of doing a correlational study is that it provides a starting point for the research, and helps to describe the relationship of the variables. Using this analysis also allows me to check the linearity of the dataset (Goodwin & Leech, 2006).

The correlation analysis will produce a value ranging from -1 and 1. If the value is close to 1, it indicates that the variables are positively correlated, meaning that an increase in one variable's value led to an increase to the other variable's value. If the value is close to -1, the variables are negatively correlated, which means an increase in value of one variable led to a decrease in value of the other variable. If the variables have a correlation of zero, a positive increase in value of one variable did not lead to an increase or decrease of value of the other variable. However, no correlation does not necessarily imply that the two variables are independent from each other. This is because linear correlation only captures the linear relationships between the two variables and does not detect any other association[5]. The two variables might exhibit a non-linear relationship, which is not captured by the correlational study.

One of the drawbacks of correlational studies is that extraneous variables might affect the variables in question. For example, in the classic case of higher crime rates and ice cream sales in the summer, I might find that there is a positive correlation between ice cream sales and crime rate. However, the two variables, ice cream sales and crime rate, are completely unrelated to each other. Thus, it is important for me to note that even if the variables are correlated, the correlational study does not imply causation between the two variables. As such, in order to draw any conclusions of the relationships of gross profit margin and stock returns, I first need to exclude other confounding variables which may be driving the relationship.

3.3 Portfolio study

In Novy-Marx's paper, the author utilized portfolio analysis to create different groups based on the values of gross profits-to-assets (Robert, 2013). This technique is useful because it allows the author to observe the stock returns distribution based on different levels of gross profits-to-assets. Portfolio analysis was also used to understand cross-sectional variation of returns based on value and glamour firms (Piotroski & So, 2012). Thus, portfolio analysis will be helpful in separating data into different logical groups to observe the relationships within isolated groups.

In my study, I will remove noise or outliers in my findings by exploring the relationships in particular groups while excluding other less important groups. Similar to how Robert performed portfolio analysis based on different levels of gross profits-to-assets to identify how stock returns differ, I will use portfolio analysis on different levels of change of gross margin in order to examine the impact on stock returns.

By using the portfolio analysis, I might be able to find a stronger relationship between gross margins and stock returns in some groups and lesser in other groups. As such, even if on the whole I do not find any relationship between the two variables, I might be able to identify some relationship across isolated groups. Lastly, with the use of the portfolio analysis, I can find out at which level of change of gross margins yields the greatest stock returns.

[5] Why zero correlation does not necessarily imply independence - Cross Validated (stackexchange.com)

To run the portfolio analysis, I divide my sample into seven different groups based on change in gross margin, from high to low, and then observe how the stock returns differ at different levels of gross margin. My initial expectations of the stock returns at different levels of gross margin are: At positive increase of gross margin, I expect the stock returns to increase. When there is a decrease in gross margin, the stock returns should decrease. This hypothesis is aligned with the expectations that when a company has higher market power or becomes more efficient, it is able to charge more or produce more cheaply, increasing the gross margin. As such, the company becomes more valuable, causing its stock price to increase.

3.4 Regression study

In my paper, I want to explore the relationship between gross margin and stock returns, which the correlational study might not be able to explain when more variables are introduced. Simple linear regression has been recognized as one of the tools to explain the relationships of different variables and stock returns (Zoran Ivanovski et al., 2016).

Linear regression identifies the linear relationship between the two variables, and is represented by a straight line. When the value of an isolated variable called an independent variable, is modified, the value of an observed variable known as the dependent variable, is expected to change as a result of the modification.

$$y = mx + c \tag{3.7}$$

The simplest linear regression is represented by the above equation, where y is the dependent variable and x is the independent variable. m is the coefficient of the corresponding variable, which represents the magnitude of the change of x as compared to y. If the magnitude of m is very small, a bigger unit change of x is required to increase per unit of y. On the other hand, if the magnitude of m is very big, a small unit change of x is required to increase per unit of y. The signs of the coefficient determine the positive or negative relationship x has with y. Lastly, c is the intercept that determines the starting bias when the independent variable has a value of zero.

