FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

Seminar: Applied Corporate Finance (F)

Christian Birk Gustafson (ldg790)

A Data-Driven Approach to Evaluating the Intrinsic Value

- Automated Trading Strategy

Submitted on: 01/06/2023

Advisor: Mikkel Godt Gregersen

ECTS: 7.5

Keystrokes: 35.926

Abstract

This paper investigates the feasibility of predicting future stock prices of publicly traded companies utilising a data-driven approach. The study uses theories from data science and finance to scrape, transform, and analyse financial data to predict a company's future return, calculate key financial ratios, and provide buy or sell recommendations based on the analysis. This paper builds a model trained on companies in the S&P 500-index and compares its predictions against actual market outcomes for the year 2019, applying data from 2013 to 2018. It finds that the discounted cash flow (DCF) variable yields a 90% successful prediction rate, whereas the other variables, price-to-earnings (P/E), price-to-book (P/B) and price-to-sales (P/S), yield poor predictive power, which diverges from the conclusions drawn in previous studies. Additionally, this paper examines the prediction power of a stock's β, which is the stock's volatility related to the overall market, and finds that it is not a good predictor for stock returns, which is in line with existing literature. The study concludes that automated trading strategies could outperform manual assessments, aiming to provide assistance to private investors and analysts, and inspire further research with an interest in integrating data science and finance in the development of such strategies.

Keywords: Automated Trading Strategy, Forecasts, Price-to-Earnings, Price-to-Sales, Price-to-Book, Discounted Cash Flow, β.

Contents

1 Introduction

In today's evolving financial market, the ability to accurately assess a company's intrinsic value and predict its potential has become increasingly important, but it can be very time-consuming if done manually. This has led to a growing interest in utilising data-driven approaches to automate the valuation process and determine whether a company's stock is over- or undervalued. This paper aims to enhance the accuracy and speed of predictions made by private investors and analysts and proposes that automated trading strategies can outperform manual assessments. It seeks to inspire individuals with a keen interest in combining data science with finance to research the feasibility of automated trading strategies. This paper answers the following question, to what extent can a data-driven approach utilising an API to scrape financial ratios and a stock's β predict future stock prices of publicly traded companies, and can it outperform manual assessments? The study employs theories from data science and finance to scrape, transform, and analyse stock and financial data, calculating the key financial ratios price-to-earnings (P/E), price-to-book (P/B), price to sales (P/S), discounted cash flow (DCF), and β to predict a company's future return. The analysis builds on the framework from *Doblas (2020)* by using financial ratios as predictors for stock returns.

My contribution to the literature is threefold: i) Expanded upon the studies conducted by *(Doblas, 2020)*, and *(Chistoph et al., Univariate Portfolio Sorts)*, in which I incorporated two additional financial predictors, namely P/S and DCF. ii) I improved the accuracy of the estimation by incorporating a larger sample size compared to Doblas. Specifically, I included more firms and expanded the time period from 2013 to 2019. iii) I approached the forecasting aspect of the study from a private investor's perspective by making concrete stock price predictions. My principal hypothesis is, automated trading strategies that utilise publicly available data yield abnormal excess stock returns. This paper aims to achieve similar results to those obtained by *(Doblas, 2020)*, i.e., that P/E, and P/B are able to predict stock returns.

This paper employs an API from *(Financial Modelling Prep, API)* to construct a model trained on companies listed in the S&P 500-index during the period spanning from 2013 to 2018. Subsequently, the model is utilised to make predictions for the year 2019. In addition, the study aims to ascertain the reliability of a stock's β as a predictive factor for future stock returns by investigating whether a portfolio exhibiting high-minus-low β generates abnormal excess returns. Overall, this paper cannot reject the principal hypothesis, with DCF being the sole variable displaying a moderately robust predictive capacity. Consequently, analysts and private investors could potentially utilise this automated trading strategy that leverages the DCF to achieve abnormal excess stock returns, surpassing the performance of manual assessment. These findings relate to several strands of research, which are presented in the following section.

Models and data

All material, including data, models, and code can be found at Github. The code contains all the steps from data cleansing, analysis to prediction.

2 Litterateur Review

Economists and financial analysts have researched various techniques to understand market behaviour and find patterns for stock prices. A comprehensive study by *(Fama and French, 1992)* is generally accepted by the academic world of finance to explain which financial information has the strongest relationship with expected stock returns. They find that the strongest predictors of expected stock returns are the book-to-market value of equity and firm size, measured by the market value of equity during the period 1963 to 1990. Furthermore, they find that a stock's β does not yield any strong predicting power and that the combination of book-to-market value of equity and firm size absorbs the role of the P/E ratio in explaining stock returns. This result is supported by *Chistoph et al., Univariate Portfolio Sorts (2023)*, which finds that a stock's β does not yield abnormal excess returns. Their study is based on the period 1964 to 2019 and uses the Center for Research in Security Prices (CRSP) as its sample universe.

A study by *O'Shaughnessy (1997)* also covered a longer period from 1951 to 1994. The paper supports the results from *Fama and French (1992)* and finds that smaller stocks with market capitalisation of less than $25 million yield the most of the excess returns. In addition, to the support of *(Fama and French 1992)*, *O'Shaughnessy (1997)* also reports that utilising portfolios with a low price-to-book ratio outperformed all other stock selection strategies except for one utilising a price-to-sales ratio. A similar study covering a time period from 1989 to 1991, *Barbee et al., (1996)* finds that a portfolio of low P/S stocks has a significant effect on expected return and even outperforms the low P/B ratio strategy. Finally, *Jensen et al., (1998)* research the financial ratios P/S, P/B, and P/E's effect on stock returns and finds that all ratios have similar predicting power. However, these approaches have been subject to criticism.

