A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

NOVA SCHOOL OF BUSINESS & ECONOMICS

"An Empirical Analysis of ChatGPT's Performance in Predicting Stock Prices: A Study of the DAX Companies"

Luca Niklas Kuhlmann
55537

Work project carried out under the supervision of:

Prof. Giorgio Ottonello

20/12/2023

Contents

List of Figures X

List of Tables X

1. Introduction 2

1.1. Background and Motivation 2

1.2. Significance of Investigating ChatGPT's Performance in Stock Price Prediction 3

1.3. Research Question and Scope of the Study 3

2. Literature Review 4

2.1. Existing Studies on NLP and LLMs in Financial Markets 4

2.2. Discussion of ChatGPT's Potential Applications in Finance and Related Research 5

3. Methodology and Data 6

3.1. Data Source/Provider 6

3.2. Methodology 6

3.2.1. Data Collection 6

3.2.2. Data Preparation and Cleansing 9

3.2.3. Regression Analysis and Backtesting Approach 10

4. Results and Analysis 15

4.1. Regression against Fama-French 3-Factor Model 15

4.2. Backtesting the Performance of HmL against DAX Benchmark Index 17

4.3. Assessing the Stability and Consistency: A Robustness Test of the HmL Strategy 20

5. Conclusion 22

5.1. Synthesis of Research Aims and Principal Discoveries 22

5.2. Answer to the Research Question **22**

5.3. Limitations and Recommendations for Future Research **24**

5.4. Implications and Future Outlook **25**

References 26

Appendix 29

List of Figures

Figure A1: HmL-performance (equal) vs. DAX30 of € 1 investing in HmL portfolio 29
Figure A2: Mean Market Capitalisation by Portfolio Allocation 29
Figure A3: Mean Return by Portfolio Allocation 29
Figure A4: Market Capitalisation-weighted Return by Portfolio Allocation.......................... 30

List of Tables

Table 1: Result of News Headline Prompt in ChatGPT... 10
Table 2: OLS Regression Results against FF3 risk factors 17
Table 3: Performance Metrics Overview 19

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CAPM	Capital Asset Pricing Model
DAX	Deutscher Aktienindex (German Stock Index)
ETF	Exchange-Traded Fund
FF3	Fama-French three-factor
GPT	Generative Pre-trained Transformer
HmL	High minus Low
LLM	Large Language Models
LSEG	London Stock Exchange Group
NLP	Natural Language Processing
OLS	Ordinary Least Squares
SR	Sharpe Ratio

Abstract

This study investigates ChatGPT's sentiment analysis for predicting the German DAX index's stock prices, examining its efficiency in processing financial texts, and forming a HmL portfolio. The approach includes regression against the FF3 model and DAX index comparison. Findings indicate ChatGPT's sentiment aligns with market and value factors but isn't a unique risk factor in European asset pricing, showing no significant alpha and correlating with the European market return and FF3 value factor. It also backtests the HmL strategy and includes a robustness test considering market capitalisation, highlighting challenges in identifying new risk factors in financial markets and the importance of balancing returns with risks and costs.

Keywords

Sentiment Analysis, ChatGPT, Stock Price Prediction, High minus Low (HmL) Portfolio, Fama-French Three-Factor Model, German DAX Index, Financial Text Processing, Regression Analysis, Market Capitalisation, Risk Factor Assessment, Backtesting Investment Strategies, Artificial Intelligence in Finance, Natural Language Processing (NLP), Asset Pricing

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

1.1. Background and Motivation

Advancements in NLP, signifying the convergence of AI and finance, have significantly altered data-driven decision-making. In this evolution, LLMs like ChatGPT stand out, as they demonstrate exceptional skill in deriving insights from extensive textual data. ChatGPT, for instance, is renowned for its effectiveness in inference, sentiment analysis, and question-answering, illustrating its advanced language comprehension (Zhong et al. 2023). Stakeholders such as investors and analysts show considerable interest in its application in financial markets, especially in sentiment analysis.

In the complex realm of financial markets, sentiment analysis , as a branch of NLP, provides a unique method for analysing and utilising market sentiment derived from financial news. This aligns with AI's trajectory in finance, where models like ChatGPT are leveraged for statistical analysis and trend prediction (Noever and McKee 2023). The use of contextual embeddings in ChatGPT enhances sentiment analysis, surpassing traditional methods even in the finance domain (Mishev et al. 2020). Thus, applying ChatGPT for sentiment analysis in financial markets could offer innovative ways to forecast stock price movements.

The influence of sentiment analysis via ChatGPT on future stock returns lies in its ability to process and interpret vast amounts of unstructured textual data rapidly and accurately. This capability reflects market emotions and anticipates shifts in investor behaviour, providing a predictive edge in stock market dynamics. Moreover, the application of ChatGPT in sentiment analysis might also intersect with behavioural finance, where investor attention and perception significantly impact market trends. Therefore, this synergy between AI-driven sentiment analysis and behavioural finance principles could reveal new insights into risk assessment and investment strategies in the volatile environment of the stock market.

1.2. Significance of Investigating ChatGPT's Performance in Stock Price Prediction

This research explores ChatGPT's use in sentiment analysis for predicting stock prices, with a focus on the German DAX index. This approach is significant, as AI's role in the DAX context remains largely unexplored. The study aims to reveal AI's impact on investment strategies and decision-making in the financial sector, highlighting the distinct features of the German market.

The research recognises AI's transformative role in finance and addresses challenges like data biases and interpretability in the DAX context. Focusing on this market, the study enriches existing knowledge and offers fresh insights into AI's role in a key European financial market.

The literature review highlights a growing interest in AI's role in finance, underlining the research's relevance and timeliness. The study also considers traditional work methods and ethical issues. It stresses the importance of addressing limitations like data biases and interpretability in AI's practical financial applications, which are key areas for further examination.

1.3. Research Question and Scope of the Study

Research Question 1: "Can ChatGPT's sentiment analysis accurately forecast stock price movements for companies in the DAX index?"

Research Question 2: "Is it profitable to trade on ChatGPT Sentiment using a cross-sectional HmL strategy? Does this strategy outperform other relevant indices, like the DAX?"

The study aims to assess ChatGPT's capability in processing financial texts and generating sentiment scores. A key focus is on constructing a HmL portfolio based on these scores. This portfolio will be evaluated in two ways: firstly, through a regression analysis against the FF3, and secondly, by comparing its performance directly with the DAX index. This dual approach offers insights into how ChatGPT's sentiment analysis stands against traditional financial predictors and the actual market index. The methodology combines qualitative and quantitative analyses using diverse financial data sources.