The purpose of the analysis is to check whether the p-value of the variable is significant or not. If the p-value is below 0.05, the variable is deemed to be significant in the regression model, and I can conclude that it is worthwhile to continue exploring the relationship between margin difference and returns. In my hypothesis, I am creating two types of regression models: The first model compares margin difference and price difference, while the second model uses the first model with an addition of industry and year effects.

In my paper, I will first run a simple linear regression of change in profit margin with change in stock price, using the formula below:

$$p = w(gm) + c \tag{3.8}$$

where p denotes the price difference, gm denotes the change in gross profit margin, w denotes the weight or the coefficient of gross profit margin, and c is the intercept. This will corroborate the results from my correlational study.

Because stock returns are influenced by a variety of variables, multiple linear regression is often run to understand and find the relationship among multiple independent variables and the dependent variable (S. Chen & Chen, 2020).

Adding additional factor variables enables me to include and observe the effects of variables such as industry and year effects on margin difference. This is because the type of industry or the year, regardless of the stock performance in an economic boom or bust, might influence the stock returns. Thus, if the margin difference variable is still significant upon adding the industry and year effect, I can conclude that margin difference is factored into stock returns irrespective of the industry or year.

The formula for the regression model is described below:

$$p = w_1(gm) + w_2(Ind) + w_3(Year) + c \tag{3.9}$$

Ind represents the industry variable, and Year represents the year variable, whereas w_n represents each coefficient of n number of variables

4 Data description

In this paper, there is a separate section on data since a large amount of time and resources was spent on finding suitable data, understanding the data, and finally cleaning the data. It is important to discuss why I chose a particular type of data over another, how I collected the data, and finally the decisions involved in cleaning the data before conducting the various studies.

4.1 Type of data

Regarding the time periods of gross margin and stock returns data, historical stock prices is available at various time frames, ranging from hourly, daily, weekly, monthly, to yearly, whereas companies are only required to produce yearly or quarterly reports of net sales and gross profit. As such, even though I had finer stock data denominated in different time frames, I was only interested in the yearly stock price data in order to match the yearly time period of the gross margin data.

After deciding on the time period of gross margin and stock returns, I decided to focus on finding these data for U.S. companies. This is because the U.S. economy is one of the most watched economies, and there is a large database of U.S. company data available publicly. Even though Wharton Research Data Services (WRDS) includes both U.S. companies and Canadian companies as a package for North American company data, I decided to filter out the Canadian companies, in case there are discrepancies that exist between U.S. and Canadian companies in representing or reporting their financial statement data.

4.2 Data collection

After deciding on using yearly data and on analyzing U.S. companies, I collected the company data from Compustat-Capital IQ available through WRDS[6]. In the database, yearly

[6] Wharton Research Data Services (upenn.edu)

data of U,S, companies from 1950 to 2019 was available. Since my study assumes that the time period is not a determining factor in the relationship between gross margin and stock returns, I collected the entire dataset of company data from 1950 to 2019. A summary of the company data and its sources is consolidated in Table 4.1, as seen below:

Table 4.1 Source of company data

Source of Company Data	
Platform	WRDS (Wharton Research Data Services)
Data Provider	Compustat – Capital IQ
Period	Yearly
Country of Company	United States

After identifying the platform and the data provider, I had to decide what type of company data to include in our dataset. I consolidated the type of company data that was queried, represented in acronyms by WRDS in the below table.