A number of researchers have proven that utilising financial ratios does not correlate to higher stock earnings *(Perez, 2017)*, and *(Rahgozar, 2005)*. *Perez (2017)* finds that both low P/E and P/B ratios do not correlate to higher stock returns. Nonetheless, the paper covers the Thai stock market, which has very different characteristics compared to western markets, meaning that the results might differ across markets. Another critique of using past market information is provided by *(Rahgozar, 2005)*. *Rahgozar (2005)* estimates the intrinsic value of all stocks included in the Dow Jones Industrial, Transportation, and Utility Indexes over five sample periods by utilising the Discounted Cash Flow model. The paper criticises the financial model and finds no evidence of it predicting stock returns. These findings are in line with *Fama (1965)*, which finds that stock market performance resembles a random walk and that using past market information will not yield abnormal excess returns.

Surprisingly, reviewing the literature reveals some important aspects. First, the results supporting or disproving the financial ratios ability to predict stock returns vary from market to market, e.g., *(Doblas, 2020)* and*(Perez, 2017)*, which find contrary results depending on the financial market. Secondly, the timing of the data seems to be an important consideration in the varying results. Stock returns seem to behave differently during a crisis compared to more stable periods, e.g., *Chi-Hsiou D. et al. (2014)*, which finds that over non-crisis periods, the market β plays a more predominant role in determining the cross section of stock returns, compared to crisis periods where most of the variables lose their explanatory power. Additionally, *Hong (2021)* discovers that during the COVID-19 crisis, there was a significant amount of instability in stock return predictability and price volatility.

From the literature and the stated results from above, I present my following hypotheses on the DCF, P/E, P/B and P/S's ability to achieve excess returns below. These hypotheses are motivated by *(Doblas, 2020)*, and the previously mentioned literature:

Hypothesis 1 (H1): *A stock with a DCF value greater than its current stock price will generate excess returns in the following period.*

Hypothesis 2 (H2): *A stock's P/E lower than its industry average yields excess returns in the subsequent period.*

Hypothesis 3 (H3): *A stock's P/B lower than its industry average yields excess returns in the subsequent period.*

Hypothesis 4 (H4): *A stock's P/S lower than its industry average yields excess returns in the subsequent period.*

Doblas (2020) finds evidence that supports the idea that both the P/E and P/B ratios are able to predict stock prices among publicly listed financial institutions in Bahrain. Doblas covers the period from 2015 to 2018, with a total sample size of 19 financial institutions. In addition, the study finds that individual time invariant variables or company and industry factors do not play a role in the ability of P/E and P/B to predict stock returns.

Furthermore, my paper takes inspiration from *(Chistoph et al., Univariate Portfolio Sorts, 2023)* to assess their null hypothesis concerning the β value:

Hypothesis 5 (H5): *Utilising a stock's β the average portfolio excess returns are equal to zero.*

3 Theory

In the following section, this paper presents the theoretical framework for the automated trading strategy. Section 3.1 outlines the theory behind the capital asset pricing model (CAPM) and derives a linear regression from CAPM to be used for testing the long-short β strategy. Section 3.2 describes the long-short strategy by utilising univariate portfolio sorts. Finally, Section 3.3 outlines the financial ratios to assess the companies' intrinsic value, which is used to predict the stock returns for the year 2019.

3.1 Capital Asset Pricing Model

The capital asset pricing model describes how the expected return of an asset depends on its systematic risk, measured by its β coefficient. The CAPM was developed by *Sharpe (1964)* and states that investors can diversify unsystematic risk by holding a well-balanced portfolio of many assets. CAPM assumes that investors are rational and risk averse and that there is a risk-free asset that pays a constant interest rate with no transaction costs or taxes.

According to CAPM, a higher β yields a higher expected return. This paper will test the CAPM theory by constructing a long-short strategy based on two portfolios: buying a high-β portfolio and, at the same time, shorting the low-β portfolio. Therefore, one should expect that the average excess returns provide a return greater than the risk-free rate, *(Chistoph et al., Univariate Portfolio Sorts, 2023).* This helps understand whether or not it is possible to achieve abnormal excess return based on a stock's β.

The CAPM can be written in the following form to illustrate that the expected returns on individual assets are linearly related to the expected return on the market portfolio.

$$E[R_{i,t} - R_f] = \beta_i \cdot (E[R_{m,t}] - R_f), \tag{1}$$

where $E[R_{i,t}]$ is the expected return for asset i in period t, R_f is the risk-free rate, and $E[R_{m,t}] - R_f$ is the market risk premium. In equation (1), $\beta > 1$ indicates that a stock is riskier than the market, and vice versa, $\beta < 1$ suggests that a stock is less risky than the market. The proportionality factor β_i is given by

$$\beta = \frac{Covariance(R_e, R_m)}{Variance(R_m)} \tag{2}$$

where R_e is the return on an individual stock and R_m is the return on the overall market, *(Verbeek, 2017).*

This paper rewrites the CAPM in equation (1) as a linear regression model and utilises certain assumption seen in the appendix (derivation for CAPM). The regression is used to test the long-short β strategy and assess its validity.

$$R_{i,t} - R_f = \alpha_i + \beta_i(R_{m,t} - R_F) + \epsilon_{i,t}. \tag{3}$$

Equation (3) is a regression model with an included intercept, α_i, for the univariate portfolio testing explained in the upcoming section. ϵ_{it} is an error term being a function of unexpected returns and has a mean zero. At last, this paper imposes that ϵ_{it} is independent of $R_{mt} - R_f$ and that ϵ_{it} does not have autocorrelation or heteroskedasticity. This implies that the computed OLS estimates, tests, and standard errors are appropriate, *(Verbeek 2017)*.

