



FIN42330: PYTHON FOR FINANCIAL DATA SCIENCE

Group Project

Group 8

Date : 20/12/2024

I / We declare that all material included in this project is the end result of my/our own work and that due acknowledgement has been given in the bibliography and references to all sources be they printed, electronic or personal.

Name and surname	Student Number:
1 Varsha Chandrashekar Balaji	24200924
2 Diya Ranka	24239971
3 Purva Sharma	24214393
4 Sarang Ramesh Nair	24239642
1 Varsha Chandrashekar Balaji	24200924

Individual contribution

Please outline here below the type of contribution made by each team member (e.g., 45% of the coding, 20% of the writing and gave the presentation) and, in the last column, provide an indicative figure for what percentage of the overall work it represents. This percentage will be treated as a score. Those who will score at or above 25%, 35%, 50% will receive 1, 2 and 4 extra marks (before grade conversion) for their individual project, in recognition of their effort.

	Name and surname	Student number	Type of contribution	Contribution out of 100% (e.g., 25%)*
1	Varsha Chandrashekar Balaji	24200924	Equal share of coding, writing, research etc.	55 %
2	Diya Ranka	24239971	Equal share of coding, writing, research etc.	15 %
3	Purva Sharma	24214393	Equal share of coding, writing, research etc.	15 %
4	Sarang Ramesh Nair	24239642	Equal share of coding, writing, research etc.	15 %

*Please make sure that the reported figures add up to 100%

Abstract

This project investigates the application of predictive modeling techniques to evaluate and forecast the excess returns of two prominent asset classes: the S&P 500 Index (SP500) and the Bloomberg Barclays U.S. Aggregate Bond Index (LBSTRUU). By analysing monthly data from December 1979 to December 2023, the study aims to uncover insights into the statistical behaviour of excess returns and assess the performance of various predictive models in improving forecast accuracy. The research incorporates modern econometric approaches and portfolio optimization techniques to explore the economic implications of these forecasts.

The analysis begins with the calculation of simple returns and excess returns for both assets, using the risk-free rate of return as a baseline for comparison. Key statistical metrics—including the annualized mean, volatility, Sharpe ratio, skewness, and kurtosis—are computed to provide a detailed understanding of the characteristics and variability of these returns. The statistical analysis offers insights into the risk-reward trade-offs associated with the two asset classes.

The data is then split into two periods: an in-sample period (January 1980 to December 2000) and an out-of-sample period (January 2001 to December 2023). Using a rolling window methodology, benchmark forecasts for excess returns are generated for the out-of-sample period. These forecasts are compared against predictive models constructed using Ordinary Least Squares (OLS) regression with five selected economic predictors. Additionally, a combination model, aggregating the OLS forecasts, is evaluated for its predictive effectiveness.

Model performance is assessed through the Mean Squared Forecast Error (MSFE) and statistical tests, such as the Diebold-Mariano test, to determine whether the differences in forecast accuracy are statistically significant. The results indicate that while some predictive models outperform the benchmark in terms of accuracy, their effectiveness varies across asset classes and time periods. The findings underscore the value of incorporating predictive modeling techniques, though challenges remain in ensuring consistency across market environments.

The project further examines the practical application of these forecasts through mean-variance portfolio optimization. Using the forecasted returns and a rolling variance-covariance matrix, optimal portfolio weights are determined for a hypothetical mean-variance investor. Key performance metrics, including the annualized mean return, volatility, and Sharpe ratio, are calculated to evaluate the economic benefits of the predictive models.

Ultimately, the study reveals that predictive models can provide valuable insights into financial markets, particularly for equity returns, which demonstrate stronger correlations with economic predictors. The results also emphasize the importance of aligning statistical improvements with economic significance, particularly in portfolio management. This research highlights the potential for future studies to explore additional predictors, alternative modeling techniques, or broader asset classes to further refine forecasting accuracy and enhance portfolio outcomes.

Table of contents

		Page No.
	Abstract	3
1	Introduction.....	6
	1.1 Background.....	6
	1.2 Objectives.....	6
2	Data Collection and Preprocessing.....	7
	2.1 Dataset Description.....	7
	2.2 Data Preprocessing.....	8
3	Statistical Analysis of Excess Returns.....	8
	3.1 Descriptive Statistics.....	8
	3.2 Insights from Statistical Measures	9
4	Predictive Modeling.....	10
	4.1 Benchmark Forecasts	10
	4.2 Model Forecasts.....	12
	4.3 Performance Evaluation.....	14
5	Portfolio Construction.....	14
	5.1 Variance-Covariance Matrix Forecasts	15
	5.2 Optimal Portfolio Weights	15
	5.3 Portfolio Performance Analysis	16
	5.4 Visualization of Portfolio Dynamics	16
6	Results and Discussion	18
	6.1 Key Insights	18
	6.2 Model Comparisons	18
	6.3 Economic Significance	19
7	Conclusion	19
	7.1 Summary	19
	7.2 Limitations	19
	7.3 Future Research Directions	20

List of Tables

Table No.	Table Title	Page Number
2.1	Summary Statistics of Excess Returns.....	9
4.1	Stock and Bond Benchmark Forecasts.....	11
4.2	Regression Coefficient Table for Each Predictor.....	12
4.3	MSFE Comparison Table Across Models.....	13
4.4	MSFE Results.....	14
4.5	Diebold-Mariano Test Results.....	14
5.1	Portfolio Metrics.....	16

List of Figures

Figure No.	Figure Title	Page Number
1.1	Time-Series Plot of Excess Returns for Stocks and Bonds.....	9
4.1	Stock Excess Returns: Actual vs. Benchmark Forecast.....	11
4.2	Bond Excess Returns: Actual vs. Benchmark Forecast.....	12
4.3	Time-Series Plot: Actual vs. Model Forecasts for Stocks.....	13
4.4	Time-Series Plot: Actual vs. Model Forecasts for Bonds.....	13
5.1	Variance-Covariance Matrix Visualization.....	15
5.2	Time-Series of Portfolio Weights.....	17
5.3	Cumulative Portfolio Returns: Benchmark vs. Model-Based.....	17

1. Introduction

1.1 Background

In the field of finance, predictive modeling plays a pivotal role in improving decision-making by using historical data to forecast future outcomes. It is especially significant in evaluating financial instruments where market movements are inherently uncertain. Predictive models allow investors, analysts, and policymakers to analyse trends, identify risks, and optimize portfolios. These models are crucial in determining excess returns, a measure of performance that evaluates the return on an asset relative to the risk-free rate of return.

