Smart Beta Strategy: Leveraging Barra Factors for Portfolio Construction

Luyuan Fan

Abstract

This study investigates factor-based portfolio construction using the Smart Beta framework and the Barra model methodology. The analysis begins with a factor validity test using the Information Coefficient (IC) to measure the predictive power of various factors. Size, Non-linear Size, and Earnings are identified as the most significant factors with |IC| > 0.02 and integrated into a composite factor score (score3) to enhance stock selection accuracy. Quintile-based portfolio grouping reveals that stocks with the highest score3 values (Quintile 5) consistently outperform others, achieving an average return of 20.47% and an excess return of 8.33%.

Accrual-based screening is applied to further refine portfolio quality. By removing high-accrual stocks often linked to weaker performance, this approach combines predictive factor selection and risk control, achieving an average annual return of 21.90% and an excess return of 9.49%. This outcome demonstrates robust portfolio performance and reliability. The study suggests future improvements, including incorporating additional factors like return on equity (ROE) and dividend yield and exploring shorter evaluation periods to identify effective short-term factors.

By leveraging the Smart Beta framework and Barra model, this research provides practical insights into factor-based investing, offering a systematic and adaptable methodology for achieving risk-adjusted returns.

Keywords: Barra Model, Smart Beta strategy, stock selection, Information Coefficient, portfolio construction, portfolio optimization.

1. Introduction

1.1 Literature Review

Portfolio optimization has been a cornerstone topic in finance, aiming to achieve higher risk-adjusted returns through effective asset allocation. Markowitz (1952) laid the foundation with the Modern Portfolio Theory (MPT), introducing the mean-variance optimization framework. Sharpe (1964) extended this with the Capital Asset Pricing Model (CAPM), which attributes asset returns to market risk premiums and individual asset beta. However, these traditional models assume market efficiency and often overlook abnormal factors that may influence asset returns.

As behavioral finance evolved, researchers recognized that asset returns are influenced not only by market risks but also by predictive factors such as size, value, and momentum. Fama and French (1992) extended CAPM by proposing a three-factor model that incorporates the book-to-market ratio (B/M) and size (Size) factors, significantly enhancing the explanatory power of asset returns. Building on this, Carhart (1997) introduced momentum as an additional factor, resulting in the four-factor model, which serves as a foundational framework for factor investing.

In recent years, Smart Beta strategies have emerged as a prominent application of factor investing. These strategies leverage systematic factors like value, momentum, and profitability to construct portfolios, aiming to outperform traditional market-cap-weighted indices. For instance, Ang et al. (2009) analyzed factor-based strategies, demonstrating how specific exposures to systematic risks drive returns. They emphasized that factor selection and portfolio construction are critical to achieving consistent performance advantages.

1.2 Overall Strategy

The Smart Beta strategy incorporates monthly prices and financial ratios from S&P 500 constituent stocks. Similar to the Barra model, it calculates ten key factors along with their Information Coefficient (IC) to establish the relationship between these factors and stock returns. Valid factors are determined based on IC values and are used to score the stocks.

Based on the strategy, an equal-weighted basic portfolio is constructed to minimize the influence of large market capitalization stocks. Furthermore, an advanced portfolio enhances this approach by initially screening stocks based on a specific financial ratio, specifically accruals, before applying the valid factors.

2 Theoretical Analysis

The Barra Multi-Factor Model, developed by Barra Inc. (now part of MSCI) in the 1970s, is a widely-used for analyzing portfolio risk and return. This model employs a systematic approach to decompose portfolio risk into contributions from various systematic factors and specific risks. The factors are categorized into style factors, such as size, momentum, and value, and industry factors, covering different sectors. The model allows for detailed attribution of portfolio performance and risk, providing investors with insights into systematic and idiosyncratic drivers of returns.