3.2 Univariate Portfoilio Sorts

To determine the CAPM's α and examine whether a high minus low β portfolio yields abnormal excess returns, this paper must sort individual stocks into portfolios based on their β values. These portfolios represent well-diversified investments that differ according to the stocks' β values and presents the long-short strategy.

Considering only one sorting variable x_{t-1} is referred to as a univariate portfolio sort. i denotes the stock, and $t-1$ indicates that the characteristic is observable by investors at time t, *(Chistoph et al., Univariate Portfolio Sorts, 2023)*. My paper aims to assess the cross-sectional relation between x_{t-1} and the excess return $r_{t,i}$ for given stocks at time t. It uses the median lagged market β as the sorting variable to sort the stocks into a high- and a low β portfolio. To estimate a stock's β for this section, this paper employs a rolling-window estimation, which utilises a total of 5 years of past data and requires at least 48 months of return data. The estimation evaluates the stability of a time-series model by observing whether the coefficients are time-invariant, *(MathWorks, 2023)*. In practice, this paper will use the estimated CAPM regression in equation (3), where it will regress the stock excess returns on the excess returns of the market portfolio. Following the initial estimation of β_i, this paper will scale its estimation to a different level and conduct rolling-window estimations for the entire CRSP sample.

To perform the test, this paper will compute the average return using the fitting linear model function. This approach was recommended by *(Whitney K. Newey and West, 1994)* and leads to the null hypothesis ($H5$) that the average portfolio excess returns are equal to zero.

3.3 Financial Ratios

The P/E ratio measures the price that investors are willing to pay for each unit of earnings from a company. When making predictions in Section 4.4, a high P/E ratio indicates that the company is overvalued (sell the stock), while a low P/E ratio implies that the company is undervalued (buy the stock). The P/E ratio will be benchmarked against other companies in the same industry. There are two common types of P/E ratios: forward P/E and trailing P/E. This paper will utilise the trailing P/E, which relies on past performance by dividing the current share price by the total earnings per share over the past 12 months. This approach is widely used and regarded as more objective, *(Investopedia, 2023)*.

The P/B ratio is used to compare a firm's market capitalisation to its book value and locate over- and undervalued companies. It is calculated by dividing the company's current stock price per share by its book value per share. Similar to the P/E ratio, a stock's P/B ratio is benchmarked against other companies in the same industry. A high P/B ratio indicates that the stock may be overvalued, while a low P/B ratio suggests that the stock may be undervalued.

The P/S ratio compares a company's stock price to its revenues, indicating how much investors are willing to pay per dollar of sales for a stock. It is calculated by dividing the current stock price by the underlying company's sales per share. Similar to the P/E and P/B ratios, a stock's P/S ratio is benchmarked against other companies in the same industry. A low P/S ratio may imply that the stock is undervalued, while a ratio higher than the industry average suggests that the stock is overvalued.

The DCF estimates a company's value by forecasting future cash flows and discounting them to their present value. This paper employs an unlevered DCF model that forecasts free cash flows before interest expense and interest income. While a levered DCF approach aims to value the equity portion of a company's capital structure directly, the unlevered DCF analysis endeavors to evaluate the firm's overall value, *(Financial Modelling Prep, "DCF", 2023)*. The DCF valuation is then compared to the current market price of the company's stock. For example, if Microsoft's DCF is calculated to be 100 while its current market price is 50, the DCF suggests a buy recommendation. To calculate the DCF, this paper employs the formula provided by Financial Modelling Prep (FMP), *DCF = Equity Value / Weigted Average Shares Outstanding Diluted*, where *Equity Value = Enterprise Value NB - Net Debt.*

4 Empirical Analysis

This section presents the empirical analysis. The first section 4.1, provides a detailed description of the data utilised in this study and highlights the most relevant variables. Section 4.1.1 outlines the data transformation procedures employed in the analysis, while Section 4.2 examines the long-short β strategy. Section 4.3 presents the results derived from the three regressions, and the final section 4.4, showcases the predictions obtained from the automated trading strategy applying the financial ratios and asserts the hypotheses stated in Section 2.

4.1 Data

This study is based on historical annual data obtained from FMP, which provides an API for stock data and financial statements. This includes the income statement, cash flow statement, balance sheet, and DCF, which this paper utilises to facilitate automation. In addition, it employs data from all companies listed in the S&P 500-index, covering the period from 2013 to the end of 2019, compared to *Doblas (2020)*, which uses the time period 2015 to 2018. As a proxy for the market index, this paper applies the SPDR S&P 500 ETF Trust (SPY), since the S&P 500-index is not directly tradable. Table 1 lists the primary variables used in this analysis.