Subsequent sections introduce a literature review framing the use of NLP and LLMs in financial markets, leading to an empirical analysis that assesses the impact of AI on financial decision-making and prediction accuracy.

2. Literature Review

2.1. Existing Studies on NLP and LLMs in Financial Markets

The integration of NLP and LLMs into financial market analysis represents a pivotal shift in the landscape of financial technologies. Recent literature in 2023 underscores this evolution, demonstrating the diverse applications and impacts of these technologies in finance.

In their ground-breaking research, Wu et al. (2023) developed BloombergGPT, a model tailored for financial analysis, highlighting the importance of bespoke datasets and model architectures in specialised domains (Wu et al. 2023). Hansen and Kazinnik (2023) delved into the utilisation of GPT-3 and GPT-4 models to interpret communications from the Federal Reserve, marking a significant advancement in qualitative financial analysis (Hansen and Kazinnik 2023). The study by Bond, Klok, and Zhu (2023) focused on the predictive capabilities of LLMs in stock market forecasting, illustrating the potential of these models in predictive analytics (Bond, Klok, and Zhu 2023). Xie et al. (2023) conducted a comprehensive assessment of ChatGPT in stock market prediction through a zero-shot analysis approach (Xie et al. 2023). Lastly, the contributions of Zhang, Yang, and Xu (2023) with the development of XuanYuan 2.0, an LLM specifically designed for the Chinese financial market, signify a notable advancement (Zhang, Yang, and Xu 2023). However, I contribute to the field by developing a HmL portfolio strategy based on ChatGPT's sentiment analysis. This approach not only leverages ChatGPT's NLP capabilities but also translates these insights into practical investment strategies. The focus is on evaluating the portfolio's performance against traditional financial benchmarks like the DAX index, showcasing the potential of AI-enhanced sentiment analysis in stock market prediction and investment decision-making. This dissertation represents a significant step in integrating AI-driven insights with financial portfolio management.

2.2. Discussion of ChatGPT's Potential Applications in Finance and Related Research

The potential applications of ChatGPT in finance, as elucidated in studies from 2023, further illustrate the intersection of AI and finance. Ko and Lee (2023) investigated the use of ChatGPT in optimising investment portfolio management, finding that it enhances asset diversification, beneficial for both retail investors and portfolio managers (Ko and Lee 2023). Bertomeu et al. (2023) analysed the economic impact of the ChatGPT ban in Italy on capital markets, revealing that reliance on ChatGPT significantly affects corporate performance during the ban period (Bertomeu et al. 2023). Lu, Huang, and Li (2023) examined ChatGPT's effectiveness in generating stock recommendations based on complex textual data, indicating its utility as a financial analysis tool (Lu, Huang, and Li 2023). The study by Chena, Wu, and Zhao (2023) reviewed the broader role of generative AI in business and finance, emphasising ChatGPT's applications in corporate sentiment analysis and market prediction (Chena, Wu, and Zhao 2023). Yang and Menczer (2023) evaluated ChatGPT's ability to assess the credibility of news sources, highlighting its potential in fact-checking and media literacy (Yang and Menczer 2023). Lastly, the research by Lopez-Lira and Tang (2023) explored the use of ChatGPT in predicting stock market returns through sentiment analysis, demonstrating a positive correlation with market performance (Lopez-Lira and Tang 2023). However, I contribute to this literature by showing a profitable trading strategy can theoretically outperform the DAX-index (before transaction costs).

Collectively, these studies underscore the expanding capabilities of LLMs, particularly ChatGPT, in the realm of financial markets. They reveal the transformative potential of these models in diverse tasks such as sentiment analysis, policy interpretation, investment decision-making, and market prediction. This array of applications points to a future where AI is increasingly integrated with financial expertise, leading to more sophisticated financial analysis and decision-making processes.

3. Methodology and Data

This section outlines the data collection, preparation, and analysis procedures, providing a comprehensive view of the empirical evaluation of ChatGPT's capability to predict stock price movements for DAX-listed companies. The aim is to detail how the data was obtained and refined to ensure its quality, and to discuss the regression and backtesting approaches used to assess ChatGPT's performance against the DAX index.

3.1. Data Source/Provider

Refinitiv, a subsidiary of the LSEG, is a leading global provider of financial data, making it an ideal choice for research focused on the German stock market, especially the historical data of DAX companies. Refinitiv, partnered with Reuters for 170 years, offers extensive financial data and news, including 7.3 million price updates per second and coverage from over 5,500 global sources, supported by 2,500 journalists. For academic inquiries into the German stock market and DAX companies' historical performance, Refinitiv is an ideal data source, enabling robust, data-driven analyses and enhancing understanding of historical market dynamics (LSEG 2023).

3.2. Methodology

3.2.1. Data Collection

Financial News Headlines

The acquisition of financial news headlines for this study was executed through a systematic process, aligned with academic rigor and precision, ensuring that each headline succinctly encapsulates critical information about financial markets, trends, corporate actions, or economic indicators in a manner that is both accurate and easily digestible for scholarly analysis. I used the Refinitiv Desktop App, a powerful tool known as News Monitor 2.0, as the primary source for gathering financial news headlines. This application allows the search and retrieval of news related to each of the 30 companies constituting the DAX index. I collected data spanning one year, starting in September 2020 and ending in September 2021.

The chosen time frame holds significance as it aligns with the temporal boundaries within which GPT models are pretrained and were kept up to date. Consequently, it ensures the relevancy and compatibility of the data collected with the GPT model used in this study. To enhance the efficiency and relevance of the data collection, a series of strategic filtering steps were meticulously employed. Initially, the data was filtered to retain only "important" headlines, a criterion predefined within the Refinitiv Desktop App. Subsequently, content type filtering was applied, encompassing categories such as "News Feeds," "Breaking News Alerts," "Research," "Transcripts," "Press Releases," and "Web News." This filtering process aimed to curate a dataset comprising only the most pertinent and impactful financial news headlines, effectively mitigating the overwhelming volume of manual data collection. Each data point within the dataset represents a unique news headline associated with a date, time, and one company in the DAX included. These data points were manually extracted from the Refinitiv Desktop App and meticulously recorded in a blank Excel file, preserving the temporal integrity of the news headlines.

Compiling the dataset with unwavering attention to detail resulted in a comprehensive collection of over 8,324 financial news headlines. This dataset acts as the foundational cornerstone of our analysis and offers a rich and extensive source of information for our research objectives.

Asset Pricing and Market Capitalisation

For empirical financial research, obtaining historical asset pricing data is a critical step. In this study, I focused on acquiring data for the thirty companies listed in the DAX index. I achieved this through a detailed and structured process using the Refinitiv Excel Add-in, part of Workspace 365. This tool provided access to the comprehensive Refinitiv Data Platform, instrumental in efficiently gathering vital financial data including market capitalisation.