The main part of the Barra Model is to decompose total portfolio risk into factor-based risks and specific risks. The formula is as follows,

$$Portfolio\ Risk = F' \Sigma_f F + \Sigma_{specific}$$

$$\begin{aligned} where, \quad & F\text{: } The\ factor\ exposure\ matrix \\ & \Sigma_f\text{: } The\ factor\ covariance\ matrix \\ & \Sigma_{specific}\text{: } The\ diagonal\ matrix\ of\ specific\ risks \end{aligned}$$

The Barra model has several limitations. It assumes that factor covariance structures and specific risks remain stable over time, which may fail during periods of market stress or structural changes. The model's accuracy heavily depends on the quality and timeliness of input data, such as factor exposures and market conditions. Additionally, the computational complexity of processing high-dimensional covariance matrices and large datasets can be a challenge. Lastly, the model has limited adaptability to emerging risks, such as ESG factors or geopolitical risks, unless explicitly updated.

Despite its limitations, its robust mathematical foundation and practical applicability make it indispensable for analyzing risk-adjusted returns, optimizing portfolios, and ensuring alignment with investment objectives.

3 Factor Selection

3.1 Data Resource

This paper uses the market data from CRSP (Center for Research in Security Prices) data, obtained from Wharton Research Data Services (WRDS), provides comprehensive stock prices, trading volumes, and market indices for all listed companies. This dataset is significant for understanding overall market dynamics, allowing analysts to track trends and evaluate market sentiment. It includes essential details such as price, volume, returns, shares outstanding, and S&P 500 index returns, offering a robust foundation for financial analysis and quantitative research.

This paper utilizes financial data from Compustat Financials, which provides key metrics such as earnings, dividends, and revenue figures. These financial signals are vital for understanding a company's internal economic condition, offering insights into its financial health and profitability. The dataset includes indicators such as book-to-market ratio, earnings growth, debt-to-asset ratio, one-year price-to-earnings (P/E) growth rate, monthly returns, alpha, idiosyncratic volatility, total volatility, and excess returns, enabling a comprehensive evaluation of corporate performance.

3.2 Factor Calculation

In financial analysis, factors are key indicators derived from large sets of market data. They are used to assess different market dynamics and predict future behavior. Below are core factors and their calculation methods:

(1) Size: Market capitalization, measured using the natural logarithm of a company's market cap (calculated as stock price multiplied by total shares outstanding), is commonly used to distinguish between large-cap and small-cap stocks. Large-cap companies typically exhibit lower risk but offer limited growth potential compared to their smaller counterparts.
(2) Beta: The correlation between a company's stock returns and overall market returns. It measures the stock's performance relative to the market, with high-beta stocks potentially yielding higher returns during market uptrends but also experiencing greater losses during downturns.
(3) Momentum: The cumulative log return over the past 12 months, reflecting performance trends. It captures the market sentiment and trends, often used to track sustained price increases or decreases.
(4) Volatility: Standard deviation of excess stock returns over the risk-free rate. A high-volatility stocks may offer higher returns but carry greater risks.
(5) Nonlinear Size: A linear regression of market cap and its cube, using residuals as the factor. It Captures the complex nonlinear relationships between market cap and other financial indicators.
(6) Book-to-Market Ratio: Ratio of a company's book value to its market value. A vital value-investing indicator; a low ratio may suggest undervaluation.
(7) Liquidity: Ratio of trading volume to shares outstanding. A high liquidity indicates ease of trading, typically associated with lower transaction costs and stable prices.

(8) Earnings Yield: Ratio of earnings per share (EPS) to stock price. A high earnings yield may suggest better long-term investment returns.
(9) Growth: Measured by the price-to-earnings (P/E) ratio divided by the projected earnings growth rate (PEG). It helps investors identify companies with potential for future earnings improvement.
(10) Leverage: Ratio of total debt to total assets. A high leverage implies greater financial risk but can enhance capital returns.