This paper constructs five new variables, namely (diff_PE), (diff_PS), (diff_PB), and (diff_dcf) to conduct the OLS regressions and making predictions. These variables are defined as the difference (diff) between the variable for the individual stock and the average for its sector. As an example, consider the construction of the variable $diff\_PE_{i,t}$,

$$diff\_PE_{i,t} = P/E_{i,t} - average(P/E)_{k,t} \tag{4}$$

where i is the individual stock, k denotes its respective industry, and t represents the given year. Lastly, this paper implements the variable $profit\_difference_{i,t}$ denoted as (PD), which is the stock return.

Furthermore, this study will employ data from the CRSP, obtained using the methodology outlined in *(Chistoph et al., WRDS, CRSP, and Compustat, 2023)*. The raw dataset from CRSP contains 3,281,726 monthly β values for 25,529 unique stocks over the period 1964-02-01 to 2021-12-01, which indicates a strong data sample yielding more robust results. The main variables in the dataset are return (ret_excess), β, and market capitalisation *mktcap*.

4.1.1 Data Transformation

Upon obtaining the primary data, this paper employs data transformations to enhance its suitability for the analysis. Specifically, it removes outliers from the financial ratios P/E, P/S, P/B, and DCF. Descriptive statistics in Table 1 illustrate the mean of the ratios: P/E (21.67), P/S (3.88), and P/B (4.36). The chosen range for these ratios is from -300 to 300, while for DCF it is from 0 to 2500. One firm, NVR, is removed due to extreme values. Additionally, the paper removes stocks with returns exceeding 1000, eliminating one stock named NVR. This stock exhibited high volatility and noise. The availability of DCF data is restricted to 307 firms and pertains solely to

Table 1: Summary statistics of FMP data

	P/E	P/S	P/B	PD	DCF
count	2524.00	2524.00	2524.00	2524.00	307.00
mean	21.67	3.38	4.36	11.29	105.65
std	32.70	5.79	17.92	30.38	86.36
min	-220.00	0.03	-268.94	-198.55	12.19
25%	13.52	1.29	1.75	-0.40	51.87
50%	19.90	2.32	3.00	6.48	82.30
75%	28.35	3.94	5.49	16.87	129.46
max	293.90	242.49	255.15	393.75	639.99

Source: FMP data and own calculations.

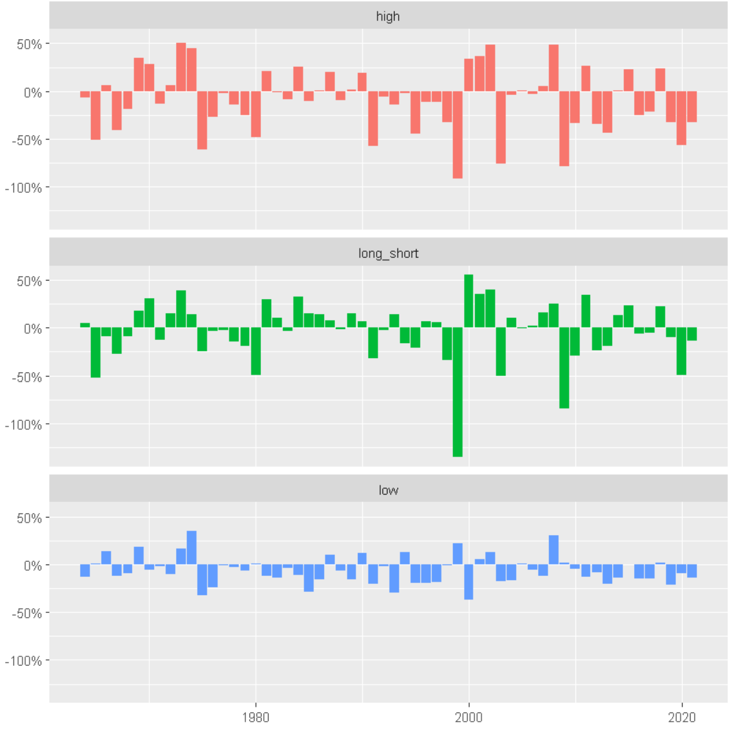
the year 2018, which subsequently serves as a predictive input for the year 2019. However, the P/E, P/S, and P/B variables spans over the years 2013 to 2018, thereby facilitating predictions not only for 2019 but also enabling a robustness assessment spanning the years 2014 to 2018.

This paper utilises a specific methodology within the CRSP dataset to conduct the long-short strategy. The median lagged market β serves as the sole breakpoint for grouping stocks into a high- and low β portfolio, allowing estimation of value-weighted returns, which are computed using lagged market capitalisation as weights. The lagged β at $t-1$ indicates that it is observable by investors at time t. A long-short strategy is formed by subtracting the high β portfolio from the low β portfolio.

4.2 Analysis of the Univariate Portfolio Sorts

Figure 1 illustrates the long-short strategy with the high portfolio at the top, the long-short portfolio in the middle, and the low portfolio at the bottom. The x-axis indicates the time period from 1964 to 2020, and the y-axis indicates the annual returns for each of the portfolios. The figure suggests that no consistent pattern emerged between the portfolios over the past years, with both the high and low portfolio exhibiting positive and negative annual returns. Therefore, the long-short strategy may not yield abnormal excess returns, implying that the β variable might not be a good predictor of future returns. Additionally, Table 2 provides a summary of statistics for the high, low, and long-short portfolios. Notably, the high β portfolio has the lowest expected return of -0.272 for a given month, which is surprising given that the CAPM suggests that the highest β stock should achieve a higher expected return. Thus, one would expect the low portfolio to contain the month with the lowest expected return, which is not the case, as it has a minimum of -0.186. However, examining the maximum value, the high β portfolio also contains the month with the highest expected return, which is consistent with the CAPM.

