6CCS3PRJ Individual Project
AI Methods for Portfolio Optimization in Taiwan's Stock Market: Analyzing Financial and Economic Data

Final Project Report

Chun-I Chien
K20022418
Supervised by: Prof. Peter McBurney
Programme of Study: BSc Computer Science

April 5, 2024

Originality Avowal

I verify that I am the sole author of this report, except where explicitly stated to the contrary. I grant the right to Kings College London to make paper and electronic copies of the submitted work for purposes of marking, plagiarism detection and archival, and to upload a copy of the work to Turnitin or another trusted plagiarism detection service. I confirm this report does not exceed 25,000 words.

Chun-I Chien
April 5, 2024

Abstract

This project employs Artificial Intelligence (AI) techniques to optimize investment portfolios within Taiwan's stock market, anchoring on the principles of Modern Portfolio Theory (MPT). The objective is to devise a portfolio that minimizes risk while guaranteeing a minimum expected return equivalent to or surpassing the Taiwan 50 Index.

The project includes the comprehensive process of economic data acquisition, cleansing, management, stock selection, and portfolio construction based on MPT principles. Specifically, the project focuses on risk reduction under defined constraints, including the attainment of a minimum expected yield that aligns with the benchmark Taiwan 50 Index. Through algorithmic implementation and analysis, this research not only delineates the portfolio construction process but also provides insightful evaluations of the outcomes, highlighting the practical utility and insights gleaned from the application of AI in achieving optimized portfolio performance in the context of Taiwan's financial market.

Acknowledgement

Firstly, I would like to express my appreciation to my supervisor Professor Peter McBurney for his helpful guidance and great support throughout the whole project.

Secondly, my heartfelt appreciation goes to my family, who offered unwavering support and patience throughout the duration of my study, making it possible for me to focus on my research.

Finally, I would like to extend a special thanks to Mayday, whose music was a constant companion through the long days and nights of this research. Their music provided me with endless spiritual support, helping me overcome numerous challenges. Their creations have not only enriched my life but also invisibly motivated me to stay positive and perseverant.

Contents

1. Introduction

In recent years, among the most significant advancements in AI is the emergence of models like ChatGPT, which has revolutionized investment strategies within stock markets. This revolution is highlighted in Romanko, Narayan and Kwon's (2023) study 'ChatGPT-Based Investment Portfolio Selection', demonstrating the effective use of generative AI models for stock selection and portfolio optimization. These models utilize sophisticated algorithms to analyze market trends, predict stock performance, and provide valuable investment insights, particularly beneficial in handling extensive datasets.

In light of these developments, my research aims to delve deeper into portfolio optimization, particularly in Taiwan's stock market. The central question is devising a portfolio focusing on stock selection of minimizing portfolio risk subject to certain constraints, including a minimum expected return at least equal to or surpassing the Taiwan 50 index return. Additionally, the study seeks to identify if certain stocks are particularly advantageous for inclusion in such portfolios. By exploring these aspects, this research aims to contribute to the evolving field of AI in financial decision-making, particularly in understanding and applying AI-driven insights to Taiwan's dynamic stock market.

1.1. Project Motivations

Since the release of ChatGPT by OpenAI in 2022 (OpenAI, 2022), an era of advanced AI applications has been ushered in, significantly affecting various sectors, including finance. To gauge the impact of this milestone on AI development, I utilized Google Scholar to observe the frequency of search results related to Taiwan's financial market across different timeframes. The analysis, as depicted in the chart below, spans from the establishment of AI in the 1950s to the present day, with the 1950s acknowledged as the dawn of AI (Cordeschi, 2006). Notably, the search result counts indicate a pronounced uptick in the number of entries within the 2022-2023 period compared to the cumulative count from 1950 to 2021. This suggests an active and growing interest in the application of AI to Taiwan's financial markets. Coinciding with the introduction of transformative AI technologies like ChatGPT, this trend implies that the financial sector is poised for significant AI-driven transformation. My research intends to contribute to this dynamic and expanding field by investigating how AI can specifically optimize investment strategies in

Taiwan's stock market, potentially enhancing risk management and yielding higher investment returns, thus shaping the future landscape of financial decision-making in Taiwan.

Research Key words Input	Search Results time period: 1950 - 2021	Search Results time period: 2022 - 2023	Total Research until 2023
AI in Taiwan's stock market	21600	12900	34500
Artificial Intelligence and Taiwan finance	20300	26900	47200
Portfolio optimization in Taiwan, AI	11400	5080	16480
Machine Learning in Taiwanese stock market	19700	18200	37900
Financial technology Taiwan, AI	40900	46300	87200
Stock market analysis Taiwan AI	19500	23600	43100
Emerging trends in Taiwan financial market AI	19000	25200	44200
Taiwan stock market AI research	178000	37900	215900
AI-driven investment strategies Taiwan	16500	6850	23350

▲ **Table 1.1: Comparative Analysis of AI Research in Taiwan's Financial Sector Before and After the Release of ChatGPT**

1.2. Scope

This study aims to find out the optimal diverse portfolio at least in Taiwan's stock market while managing the expected returns and minimizing risks. It looks at data from January 1, 2010, to January 1, 2020, a time chosen to avoid the confusion caused by big economic events like the 2008-2009 Global Recession and the 2020-2021 Global Pandemic. By focusing only on Taiwan's market, the research can really dive into what makes it special. The information for the study mainly comes from the Taiwan Stock Exchange Official Website, including Taiwan 50 Index, listed stocks, giving a full picture of the market over ten years. Modern Portfolio Theory and Monte Carlo Simulation, two well-known financial methods, will be used to create and check different portfolios. The heavy lifting for calculations will be done with Python and its powerful tools like Quantlib, NumPy, and Pandas.

The goal is to make sense of how AI can help build financial strategies and to give insights that could guide future money-making choices in Taiwan's lively stock market.

2. Background

Between 2010 and 2020, Taiwan's stock market experienced significant evolution, shaped by a confluence of global economic events, domestic policy shifts, and technological advancements. The market was initially influenced by the aftermath of the 2008 global financial crisis, impacting investor confidence and financial stability (Chung-Hua, 2010). The period saw Taiwan navigating the complexities of the European Sovereign Debt Crisis and the repercussions of the US Credit Rating Downgrade, which sent ripples across international markets, including Taiwan.

Domestically, Taiwan's economic landscape was reshaped by key policy initiatives, notably the Economic Cooperation Framework Agreement (ECFA) with China (The International Trade Administration, MOEA, 2018), fostering closer trade and economic ties. This era also witnessed substantial growth in Taiwan’s technology sector, particularly in semiconductors and electronics, driven by global demand and technological innovation. The government’s push towards renewable energy sources further diversified the market, attracting new investments and changing the market composition.

The US-China Trade War (Chow, 2020), spanning 2018-2019, presented both challenges and opportunities for Taiwan's stock market, impacting supply chains and trade dynamics. Throughout the decade, regulatory changes aimed at enhancing market transparency and corporate governance were implemented, aligning Taiwan's financial market with international standards and attracting foreign investment.