By integrating these factors, investors and analysts gain a comprehensive view of financial markets and individual securities from both macro and micro perspectives. Accurate data analysis and factor calculation enable financial professionals to better manage assets and risks, optimizing portfolio performance.

3.3 Data Cleaning and Standardization

A systematic data preprocessing framework was applied to ensure robustness and reliability in the analysis. Lagging was implemented to avoid look-ahead bias and reflect real-world investment scenarios. Financial factors, including size, non-linear size, leverage, and book-to-price ratio, were lagged by 3 months to account for reporting delays. In contrast, market factors such as liquidity, momentum, earnings, growth, beta, and volatility were lagged by 1 month due to their immediate availability.

Outliers were addressed using the Interquartile Range (IQR) method, where outliers were defined as values outside the range

$$[Q_1 - 3 \times IQR,\ Q_3 + 3 \times IQR],$$

$$where\ Q_1\ is\ the\ first\ quartile,\ Q_3\ is\ the\ third\ quartile,\ IQR = Q_3 - Q_1$$

A multiplier of 3 was selected to avoid excessive data removal compared to the conventional 1.5. Rows containing outliers were removed entirely, resulting in a reduction of the dataset size by 17.38% (from 125,629 rows to 103,790 rows), yielding a cleaner and more stable dataset.

This paper used the Z-score normalization method to standardize the factor data, ensuring comparability across factors with varying scales. The formula is:

$$Z = \frac{X - \mu}{\sigma}$$

$$where\ X:\ Original\ value,\ \mu:\ Mean\ of\ the\ data,\ \sigma:\ Standard\ deviation\ of\ the\ data.$$

This transformation centered each factor’s distribution around a mean of 0 and a standard deviation of 1, minimizing the influence of larger-scaled factors on the analysis. These preprocessing steps provided a robust foundation for subsequent factor evaluation and portfolio construction.

3.4 Factor Validity Test

The Information Coefficient (IC) measures the predictive power of a single factor in relation to the target variable, typically returns. It quantifies the rank correlation between the factor values and the corresponding returns. The IC value ranges from [-1,1], with absolute values closer to 1 indicating stronger predictive power.

In the table below are the average IC values for the factors.

Table: Factors and IC Values

Factors	IC value
Liquiditily	0.018442
Size	-0.024664
Momentum	-0.010018
Non-linear size	-0.025063
Book-to-Price	-0.007384
Earning	0.023869
Growth	0.002953
Leverage	-0.000421
Beta	0.011993
Volatility	-0.009266

Considering |IC|>0.02 as valid values, **Size**, **Non-linear size** and **Earning** are valid factors to construct the portfolio.

4 Basic Portfolio Construction

4.1 Introducing the Composite Factor Score

To integrate the predictive power of multiple factors and better reflect their combined influence on stock returns, we introduce a composite factor score, denoted as score3. The formula for score3 is defined as:

$$score3 = EARNING - SIZE - NON—LINEAR\ SIZE$$

The IC values of Size and Non-linear size are negative, indicating that a higher value of these factors corresponds to lower returns. Therefore, their weights are subtracted. By combining these factors into a single score, this paper aims to enhance the overall predictive ability of the factors and provide a more robust ranking for stock selection.

4.2 Portfolio Grouping and Return Analysis

To evaluate the predictive power of the composite factor score3, stocks are grouped into five quintiles (Q1 to Q5) based on their score3 values. The grouping process categorizes stocks into Q1 (lowest score3) to Q5 (highest score3), with each quintile containing an equal number of stocks.