Figure 1: Long-short strategy, annual returns of β portfolios



Source: CRSP data and own calculations.

Table 2: Summary of long-short strategy with CRSP data

	high	low	long-short
Min.	-0.272	-0.186	-0.324
1st Qu.	-0.027	-0.015	-0.016
Median	0.009581	0.007	0.000
Mean	0.006	0.006	0.000
3rd Qu.	0.041	0.028	0.018
Max.	0.215	0.163	0.159

Source: CRSP data and own calculations.

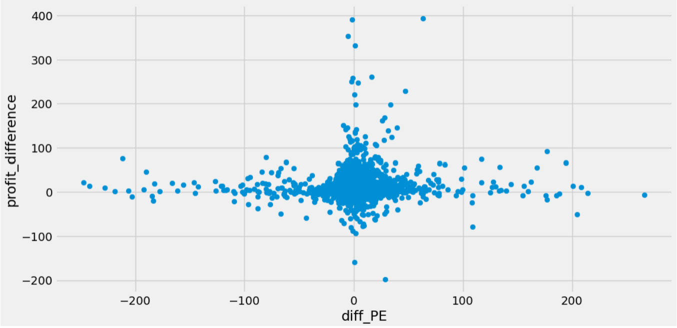
4.3 Empirical Strategy: OLS

This paper performs two OLS regressions to ascertain any correlations between the dependent variable, PD, and the explanatory variables: diff_PE, diff_PB, diff_PS, and DCF. Subsequently, this paper will conduct a third regression on the long-short strategy to assess the null hypothesis of average returns being equal to zero.

The first OLS regression is based on the following multiple linear equation:

$$PD_{i,t} = \beta_1 + \beta_2 \cdot diff\_PE_{i,t} + \beta_3 \cdot diff\_PB_{i,t} + \beta_3 \cdot diff\_PS_{i,t} + \epsilon_{i,t}, \tag{5}$$

Figure 2: diff_PE & PD



Source: FMP data and own calculations.

where i represents the stock, t indicates the given year, β denotes unknown population parameters, and ϵ is the error term. The independent variables are diff_PE, diff_PB, and diff_PS. The dependent variable is PD. The regression results are summarised in Table 3, showing statistical significance at a 5% level with an F-statistic of 6.206 and a p-value of 3.28e-05. Nonetheless, the R-squared value is low, indicating that only 0.9% of the variation in PD can be explained by the three independent variables. The independent variables exhibit statistical significance at a 5% level of significance. Notably, all ratios have a positive effect on PD, and with one unit increase in P/S resulting in a 0.248 unit increase in PD, P/S has the largest effect on PD. However, Figure 2 illustrates the absence of a linear correlation between diff_PE and PD, suggesting that the P/E ratio may not influence PD as expected. This discrepancy might imply the existence of unaccounted factors in this model.

Table 3: OLS regression (1), PD as dependent variable

	coef	**std err**	**t**	**P> |t|**
const	11.296	0.603	18.731	0.000
diff_PE	0.049	18.731	2.177	0.030
diff_PB	0.089	0.034	2.259	0.008
diff_PS	0.248	0.111	2.975	0.024

Source: FMP data and own calculations.

Note: F-statistic = 6.206, with Prob (F-statistic) = 3.28e-05, and R-squared = 0.009.

The second OLS regression uses the following equation:

$$PD_{i,t} = \beta_1 + \beta_2 \cdot DCF_{i,t} + \epsilon_{i,t}, \tag{6}$$

with the variable diff_DCF as an explanatory variable for PD. The resulting regression coefficients are presented in Table 4, indicating that the model is statistically significant at a 5% level of significance, with an F-statistic of 20.93 and a small p-value of 6.94e-06. Regardless of that, the model's explanatory power, as measured by the low R-squared value of 0.065, suggests that the correlation between the two variables may be weak.

The coefficient for diff_dcf is statistically significant at a 5% level of significance, with a p-value of less than 0.001. This finding implies that a unit increase in diff_dcf leads to a 0.322 unit increase in PD. To further investigate the relationship between the two variables, Figure 3 is employed. The figure suggests that it is challenging to discern a linear relationship between the variables, which suggests that diff_dcf may not be capable of explaining the variations in PD. Therefore, caution should be exercised when interpreting the results, and other factors that may be influencing PD should be considered.

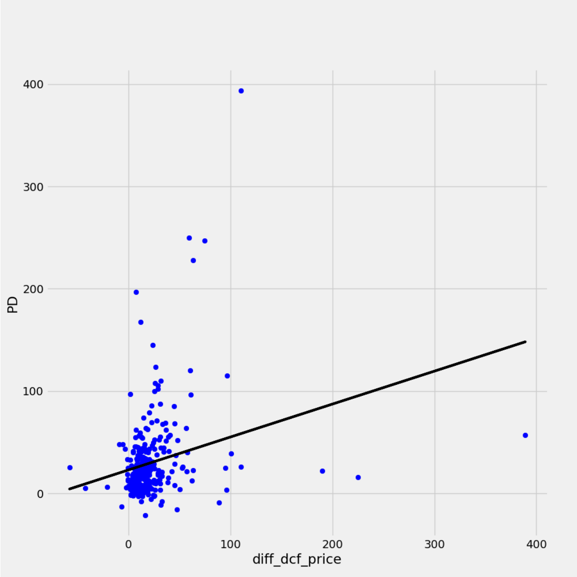
Table 4: OLS regression (2), PD as dependent variable

	coef	**std err**	**t**	**P> |t|**
const	22.888	2.723	8.404	0.000
diff_dcf	0.322	0.070	4.575	0.000

Source: FMP data and own calculations.