These factors collectively contributed to a dynamic period of growth and transition for Taiwan's stock market, reflecting its adaptive response to both internal developments and global economic trends. This background sets the stage for examining how AI technologies like ChatGPT can be integrated into financial strategies to navigate and capitalize on such a complex market environment.

2.1 Literature Review

Modern Portfolio Theory (MPT), a concept introduced by Harry Markowitz, continues to be a fundamental framework in portfolio management, emphasizing the crucial balance between risk and return. Pedersen (2014) explores the relevance of MPT in contemporary settings, comparing it with the Kelly criterion to emphasize the significance of risk-return optimization across diverse financial environments. Similarly, Fajartama and Faturohman (2021) demonstrate the adaptability of MPT in sector-specific scenarios, particularly in the Indonesian automotive financing industry, showing its versatility under different economic conditions. While these studies do not specifically focus on Taiwan, their findings and methodologies offer valuable insights for various global contexts, including the Taiwanese market.

The integration of Monte Carlo simulations into portfolio optimization represents a significant advancement, especially in volatile markets. Sihombing (2013) illustrates how combining these simulations with linear programming within the MPT framework refines investment strategies, demonstrating its effectiveness in markets such as the Indonesian capital market. Additionally, the application of Monte Carlo simulations in risk assessment, particularly in calculating Value at Risk (VaR), is exemplified in Ghodrati and Zahiri (2014), further underscoring the methodology's robustness.

Beyond traditional MPT, innovative approaches like classifier systems in portfolio management have emerged, indicating new strategies in asset selection. The study conducted by Chen, Lin, and Chen (2007) on the Taiwan 50 Index exemplifies this, providing insights into alternative asset selection strategies. Moreover, Ai et al. (2011) broaden the scope of financial theories by integrating strategic enterprise risk management into financial decision-making, highlighting its importance in risk-optimized business strategies.

Complementing these traditional models, the advent of AI and predictive analytics in finance, as discussed in Broby (2022), has brought about transformative changes in financial decision-making and risk management. Broby's comprehensive overview of predictive analytics techniques in finance illustrates the application of methods such as classification, regression, and time series models in economic prediction, stock price volatility prediction, and optimal portfolio prediction. Although not solely focused on AI, this study underscores the increasing importance of data-driven predictive techniques in finance.

AI's role in asset selection and risk assessment is increasingly significant, as various studies demonstrate its effectiveness in analyzing financial news, market sentiments, and quantitative data. This integration of AI and traditional financial theories like MPT and Monte Carlo simulations is particularly relevant in dynamic markets such as Taiwan's, offering novel perspectives and methods to optimize investment strategies and manage risks.

In summary, the evolving field of AI and predictive analytics in finance, alongside traditional theories like MPT and Monte Carlo simulations, represents a significant area of growth and innovation. The integration of these approaches offers promising avenues for enhancing investment strategies and financial analysis, particularly in dynamic markets like Taiwan.

3. Requirements Specification

3.1 Requirements

The successful execution of this research project hinges on meeting specific requirements. Firstly, accurate and comprehensive financial data from January 1, 2010, to January 1, 2020, is crucial, to be mainly sourced from Taiwan Stock Exchange Corporation (TWSE). TWSE is a website which contains a wealth of real-time information, such as "market information," "indices information," "listed companies" with constant updates on market conditions to provide a comprehensive range of information and services. This website also includes the historical stock prices, trading volumes, and fundamental company data. Technologically, the project requires the integration of Python for data processing and analysis, utilizing libraries such as NumPy and Pandas, along with Quantlib for financial computations. Methodologically, the application of Modern Portfolio Theory and Monte Carlo Simulation is necessary to construct and evaluate the investment portfolios. These requirements are critical to ensure the project's rigor and reliability, providing a robust foundation for exploring AI's impact on portfolio optimization in Taiwan's stock market.

3.2 Specification

The specifications of the research project include a carefully curated integration of cutting-edge artificial intelligence technology, ChatGPT, with well-established financial methodologies, Modern Portfolio Theory (MPT), and Monte Carlo simulations, in order to achieve optimal portfolio optimisation of the Taiwan stock market landscape in a complex environment.

Ensuring the integrity of the project's data foundation is paramount. A systematic approach to data collection and processing will be rigorously executed, focusing on key financial metrics, and drawing from reputable sources, and which is mainly from Taiwan Stock Exchange Official Website (TWSE). This source offers a comprehensive spectrum of data, including historical stock prices, trading volumes, financial statements, market listings, and indices. This wealth of data is indispensable for conducting granular market trend analysis, evaluating long-term company performance, and facilitating quantitative assessments of factors like price fluctuations and trading volumes.

Methodologically, the project will employ a dual-pronged approach, with MPT and Monte Carlo Simulation as its primary pillars. Modern Portfolio Theory will be leveraged to construct efficient investment portfolios. Through the calculation of expected returns, variance, and covariance of stock returns, the project will strategically determine optimal asset allocations that minimise the risk while managing expected returns. Complementing this, Monte Carlo Simulation will be employed to simulate a multitude of portfolio scenarios under diverse market conditions, providing valuable insights into risk profiles and potential returns.

Supporting these methodologies are a suite of auxiliary tools and technologies designed to enhance research accuracy and efficiency. Quantlib, alongside Python libraries like NumPy and Pandas, will form the bedrock for financial calculations and data analysis. Python's adaptability will be instrumental in data collection, processing, and the implementation of MPT and Monte Carlo Simulation models, allowing for a comprehensive, dynamic, and responsive approach to portfolio optimization within the dynamic context of Taiwan's stock market. Incorporating ChatGPT as an auxiliary tool offers real-time insights, augmenting the project's adaptability and ongoing evolution. This advanced capability ensures the project remains at the forefront of Taiwan's changing stock market environment. With the multitude of portfolio scenarios generated by Monte Carlo simulations, ChatGPT assists in interpreting the underlying dynamics and potential impacts of these simulations, offering a deeper understanding of risk profiles and potential returns.

Throughout this project, meticulous documentation will be upheld, ensuring transparency and facilitating future analysis. The project's adaptability and continuous evolution, driven by the integration of ChatGPT's real-time insights, will further enhance its ability to navigate and excel in the ever-evolving landscape of Taiwan's stock market.

4. Design and Implementation

The design of this research project is an essential aspect that outlines the systematic approach to achieving the project's objectives and the methodologies for implementing the portfolio optimization strategy. It ensures that the project's execution aligns with the defined specifications while emphasizing the practical aspects of implementation.

4.1 Data Collections and Cleaning

- Objective

The objective of this process is to obtain and refine the required financial dataset, primarily sourced from the Taiwan Stock Exchange (TWSE) and supplemented by other credible economic databases. The accuracy and integrity of all the data are crucial, as this would be used for further process of portfolio optimization. Therefore, ensuring data's precision, relevance, and consistency, involving comprehensive steps for filtering, normalization and handling missing values are imperative.