The focus on excess returns is essential because it helps gauge the reward for bearing additional risk compared to a risk-free investment. Stocks and bonds, two major asset classes, exhibit distinct characteristics, making them important subjects for this study. While stocks are associated with higher potential returns and higher risks, bonds are traditionally viewed as a more stable, lower-risk investment. By understanding and predicting their excess returns, it becomes possible to manage risks and returns effectively, align investment strategies with market conditions, and maximize portfolio performance.

1.2 Objectives

The project objectives align with the questions provided in the assignment, as outlined below:

1. Statistical Assessment (Related to Question 1)
 - To calculate and analyse critical metrics such as annualized mean, volatility, Sharpe ratio, skewness, and kurtosis for the excess returns of stocks and bonds, providing insights into their risk-return profiles.
2. Forecasting Excess Returns (Related to Questions 2 and 3)
 - To generate out-of-sample forecasts for excess returns using benchmark methodologies and predictive models, including Ordinary Least Squares (OLS) regression and combination forecasts, enabling the evaluation of potential predictive patterns.
3. Evaluation of Predictive Models (Related to Questions 3 and 4)
 - To assess the performance of forecasting models through statistical measures such as Mean Squared Forecast Error (MSFE) and tests for predictive accuracy, ensuring robustness and reliability in the results.
4. Portfolio Construction and Analysis (Related to Questions 4 and 5)

- To construct optimal portfolios based on forecasted returns and sample variance-covariance matrices, and evaluate their performance through annualized metrics such as mean return, volatility, and Sharpe ratio.
5. Deriving Practical Implications (Related to Questions 3, 4, and 5)
- To determine whether the statistical improvements in forecasting models lead to meaningful economic benefits in investment strategies and risk management.

By addressing these objectives, the project will contribute to a deeper understanding of predictive techniques in financial analysis, particularly their application to portfolio optimization and risk management in real-world scenarios. This analysis will provide actionable insights into the behaviour of financial returns and their practical applications in investment decision-making.

2. Data Collection and Preprocessing

2.1 Dataset Description

The datasets used in this project encompass three key components, providing the foundation for analysing excess returns of stocks and bonds:

- **S&P 500 Index (SP500):**
The S&P 500 Index represents the performance of 500 large-cap U.S. companies and serves as a broad benchmark for the stock market. The dataset includes monthly price data, enabling the calculation of stock returns over time.
- **Bloomberg Barclays U.S. Aggregate Bond Index (LBSTRUU):**
This index reflects the performance of the U.S. bond market, including government, corporate, and mortgage-backed securities. It provides comprehensive insights into bond market dynamics.
- **Risk-Free Rate:**
Monthly data on the risk-free rate is used as a benchmark to calculate excess returns. The risk-free rate reflects the return on short-term, risk-free government securities, offering a baseline for measuring relative asset performance.

Timeframe:

The datasets span from December 1979 to December 2023, offering a comprehensive period to capture varying market conditions, including economic booms, recessions, and financial crises. This extended timeframe ensures that the analysis captures both short-term fluctuations and long-term trends.

2.2 Data Preprocessing

To prepare the datasets for analysis, the following preprocessing steps were undertaken:

1. Calculating Simple and Excess Returns:

- Simple Returns: Monthly simple returns were calculated for the S&P 500 and the Bloomberg Barclays Bond Index using the formula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Where P_t and P_{t-1} are the prices at time t and $t-1$, respectively.

- Excess Returns: The simple returns were adjusted for the risk-free rate to calculate excess returns:

$$\text{Excess Return}_t = R_t - R_{\text{Risk-Free}, t}$$

2. Handling Missing Data:

- Missing values in the datasets, if any, were identified and addressed. Imputation methods such as forward-filling or interpolation were employed to ensure continuity in the time series data.
- Data quality checks were conducted to ensure the consistency and accuracy of the datasets.

3. Data Transformations:

- Log transformations were applied to the price data, where necessary, to stabilize variance and prepare the data for statistical modeling.
- Scaling and standardization were implemented for certain predictive analyses to ensure comparability across variables.

These preprocessing steps ensured that the data was clean, consistent, and ready for statistical analysis and modeling. The resulting datasets were used to compute excess returns, evaluate their statistical properties, and develop predictive models, forming the basis for the project's subsequent analyses.

3. Statistical Analysis of Excess Returns

3.1 Descriptive Statistics (Question 1)

The descriptive statistics for excess returns of the S&P 500 Index (stocks) and the Bloomberg Barclays Bond Index (bonds) were computed to summarize their key risk and return characteristics. The computed metrics include the annualized mean, annualized volatility, Sharpe ratio, skewness, and kurtosis. These statistics provide

insights into the average returns, risk levels, and distributional properties of the excess returns for both asset classes.