Note: F-statistic = 20.93, with Prob (F-statistic) = 6.94e-06, and R-squared = 0.065.

Figure 3: diff_DCF & PD



Source: FMP data and own calculations.

The third regression is based on the following equation:

$$long\_short_i = \alpha_1 + \epsilon_i, \tag{7}$$

where α_1 represents the CAPM-adjusted alpha intercept. It applies the CRSP dataset to examine the efficiency of the long-short strategy. Following the methodology outlined in Section 2.2, the dependent variable in equation (3) is long_short, and a constant term α_1 is included to estimate the regression intercept. The regression results are presented in Table 5, revealing a p-value of 0.83, which exceeds the 5% significance level. Thus, the null hypothesis *(H1)* of average returns being equal to zero cannot be rejected. It implies that the β value is a poor predictor, and applying the median as a breakpoint does not generate abnormal excess returns. This outcome is surprising because, according to the CAPM, high β stocks should produce higher expected returns than low β stocks. Given that the strategy focuses on purchasing high β stocks while shorting low β stocks, one would expect the average excess returns to be greater than the risk-free rate.

Table 5: Long-short regression (3)

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	0.00029	0.00131	0.22	0.83

Source: CRSP data and own calculations.

In summary, the regression results suggest that investors might be unable to predict the variations in stock returns, which is in line with *(Fama, 1965)*. *Fama (1965)* explains that by utilising past stock market data, one will not be able to predict future stock returns. In addition, these findings moderately contradict the studies by *(O'Shaughnessy, 1997)* and *(Doblas, 2020)*. Despite the statistical significance observed in the variables P/E, P/B, and P/S, the low R-squared value of 0.9% suggests a negligible correlation with stock returns, thus indicating their poor predictive power. To enhance the reliability of the findings, this paper will proceed to make predictions of stock returns.

4.4 Predictions and Robustness

In this section, this paper conducts three predictions using the same variable selection as in the OLS regressions. The first prediction, presented in Table 7, employs diff_PE, diff_PS, and diff_PB. The prediction is based on the years 2014-2019, allowing for a more robust result. For instance, the prediction for diff_PE is constructed as follows: when $diff\_PE_{i,t} > 0$ and $PD_{i,t} > 0$, or when $diff\_PE_{i,t} < 0$ and $PD_{i,t} < 0$, the model predicts correctly. Correct predictions are assigned a value of 1, while incorrect predictions receive a value of 0.

Table 6 illustrates that predicting whether a stock will be positive or negative in the following

Table 6: prediction, diff_PE, diff_PS, diff_PS & PD

	pred_diff_PE	pred_diff_PS	pred_diff_PB
0	52.021	57.924	55.349
1	47.979	42.076	44.651

Source: FMP data and own calculations.

year is extremely challenging. For instance, the worst variable predictor is diff_PS, which makes incorrect predictions 58% of the time, followed by diff_PB, which makes incorrect predictions 55% of the time, and diff_PE, which makes 52% incorrect predictions. Consequently, these variables are not effective in generating accurate stock price predictions. This outcome aligns with the results from the OLS regression, which indicate that these ratios only explain a small proportion of the variation in PD. Thus, this paper rejects hypothesis *(H2)*, *(H3)*, and *(H4)* indicating that these financial ratios do not yield excess returns.

The final prediction, displayed in Table 7, shows a surprising level of accuracy, with the diff_DCF variable achieving a success rate of 90%. This result appears to be in line with the significant regression results but contradicts the findings by *Rahgozar (2005)*, which show that the DCF model is unable to predict future stock returns. In addition, this paper cannot reject the hypothesis *(H1)* since the OLS regression did find a statistical correlation between the DCF and stock return at a 5% level of significance and the prediction shows that utilising the DCF yields excess returns. However, the results must be evaluated carefully since both Figure 3 and *Rahgozar (2005)* find no correlation between DCF and stock returns. One possible explanation for this inconsistency might be due to the year 2019, which was highly favourable for individual firms' stock returns, with no firms experiencing negative returns among those with available DCF data. As a result, it is possible that the accuracy of the DCF variable's predictions is due to an overly optimistic outlook. On the other hand, the high success rate may be due to the DCF method's ability to accurately capture a firm's intrinsic value.

Table 7: Prediction, diff_DCF & PD

	diff_DCF
1	89.508
0	10.492

Source: FMP data and own calculations.