Table 4.2 Company data type

Company Data Type	
FYEAR	Fiscal Year
TIC	Ticker Symbol
CURCD	Currency Code
GIND	Global Industry Classification - Industries
GP	Gross Profit (Loss)
REVT	Net Sales
PRCC_F	Annual Price Close - Fiscal

In table 4.2, I also included GIND[7], which is the Global Industry Classification Standard (GICS) industry code for the companies. The GICS is divided into 11 economic sectors, 24 industry groups, 68 industries and 157 sub-industries[8]. At each level, the GICS is represented by GSECTOR, GGROUP, GIND, and GSUBIND respectively. Since GIND ensures enough number of groupings of companies into different logical industries without resulting in too many independent factor variables when running the regressions, GIND is the most appropriate choice for my study.

Furthermore, I was given a choice between PRCC_F and PRCC_C from t Computstat Capital IQ[9]. PRCC_F refers to the price close at the fiscal year whereas the PRCC_C refers to the

[7] Master, Header, and Header History | CRSP - The Center for Research in Security Prices
[8] Global Industry Classification Standard (GICS) Definition (investopedia.com)
[9] Compustat Daily Updates - Fundamentals Annual (upenn.edu)

price close at the calendar year. I utilized PRCC_F to keep the company data from the financial statements and the stock price data consistent.

4.3 Data cleaning

Understanding and cleaning the data plays an important role in this study. As the common saying goes, “Garbage in, garbage out”. Without proper handling of raw data, the models might lead me to the wrong analysis and the wrong conclusion (Thomas, 2018).

One main advantage in keeping the number of variables small, choosing three variables (net sales, Gross Profit, and Price Close) for U.S. companies is that the amount of data correction and cleaning was kept at a manageable level. If more variables were added, I would observe more problematic observations and lose a significant amount of data after cleaning.

Before cleaning data, I had about 469,028 rows of yearly company data from 1950 to 2019. With a large amount of data to start with, I chose to remove the entire row if there were any spurious or sparse data in the row.

4.3.1 Spurious data

Under most rows, there was little spurious data on net sales, gross profit, or closing price. However, in some cases, there was incorrect data that needed cleaning. The error data occurred far more often in gross profit and set sales as compared to closing price of the companies.

Table 4.3 contains a summary of the spurious data. Firstly, I removed any total revenue that was either negative or smaller than gross profit. Gross profit is the result of deducting COGS (Cost of Goods Sold), a positive value, from total revenue. Hence, gross profit has to be smaller than net sales, or conversely, net sales have to be larger than gross profit. Lastly, it is impossible to have negative values in stock price since the smallest stock price amount is zero, representing that the company has zero value (Hartman, 2019).

Table 4.3 Removal of spurious data

Removal of Spurious Company Data			
Data	**Data for removal**	**Removal Reason**	**Observations**
Starting data			469,028
	Net sales	Negative values, zero values, empty values or smaller than gross profit	21,028
	Gross profit	Larger than total revenue and empty values	49,775
	Closing price	Negative values or empty values	134,079
Ending Data			264,146

I removed all observations with spurious data as defined above. I also removed all observations with empty or "not applicable" values among total revenue, gross profit and closing price variable. After removing the rows of spurious and sparse data, I was left with 264,146 rows of yearly company data.

4.3.2 Outliers and invalid data

Theoretically, it is possible to have a closing price of zero. However, I removed any stock price that had a value of less than a $1 and price difference of more than 300% to prevent over-representation of outliers on the mean returns under the portfolio studies.

The formulas of margin difference (3.2) and price difference (3.4) were performed on the raw data. If the current row of data could not find the preceding row of data under the same company and exactly 1 year before, there would be empty values for margin difference and price difference, which were removed. The results upon removal of outliers and invalid data for current price and forward price can be seen in table 4.4 and table 4.5 respectively.