For each quintile, the average return and cumulative return are calculated. The cumulative return is computed as follows, and the results are illustrated in the figure 4.1:

$$Cumulative\ Return\ (t)\ =\prod_{i=1}^{t}(1\ +\ RET_i)\ -\ 1$$

$$where\ RET_i\ represent\ the\ average\ return\ for\ a\ special\ quintile\ at\ time\ i$$

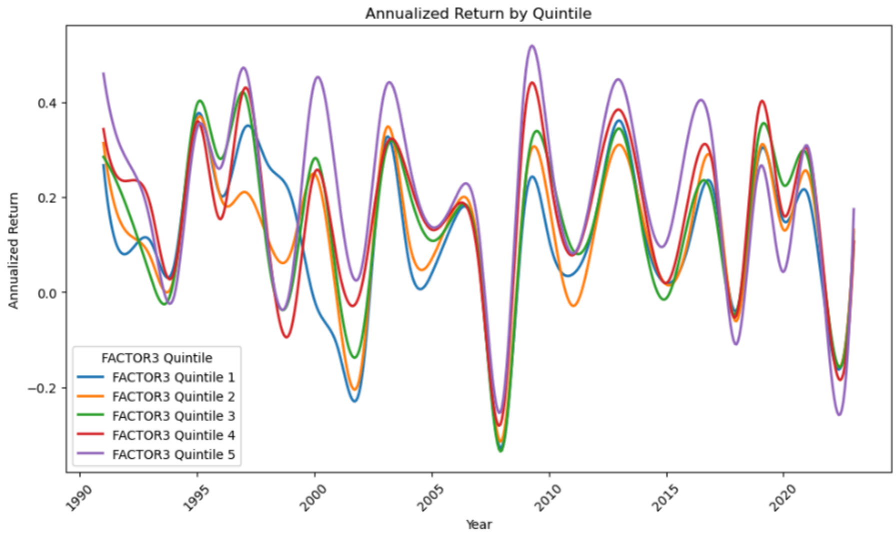

Figure 4.1 Basic Annualized Return by Quintile

Q5, representing stocks with the highest score3 values, outperforms other quintiles in most cases. While Q5's performance dips below other quintiles in certain years, its overall trajectory demonstrates superior cumulative returns over the long term, highlighting its robustness and predictive power. Additionally, the average return for Q5 is **20.47%**, which is a strong and favorable result.

4.3 Excess Return Analysis

To assess the performance of our portfolio against a widely recognized benchmark, we use the S&P 500 index (SPY) as the benchmark. The annualized returns of the portfolio and SPY are calculated and illustrated in the following figure,

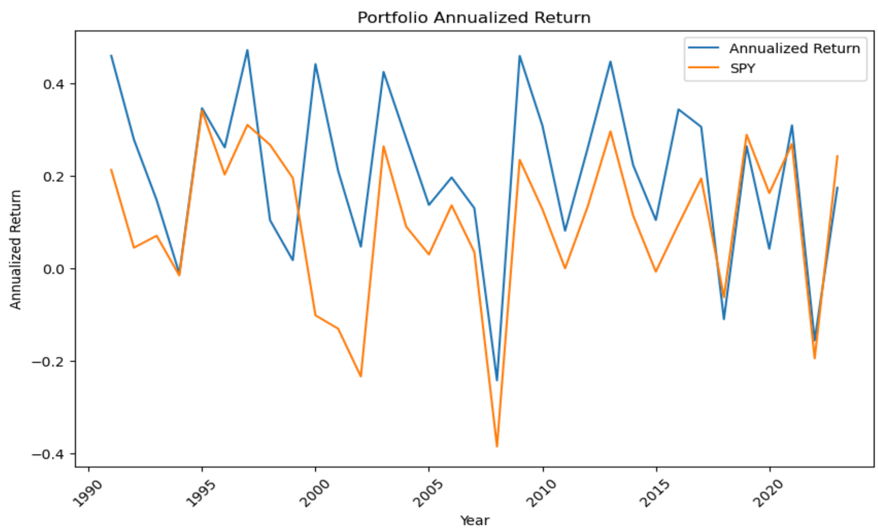

Figure 4.2 Basic Portfolio Annualized Return

From the figure, it is evident that Q5 consistently outperforms the SPY in most years, with higher annualized returns. Although the portfolio exhibits greater volatility in some periods, its overall trajectory demonstrates superior performance compared to the market benchmark.