- Tools & Technologies

Python, renowned for its wide-range and powerful functions and the extensive support of its libraries, served as the primary tool for data manipulation and cleaning. Particularly, the Pandas library was helpful because of its solid data structures and methods made for easy manipulate of structured data. The main procedures and tools utilized in the data collection and cleaning process are described below.

4.1.1. Data Sourcing

Data initially downloaded from Taiwan Stock Exchange (TWSE) website, including stock information such as stock code, stock name, monthly stock price. Taiwan 50 Index, and Taiwan 50 Return Index were downloaded from Taiwan Open Government Data (OGD) Platform. All the acquisition of financial data is within 10 years and between Jan. 2010 and Dec. 2019.

The table below includes the original downloaded datasets and their details.

File Name	Data Type	Description	Data Source
chk201001.xls chk201002.xls … chk201912.xls	Listed Companies Monthly Statistics for Status of Securities Listed on TWSE	The Status of Securities Listed on Taiwan Stock Exchange between Jan. 2010 and Dec.2019, for ten-year period.	TWSE website
2010-01.xls 2010-02.xls … 2019-12.xls	Listed Companies Monthly Statistics for P/E Ratio & Yield & P/B Ratio of Listed Stocks	Latest Price, PER, Yield, and PBR for listed stock between Jan. 2010 and Dec. 2019, for ten-year period.	TWSE website
TAI50I-2010-01.csv TAI50I-2010-02.csv … TAI50I-2019-12.csv	Taiwan 50 Index and Taiwan 50 Total Return Index	Taiwan 50 Index and Taiwan 50 Total Return Index between Jan. 2010 and Dec. 2019, for ten-year period.	Taiwan Open Government Data (OGD) Platform

▲ **Table 4.1: original downloaded datasets and their details**

4.1.2. Data Cleaning

After retrieving all the required data, filtering procedure involved using Python to extract necessary data from original Excel files, removing irrelevant columns that did not contribute to the analysis. The obtained financial data was then normalized to csv files, in order to better process the data while Python analysis. Missing data points were handled using different strategies based on their nature and impact on future analysis.

Below is a table summarizing the specific data handling actions performed during this process:

Action	Description
Data Extraction	Relevant data was extracted from the original Excel files, focusing on the required financial metrics.
Column Removal	Unnecessary columns were identified and removed to ensure the dataset only contained variables imperative for analysis.
File Conversion	The data was converted from Excel (.xls) format to Comma-Separated Values (CSV) format for better applicability with Python's data analysis libraries.
Missing Data Handling	Strategies were implemented to deal with missing values, including data interpolation or removal based on their impact on the integrity of the dataset.
Data Translation	Chinese column titles and content were translated into English to align with international research standards and improve the accessibility of the data.

▲ **Table 4.2: process actions**

4.1.3. Data Accuracy Confirmation

Checking for the precision of the financial data was mainly based on manual check. For example, a subset data was manually reviewed for consistency with TWSE published record.

Through these several processes, including application of Python and its libraries, this project established a solid foundation of financial data collection which would be further applied to Modern Portfolio Theory (MPT) and Monte Carlo simulations in order to explore and examine the portfolio optimization in Taiwan's stock market between Jan. 2010 and

Dec. 2019.

4.2 Data Selection

- Objective

The objective of this stage is to filter out stocks with poor performance from the dataset, ensuring the preliminary stocks had the potential to contribute to the portfolio optimization. Dataset was refined by selecting stocks with exhibited stability, liquidity, and positive expected returns over a decade. This approach aims to ensure the subsequent portfolio optimization is built based on consistent performance and reliability.

- Tools & Technologies

Pandas and NumPy are the main libraries used in Python while data selection process. Pandas is used for data manipulation, allowing efficient implementation of filtering criteria; and NumPy is used for data statistical calculations, including computation of the average trading volume over a decade, monthly expected returns for each stock, annualized expected returns, volatility, and annualized volatility.

4.2.1. Longevity

The first step of stock selection is to ensure that only include stocks that have been continuously listed on TWSE, in other words, have continuously existed in Taiwan's stock markets for the entire ten-year period under this research. This condition is aimed at selecting stocks with proven stability in the market which is contributing to portfolio optimization.

4.2.2. Liquidity (Average Trading Volume)

Evaluate the liquidity of stocks after checking the longevity. The liquidity is based on stocks' average trading volume over the ten-year period. The stock which with a lower average trading volume than the ten-year average volume in all stocks would be filtered. This step confirms that selected stocks have enough market activity to facilitate buying and selling without significant price impacts. The formula of this calculation is shown below.

$$\text{Average Trading Volume} = \frac{\sum_{i=1}^{n} \text{Monthly Volume}_i}{n}$$

- where n is the total number of months

- $MonthlyVolume_i$ represents the trading volume of month i

4.2.3. Expected Return

Compute the expected return value for each stock based on the monthly historical price data. The formula is shown below.

$$\text{Expected Return} = \frac{P_{\text{current month}} - P_{\text{previous month}}}{P_{\text{previous month}}}$$

- $P_{\text{current month}}$ represents the closing price of the current month
- $P_{\text{previous month}}$ represents the closing price of the previous month

This computation facilitates the assessment of each stock's performance from one month to the next, providing a standardized metric to compare across different time periods and stocks.

4.2.4. Data Segregation

The data was further processed to segregate and store each stock's information individually after computing the monthly expected returns for each stock and add them to a new DataFrame and save as a CSV file to organize and facilitate more focused analysis. The individual CSV files were names using combination of the stock code and stock name to ensure clarity and consistency. The file includes stock code, stock name, average trading volume, monthly expected returns, and the date.

Below is an example table illustrating the process for a subset of the stocks (after the stock selection for longevity).

Stock Code	Stock Name	Data File Created
1102	亞洲水泥 ACC	1102_亞洲水泥 ACC
1103	嘉新水泥 CHC	1103_嘉新水泥 CHC

▲ **Table 4.3: showing how the name of data file being created**

4.2.5. Annualized Expected Return & Volatility

For each stock, calculate the monthly expected return value, annualized expected return and then calculate the volatility (standard deviation) over this ten-year period. These measurements help choose a diverse group of securities that strikes a balance between allowable risk and potential rewards by offering insights into each stock's risk return profile. This data is saves as 'MPT_algorithm_data.csv' to be further applied for stock selection. The formulas of annualized expected return and volatility are shown below.

4.2.6. Filtering Based on Liquidity & Expected Returns

While the liquidity and annualized expected return is obtained, eliminate stocks that have an average trading volume lower than the ten-year average across all stocks, ensuring adequate liquidity. Additionally, only reserving the stocks with a positive annualized expected return over the period, preliminarily guaranteeing the rate of returns for portfolio optimization.