Table 4.4 Removal of outliers and invalid data for current price

Removal of Outliers and Invalid data for current Price			
Data	**Data for removal**	**Removal Reason**	**Observations**
Starting data			264,146
	Forward & current price	Value below 1 result in over-representation of outliers	51,083
	Price difference	Price difference more than 300%	1,451

	Current price difference vs margin Change	Empty current price difference or empty margin change	23,023
Ending data			188,589

Table 4.5 Removal of outliers and invalid data for forward price

Removal of Outliers and Invalid data for forward Price			
Data	**Data for removal**	**Removal Reason**	**Observations**
Starting Data			264,146
	Forward & current price	Value below 1 result in over-representation of outliers	51,083
	Price difference	Price difference more than 300%	1,451
	Forward Price Difference Vs Margin Change	Empty forward price difference or empty margin change	42,291
Ending Data			169,321

4.4 Final data

Lastly, I divided the dataset into two types. Firstly, I compared margin difference with the current price difference in order to understand how the market reacts to the margin difference in the current year. Secondly, I also compared margin difference with the forward price difference to test if there is a relationship between gross profit margin and the next year's price. The final table of cleaned and processed data is shown below:

Table 4.4 Cleaned and processed data

Margin Difference vs Price Difference	
Company Data	**No. of Observations**
Margin Difference vs current price difference	188,589
Margin Difference vs forward price difference	169,321

After cleaning and processing the data, I conduct the different studies described in the hypothesis section and analyze the results. The distribution of the two types of processed data corresponding to current price difference and forward price difference is shown in the tables below.

Table 4.5 Statistics of Margin Difference vs Current price difference

Margin difference vs Current price difference				
Data	**Mean**	**Median**	**Low**	**High**
Margin difference	-0.0011866	-0.0002024	-1.8299315	1.1670562
Current price difference	0.08026	0.02158	-0.99919	3.00000

Table 4.6 Statistics of margin difference vs forward price difference

Margin difference vs Forward price difference				
Data	**Mean**	**Median**	**Low**	**High**
Margin Difference	-0.0002303	0.0000000	-1.8299315	1.1670562
Forward price difference	0.08037	0.02353	-0.99919	3.00000

5 Results

5.1 Correlational study

The correlation between margin difference and current price difference and forward price difference is summarized in the following table.

Table 5.1 Correlation matrix

Correlational study		
Correlation variables	**Correlation**	**p-value of margin difference**
Margin difference vs current price difference	0.1447614	< 2.2e-16
Margin difference vs forward price difference	0.02766571	< 2.2e-16

The correlation values for both comparisons are quite significant since there are many factors affecting stocks, especially when compared with current price difference. These results indicate that there is some linear relationship between gross margin and stock returns, especially in the current period. Though margin difference is significant, the R-squared value of the linear regression between margin difference and price difference is not going to be higher than 0.7. This is because stock prices are very noisy and there are many factors affecting stock prices. Thus, a good regression model will have to include other variables to have a reasonable prediction on stock prices.

In conclusion, the correlational study tells me that the margin difference is factored into the current price difference and to some extent into the forward price difference. The margin difference is significant in both cases.

5.2 Portfolio study

In order to separate the different levels of margin difference to conduct the portfolio study, it is important to understand the distribution of this variable. From initial inspection of the below histogram, Figure 5.1, the number of observations above 0.2 and below -0.2 were negligible.