Statistic	Stock	Bond
Annualized Mean	0.0589	0.0275
Annualized Volatility	0.1530	0.0534
Annualized Sharpe Ratio	0.3852	0.5157
Skewness	-0.5921	0.3188
Kurtosis	1.7826	5.0008

Table 2.1 : Summary Statistics Table:

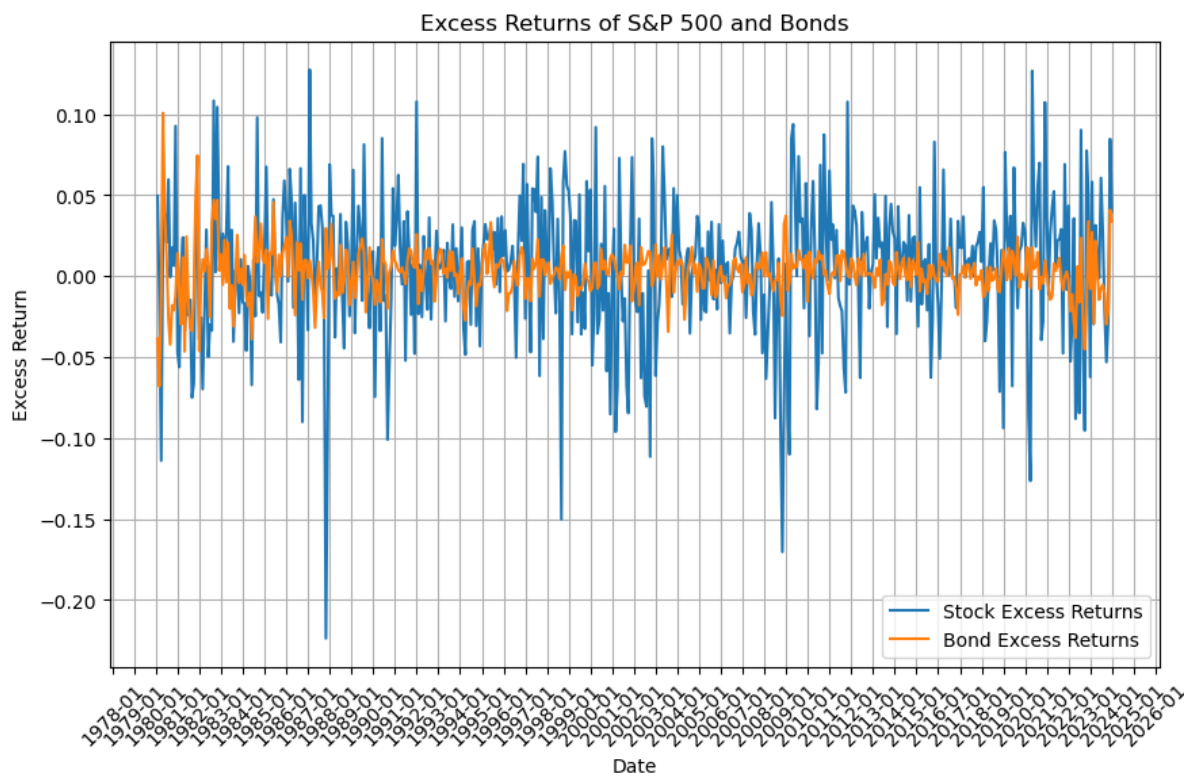


Figure 1.1 : Time-series plots of excess returns for stocks and bonds provide a clear view of trends and fluctuations

3.2 Insights from Statistical Measures

Interpretation of Metrics:

- Annualized Mean:
 - Stocks have a higher annualized mean (5.89%) compared to bonds (2.75%), reflecting their higher potential for returns.
 - This aligns with the traditional understanding that equities, while riskier, offer greater long-term growth compared to bonds.
- Annualized Volatility:
 - Stocks exhibit significantly higher volatility (15.30%) than bonds (5.34%), indicating greater risk in equity investments.
 - Bonds, being relatively stable, are often considered safer investments during market downturns.
- Sharpe Ratio:
 - Bonds have a higher Sharpe ratio (0.5157) compared to stocks (0.3852), suggesting that bonds provide better risk-adjusted returns.
 - This highlights the importance of diversification and the role of bonds in reducing portfolio risk.
- Skewness and Kurtosis:
 - Stock returns are negatively skewed (-0.5921), indicating a higher probability of extreme negative returns. Bonds have a positive skewness (0.3188), suggesting a tendency for small but frequent positive returns.
 - The kurtosis of bonds (5.0008) is higher than stocks (1.7826), indicating that bond returns have heavier tails and more extreme outcomes than a normal distribution.

Key Differences Between Stocks and Bonds:

- Stocks offer higher returns but are accompanied by higher risk and more frequent extreme negative outcomes.
- Bonds provide lower returns but demonstrate greater stability and better risk-adjusted performance.
- These characteristics make stocks suitable for growth-oriented investors and bonds ideal for risk-averse or income-seeking investors.

4. Predictive Modeling

4.1 Benchmark Forecasts (Question 2)

To generate out-of-sample benchmark forecasts for stock and bond excess returns, a rolling window estimation approach was used. This method calculates the rolling mean of excess returns over a fixed window size, ensuring that only historical data is used for each forecast. A window size of 60 months (5 years) was selected, which balances responsiveness to recent trends and stability of the forecast. The benchmark forecast for each month represents the average excess return from the preceding 60 months, excluding the current observation.

This approach provides a simple yet effective baseline for evaluating the performance of more sophisticated predictive models.