Fama French Three Factor Model

Furthermore, the data collection for the FF3 model is meticulously sourced from the renowned repository hosted by Dartmouth College (French 2023). Specifically, the European factors of the FF3 model are employed, chosen for their robustness and relevance to the European market context. This data is essential for conducting the regression analysis, aimed at understanding the relationship between stock returns and the three factors: market risk, size, and book-to-market value in the European context.

Challenges in API-Based Data Retrieval

In my research, initially using the Refinitiv Eikon API to gather financial news headlines marked a crucial stage in data collection. Yet, this approach encountered significant challenges, necessitating a shift to a different method. The primary tool intended was the ek.get_news_headlines function in Python. Unfortunately, this method faced substantial issues.

A critical problem was the function's inability to fetch headlines within a certain timeframe, an unexpected issue considering these headlines were available via the Refinitiv Desktop App. Despite extensive efforts to fix this, the issue persisted.

Additionally, the API's limitation of providing only 100 headlines per request presented a hurdle. Addressing this meant considering a complex, loop-based strategy for collecting headlines incrementally, significantly complicating and prolonging the process.

Confronted with these challenges, it became evident that the Eikon API was unsuitable for the research's objectives. This led to the decision to explore an alternative approach to ensure comprehensive acquisition of financial news headlines.

3.2.2. Data Preparation and Cleansing

Data preparation and cleansing are critical steps in any data analysis process, particularly when dealing with complex datasets such as financial news headlines and asset pricing data. In this section, I describe the steps taken to prepare the dataset for evaluating financial news headlines using ChatGPT to provide a sentiment score. This sentiment score serves as an important factor in understanding the impact of news on asset prices. I detail the procedures involved in data collection, cleaning, and the formulation of prompt for ChatGPT.

Data Cleansing

Before conducting sentiment analysis, it was imperative to clean the dataset to ensure its quality and relevance. Duplicate or highly similar news headlines were identified and removed from the dataset. This step helped prevent redundancy and bias in the analysis.

Prompt for Evaluating GPTScore

To facilitate the sentiment analysis, a standardised prompt was evaluated within ChatGPT 3.5 version. This prompt was structured as follows:

"Please disregard all prior instructions and assume the role of a financial expert to evaluate the sentiment of the following financial news headlines related to {company_name}: '{headline}.' Provide a score between -1 (strongly negative) to 1 (strongly positive) to indicate whether the news is positive, neutral, or negative, and generate an output containing the following elements: 'Sentiment,' 'Score,' and 'Reasoning' in tabular form."

The variables *{company_name}* and *{headline}* were substituted with the headlines along with their corresponding company names after the prompt inserted. This substitution allowed us to evaluate the sentiment of each headline systematically. To obtain sentiment scores for the financial news headlines, ChatGPT was instructed using the defined prompt. This prompt played a pivotal role in shaping ChatGPT's responses, providing the initial context and directives necessary for generating meaningful replies. The prompt structure allowed ChatGPT to mimic the role of a financial expert in assessing news sentiment. Within the prompt, ChatGPT was guided to assign sentiment scores between -1 (strongly negative) and +1 (strongly positive) or 0 (neutral) to each news headline, accompanied by a brief reasoning. The reasoning provided additional context for the sentiment score, making the analysis more interpretable.

It was implicitly assumed that the financial news headlines contained sufficient information for an expert to reasonably assess their impact on stock prices. This assumption underscores the importance of informative and well-structured headlines in the financial domain.

For instance, consider the following headline:

Headline	***GPTScore***	***Sentiment***	***Reasoning***
Daimler AG: 'Daimler Truck says batteries, hydrogen are the future.	*0.6*	*Positive*	*Reiteration of Daimler Truck's belief that batteries and hydrogen are the future is consistent with a positive stance on sustainable technology.*

Table 1: *Result of News Headline Prompt in ChatGPT*

Data Enrichment and Preparation

After obtaining sentiment scores and rationale for each news headline, the dataset was subsequently augmented with "GPTScore." An average GPTScore per day was computed for each company. Redundant news headlines were eliminated to guarantee the presence of a unique set of headlines for analysis each day. Subsequently, the asset prices were integrated with financial news headlines into a comprehensive dataset suitable for evaluating the relationship between news sentiment and asset price movements.

3.2.3. Regression Analysis and Backtesting Approach

HmL Portfolio Strategy

The HmL portfolio strategy, commonly referred to as a market anomaly-based investment approach, is characterised by a dual-pronged strategy of acquiring assets exhibiting high returns while simultaneously divesting from assets with lower returns. This investment methodology centres on establishing long positions in assets anticipated to exhibit strong performance (high sentiments) and short positions in assets expected to demonstrate weaker performance (low sentiments) within a specified investment horizon. The fundamental premise underlying the HmL strategy is the exploitation of return differentials between assets grouped according to their historical return patterns (Boynton, Blosick, and Rainish 2015).

Sentiment scores have emerged as a pertinent tool for evaluating the effectiveness of HmL strategies due to their ability to offer valuable insights into prevailing market sentiment and investor sentiment toward specific assets or market segments. The analysis of sentiment scores enables investors to gain a nuanced understanding of market sentiment, facilitating informed decisions regarding the composition of their long and short positions within the HmL strategy. Sentiment scores, generated from diverse sources such as public perception, news sentiment, and social media sentiment, among other factors, can be instrumental in identifying assets that are likely to outperform or underperform relative to their peers (He and Casey 2018).

In the context of the dataset, each GPTScore for every company at any given point in time has been allocated into one of five predefined portfolios based on sentiment strength. These portfolios are delineated by sentiment boundaries, which were established as follows: sentiment scores less than or equal to -0.5 are allocated to Portfolio 1, representing the most negative sentiments; scores ranging from -0.5 to 0.0 are assigned to Portfolio 2; sentiment scores equal to 0.0 are categorised under Portfolio 3, representing neutral sentiments; scores between 0.00 and 0.5 are allocated to Portfolio 4; and scores greater than or equal to 0.5 are assigned to Portfolio 5, signifying predominantly positive sentiments. This allocation ensures a clear correlation between portfolio assignments and the underlying sentiment values, forming the foundation for the subsequent calculation of our HmL portfolio. However, the number of shares in the portfolio fluctuates daily due to changes in the allocation of companies under each sentiment category.