5 Discussion

5.1 Data Suitability and the Impact of Transaction Cost

The careful evaluation of data suitability is a fundamental aspect of statistical modeling. In this regard, this paper employs simple- and multiple linear regressions using OLS. However, a notable shortcoming of this paper is the lack of verification of the assumptions made in Section 3.1. While it is true that this paper demonstrates that the error term has a mean of zero and is uncorrelated with the regressor, thus yielding a consistent OLS estimator, the absence of verification of the error term's independence and the lack of testing for autocorrelation and heteroskedasticity pose questions about the validity of the regressions. Admittedly, given that the primary objective of this paper is not to assess the CAPM using financial data, the omission of misspecification tests is understandable. Nonetheless, should the models be employed for further research purposes, it would be prudent to undertake misspecification tests to enhance the validity of the findings.

Transaction costs are another important aspect to consider when implementing a trading strategy. Excessive transaction costs have the potential to exert a significant influence on trade profitability, thereby having a consequential effect on the overall performance of the strategy. Nevertheless, assuming the transaction costs are reasonably assessed, their impact on the strategy appears to be limited. This is attributable to the strategy's emphasis on purchasing the recommended stock at the beginning of a year and retaining it until the end of the same year, thereby minimizing any transaction costs during the period.

5.2 Optimise the Long-Short Strategy and Asserting the Forecasts

To enhance the efficacy of the long-short strategy, a viable alternative would be to sort stocks into an arbitrary number of portfolios instead of using the median lagged market β to sort the stocks into a high- and a low β portfolio. This approach can be achieved using the curly-curly operator in R, which confers significant flexibility regarding the variables to employ for sorting, as espoused by *(Henry L 2023)*. In particular, one could opt to sort stocks into N portfolios each month based on lagged β values, a methodology that can yield more robust results. Indeed, using the median as a breakpoint is often deemed a naive strategy and might be the reason why I cannot reject the null hypothesis ($H5$). In addition, it does not permit any flexibility, thus underscoring the need for more nuanced and adaptable techniques in the sorting of stocks for long-short strategies.

Despite the statistically significant findings in OLS regression 1, where variables P/E, P/B, and P/S demonstrate significance at a 5% level, the forecasts depicted in Table 7 and 8 align with the negligible correlation illustrated in Figure 2. Moreover, the small value of $R^2 = 0.009$ further supports the assertion of a lack of correlation. In order to strengthen the robustness of these findings, this study incorporates different time periods for prediction. For instance, the time period is altered from one year to exclusively predict stock returns after the first week, month, and six months of a given year. The results from these variations prove to be negligible for all the regressions. Hence, it is reasonable to reject the hypotheses ($H2$), ($H3$), and ($H4$), which conclude that the variables are ineffective in predicting stock returns.

Furthermore, this paper's finding that the diff_DCF variable has a correct prediction rate of 90%, despite the low $R^2 = 0.065$ in the regression results, is an optimistic outcome that warrants further scrutiny. One important aspect to consider is the sample size, which is relatively small, containing only 307 observations. Such a sample size may not be sufficient to provide a reliable estimate of the variable's predictive power. Moreover, the time period analysed is relatively short, which could also have impacted the results. An improvement to the model would be to include more years of data, i.e., 2020 and 2021, and increase the number of observations for the DCF variable. This would provide more robust results, particularly in the years 2020 and 2021, which will provide an unstable period, thereby yielding a more accurate measure of the variable's effectiveness in predicting stock returns. Furthermore, predicting firms outside the S&P 500-index, which represents a relatively small sample of the overall market, could produce a more generalised result. This observation could explain the divergence in results compared to *Doblas (2020)*, and *O'Shaughnessy, (1997)*, suggesting that the predictive variables vary across different time periods and stock markets.

Nevertheless, it is important to note that the accuracy of the predictions may be influenced by various factors, such as market conditions, industry trends, and macroeconomic factors, which is not captured in the analysis. Further research conducted during a market downturn may be necessary to determine whether the variables will provide similar results in such conditions.

Despite the caveats above, the OLS regression did find that the DCF variable has a significant effect on PD at a 5% level of significance, indicating that the predictions may have some validity. Hence, the hypothesis ($H1$) cannot be rejected. But, it is relevant to explore other regression models, such as the Ridge or LASSO regression, which may offer more robust results, especially when working with small sample sizes.

6 Conclusion

In conclusion, this paper assesses the feasibility of utilising an API to construct an automated trading strategy for achieving abnormal excess returns. It finds that the DCF variable exhibits strong predicting power, accurately forecasting positive and negative returns for firms listed in the

S&P 500-index with a 90% accuracy rate, supporting *(H1).* However, these findings contradict the study by *(Perez, 2017)*, and may be limited due to the small sample size. Future research should expand the sample size and time frame for more robust results. The study rejects *(H2), (H3),* and *(H4)* indicating that P/E, P/B, and P/S ratios lack significant predicting power, contradicting previous studies such as *(Doblas, 2020)*, and *(O'Shaughnessy, 1997).* The long-short strategy based on β sorting also fails to yield noteworthy results, rejecting *(H5)* and aligning with the results by *(Fama and French 1992)*, and *(Chistoph et al., Univariate Portfolio Sorts, 2023).* In light of these findings, this paper cannot reject the principal hypothesis and recommends utilizing the DCF variable in this automated trading strategy for improved forecasting efficiency, surpassing manual assessments, and benefiting analysts and private investors to achieve abnormal excess stock returns.