To further evaluate the predictive power of the composite factor score3, calculating the excess return for each stock as:

$$Excess\ Return = RET - Benchmark\ Return$$

Using the same quintile grouping based on score3, the annualized excess return for each group is computed. The results are illustrated in the following figure,

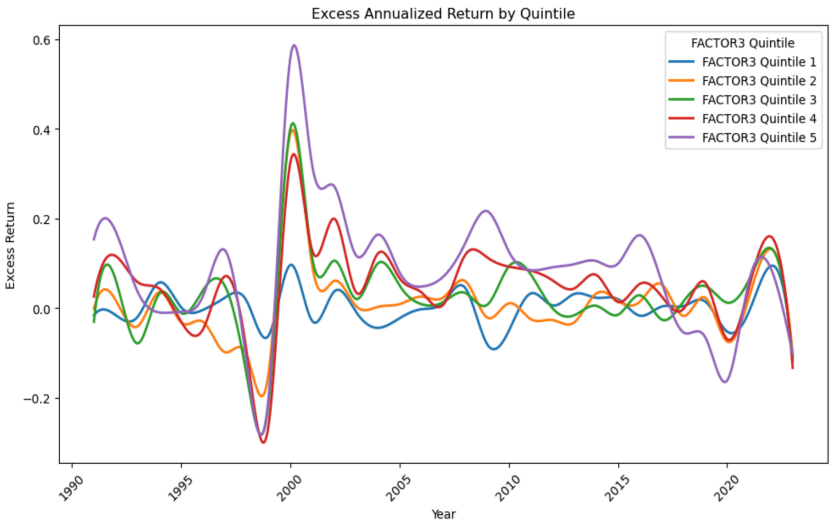

Figure 4.3 Basic Excess Annualized Return by Quintile

Similar to the previous cumulative return analysis, Q5 shows strong performance in most periods. While it dips below other quintiles in a few individual years, Q5's overall trajectory remains the best, demonstrating its ability to deliver the highest cumulative excess returns over time.

Q5 achieves an average excess return of **8.33%**, with the sharpe ratio **0.58**, which is a solid result, further supporting the effectiveness of the score3-based stock selection strategy.

5 Advanced Portfolio Construction

5.1 Screening Indicator in Enhancing Portfolio Performance

Found that the portfolio drawdown is too fast in some periods, to reduce this risk, this paper decides to look for further screening of the stock pool through some indicators and construct further portfolios. Through testing different indicators and research on papers in the field of quantitative investment, Cai (2022) demonstrated that accrual as a screening indicator can significantly improve the performance of the investment portfolio.

Accrual refers to the practice of a company recording revenue in advance or recognizing expenses in a delayed manner. Unlike cash flow, accrual can be affected by management's accounting policies and therefore can be used to adjust or even manipulate a company's earnings. If a company's accrual value is high, it means that there are more non-cash components in the company's earnings. This is very unfavorable for investors to judge the financial status of a company. For example, the most famous financial fraud by Enron was the use of accrual operations to confirm the expected profits for the next 20 years in advance and conceal the huge debts, which eventually led to the company's bankruptcy.

5.2 Advanced Portfolio Performance Analysis

Based on the above information, the higher the accrual value, the more uncertain the company's financial situation is. This paper sorts the stocks according to the size of the accrual value, selects half of the stocks with lower accrual values in the original stock pool, and rebuilds the investment portfolio. The new portfolio performance is shown below. The order of the images is placed in the order of the basic portfolio chapter images for comparison.