Subsequently, the calculation of the long and short portfolios ensued, representing the Portfolio 1 (HmL Short) and Portfolio 5 (HmL Long). Initially, the number of assets (tickers) on each date was tallied for each portfolio, considering that not all companies necessarily have related news available every day, particularly following the new sentiment score assignment. For return calculation, equal weighting was then applied to each ticker within each portfolio on each date. This weighted approach, combining the prediction of cross-sectional returns via the sentiment scored HML portfolio, generates daily portfolio returns for the next period (t+1). Further analysis involved the consolidation of the long and short portfolios, resulting in the derivation of the HmL portfolio, characterised by the difference between the long and short positions. An additional step in the analysis incorporated the integration of European FF3 into a comprehensive dataset for regression analysis, facilitating a more nuanced understanding of the factors influencing the performance of the HmL portfolio.

Empirical Design: Regression against Fama-French 3-Factor Model

In the realm of financial asset pricing, the FF3 Model, conceptualised by Eugene Fama and Kenneth French in 1993, emerges as a seminal framework. This model represents an advancement over the conventional CAPM by integrating three additional factors that more comprehensively capture the essence of risk and expected returns in the stock market.

Firstly, the Market Risk factor (mktrf) elucidates the concept of excess returns generated from investments in the aggregate market, often represented by benchmark indices like the EURO STOXX. This factor embodies the systematic risks associated with investing in a broad array of stocks, highlighting the inherent volatility and uncertainty prevalent in the market.

Secondly, the Size factor (smb) considers the historical observation that small-capitalisation (small-cap) stocks tend to outperform their larger counterparts over extended periods. This factor aims to quantify the additional expected return that investors might anticipate from portfolios predominantly consisting of small-cap equities, recognising the disproportionate growth potential inherent in these stocks.

Lastly, the HmL factor (hml_ff) underscores the performance differential between value and growth stocks. Value stocks, typically characterised by lower price-to-book ratios, contrast with growth stocks, which exhibit higher ratios. This factor implies that investing in a portfolio skewed towards value stocks can potentially yield higher returns compared to a portfolio heavily weighted in growth stocks (Fama and French 2004).

Sentiment scores into the HmL portfolio regression analysis adds an intriguing dimension to financial research, enhancing our comprehension of the factors that drive portfolio returns. These scores act as indicators of investors' emotional responses and perceptions towards specific stocks or the broader market environment. By including them in the regression model, the study aims to explore the impact of sentiment-driven factors on HmL portfolio returns, potentially surpassing the explanatory power of the traditional FF3 model. The regression equation provided for this analysis is expressed as follows:

$$\boldsymbol{HmL_{t+1} = \beta_0 + \beta_1 \times mktrf_{t+1} + \beta_2 \times smb_{t+1} + \beta_3 \times hml\_ff_{t+1} + \varepsilon}$$

This incorporation of sentiment scores is predicated on the hypothesis that investor sentiment significantly sways market dynamics and, consequently, the returns of the HmL portfolio. These scores, representing the prevailing mood and expectations of investors, might challenge the rational market assumptions commonly held in behavioural finance.

To implement this model, sentiment scores are examined alongside the established factors of the Fama-French model. This methodology tests if the sentiment variable adds extra explanatory value above established risk factors in asset pricing. It entails expanding the regression model to integrate sentiment scores as additional independent variables.

This study seeks to bridge the divide between behavioural finance and empirical asset pricing by incorporating investor sentiment into the conventional Fama-French framework. It aims to shed light on how the emotional and perceptual factors of market participants influence portfolio returns, an aspect often neglected in standard asset pricing models.

The novel aspect of this research lies in its holistic approach, considering both rational economic factors and the behavioural aspects of market participants. By extending the regression model to include sentiment scores along with Fama-French factors, it offers insights into the role of investor sentiment in determining stock returns. The theoretical underpinning of behavioural finance theory suggests that higher sentiment levels, indicative of a more positive investor outlook, should lead to higher future returns. This hypothesis stems from the understanding that investor sentiment can lead to over- or under-reaction in the market, thus temporarily mispricing securities. When sentiment is high, it often reflects a widespread optimism among investors about future economic conditions and corporate performance. This optimism can lead to increased buying activity, pushing up the prices of securities, especially those in the HmL portfolio, which are often more sensitive to market sentiments (López-Cabarcos et al. 2019).

Utilising Python for OLS regression provides a robust statistical framework to quantify the impact of sentiment on HmL portfolio returns, while accounting for established risk factors. This methodological rigor enhances the study's relevance in understanding market dynamics and shaping investment strategies.

Evaluating Performance: Backtesting Approach of the HmL Strategy

Firstly, the HmL strategy was merged with the benchmark data of the DAX companies. Subsequently, the cumulative returns for each portfolio were computed. These cumulative returns were then graphically depicted. The resulting plot illustrated the performance of various portfolios, including HmL, HmL long, HmL short, and the benchmark DAX Index, facilitating a visual assessment of their comparative performance.

In addition to the visual analysis, a more detailed examination of the portfolios was carried out through a rigorous backtesting process. This involved the calculation of several key performance metrics, namely annualised returns, standard deviation, annualised volatility, and the Sharpe ratio. These metrics were computed to provide a comprehensive evaluation of each portfolio's performance.

After this analysis, a robustness test against size effects was conducted by analysing the impact of market capitalisation on portfolio allocation. This involved segmenting the portfolios based on their sentiment and evaluating the market capitalisation across these segments. This step ensures that the performance of the HmL strategy is not disproportionately influenced by the size of the companies within the DAX, providing a more nuanced understanding of its effectiveness across different market caps.

The utilisation of these performance metrics, along with the robustness test against size effects, enabled a quantitative assessment of the risk and return characteristics of the HmL strategy, as well as its variants. This systematic approach provided valuable insights into the effectiveness and suitability of the HmL strategy within the context of the DAX company's benchmark data, aiding in the formulation of informed investment decisions.

4. Results and Analysis

4.1. Regression against Fama-French 3-Factor Model

In this regression analysis, the potential of a sentiment-informed strategy, denoted as HmL_{t+1}, within the European asset pricing market is explored as shown in Table 2. The primary objective was to determine if HmL_{t+1} is adequately priced in the cross-section and whether its performance could be explained by the established FF3 model. The approach involved using a regression model where HmL_{t+1} served as the dependent variable, while the FF3 factors - market, size, and value - acted as independent variables. The choice of the FF3 model is based on its widespread acceptance and proven effectiveness in asset pricing contexts.

An ideal outcome in this study would have been a significant positive alpha (intercept) for HmL_{t+1}, with all coefficients and t-statistics for the FF3 factors being insignificant and trending towards zero. Furthermore, a low R-squared value would indicate the limited explanatory power of FF3 factors over the returns of the HmL_{t+1} strategy. Such a result would imply that HmL_{t+1} captures a unique risk factor not explained by the traditional FF3 model, thereby demonstrating the strategy's effectiveness in generating significant alpha.