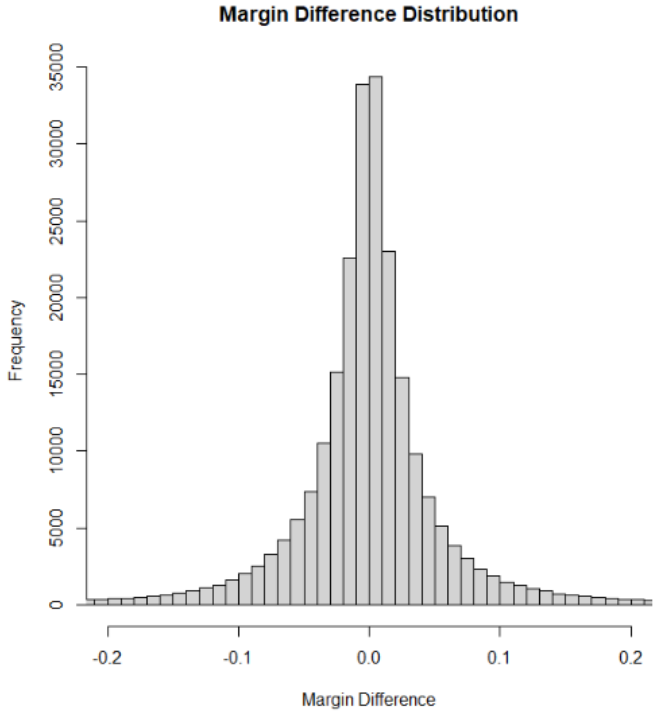

Figure 5.1 Margin difference distribution between -0.2 and 0.2

From the margin difference distribution, it can be noticed that the shape of the distribution resembles a normal distribution, which is consistent with our expectations. Firstly, there is approximately equal probability to have a positive or negative margin difference. Hence, the centering of the histogram of the margin is at zero. Furthermore, at any period of time, the margin difference for a particular company is close to zero. Thus, the highest frequencies are centered around zero.

However, it was difficult to divide the intervals based on the current scale. As such, another margin difference distribution based on the scale of between -0.1 and 0.1 is plotted below.

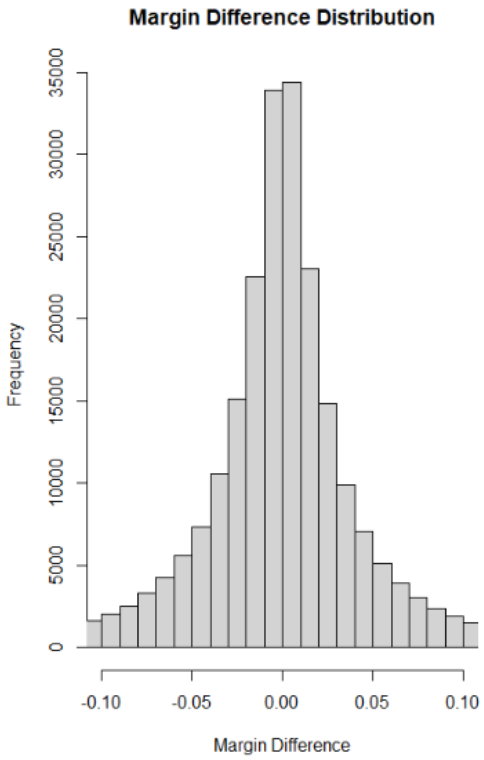

Figure 5.2 Margin difference between -0.1 and 0.1

The above histogram, Figure 5.2, shows that the highest values are centered between -0.02 and 0.02. At 2% change in margin difference, there is a possibility that some random rounding errors caused the margin to change by an immaterial amount. Hence, the intervals used in the portfolio study are as follows:

- Less than -0.1
- Between -0.02 and -0.1
- Between -0.02 and 0
- Between 0 and 0.02
- Between 0.02 and 0.1
- Greater than 0.1

Table 5.1 Portfolio of margin difference observations

Portfolio Study		
Margin Difference	**Returns observations**	**Forward Returns observations**
Lesser than -0.1	7,392	5,808
Between -0.1 and -0.02	39,646	34,777
Between -0.02 and 0	48,475	44,144
Between 0 and 0.02	49,344	45,428
Between 0.02 and 0.1	36,913	33,351
Greater than 0.1	6,819	5,813

Table 5.1 shows that most of the observations are centered on a very small range between 0 and 0.02 and between -0.02 and 0. This result is consistent with the normal distribution shape in the earlier histograms. Table 5.2 shows the mean current and forward returns by portfolio.

When the margin difference increases, the mean returns and mean forward returns also increase. The mean returns are affected on a larger scale than the mean forward returns when the margin difference increases or decreases. For example, when the margin difference is >0.1, the mean return is 0.22, or 22%, and the mean forward return is 0.13, or 13%. On the other hand, when the margin difference is <-0.1, the mean return is -0.11, or -11%, and the mean forward return is 0.053, or 5.3%, which is much less of a decrease than the mean return. This suggests that the margin difference factors more heavily in the mean returns than in the mean forward returns.