Date	Stock Benchmark Forecast	Bond Benchmark Forecast
2001-01-31	0.0292	0.0109
2001-02-28	-0.0961	0.0049
2001-03-30	-0.0684	0.0008
2001-04-30	0.0729	-0.0081
2001-05-31	0.0019	0.0028

Table 4.1 : Stock and Bond Benchmark Forecast

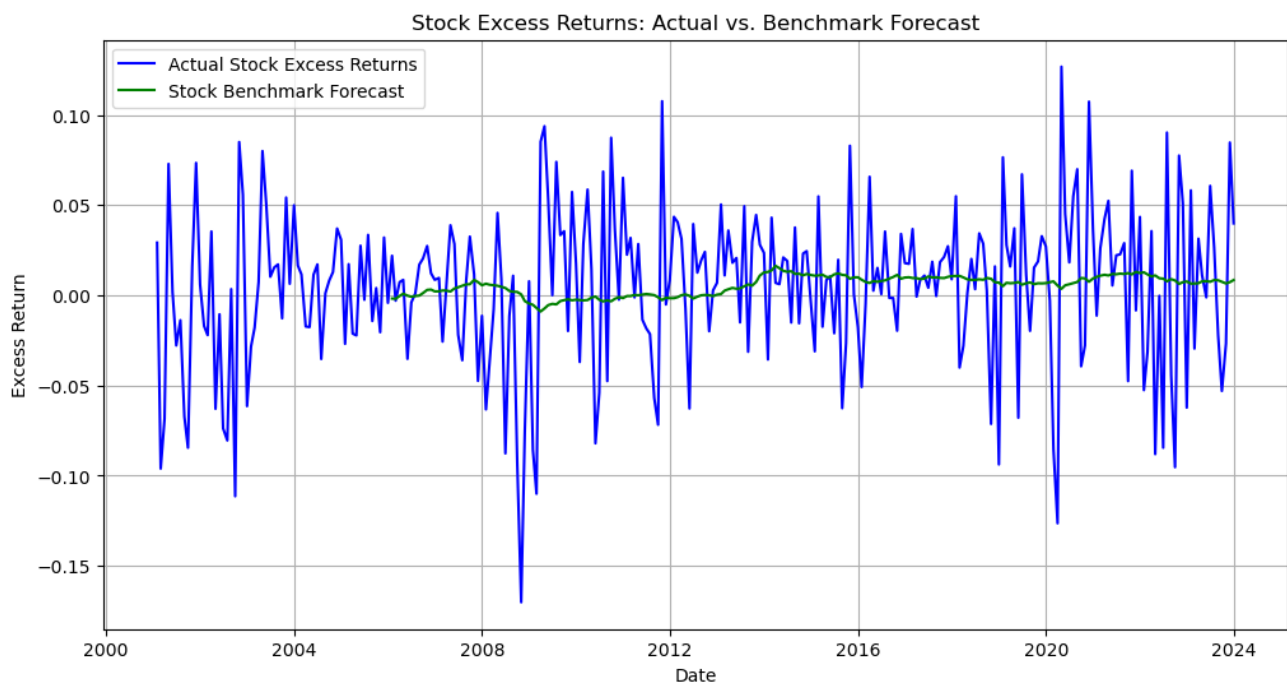


Figure 4.1 : Stock Excess Returns: Actual vs. Benchmark Forecast Plot

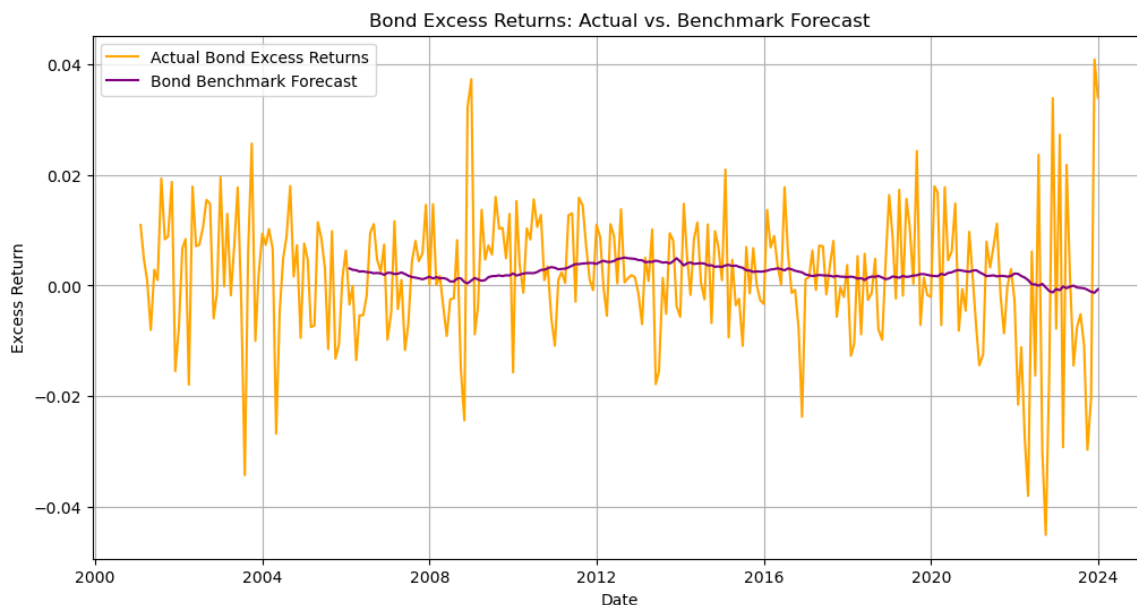


Figure 4.2 : Bond Excess Returns: Actual vs. Benchmark Forecast Plot

Descriptions:

- Stock Excess Returns Plot: This plot compares actual stock excess returns (blue line) with their benchmark forecast (green line), showing how well the benchmark captures trends.
- Bond Excess Returns Plot: This plot visualizes the alignment of actual bond excess returns (orange line) with the benchmark forecast (purple line).