Regarding the regression outcomes for the HmL_{t+1} strategy, the following results were observed: The market factor, $mktrf_{t+1}$, showed a coefficient of 0.0076 and a P-value of less than 0.000. This significant positive correlation with a high t-statistic suggests that the HmL_{t+1} strategy tends to perform well when the market rises, indicating a significant positive dependency of the strategy on market movements.

In contrast, the value factor, hml_ff_{t+1}, exhibited a positive, though weaker, correlation. With a coefficient of 0.0032 and a P-value of 0.078, the strategy showed a correlation with the value factor, albeit significant only at the 10% level. This implies a partial alignment of the strategy with the value component of the FF3 model, although not as pronounced as its alignment with the market factor.

Regarding the size factor, smb_{t+1}, the results indicated an almost negligible correlation. The coefficient was -8.232e-05 with a P-value of 0.972. This lack of correlation with size aligns with observations from equally weighted returns and serves as a control for size in the analysis, which will be discussed in full depth in the robustness test of the analysis.

The sentiment strategy HmL_{t+1} itself, represented by the intercept, had a coefficient of 0.0008 and a P-value of 0.354. Further, the annualised intercept, equivalent to the annualised return of the strategy, showing a return of 20.16%, accounting for established risk factors. However, the lack of statistical significance in this intercept implies that the strategy does not yield a significant excess return over the market, given the current data.

The findings indicated that the HmL_{t+1} strategy exhibits a strong correlation with the market and value factors. However, this did not translate into a significant alpha, suggesting that the

returns of the strategy are largely explained by these established risk factors. The R-squared value of 0.247 (respectively 24.7%) reflects the substantial explanatory power of the FF3 factors for HmL returns. Furthermore, the strong correlation with market and value factors, coupled with the insignificant intercept, suggests that the sentiment-based risk factor is not uniquely priced in the cross-section of European markets, particularly within the DAX.
However, the absence of a significant alpha indicates that the strategy does not capture a distinct risk factor beyond the traditional market and value factors.

In conclusion, the regression analysis demonstrates that while the sentiment-informed strategy HmL_{t+1} aligns significantly with market and value factors, it fails to emerge as a unique, priced risk factor in the cross-section of European asset pricing. The absence of a significant alpha and the strategy's correlation with established risk factors suggest that it does not offer an advantage over traditional models in explaining asset returns, underscoring the challenge of identifying new risk factors.

Dep. Variable:		HmL_{t+1}	**R-squared:**			0.247
Model:		OLS	**Adj. R-squared:**			0.237
	coef	**std err**	**z**	**P>[z]**	**[0.025**	**0.975]**
Annualised Intercept	0.2016	0.001	0.926	0.354	-0.001	0.002
Intercept	0.0008	0.001	0.926	0.354	-0.001	0.002
$\boldsymbol{mktrf_{t+1}}$	0.0076	0.001	6.700	0.000	0.005	0.010
$\boldsymbol{smb_{t+1}}$	-8.232e-05	0.002	-0.035	0.972	-0.005	0.005
$\boldsymbol{hml\_ff_{t+1}}$	0.0032	0.002	1.762	0.078	-0.000	0.007

Table 2: *OLS Regression Results against FF3 risk factors*

4.2. Backtesting the Performance of HmL against DAX Benchmark Index

The meticulous process of backtesting investment strategies is essential for investors, offering valuable insights from historical data to forecast the potential performance of an investment strategy. This section presents the backtesting results for various investment portfolios, including the HmL portfolio, along with its subdivisions into long (HmL Long) and short (HmL Short) strategies, and compares their performance with the established DAX benchmark. The detailed performance in Figure A1 underscores the temporal dynamics of these portfolios' returns, offering a clear visual representation of their performance over the selected period.

A detailed examination of the performance chart indicates a consistent outperformance by the HmL and HmL Long portfolios compared to the DAX benchmark. This trend highlights the effectiveness and resilience of these strategies, even in fluctuating market conditions. The red line shows the performance of an investment of € 1 in the HmL portfolio with daily rebalancing of the portfolio. This investment would grow by € 1.42 after one year (excluding transaction costs). Conversely, the HmL Short portfolio exhibits an expected underperformance, reflecting the inherent risks and volatilities associated with short-selling strategies. Interestingly, this negative performance of the short portfolio suggests that implementing a short position in this portfolio as part of the HmL strategy would likely have a positive impact on the overall HmL portfolio, contributing to its resilience and effectiveness.

The chart also shows a notable correlation between the movements of the HmL and HmL Long portfolios, suggesting shared market influences and possibly a similar strategic foundation. In contrast, the DAX benchmark shows more stable variations, indicating a less volatile but also less aggressive market approach.

The backtesting results, illustrated through the performance chart, provide a thorough perspective on the performance and behaviour of the HmL, HmL Long, and HmL Short portfolios in comparison to the DAX benchmark. In the following sections, I will delve into key metrics to further deepen our understanding of these strategies. This detailed examination will enhance our comprehension of each portfolio's strengths and limitations, providing a more nuanced view of their performance and potential applicability in various market conditions.

In the intricate realm of financial portfolio management, striking a balance between anticipated gains and attendant risks stands paramount. This scholarly exposition delves into the performance metrics of various investment strategies, with a specific focus on the HmL strategy in comparison with the DAX. The analysis is grounded in a detailed examination of key financial indicators, including annualised return, volatility, and the SR, as encapsulated in Table 3. These metrics serve as critical instruments in unravelling the complexities of the risk-reward equation inherent in each investment strategy. They have been calculated as commonly used in asset pricing as follows: The annualised return is determined by multiplying the daily return by 252, representing the number of trading days in a year. Annualised volatility is calculated as the standard deviation of the daily returns, then multiplied by the square root of 252, to reflect the asset's price fluctuations over time. Lastly, the annualised Sharpe Ratio, a measure of return versus risk, is computed by dividing the difference between the annualised return and the risk-free rate by the annualised volatility.

	Annualised Return	Annualised Volatility	Annualised SR
HmL	44.95%	26.17%	1.72
HmL Long	23.02%	25.50%	0.90
Dax	20.08%	17.40%	1.20
HmL Short	-21.93%	15.61%	-1.40

Table 3: *Performance Metrics Overview*

When considering the annualised return as a prognostic tool for future performance, the HmL strategy emerges as particularly notable, achieving a substantial return of 44.95%. This figure underscores the strategy's effectiveness within the evaluated period. The HmL Long variant, while displaying a notable profitability with a 23.02% return, does not match the overarching success of the HmL approach. In contrast, the DAX index, with a more conservative growth of 20.08%, reflects a strategy characterised by lower volatility and a less aggressive investment stance.