Table 5.2 Portfolio of margin difference vs mean returns

Portfolio study		
Margin difference	**Mean returns**	**Mean forward returns**
Lesser than -0.1	-0.11407123	0.05307547
Between -0.1 and -0.02	-0.03675046	0.07363751
Between -0.02 and 0	0.04719906	0.07142711
Between 0 and 0.02	0.12601060	0.07744147
Between 0.02 and 0.1	0.20117299	0.09858805
Greater than 0.1	0.22066167	0.13423212

As the mean is more heavily influenced by outliers in the data, looking at the median of the data is important. Table 5.3 shows that when the margin difference increases, the median returns and median forward returns increase as well. Similarly, when the margin difference decreases, the median returns and median forward returns also decrease. The trend in Table 5.2 shows the same positive relationship between margin difference and returns as in Table 5.3. Thus, we can conclude that the margin difference factors more in the current returns than in the forward returns.

Table 5.3 Portfolio of margin difference vs median returns

Portfolio study		
Margin difference	**Median returns**	**Median forward returns**
Lesser than -0.1	-0.1819074542	-0.028517810
Between -0.1 and -0.02	-0.0791578947	0.008390542
Between -0.02 and 0	0.0007324219	0.022058824
Between 0 and 0.02	0.0708661417	0.027897442
Between 0.02 and 0.1	0.1188118812	0.036363636
Greater than 0.1	0.1028996664	0.054487179

5.3 Regression study

I ran regression analyses on two sets of data, the current price difference and the forward price difference. First, I ran a simple linear regression between margin difference and returns. The second test added industry codes and years as control variables.

5.3.1 Regression of margin difference vs current returns

The simple regression model shows that the p-value of the margin difference in Table 5.4 is below 2.2e-16, which is significant at the 1% significance level. This result is consistent with the correlational study. Including industry and year as control variables did not change the significance of the current price difference variable. The p-value of the margin difference remains below 2.2e-16. As seen in table 5.4, the margin difference coefficient is positive whether the independent variable is margin difference or when the independent variables are margin difference, industry, and year. This result shows a positive relationship between margin difference and current price difference irrespective of industry and year. Thus, the relationship between the margin difference and the current price difference of a company is not affected by the industry or year.

Table 5.4 Simple regression on current price difference

Simple regression on current price difference		
Independent variables	**Margin difference coefficient**	**p-value of margin difference**
Margin difference	0.9975	< 2.2e-16
Margin difference, industry, year	0.9957	< 2.2e-16

5.3.2 Regression of margin difference vs future returns

In the first test on future price difference, I included only margin difference. In the second test, I included margin difference, industry, and years as independent variables.

In both tests, the p-values of margin difference are below 2.2e-16, regardless of whether industry and year were included as additional factor variables. At the same time, the margin difference coefficient remains positive irrespective of control variables such as industry and year. Thus, the results are quite consistent with the previous regression with current price difference that margin difference is an important variable in explaining forward price difference regardless of industry or year, and margin difference has a positive relationship with future returns.

Table 5.5 Simple regression on forward price difference

Simple regression on forward price difference		
Independent variables	**Margin difference coefficient**	**p-value of margin difference**
Margin difference	0.1992	< 2.2e-16
Margin difference, industry, year	0.1956	< 2.2e-16

6 Conclusion and suggestions for future work

From the correlational study, it is evident that there is a positive correlation between margin difference and both current and future stock returns. Looking at the portfolio study, I observed a clear trend in which a more negative margin difference results in a lower mean stock returns whereas a more positive margin difference results in a higher mean stock return. This phenomenon is not just limited to the current price difference but also extends to the forward price difference. Lastly, from the regression study, the margin difference is a significant variable to explain the current or future price difference, whether or not I include industry and year as control variables. Overall, these results tell me that margin difference is factored in current price difference but at the same time not fully factored into future price difference.