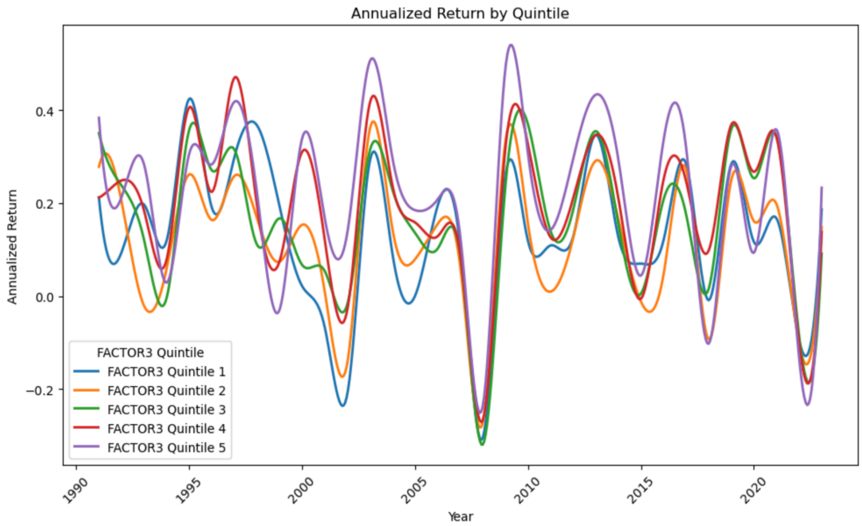

Figure 5.1 Advanced Annualized Return by Quintile

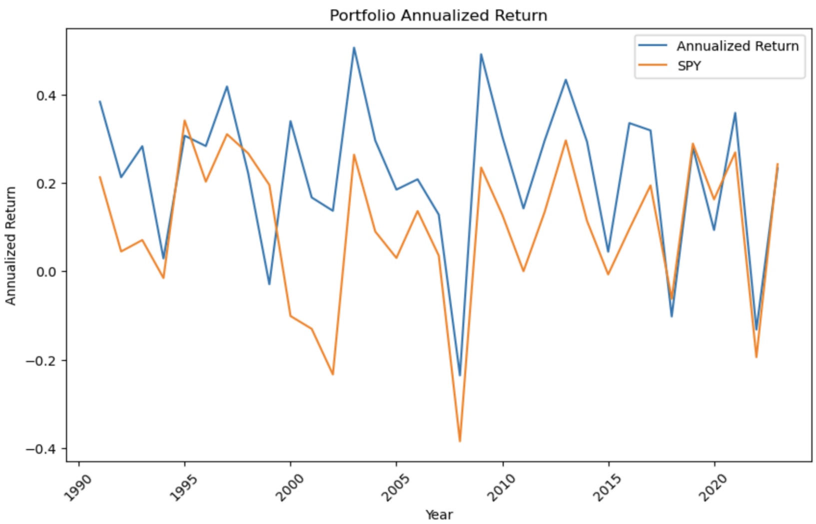

Figure 5.2 Advanced Portfolio Annualized Return

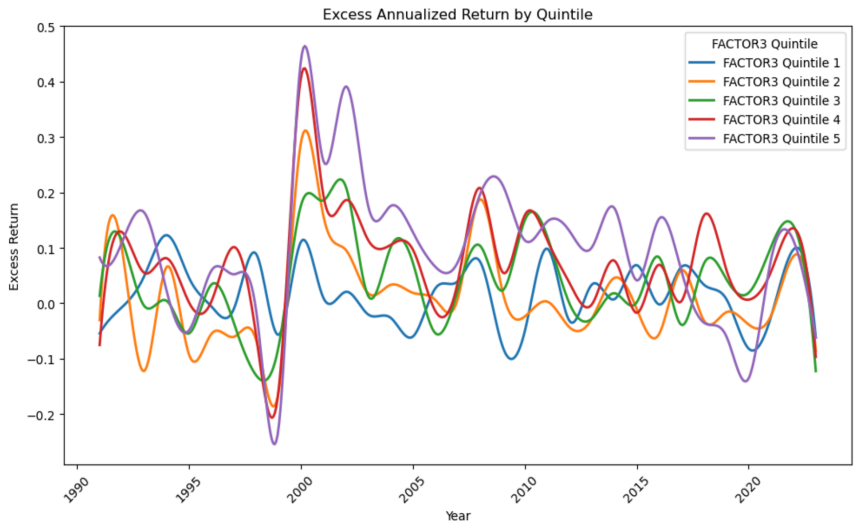

Figure 5.3 Advanced Excess Annualized Return by Quintile

After deleting half of the stocks with higher accruals, the overall performance of the reconstructed investment portfolio is greatly improved. Its annualized return and excess annualized return increases by nearly 1.5%, reaching **21.90%** and **9.49%** respectively, and the drawdown rate has been reduced to a certain extent, meaning the risk of the portfolio has been reduced.

In addition, the sharpe ratio of the advanced portfolio, **0.72**, is much higher than that of the basic portfolio, indicating that the new portfolio has a lower risk-to-return ratio. During the period from 1990 to 2023, the US economy experienced several growths and recessions, and the new investment portfolio curve almost performs better than the S&P 500 in both growth and recession economic environments.

The statistics prove that the advanced investment portfolio is very successful. However, in recent years, the portfolio and the S&P 500 yield curve are almost fitted, which may indicate that this quant strategy has been used by many quantitative companies, and its return will slowly decrease back to the market benchmark level.

6 Conclusion

This paper examines the factor-based portfolio construction using a structured approach inspired by the Smart Beta framework and the Barra model. The IC values are used to measure the predictive power of factors, identifying Size, Non-linear Size, and Earnings as the most important factors with |IC| > 0.02. These factors are combined into a composite factor score (score3) to improve accuracy and provide a solid foundation for stock selection.

The results show that score3 effectively identifies high-performing stocks. Stocks in Quintile 5 (Q5), with the highest score3 values, deliver strong cumulative returns and excess returns over the long term. The average return of 20.47% and excess return of 8.33% for Q5 confirm the success of the composite factor score approach.