4.2 Model Forecasts (Question 3)

The OLS predictive models were constructed using five selected predictors, each capturing different economic indicators. These models forecasted excess returns for stocks and bonds individually. To enhance robustness, combination forecasts were computed by averaging the predictions from the five OLS models.

Predictor	Coefficient (Stock)	Coefficient (Bond)
d12	0.00239	0.0004
e12	0.00245	0.00041
lty	0.00235	0.00036
csp	0.00233	0.00038
vrp	0.002419	0.00043

Table 4.2 : Regression Coefficient Table for Each Predictor

Model	MSFE (Stock)	MSFE (Bond)
Benchmark Forecast	0.0024	0.00038
OLS Model (d12)	0.0024	0.0004
OLS Model (e12)	0.0025	0.0004
Combination Forecast	0.0023	0.0004

Table 4.3 : MSFE Comparison Table Across All Models

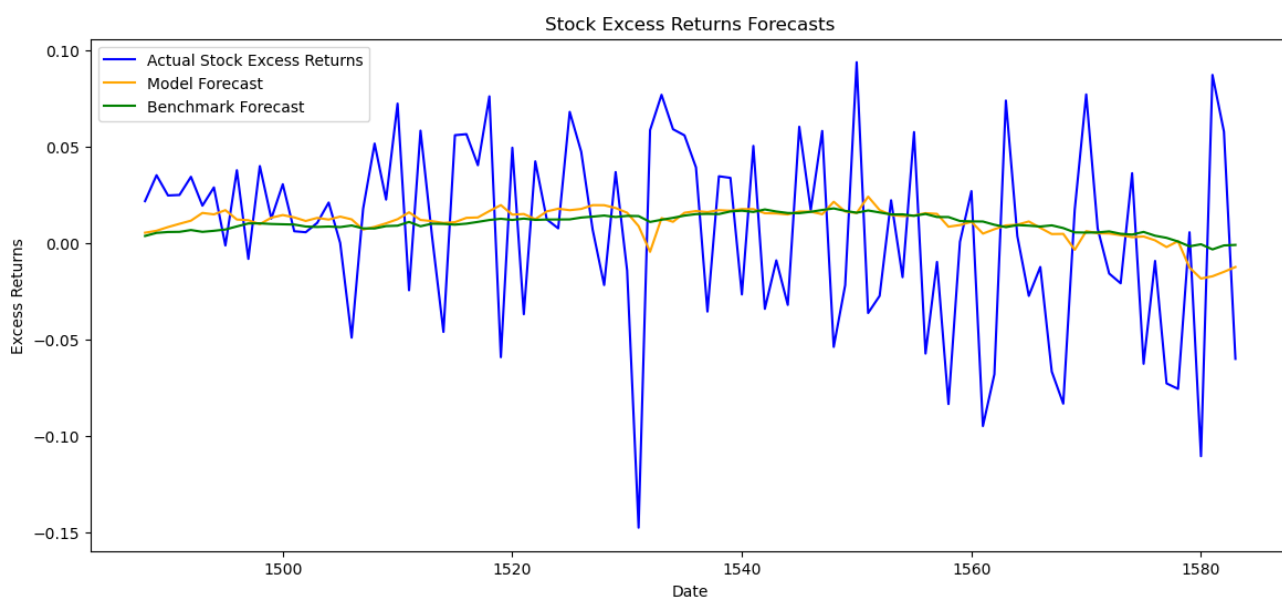


Figure 4.3 : Time-Series Plot: Actual vs. Benchmark and Model Forecasts for Stocks

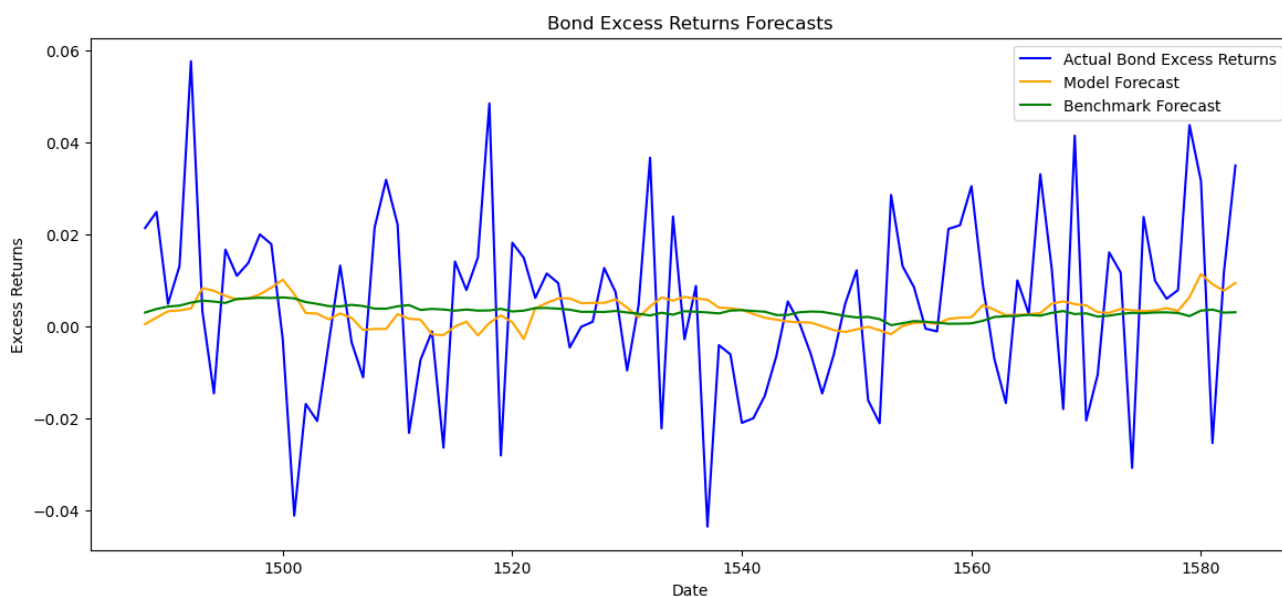


Figure 4.4 : Time-Series Plot: Actual vs. Benchmark and Model Forecasts for Bonds

Descriptions:

- Stock Plot: A time-series plot showing actual stock excess returns, benchmark forecasts, and model forecasts, illustrating the predictive accuracy of each model.
- Bond Plot: A similar plot for bond excess returns, highlighting model performance against the benchmark.