Furthermore, the results have cast some doubts on the efficient market hypothesis. This is because the forward returns have some relationship with margin difference as shown in the correlational study, portfolio study and regression study. Hence, it takes up to a period of a year for the stock prices to factor in the margin difference effect.

Though I used many different models to test the validity of the hypothesis, I believe that more can be done to understand the relationship between margin difference and stock returns. Future work can include the use of CART models, Random Forest or XGBOOST in combination with other identified variables in order to understand the significance of margin difference as an independent variable contributing to stock returns.

Even though I began with a large data set, many observations had to be removed because of missing or inconsistent data. The removal of the data might have contributed to a "survivorship bias" phenomenon. As such, it will be interesting to fill in the sparse data or conduct additional statistical tests to check for the significance of survivorship bias on the results.

This paper uses data from 1950 until 2019. Though I have considered year as an important control variable to dissociate its effects from margin difference, breaking the time frame into different logical period could help account for structural changes in the market over time. Running a new study while considering different time frames may allow further investigation of the relationship between margin difference and stock returns.

References

Adkins W. (2018). *Does Net Income Increase With Market Capitalization?* https://smallbusiness.chron.com/net-income-increase-market-capitalization-35843.html

Anderson, R., Eom, K., Hahn, S., & Park, J. (2008). Stock return autocorrelation is not spurious. *Working Paper*. http://escholarship.org/uc/item/9s35b82c.pdf

Bloomenthal A. (2021). *Gross Profit Margin Definition*. https://www.investopedia.com/terms/g/gross_profit_margin.asp

Brainard, W. C., Shoven, J. B., Weiss, L., Cagan, P., & Hall, R. E. (1980). The Financial Valuation of the Return to Capital. *Brookings Papers on Economic Activity*, *1980*(2), 453. https://doi.org/10.2307/2534328

Buda, A. (2010). *Life time of correlation between stocks prices on established and emerging markets*. *September*, 1–17. http://th.if.uj.edu.pl/~gulakov/life_corr/

CFI. (2021). *Public Companies - Overview, Advantages and Disadvantages*. https://corporatefinanceinstitute.com/resources/knowledge/finance/public-companies/

Chen, J. (2021). *Stock Market Definition*. https://www.investopedia.com/terms/s/stockmarket.asp

Chen, S., & Chen, S. (2020). *Forecasting Daily Stock Market Return with Multiple Linear Regression Forecasting Daily Stock Market Return with Multiple Linear Regression*. 1–10.

Connolly B. (2021). *Amazon Return Policy for Sellers: 2021 Changes that Affect FBAs*. https://www.junglescout.com/blog/amazon-return-policy-for-sellers/

DiLallo M. (2021). *Best Cyclical Stocks to Buy in 2021 | The Motley Fool*. https://www.fool.com/investing/stock-market/types-of-stocks/cyclical-stocks/

Downey, L. (2021). *Efficient Market Hypothesis (EMH) Definition*. https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp

Fama, E. F. (1970). American Finance Association Efficient Capital Markets : A Review of Theory and Empirical Work Author (s): Eugene F . Fama Source : The Journal of Finance , Vol . 25 , No . 2 , Papers and Proceedings of the Twenty- Eighth Annual Meeting of the American. *The Journal of Finance*, *25*(2), 383–417.

Gee-Jung, K. (2009). The Value Relevance of Book Values , Earnings and Cash Flows : Evidence from Korea. *International Journal of Business and Management*, *4*(10), 28–42.