4.3 Implementation of Modern Portfolio Theory (MPT)

- Objective

  The aim of this process is to apply Modern Portfolio Theory (MPT) using Covariance Matrix Calculation and Quadratic Programming to minimize the risk, while maximizing portfolio returns, relative to the Taiwan 50 Total Return Index. This stage includes several statistical and computation methods to construct a portfolio optimization which attaches to the MPT principles of diversification and risk management.
- Tools & Technologies

  A number of analytical tools are used from Python environment. Python and NumPy are pivotal for data manipulation, statistical calculations, and handling time-series data. The covariance matrix is regularized using the LedoitWolf estimator from the sklearn.covariance module, ensuring its positive definiteness. The final optimization applying quadratic programming to solve the problem, and the required functions are provided by the CVXOPT package.

4.3.1. Annualized Taiwan 50 Return Index Calculation

Compute annualized Taiwan 50 Return Index based on ten-year data to be the target return in the following computation. The formula and procedure are shown below.

$$\text{Calculate Monthly Log Returns: } r_{\log}(t) = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

$$\text{Compute Average Monthly Log Return: } \overline{r_{\log}} = \frac{1}{n}\sum_{t=1}^{n} r_{\log}(t)$$

$$\text{Annualized Average Log Return} = \overline{r_{log}} \times 12$$

- P_t represents the index value at the end of month t
- P_{t-1} $represents$ the index value at the end of month $t-1$
- $r_{\log}(t)$ represents the log return for month t
- where n is number of monthly log returns

4.3.2. Covariance Matrix of Expected Returns

Expected returns data is prepared for selected stocks, and a covariance matrix is generated using NumPy, highlighting the relationships between each stock return in order to manage the risk of the portfolio. Matrix represents the variance of each stock's returns, indicating the level of individual stock's risk. By analyzing this matrix, this could be used to create the portfolio optimization which balances risk and return.

4.3.3. Regularization of Covariance Matrix

Check if the covariance matrix of the selected stock returns is at least semi-positive definite or positive definite. As the covariance matrix is not semi-positive definite, apply LedoitWolf shrinkage estimator to regularize the covariance. This step ensures the matrix is positive definite, making it suitable for further optimization algorithm. After regularization, check if the regularized covariance matrix is positive definite again.

4.3.4. Quadratic Programming Optimization Model

Implement MPT by applying the quadratic programming optimization model to determine the optimal weights of each selected stocks in the portfolio. Specifically, this approach focuses on minimizing portfolio risk while ensuring the expected return at least equal to a particular target based on the Annualized Taiwan 50

Return Index. This stage mainly utilized Python libraries including Pandas for data handling, NumPy for numerical computations, and 'cvxopt' for solving the optimization problem using quadratic programming.

Constraints are listed below:

- The sum of the asset weights must equal to 1.
- All assets wights are non-negative.
- The portfolio's return is at least equal to a specified target based on the Annualized Taiwan 50 Return Index.

These constraints ensure to maintain a minimum expected return, and all investments are non-negative with the total allocation not exceeding the available capital.

To set up the optimization problem:

$$\text{Objective Function: minimize} \quad \frac{1}{2}x^T P x$$

- where P is the covariance matrix of asset returns, here is the regularized covariance matrix
- x represents the vector of asset weights in the portfolio, this is the problem needed to be solved

Constraints:

$$Gx \leq h, \quad \text{where } G = -I \text{ and } h = 0$$

- G is a matrix to ensure that all weights are non-negative
- h is a zero vector of suitable dimensions

$$Ax = b, \quad \text{where } A = [1,1,\ldots,1] \text{ and } b = 1$$

- A is a row vector of ones, representing the sum of weights
- b is a scalar value 1

$$x^T \text{expected\_returns} \geq r_{\text{target}}$$

- this ensures portfolio's expected return must meet or exceed the target value

Optimization execution:

The quadratic programming problem is solved by 'cvxopt' solver, which is sufficient to handle convex optimization problems. The output will be the optimal set of asset weights that achieve the goal.

After obtaining the optimal weights, the overall risk and expected return of the portfolio are computed and saved.

4.3.5. Secondary Optimization Model

As the first quadratic programming portfolio optimization is based on 127 selected stocks, considering the complexity of managing a large number of stocks and the , filtering selected stocks is essential. Research indicates that the number of stocks in a portfolio is crucial for risk reduction. According to Zhou (2014), examining the correlation between risk reduction and portfolio size indicates that as the number of stocks increases, risk will asymptotically reduce, with returns decreasing at a certain point. The research suggests that an ideal portfolio size could be around ten stocks, which balances manageability with risk reduction (Zhou, 2014). Similarly, Stotz and Lu (2014) conducted a study according to portfolio size across 13,000 stocks in Asia, finding that ten stocks are sufficient to remove 64% of unsystematic risk; they further found that increasing the stock number to twenty only increases 10% of risk mitigation. This suggests little gain in further diversification at the cost of increased management complexity.

Drawing the insights from existing research, the initial portfolio of 127 stocks decided to reduce by concentrating on the top 20 stocks with the highest weights from the previous optimization model. This strategy aligns with the findings from Zhou (2014) and Stotz & Lu (2014), where a smaller number of well-selected stocks can significantly reduce unsystematic risk with the reduction of burden that managing a large portfolio.

Secondary optimization:

The secondary optimization model still uses the same approach as the first optimization by applying the quadratic programming but with the refined list of 20 stocks. The procedure is exactly the same except the difference of stock selection. The model recalculates the optimal weight distribution among these stocks to achieve the objective of minimizing the risk while maintaining the same target value based on the Annualized Taiwan 50 Return Index. This expects to enhance portfolio performance while remaining manageability and aligned with invertor’s risk tolerance.

4.4 Implementation of Monte Carlo Simulation

- Objective

  The objective of this stage is to use Monte Carlo simulation to evaluate the risk and variability of the optimized portfolio based on historical data. The simulations supposed to help better understanding of how the portfolio might have performed under different market conditions depending on historical data, providing insights into its solidity and potential volatility. Two optimal portfolios are both simulated to compare the possible outcomes for the portfolios.
- Tools & Technologies

  The powerful library in Python is employed in this stage including Pandas and NumPy. Pandas is used for handling historical dataand temporal analysis. The simulations are executed using NumPy for generating and analyzing a multitude of possible historical performance scenarios, providing a solid framework for assessing portfolio risk based on historical data trends.

4.4.1. Data Integration

Combine the historical stock price data within optimized portfolio. This data forms the basis for simulation variations in stock prices and calculating the resulting portfolio returns.

4.4.2. Simulation Parameter

Define the number of simulations runs to provide a wide range of probable results, which helps in capturing a large variety of potential scenarios that the portfolio might experience.