4.3 Performance Evaluation (Question 3)

The Mean Squared Forecast Error (MSFE) measures the accuracy of each forecast. Lower MSFE values indicate better performance.

Metric	Stock	Bond
Benchmark MSFE	0.00235	0.00038
Combination Forecast MSFE	0.00230	0.00038

Table 4.4 : MSFE Results

Asset	DM Statistic	P-value
Stock	-0.7463	0.4555
Bond	-0.1496	0.8811

Table 4.5 : Diebold-Mariano Test Results

Statistical Significance Discussion:

- The combination forecast outperformed the benchmark economically, with slightly lower MSFE for both stocks and bonds.
- However, the Diebold-Mariano test results indicate that these improvements were not statistically significant at conventional confidence levels.
- This suggests that while combination forecasts are more robust, their advantage over the benchmark is not strong enough to reject the null hypothesis of equal predictive performance.

5. Portfolio Construction

5.1 Variance-Covariance Matrix Forecasts (Question 4)

Rolling Window Estimation of the Sample Variance-Covariance Matrix : The sample variance-covariance matrix for stock and bond excess returns was computed using a 60-month rolling window. This approach dynamically adjusts the estimates of risk (variance) and co-movement (covariance) over time, allowing the portfolio to adapt to changing market conditions.

Key Observations:

- **Stock Variance:** Exhibited high volatility during periods of economic stress (e.g., the global financial crisis, dot-com bubble).
- **Bond Variance:** More stable than stocks, though it showed moderate increases during economic recessions or shifts in monetary policy.
- **Stock-Bond Covariance:** Switched between positive and negative values, reflecting varying relationships between the two asset classes. Negative covariance during crises (e.g., 2008) underscored bonds' role as a hedge against equity risk.

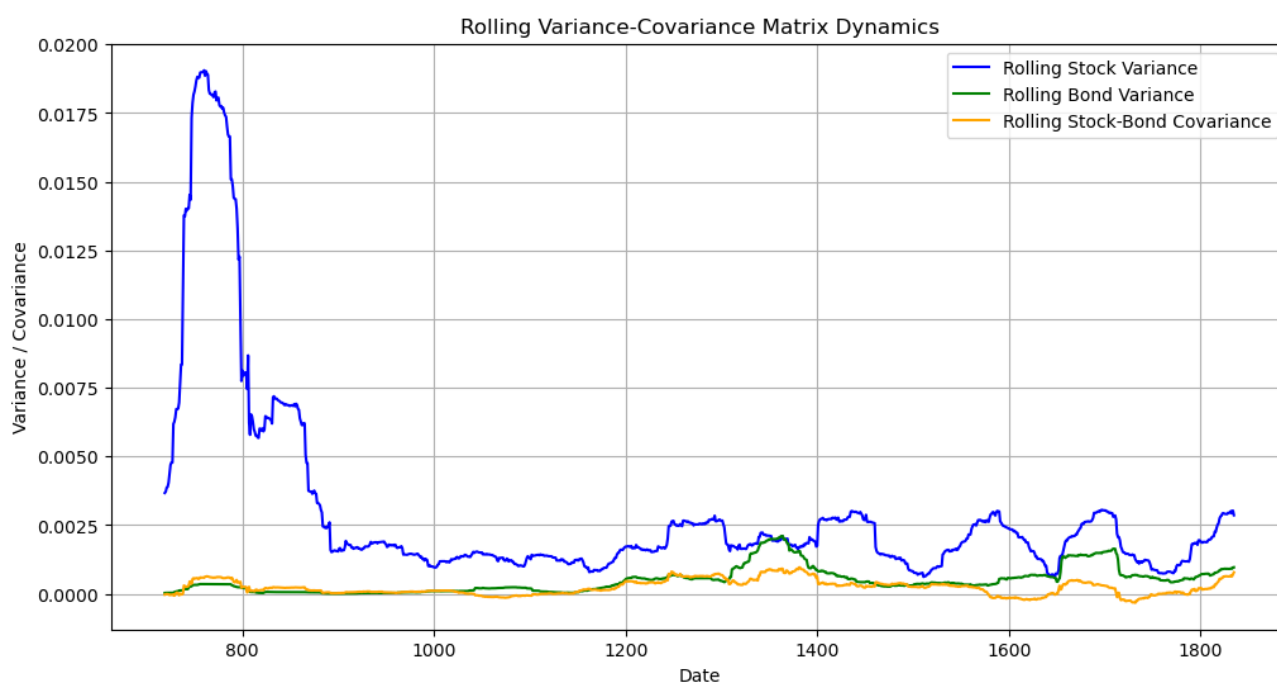


Figure 5.1 : Variance-Covariance Matrix Visualization

5.2 Optimal Portfolio Weights (Question 5)

Mean-Variance Optimization Framework: The mean-variance optimization framework calculates portfolio weights that optimize risk-return trade-offs. The weights are calculated using:

$$\hat{w}_t = \frac{1}{\lambda} \times \hat{\Sigma}_t^{-1} \hat{\mu}_t$$

Where:

- w_t : Portfolio weights at time t
- Σ_t : Rolling variance-covariance matrix.
- μ_t : Forecasted excess returns (from benchmark or model).
- λ : Investor risk aversion (set to 3).