Glezakos, M., Mylonakis, J., & Kafouros, C. (2012). The Impact of Accounting Information on Stock Prices: Evidence from the Athens Stock Exchange. *International Journal of Economics and Finance*, *4*(2), 56–68. https://doi.org/10.5539/ijef.v4n2p56

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *Journal of Experimental Education*, *74*(3), 249–266. https://doi.org/10.3200/JEXE.74.3.249-266

Hartman, D. (2019). *Can Stock Value Be Negative?* https://finance.zacks.com/can-stock-value-negative-9119.html

McDermott, J. (2021). *GameStop Stock, Explained: How Reddit Traders Manipulated the Stock Market*. https://www.esquire.com/news-politics/a35339535/game-stop-stock-short-squeeze-explained/

Murphy, C. (2021). *Using the Price-to-Earnings Ratio and PEG to Assess a Stock*. https://www.investopedia.com/investing/use-pe-ratio-and-peg-to-tell-stocks-future/

Murphy C. (2021). *Why Do Companies Care About Their Stock Prices?* https://www.investopedia.com/investing/why-do-companies-care-about-their-stock-prices/

Nguyen T. (2021). *How Often Do Publicly Traded Companies Prepare Financial Statements for External Reporting Purposes?* https://smallbusiness.chron.com/publicly-traded-companies-prepare-financial-statements-external-reporting-purposes-63331.html

Ohlson, J. A. (1995). *Eamings, Book Values, and Dividends in Equity Valuation*. *11*(2), 661–687.

Omokhudu, O. O., & Ibadin, P. O. (2015). The Value Relevance of Accounting Information: Evidence from Nigeria. *Accounting and Finance Research*, *4*(3). https://doi.org/10.5430/afr.v4n3p20

Piotroski, J. D., & So, E. C. (2012). Identifying expectation errors in value/glamour strategies: A fundamental analysis approach. *Review of Financial Studies*, *25*(9), 2841–2875. https://doi.org/10.1093/rfs/hhs061

Robert, N.-M. (2013). The Other Side of Value: The Gross Profitability Premium. *CFA Digest*, *43*(2), 105–107. https://doi.org/10.2469/dig.v43.n2.56

Sharma, C., & Banerjee, K. (2015). A study of correlations in the stock market. *Physica A: Statistical Mechanics and Its Applications*, *432*, 321–330. https://doi.org/10.1016/j.physa.2015.03.061

Shiller, R. J. (1981). *Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends ? Published by : American Economic Association REFERENCES Linked references are available on JSTOR for this article : You may need to log in to JSTOR to access the linked ref*. *71*(3), 421–436.

Sofi. (2008). *A Brief History of the Stock Market | SoFi*. https://www.sofi.com/learn/content/history-of-the-stock-market/

Summers, L. H. (2010). Do we really know that financial markets are efficient? *Recent Developments in Corporate Finance*, *99*, 13–24. https://doi.org/10.1017/cbo9780511628610.003

Thomas, S. (2018). *Data Cleaning in Machine Learning: Best Practices and Methods*. https://www.einfochips.com/blog/data-cleaning-in-machine-learning-best-practices-and-methods/

Velasco, S. (2013). *"It's far better to buy a wonderful company at a fair price than a fair company at a wonderful price." - CSMonitor.com*. https://www.csmonitor.com/Business/2013/0830/Warren-Buffett-10-pieces-of-investment-advice-from-America-s-greatest-investor/It-s-far-better-to-buy-a-wonderful-company-at-a-fair-price-than-a-fair-company-at-a-wonderful-price.

Wiley. (2021). *Momentum trading strategies - Fidelity*. https://www.fidelity.com/learning-

center/trading-investing/trading/momentum-trading-strategies

Zhang, Y., Uchida, K., & Bu, H. (2013). How do accounting standards and insiders' incentives affect earnings management? Evidence from China. *Emerging Markets Review*, *16*, 78–99. https://doi.org/10.1016/j.ememar.2013.04.002

Zoran Ivanovski, Zoran Ivanovski, & Zoran Narasanov. (2016). The Regression Analysis of Stock Returns at MSE. *Journal of Modern Accounting and Auditing*, *12*(4), 2006–2009. https://doi.org/10.17265/1548-6583/2016.04.003