4.4.3. Risk and Return

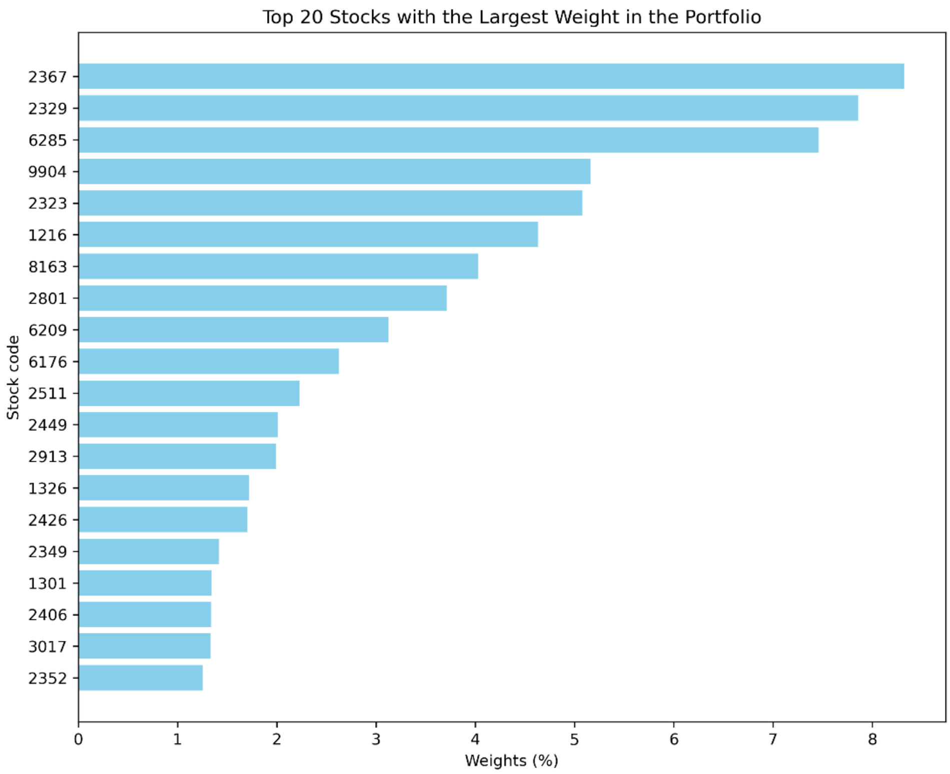
Examine the distribution of simulation returns to compute the potential risk and expected returns. This analysis helps in verification of portfolio risk and performance in obtaining an ideal expected return.

5. Experimental Results

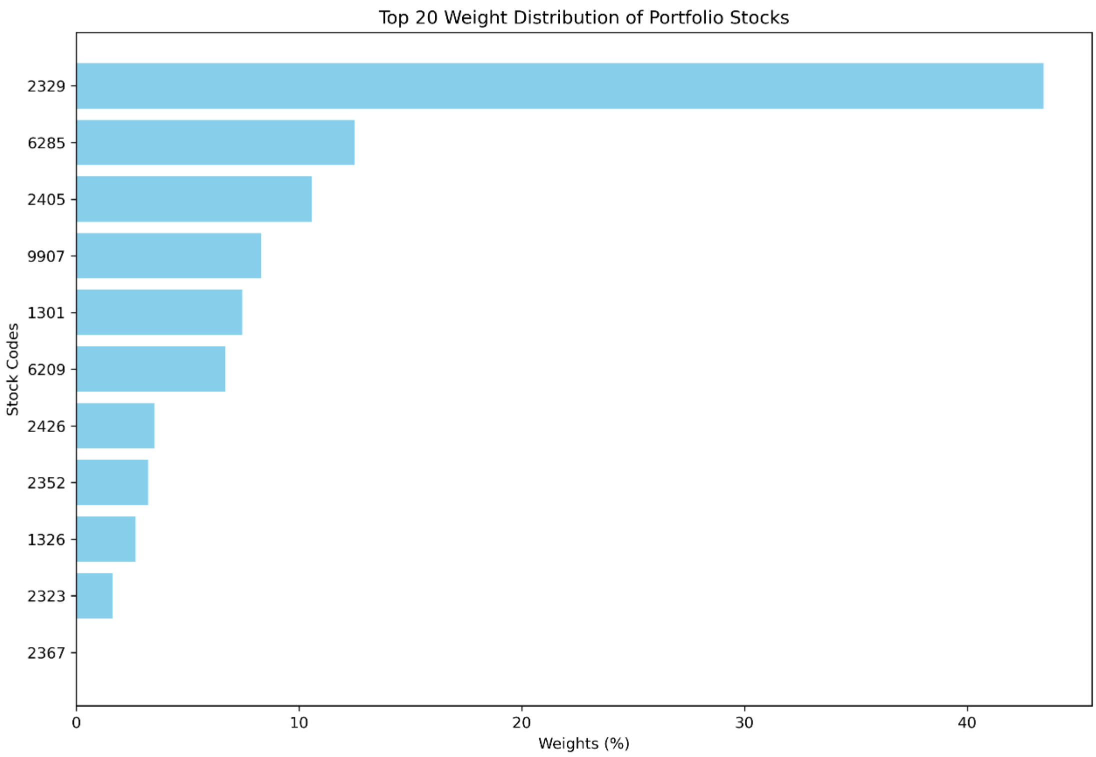
The implementation of Modern Portfolio Theory (MPT) and subsequent optimizations have yielded significant insights into portfolio structuring and performance. The process included the derivation of optimized weights for a selection of 127 stocks and a refined analysis focusing on the top 20 stocks. Furthermore, the portfolio allocations by industry and the distribution of simulated annualized returns were examined to assess the outcomes of the portfolio optimizations.

5.1. Optimized Weights for Selected Stocks

The analysis commenced with the determination of the optimized weights for 127 stocks, culminating in a focused evaluation of the top 20 stocks that hold the most substantial weights in the portfolio. These top constituents represent the core of the optimized portfolio, reflecting a strategic emphasis on specific sectors deemed to have favorable risk-return profiles.



▲ Table 5.1: the weights of the top 20 stocks in the comprehensive portfolio of 127 stocks



▲ **Table 5.2: the weight distribution among these top 20 portfolio stocks after secondary optimization**