References

Bali, Turan G, Robert F Engle, and Scott Murray. (2016). ”Empirical asset pricing: The cross section of stock returns”.
https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat07954

Chistoph Scheuch, Stefan Voigt & Patrick Weiss. (2023). ”Univariate Portfolio Sorts”, accessed 11-03-2023. https://www.tidy-finance.org/univariate-portfolio-sorts.html

Chi-Hsiou D. Hung, A.S.M. Sohel Azad, Victor Fang. (2014). ”Determinants of stock returns: Factors or systematic co-moments? Crisis versus non-crisis periods”, Journal of International Financial Markets, Institutions and Money.

Chistoph Scheuch, Stefan Voigt & Patrick Weiss. (2023). ”WRDS, CRSP, and Compustat”, accessed 11-03-2023. https://www.tidy-finance.org/wrds-crsp-and-compustat.html

Doblas, M.P, Lagaras, M.C.P., and Enriquez, J.A. (2020). ”Price to Earnings and Price to Book Ratios as Determinants of Stock Return: The Case of Financial Institutions Listed in Bahrain Bourse”. Journal of Applied Economic Sciences, Volume XV, Fall, 3(69): 532-539.

Fama, Eugene F. (1965). ”The Behavior of Stock Market Prices. Journal of Business, Vol. 38, No. 1, 34-105.

Fama, Eugene F., and Kenneth R. French. (1992). The Cross-Section of Expected Stock Returns. Journal of Finance June: 427-465.5.

Financial Modelling Prep, ”API”, accessed 02-05-2023. https://site.financialmodelingprep.com

Financial Modelling Prep, ”DCF”, accessed 02-05-2023. https://site.financialmodelingprep.com/discounted-cash-flow

FRED Economic data, ST. LOUIS FED, ”1-Year Treasury Bill Secondary Market Rate, Discount Basis”, accessed 24-03-2023. https://fred.stlouisfed.org/series/TB1YR

Github. ”Seminar-Applied-Corp.-FInance”. https://github.com/Birk54/Seminar-Applied-Corp.-FInance

Henry Lionel, Tidyverse. ”rlang 0.4.0”, accessed 11-03-2023. https://www.tidyverse.org/blog/2019/06/rlang-0-4-0/#a-simpler-interpolation-pattern-with-

Hong Hui, Bian Zhicun, Lee Chien-Chiang. (2021). ”COVID-19 and instability of stock market performance: evidence from the U.S.”, Financial Innovation.

Jason Fernando. ”P/E Ratio - Price-to-Earnings Ratio Formula, Meaning, and Examples”, accessed 02-04-2023. https://www.investopedia.com/terms/p/price-earningsratio.asp

Jensen, Gerald R., Robert R. Johnson, and Jeffrey M. Mercer. (1998). ”The Inconsistency of Small-Firm and Value Stock Premiums”. The Journal of Portfolio Management Winter: 27-36.

Manuel Bleve (2020). ”WACC: Weighted Average Cost of Capital Explained”. https://finbox.com/blog/wacc-weighted-average-cost-of-capital-explained/

MathWorks. ”Rolling-Window Analysis of Time-Series Models”, accessed 05-04-2023. https://www.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html

Newey, Whitney .K, and Kenneth D. West. (1994). “Automatic lag selection in covariance matrix estimation.” The Review of Economic Studies 61 (4): 631–53. https://www.jstor.org/stable/2297912

O’Shaughnessy, James. P. (1997). What Works on Wall Street New York: McGraw-Hill.

Rahgozar Reza. (2005). ”Comparing Stock Prices to their Estimated Intrinsic Values Using Valuation Model”. Journal of Business and Behaviroal Sciences, Vol. 13.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. The Journal of Finance, 19 (3), 425–442. https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1964.tb02865.x

Stern-NYU-EDU. ”Ratings, Interest Coverage Ratios and Default Spread”, accessed 02-04-2023. https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/ratings.html

Verbeek Marno. (2017). ”A Guide to Modern Econometrics”, 5th Edition, Chapter 2.

Appendix

Derivation for CAPM

To begin, this paper assumes that expectations are rational, so that the expectations of economic agents correspond to mathematical expectations, meaning it is possible to derive a relationship from equation (1) involving actual return. Furthermore, the OLS estimator is consistent and therfore the error term is uncorrelated with the regressor.

First, define the unexpected return on asset i as

$$u_{it} = R_{it} - E\{R_{it}\}$$

and for the market portfolio as

$$u_{mt} = R_{mt} - E\{R_{mt}\}.$$

Rewriting equation (1) as

$$R_{it} - R_f = \beta_i(R_{mt} - R_F) + \epsilon_{it} \tag{8}$$

where

$$\epsilon_{it} = u_{it} - \beta_i u_{mt}.$$

Equation (8) is a regression model, without an intercept. ϵ_{it} is an error term being a function of unexpected returns and has a mean zero, that is,

$$E\{\epsilon_{it}\} = E\{u_{it}\} - \beta_i E\{u_{it}\} = 0$$

which shows that the OLS estimator is consistent, *Verbeek, (2017).* Furthermore, the error term is uncorrelated with the regressor $R_{it} - R_f$, using the definition of β_i, which can be written on the form

$$\beta_i = \frac{E\{u_{it}u_{mt}\}}{Variance\{u_{mt}\}}$$

noting that r_f is not stochastic, it provides the result that

$$E\{\epsilon_{it}(R_{mt} - R_f)\} = E\{(u_{it} - \beta_i u_{mt})u_{mt}\} = Eu_{it}u_{mt} - \beta_i Eu_{mt}^2 = 0. \tag{9}$$