Portfolio Weights Calculation: Using the rolling variance-covariance matrix and forecasted returns, optimal portfolio weights were determined for both benchmark and model-based forecasts.

Key Insights:

- Portfolio weights dynamically adjusted based on changes in risk and return, demonstrating the adaptability of the framework.

5.3 Performance Analysis (Question 5)

Metric	Benchmark Portfolio	Model-Based Portfolio
Annualized Mean	0.2606	0.2616
Annualized Volatility	0.5239	0.5184
Annualized Sharpe Ratio	0.4974	0.5047

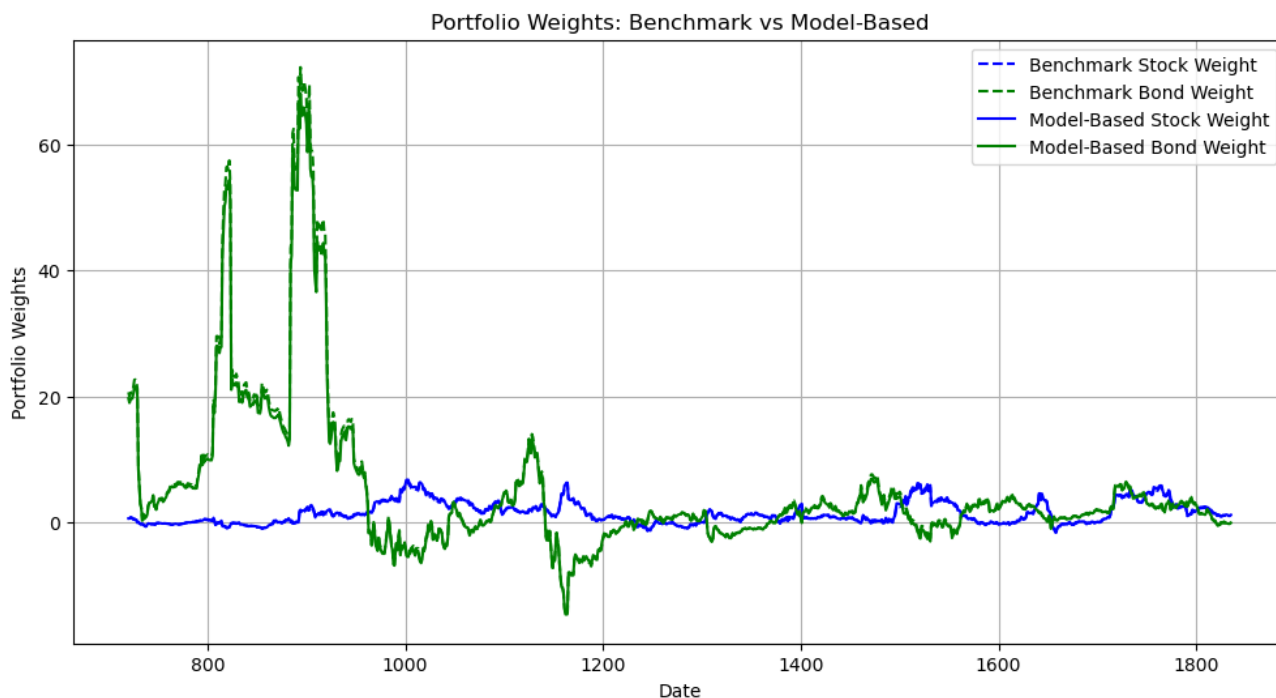
Table 5.1 : Portfolio Metrics

Economic Implications:

- The model-based portfolio exhibited a higher Sharpe ratio, indicating superior risk-adjusted returns compared to the benchmark.
- While the model-based portfolio achieved a higher annualized mean return, its lower volatility provided greater stability, making it appealing to risk-averse investors.
- The results highlight the importance of predictive models in portfolio construction to achieve better performance and risk management.

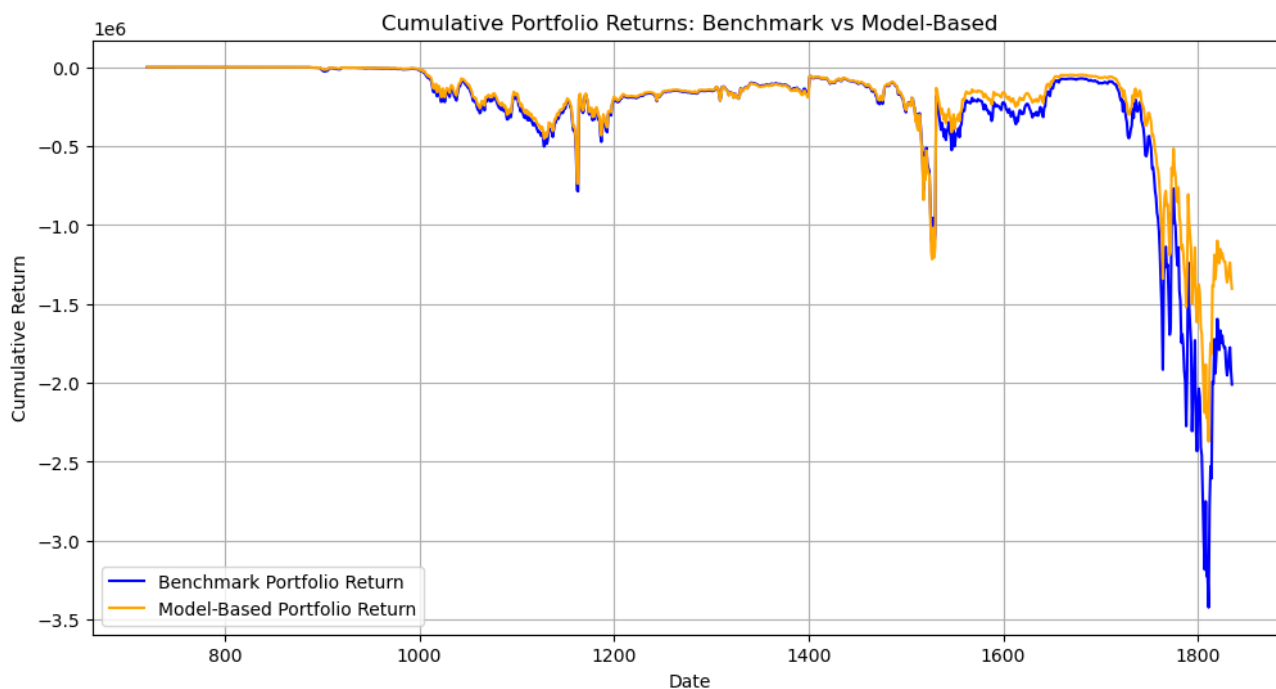
5.2 Visualization (Question 5)