5.2. Portfolio Allocation by Industry

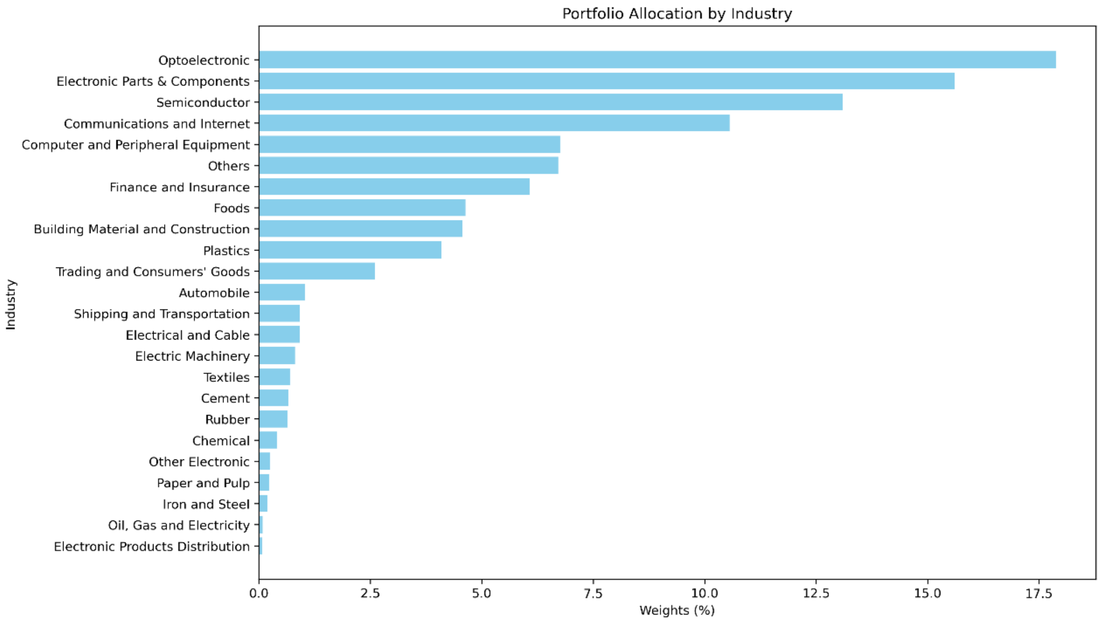
This analysis offers a lens through which to view the focus of the investment strategy and to understand the specific industry might have a potential performance in the stock market with portfolio optimization.

The bar chart showing the top 20 stocks with the largest weight in the portfolio reveals a relatively balanced distribution among the top contenders, with no single stock overwhelmingly dominating the portfolio. The highest weight falls just below 8%, suggesting a strategy that avoids over-reliance on any individual stock and instead spreads risk across multiple assets. This approach is typically employed to mitigate the impact of any single stock's adverse performance on the overall portfolio.

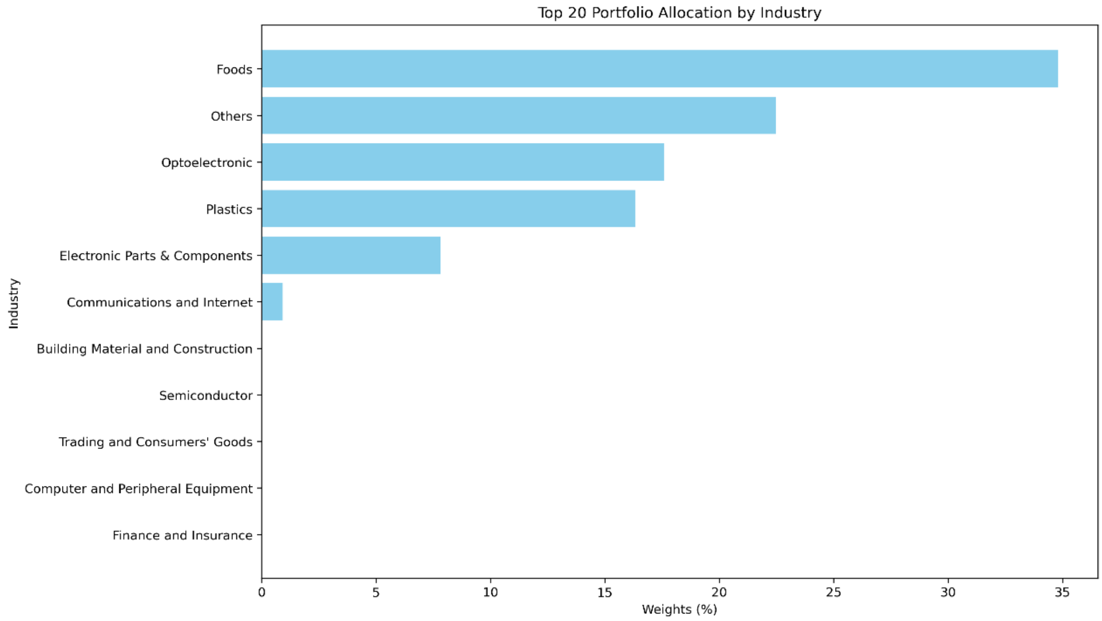
In contrast, the second bar chart showing the top 20 portfolio stock weightings shows a more pronounced allocation to certain stocks, with weightings up to nearly 40%. This suggests a strategic decision to concentrate investments in stocks that are considered high potential or low risk. This allocation may reflect a conviction strategy, where a greater proportion of capital is allocated to stocks with strong fundamentals or the prospect of outperforming the market.

The variation in strategy from diversified to focused reveals an adaptive investment philosophy within the portfolio. It shows a nuanced approach to portfolio construction, where certain segments receive a more considerable investment based on their perceived potential for yield or stability, while still maintaining an overall diversification to hedge against market uncertainties.

In essence, the investment strategy encapsulates both conservative diversifications to protect the portfolio against market volatilities and a selective concentration in equities with the potential for higher returns. This strategic blend aims to balance the pursuit of performance with the management of risk, catering to a sophisticated investment mandate that seeks to optimize returns without exposing the portfolio to undue volatility.



▲ **Table 5.3: the portfolio allocation by industry within the broader portfolio of 127 stocks, highlighting the predominance of certain sectors such as 'Optoelectronics' and 'Electronic Parts & Components' which command significant portions of the portfolio**



▲ **Table 5.4: the allocation by industry for the top 20 stocks, where the 'Foods' and 'Optoelectronics' sectors appear to dominate, suggesting a strategic concentration in these areas**