The discussion on volatility reveals the HmL strategy's heightened risk profile, evidenced by a volatility of 26.17%. This elevated level of risk is mirrored in the potential for higher returns but also indicates a significant exposure to market fluctuations. The HmL Long strategy, with a slightly lower volatility of 25.50%, suggests a more stable approach within the HmL framework. The DAX, with its reduced volatility of 17.40%, stands out as a more balanced and steady option.

The SR highlights the strength of the HmL strategy over the DAX. With an SR of 1.72, the HmL strategy offers high returns with a favourable balance of risk, outperforming the DAX's SR of 1.20. This demonstrates the HmL strategy's superior ability to deliver higher returns at a relatively lower risk compared to the DAX.

However, a crucial aspect to consider in this analysis is the impact of transaction costs, particularly for the HmL strategy. The HmL, being a daily rebalancing portfolio, incurs significant transaction costs, primarily due to the bid-ask spread involved in frequently going long or short on shares. These costs can substantially diminish the net profitability of the strategy. In stark contrast, an investment in an ETF replicating the DAX involves transaction costs primarily at the points of buying and selling the ETF, making it a more cost-effective option in the long run. Therefore, while the HmL strategy may outperform the DAX in terms of gross returns, the nett profitability, when accounting for transaction costs, could render it less advantageous than initially perceived.

In sum, these insights affirm the fundamental principle in finance that higher returns are often accompanied by elevated risks. The HmL strategy, despite its high returns and integration of advanced analysis techniques, is not immune to this rule, as evidenced by its significant volatility. Conversely, the DAX, with its more conservative returns and lower risk, exemplifies the trade-offs investors face, balancing potential profitability against risk exposure. This analysis highlights the importance of a nuanced understanding of investment strategies, where the allure of high returns must be weighed against the potential costs and risks involved.

4.3. Assessing the Stability and Consistency: A Robustness Test of the HmL Strategy

In the realm of financial asset pricing, the HmL investment strategy stands as a pivotal methodology, necessitating comprehensive examination to ascertain its efficacy and resilience amidst varying market conditions. This segment of analysis is devoted to a robustness test, particularly tailored to determine if the returns yielded by this strategy are swayed by the size of the shares involved.

The methodology adopted for this robustness test involves a detailed examination of various portfolios within the HmL strategy framework. This examination is crucial for understanding the influence of market capitalisation on the returns generated by these portfolios.

The initial phase of the analysis, represented by Figure A2, demonstrates the mean market capitalisation across different portfolios. A significant observation is the similarity in market capitalisation between Portfolio 1, representing the short portfolio, and Portfolio 5, representing the long portfolio. This similarity is instrumental in deciphering the underlying dynamics of the HmL strategy, as it shows that market capitalisation has no influence on the HmL strategy since it is not size-driven. The analysis further extends to include the size of stocks in Portfolios 3 and 4, where Portfolio 3 primarily comprises smaller stocks, and Portfolio 4 includes larger stocks. Advancing to Figure A3, the focus shifts to the mean returns per portfolio. This phase corroborates the findings from the market capitalisation analysis, illustrating a notable alignment. It reveals that long portfolios register significantly higher returns compared to the short portfolios, which exhibit lower and often negative returns. This pattern endorses the effectiveness of a positive HmL strategy, crucial for asset pricing and strategy evaluation. The third phase, depicted in Figure A4, delves into the market capitalisation-weighted returns for each portfolio. This analysis is vital as it assesses the returns on a market value-weighted basis, a common practice in asset pricing. The positive correlation observed here is pivotal for the calculation of both equally weighted and market cap weighted returns. It reflects the dependency of the market on larger stocks, as seen in significant indices like the DAX. Furthermore, adopting market capitalisation-weighted returns makes this strategy even more risk-stable because the short portfolio would realise even greater gains.

The crux of this robustness test lies in addressing the question: "Are the results predominantly driven by market capitalisation?" The findings decisively indicate that the returns in the HmL strategy are not primarily driven by market capitalisation. Rather, the risk factor in this strategy is more significantly influenced by sentiment measures, showcasing resilience against size effects. This conclusion not only underscores the robustness of the HmL strategy but also highlights its capability to yield reliable results, independent of market capitalisation dynamics, thereby marking it as a robust trading strategy.

5. Conclusion

5.1. Synthesis of Research Aims and Principal Discoveries

This study embarked on a comprehensive investigation into the efficacy of ChatGPT's sentiment analysis in forecasting stock price movements for companies listed on the DAX index. The primary research objectives were twofold: first, to evaluate if ChatGPT's sentiment analysis is capable of effectively predicting stock price movements in a cross-sectional manner; and second, to ascertain whether utilising sentiment scores from ChatGPT to construct a HmL portfolio represents a profitable strategy that can surpass the performance of benchmark indices like the DAX.

Throughout the course of this research, I conducted an in-depth analysis, which encompassed regression analysis against the FF3 Model and backtesting the performance of the HmL portfolio against the DAX benchmark. The key findings of this study shed light on the intersection of sentiment analysis and financial markets.

5.2. Answer to the Research Question

Considering the extensive analysis conducted in this master dissertation, there will be now conclusive answers provided to the two research questions that have guided this research:

Research Question 1: "Can ChatGPT's sentiment analysis accurately forecast stock price movements for companies in the DAX index?"

The regression analysis indicates that ChatGPT's sentiment-informed strategy, HmL_{t+1}, exhibits significant correlations with market and value factors. However, it fails to generate a significant alpha beyond these established risk factors. The substantial explanatory power of the FF3 for HmL_{t+1} returns suggests that the strategy's returns are primarily explained by traditional market factors. Thus, ChatGPT's sentiment analysis, in this context, does not provide accurate forecasts of stock price movements beyond established factors.

Research Question 2: "Is it profitable to trade on ChatGPT-Sentiment using a cross-sectional HmL strategy? Does this strategy outperform other relevant indices, like the DAX?"

Our backtesting analysis reveals that the HmL and HmL Long portfolios consistently outperform the DAX benchmark over the selected period, highlighting their effectiveness and resilience, even in fluctuating market conditions. However, it's essential to consider the impact of transaction costs, particularly for the HmL strategy, which may affect its net profitability. Therefore, while the HmL portfolio may outperform the DAX in terms of gross returns, careful evaluation of transaction costs is necessary when assessing its overall advantage.