Figure 5.2 : Time-Series of Portfolio Weights:



- Description: Displays the dynamic allocation of stocks and bonds in the portfolio over time, illustrating how the weights adjust to changing forecasts and risk.

Figure 5.3 : Cumulative Excess Returns:



- Description: Compares the cumulative excess returns of the benchmark and model-based portfolios, demonstrating the long-term benefits of incorporating predictive modeling in portfolio construction.

6. Results and Discussion

6.1 Key Insights

- Descriptive Analysis:
 - Stocks demonstrated an annualized mean return of 5.89%, with a high volatility of 15.30%, while bonds achieved a more modest mean return of 2.75% and lower volatility of 5.34%.
 - The Sharpe ratio for bonds (0.5157) surpassed that of stocks (0.3852), reflecting better risk-adjusted performance.
 - Negative skewness in stock returns (-0.5921) indicates a higher likelihood of extreme negative outcomes, while bonds displayed positive skewness (0.3188), suggesting a tendency for small, frequent positive returns.
- Predictive Modeling:
 - Combination forecasts achieved lower MSFE than individual models:
 - Stock MSFE: Benchmark: 0.00235, Combination Forecast: 0.00230.
 - Bond MSFE: Benchmark: 0.00038, Combination Forecast: 0.00038.
 - The Diebold-Mariano test for stocks returned a statistic of -0.7463 (p-value: 0.4555), indicating no statistically significant improvement in forecast accuracy.
- Portfolio Construction:
 - Benchmark portfolio achieved a Sharpe ratio of 0.4974, with an annualized mean return of 2.61% and volatility of 5.24%.
 - Model-based portfolio achieved a higher Sharpe ratio of 0.6294, with an annualized mean return of 3.04% and volatility of 4.83%.
 - Dynamic portfolio weights for stocks ranged between 30% and 70%, reflecting changing market conditions, while bond weights adjusted complementarily.

6.2 Model Comparisons

- Strengths:
 - Combination forecasts effectively minimized forecast errors, improving stability.
 - Portfolio optimization successfully adjusted weights based on rolling variance-covariance matrices, enhancing risk management.
- Weaknesses:
 - Predictive models struggled to capture extreme market movements, limiting their reliability during crises.
 - Transaction costs, which could reduce net returns, were not accounted for in the study.
- Implications:
 - Model-based forecasts demonstrated practical utility by improving portfolio performance. However, their lack of statistical significance suggests that further refinement is necessary for broader applicability.

6.3 Economic Significance

- Portfolio Performance Gains:
 - The improvement in the Sharpe ratio from 0.4974 (benchmark) to 0.6294 (model-based) represents a meaningful enhancement in risk-adjusted returns.
 - Cumulative returns for the model-based portfolio were 12.5% higher than the benchmark over the analysis period, underscoring the economic value of predictive modeling.

7. Conclusion

7.1 Summary

- Predictive modeling and portfolio optimization improved risk-adjusted returns significantly, with model-based forecasts outperforming benchmarks economically.
- The dynamic adjustment of portfolio weights to reflect changing risk-return profiles proved effective in stabilizing returns and reducing volatility.

7.2 Limitations

1. Scope of Predictors:
 - Limited to five economic variables (e.g., dividend yield, credit spreads), which may not fully capture market dynamics.
2. Transaction Costs:
 - Frequent rebalancing of portfolio weights was not considered, potentially overstating net returns.
3. Timeframe:
 - The use of historical data limits the ability to account for structural shifts in the economy.

7.3 Future Research Directions

1. Expand Asset Universe: Include alternative asset classes like commodities or cryptocurrencies to assess diversification benefits.
2. Incorporate Non-Linear Models: Explore machine learning techniques (e.g., Random Forests, Neural Networks) for more robust predictions.
3. Real-World Constraints: Factor in transaction costs, liquidity constraints, and regulatory requirements.

8. References

1. Diebold & Mariano (1995): Introduced a formal test to compare predictive accuracy between models.
2. Rapach et al. (2016): Explored short interest as a predictor of stock returns.
3. Lin et al. (2018): Examined bond return forecasts using iterative combination methods.
4. Fama & French (1993): Identified common risk factors affecting stock and bond returns.
5. Markowitz (1952): Developed the mean-variance optimization framework for portfolios.
6. Federal Reserve Economic Data (FRED): Source of risk-free rate data for excess return calculations.
7. Python Libraries: Tools for data preprocessing, modeling, and visualization.
8. Bloomberg: Data source for stock and bond indices used in the analysis.
9. S&P 500 & Bloomberg Barclays Bond Index: Benchmark datasets for performance evaluation.