5.3. Simulated Portfolio Annualized Returns Distribution

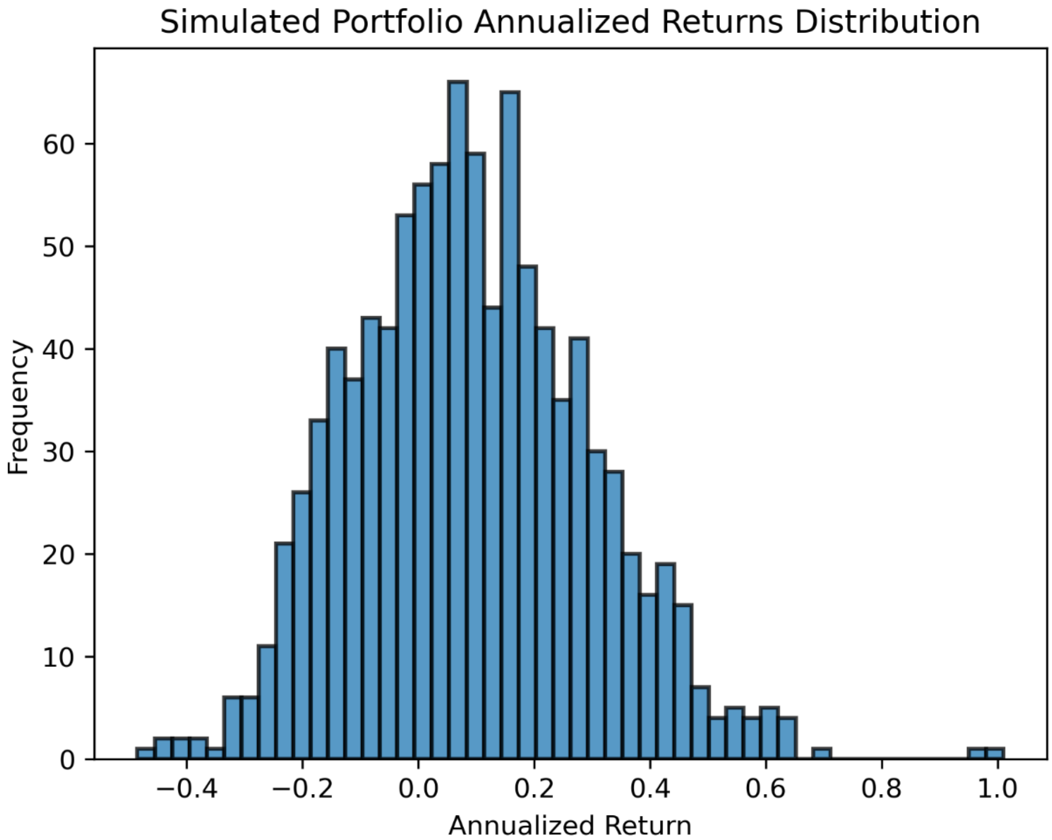
The performance of the portfolio optimizations was further examined through the lens of simulated annualized return, providing an estimation of how the optimized portfolio might behave under different historical market conditions. The simulations try to verify the performance and stability of the portfolio.

The portfolio with 127 stocks yielded an average annualized return of approximately 0.0911 (9.11%) with a standard deviation is 0.20. On the other hand, the portfolio optimization which only focuses on the top 20 stocks showed an average annualized return 0.085 (8.5%) with a lower standard deviation is 0.16.

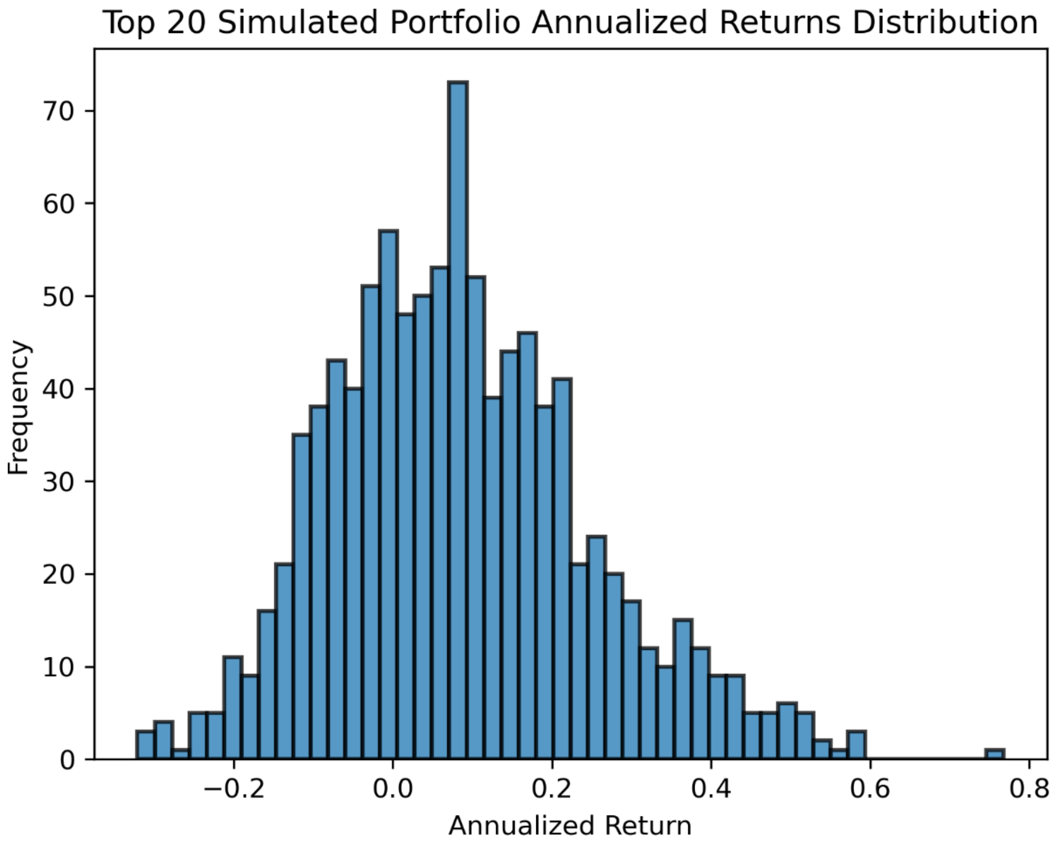
The reduced standard deviation in the top 20 stock portfolio suggests a lower level of risk or volatility compared to the broader portfolio. This can be attributed to the concentration of investment in the highest-weighted stocks, which may have been more stable or had less variable historical returns. Although the average annualized return is slightly lower in the top 20 stock portfolio, the decrease in risk, as indicated by the lower standard deviation, could be a favorable trade-off for risk-averse investors seeking stability.

The results from the simulations corroborate the principles of Modern Portfolio Theory, which advocates for diversification to minimize risk. However, they also highlight the diminishing marginal benefits of diversification beyond a certain point, as reflected by the modest reduction in returns coupled with a notable decrease in risk when moving from a broad portfolio to one concentrated in top-performing stocks.

In conclusion, the Monte Carlo simulation results suggest that a more focused portfolio of top-performing stocks may offer a balance between performance and stability, potentially making it a suitable strategy for investors who prioritize a conservative risk profile without significantly compromising on returns.



▲ **Table 5.5: the simulated annualized returns distribution for the 127-stock portfolio, indicating a diverse range of outcomes and illustrating the portfolio's response to market fluctuations.**



▲ Table 5.6: focuses on the top 20-stock portfolio, revealing a more concentrated distribution of simulated returns

6. Evaluation

During the execution of this research, Modern Portfolio Theory and AI techniques were successfully applied to optimize portfolios in Taiwan's stock market. However, the study encountered several challenges and limitations that might impact the broader applicability and depth of the findings.

Initially, an attempt was made to incorporate sentiment analysis of historical financial news into the evaluation criteria for the investment portfolios, considering the potential impact of media on market sentiments. Unfortunately, due to regional restrictions and the availability of historical news data, sufficient data for this analysis could not be acquired. This limitation restricted the ability to understand market dynamics from a broader perspective and may have impacted the accuracy of the portfolio risk assessments.