Additionally, our robustness test underscores the HmL strategy's resilience against size effects and confirms that the returns in the HmL strategy are not primarily driven by market capitalisation but are more significantly influenced by sentiment measures. This reaffirms the HmL strategy's robustness and reliability in delivering results, irrespective of market capitalisation dynamics. Furthermore, this robustness test demonstrates that using market capitalisation-weighted return results in even higher gains due to its more negatively oriented short portfolio. However, the strategy leans towards value, so they primarily invest in stocks with a high book-to-market ratio (BE/ME) and tend to short stocks with a low book-to-market ratio (BE/ME).

In conclusion, ChatGPT's sentiment analysis, while not significantly improving upon established factors, provides insights into stock price movements. The HmL portfolio constructed using ChatGPT's sentiment scores shows promise in outperforming the DAX benchmark, with transaction costs as a consideration. The strategy's robustness is evident, as it consistently delivers results regardless of market capitalisation dynamics.

5.3. Limitations and Recommendations for Future Research

The study's primary limitation was its reliance on manually collected data from just one year, which limited the scope and generalisability of the findings. This limitation was particularly true in assessing sentiment analysis under varying economic conditions. Future research should broaden data collection across multiple years to better understand sentiment analysis' effectiveness through different economic cycles, which would require advanced, automated data retrieval methods.

Additionally, the study only utilised data available until September 2021, constrained by the limitations of pre-trained datasets for ChatGPT. However, with Open Ai's recent announcement that ChatGPT can now browse the internet to provide current and authoritative information, including direct links to sources, there's no longer a restriction to data before September 2021 (2023 Open AI). This advancement opens new possibilities for research, enabling the development of real-time data integrations with APIs for tools like portfolio recommendations, advising whether to go long or short on shares.

Moreover, the study encountered challenges with the varying levels of detail in news headlines, impacting the accuracy of sentiment analysis for predicting stock prices. To enhance this, future research should include analyses of both headlines and full articles, employing advanced natural language processing tools for more precise sentiment analysis.

Finally, the study did not adequately account for transaction costs in its backtesting methodology. This omission could lead to an overly optimistic view of the trading strategy's performance. Future studies should include a detailed assessment of transaction costs for a more realistic evaluation of the strategy's effectiveness, thereby enhancing the research's credibility and applicability. However, I can observe that ChatGPT sentiment theoretically is crucial for stock returns.

5.4. Implications and Future Outlook

The integration of AI in financial markets, particularly through sentiment analysis, represents a significant advancement in financial economics. The study's investigation into the HmL strategy within European asset pricing highlights both the potential and limitations of AI in this area. It reveals that while AI-driven methods align with certain market factors, they have yet to establish themselves as revolutionary tools that can consistently outperform established financial models. This suggests that AI's role in financial markets is currently complementary, helping to process and analyse large datasets but not yet overcoming the complexities of financial markets entirely.

The challenge lies in developing AI strategies that can reliably surpass traditional models, indicating the complexity and unpredictability of financial markets. This finding paves the way for future research, emphasising the need for more advanced AI techniques and broader data analysis.

Furthermore, the growing use of AI in financial strategies raises important ethical and labour market considerations. The potential displacement of traditional analysis jobs necessitates shifting skill sets towards data science and AI competencies. Ethically, there's a need for transparency and accountability in AI-driven decisions to avoid market imbalances. Regulatory bodies must also adapt, creating frameworks to manage the complexities of AI-driven financial strategies while ensuring market fairness.

In summary, the study illuminates AI's role in financial markets, highlighting its current limitations and the need for further advancements. It underscores the importance of a balanced approach to AI integration, considering potential risks and benefits. As AI evolves, continuous assessment and adaptation are essential for its effective and responsible integration into financial markets.

References

Bertomeu, Jeremy, Yupeng Lin, Yibin Liu, and Zhenghui Ni. 2023. "Capital Market Consequences of Generative AI: Early Evidence from the Ban of ChatGPT in Italy." SSRN. https://doi.org/10.2139/ssrn.4452670.

Bond, Shaun Alexander, Hayden Klok, and Min Zhu. 2023. "Large Language Models and Financial Market Sentiment." SSRN. https://doi.org/10.2139/ssrn.4584928.

Chen, Boyang, Zongxiao Wu, and Ruoran Zhao. 2023. "From Fiction to Fact: The Growing Role of Generative AI in Business and Finance." Journal of Chinese Economic and Business Studies 21 (4): 471-496. https://doi.org/10.1080/14765284.2023.2245279.

David Noever and Forrest McKee. "Numeracy from Literacy: Data Science as an Emergent Skill from Large Language Models." ArXiv, abs/2301.13382 (2023). https://doi.org/10.48550/arXiv.2301.13382.

Fama, Eugene F., and Kenneth R. French. 2004. "The Capital Asset Pricing Model: Theory and Evidence." Journal of Economic Perspectives 18 (3): 25-46. https://doi.org/10.1257/0895330042162430.

French, Kenneth R. n.d. "Data Library." Accessed 2023, December 16. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Gürsoy, S., and M. Doğan. 2023. "Examining The Use of ChatGPT in Financial Markets with Swot Analysis." TroyAcademy 8 (3): 296-305. https://doi.org/10.31454/troyacademy.1363366.

Hansen, Anne Lundgaard, and Sophia Kazinnik. 2023. "Can ChatGPT Decipher Fedspeak?" SSRN. https://doi.org/10.2139/ssrn.4399406.

Ko, Hyungjin, and Jaewook Lee. 2023. "Can Chatgpt Improve Investment Decision? From a Portfolio Management Perspective." SSRN. https://doi.org/10.2139/ssrn.4390529.

Korinek, Anton. 2023. "Language Models and Cognitive Automation for Economic Research." NBER Working Paper No. 30957. National Bureau of Economic Research.

Ling T. He and K. Casey. "Improvements in forecasting of bank stock excess returns using the investor sentiment endurance index: a comparison with CAPM and Fama-French models." International Journal of Financial Markets and Derivatives, 6 (2018): 210-224. https://doi.org/10.1504/IJFMD.2018.10012859.

Lopez-Lira, Alejandro, and Yuehua Tang. 2023. "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models." SSRN. https://doi.org/10.2139/ssrn.4412788.

López-Cabarcos, M. Ángeles, Ada M. Pérez-Pico, Paula Vázquez-Rodríguez, and M. Luisa López-Pérez. 2019. "Investor Sentiment in the Theoretical Field of Behavioural Finance." Economic Research-Ekonomska Istraživanja 32 (1): 2101-2119. https://doi.org/10.1080/1331677X.2018.1559748.

Lu, Fangzhou, Lei Huang, and Sixuan Li. 2023. "ChatGPT, Generative AI, and Investment Advisory." SSRN. https://doi.org/10.2139/ssrn.4519182.

LSEG. 2023. 'Home.' Accessed December 14, 2023. https://www.lseg.com/en/

OpenAI (@OpenAI). 2023. "ChatGPT can now browse the internet to provide you with current and authoritative information, complete with direct links to sources. It is no longer limited to data before September 2021." Twitter, September 27, 6:00 PM. https://twitter.com/OpenAI/status/1707077710047216095.

Qihuang Zhong, Liang Ding, Juhua Liu, Bo Du and Dacheng Tao. "Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT." ArXiv, abs/2302.10198 (2023). https://doi.org/10.48550/arXiv.2302.10198.

Wentworth Boynton, Gregory Blosick and Robert Rainish. "The principal–agent problem, tracking error, and the optimal investment portfolio." Applied Economics Letters, 22 (2015): 239 - 246. https://doi.org/10.1080/13504851.2014.934429.

Wu, Shijie, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. "BloombergGPT: A Large Language Model for Finance." arXiv. https://doi.org/10.48550/arXiv.2303.17564.

Xie, Qianqian, Weiguang Han, Yanzhao Lai, Min Peng, and Jimin Huang. 2023. "The Wall Street Neophyte: A Zero-Shot Analysis of ChatGPT Over Multimodal Stock Movement Prediction Challenges." arXiv. https://doi.org/10.48550/arXiv.2304.05351.

Yang, Kai-Cheng, and Filippo Menczer. 2023. "Large Language Models Can Rate News Outlet Credibility." arXiv. https://doi.org/10.48550/arXiv.2304.00228.

Zhang, Xuanyu, Qing Yang, and Dongliang Xu. 2023. "XuanYuan 2.0: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters." arXiv. https://doi.org/10.48550/arXiv.2305.12002.

Kostadin Mishev, Ana Gjorgjevikj, I. Vodenska, Lubomir T. Chitkushev and D. Trajanov. "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers." IEEE Access, 8 (2020): 131662-131682. https://doi.org/10.1109/ACCESS.2020.3009626.

Appendix

Figure A1: *HmL-performance (equal) vs. DAX30 of € 1 investing in HmL portfolio*

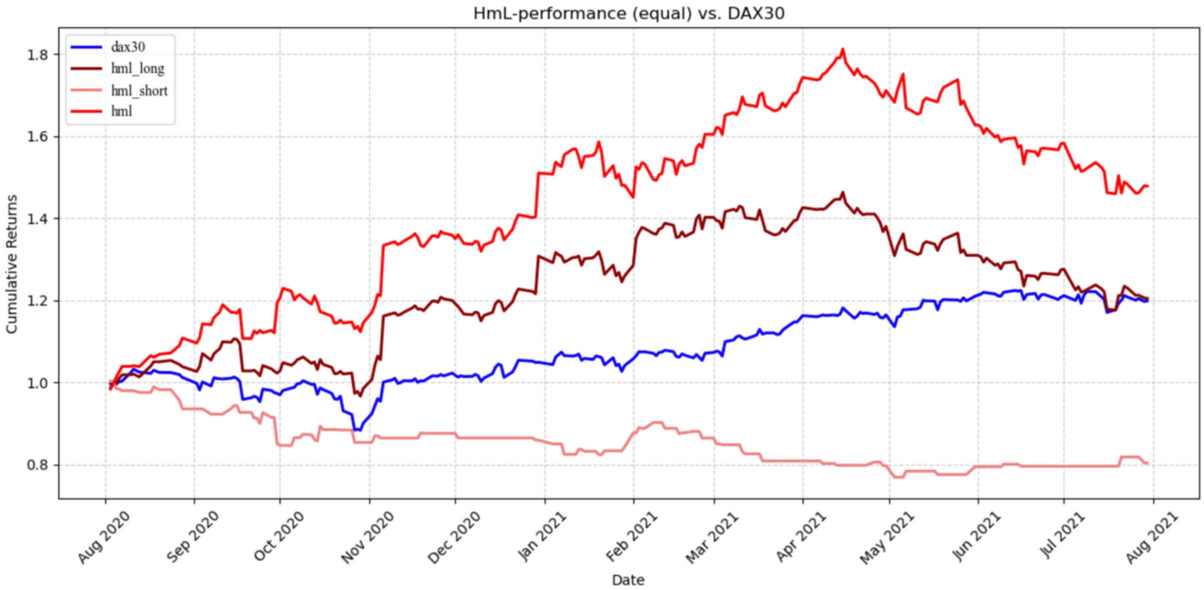

Figure A2: *Mean Market Capitalisation by Portfolio Allocation*

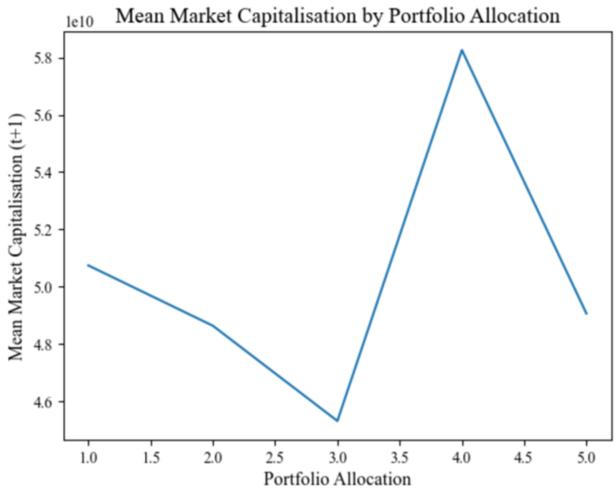

Figure A3: *Mean Return by Portfolio Allocation*

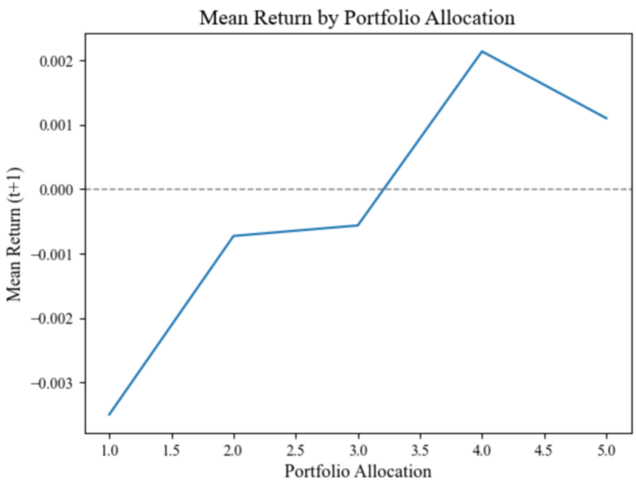

***Figure A4**: Market Capitalisation-weighted Return by Portfolio Allocation*