Moreover, the data used in this study predominantly came from the Taiwan Stock Exchange, covering the period from 2010 to 2020. Although this period includes recovery phases post the global financial crisis and other significant economic events, it does not encompass the impacts of the COVID-19 pandemic on the markets. Thus, the research findings might not fully reflect portfolio performances under extreme market conditions.

Additionally, despite efforts to ensure data accuracy and rigor in analysis, potential human errors in data collection and processing could have occurred. For example, manual checks for data consistency might lead to misjudgments. Future research could benefit from introducing more automated data checks and sophisticated data processing techniques to enhance the reliability and efficiency of the study.

In summary, while this research has made preliminary achievements in optimizing portfolios in Taiwan's stock market, the results should be interpreted cautiously due to the aforementioned limitations. Future studies could overcome these limitations by expanding data sources, incorporating new analytical tools and techniques, and including more market influence factors such as news sentiment in the analysis to enhance the breadth and depth of the research.

7. Professional Issues

There were a number of issues to consider when completing this project.

7.1 Compliance with Regulatory Frameworks

The integration of AI in portfolio optimization must adhere strictly to the regulatory frameworks established by financial authorities in Taiwan. This includes compliance with guidelines set by the Taiwan Securities and Exchange Commission (TWSE), which mandates transparency in AI operations and decisions. It is essential that AI systems are designed to comply with these regulations to prevent potential market manipulation or unfair trading advantages.

7.2 Data Privacy and Security

According to Taiwan's data privacy regulations, it is important that AI systems managing investor data uphold the highest standards of security. AI systems must properly follow both local and international data protection rules. This guarantees that data collection, use, and management follow legal and ethical guidelines. It is critical to develop protocols that ensure that all data is collected with valid consent and managed in a way that respects individuals' privacy rights. Furthermore, data handlers must stay current on legislative developments and modify their processes accordingly to remain compliant with new rules.

7.3 Ethical Considerations

The transparency of AI applications in financial markets and maintaining market integrity and public trust is crucial in this project. Clearly explaining processes to users and stakeholders to keep the transparency and ensuring that AI models must be free from bias that could affect investment decision.

8. Conclusion and Future Work

8.1. Conclusion

The optimized portfolio built on the MPT basis demonstrates a clear risk diversification strategy, while still ensuring a minimum expected return in line with the benchmark Taiwan 50 Index. The top 20 stocks were invested in a more concentrated manner, with a lower standard deviation of simulated annualized returns and reduced risk. The strategy focuses on a tailored investment philosophy, with significant allocations to sectors such as 'Optoelectronics' and 'Food', reflecting confidence in the performance of their markets.

The juxtaposition of diversification and concentration strategies in the project highlights the adaptability of portfolio management. It suggests that while diversification remains a key principle in reducing risk, there is a threshold beyond which additional diversification may not yield commensurate benefits in terms of risk reduction. Conversely, well researched and concentrated investments in well performing stocks can potentially improve portfolio performance without incurring excessive risk, which is in line with the preferences of investors with a more conservative risk appetite.

8.2. Future Work

8.2.1. Comprehensive Market Indicators

Future research could incorporate a broader range of market indicators in order to better understating the stock market, including indicators such as Taiwan's GDP, unemployment rates, and financial news to provide a more detailed analysis. These indicators could provide valuable insights into the stock market and economic conditions, potentially leading to a more concrete portfolio optimization strategy.

8.2.2. Expansion of Data Scope

Including the data during the COVID-19 pandemic would provide a more comprehensive view of portfolio optimization under extreme market conditions. Therefore, this would enhance the relevant and applicable research findings.

8.2.3. Comparative Analysis with Economic Sectors

Analysis comparing the performance between the portfolio and specific economic sectors could further understand the impact of economic developments on stock selection and portfolio optimization.

9. Bibliography

- Ai, J., Brockett, P. L., Cooper, W. W., & Golden, L. L. (2011). Enterprise Risk Management Through Strategic Allocation of Capital. *Journal of Risk and Insurance*, *79*(1), 29–56. https://doi.org/10.1111/j.1539-6975.2010.01403.x
- Broby, D. (2022). The use of predictive analytics in finance. *The Journal of Finance and Data Science*, *8*, 145–161. https://doi.org/10.1016/j.jfds.2022.05.003
- Chen, M., Lin, C.-T., & Chen, A.-P. (2007). Constructing a dynamic stock portfolio decision-making assistance model: using the Taiwan 50 Index constituents as an example. *Soft Computing*, *11*(12), 1149–1156. https://doi.org/10.1007/s00500-007-0158-y
- Chow, P. (2020). The Spillover Effect of the US–China Trade War on Taiwan's Economy*. *Advances in Pacific Basin Business, Economics, and Finance*, *8*. https://doi.org/10.1108/s2514-465020200000008011

- Chung-Hua, S. (2010). *Taiwan's Policy Responses to the Financial Tsunami in 2008*. Cambridge University Press; ISEAS–Yusof Ishak Institute. https://www.cambridge.org/core/books/abs/managing-economic-crisis-in-east-asia/taiwans-policy-responses-to-the-financial-tsunami-in-2008/25CF8384ECC27787606B8067943ADB01
- Fajartama, R., & Faturohman, T. (2021). *Modern Portfolio Theory and Forecast Implementation Using Monte Carlo Simulation in Automotive Financing Company*. https://sibresearch.org/uploads/3/4/0/9/34097180/riber_10-s3_22_s21-049_306-323.pdf
- Ghodrati, H., & Zahiri, Z. (2014). A Monte Carlo simulation technique to determine the optimal portfolio. *Management Science Letters*, *4*(3), 465–474. https://m.growingscience.com/beta/msl/1336-a-monte-carlo-simulation-technique-to-determine-the-optimal-portfolio.html
- OpenAI. (2022, November 30). *Introducing ChatGPT*. OpenAI; OpenAI. https://openai.com/blog/chatgpt
- Pedersen, M. E. H. (2014). Portfolio Optimization and Monte Carlo Simulation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2438121

- Romanko, O., Narayan, A., & Kwon, R. H. (2023). ChatGPT-Based Investment Portfolio Selection. *Operations Research Forum*, *4*(4). https://doi.org/10.1007/s43069-023-00277-6
- Sihombing, T. (2013). Applying Monte Carlo Concept and Linear Programming in Modern Portfolio Theory to Obtain Best Weighting Structure. *Indonesian Capital Market Review*, *5*(1). https://doi.org/10.21002/icmr.v5i1.1582
- Stotz, A., & Lu, W. (2014). Ten Stocks are Enough in Asia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2461115
- The International Trade Administration, MOEA. (2018). *Cross-Straits Economic Cooperation Framework Agreement*. https://www.trade.gov.tw/Files/PageFile/651641/ECFA.pdf
- Zhou, Y. (2014). *Number of Stocks in Portfolio and Risk Reduction*. Escholarship.org. https://escholarship.org/uc/item/78v69630