Additionally, stocks are screened using accrual values, as accruals are often seen as a negative signal for stock performance. By removing the top 50% of stocks with higher accruals and keeping those with lower accruals, this paper adds an extra layer of risk control to the portfolio construction process. The cumulative return for the accrual-screened portfolio demonstrates consistent outperformance relative to the benchmark, with a Sharpe ratio of 0.72 and an average return of 21.90%. This combination of factor-based and accrual-based strategies enhances the portfolio's overall quality and reliability.

Compared to the S&P 500 benchmark, with an average return of 9.78% during the same period, the proposed portfolio demonstrates superior performance in both returns and risk-adjusted metrics. This demonstrates the effectiveness of combining factor analysis with accrual screening in constructing a high-performing portfolio.

In conclusion, this study shows that combining factor analysis with accrual screening improves portfolio performance by integrating predictive strength and risk management. This approach provides a practical and flexible framework for constructing better-performing portfolios. To improve upon this study, future work could include exploring additional factors like return on equity (ROE) and dividend yield, as well as shortening evaluation periods to identify effective short-term factors. These enhancements would make the strategy more adaptable to rapidly changing market conditions.

Reference:

Markowitz, Harry. “Portfolio Selection.” *The Journal of Finance* 7, no. 1 (1952): 77–91. https://doi.org/10.2307/2975974.

Sharpe, William F. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance* 19, no. 3 (1964): 425–42. https://doi.org/10.2307/2977928.

Fama, Eugene F., and Kenneth R. French. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance* 47, no. 2 (1992): 427–65. https://doi.org/10.2307/2329112.

Carhart, Mark M. “On Persistence in Mutual Fund Performance.” *The Journal of Finance* 52, no. 1 (1997): 57–82. https://doi.org/10.2307/2329556.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. "High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence." *Journal of Financial Economics* 91, no. 1 (2009): 1–23. https://doi.org/10.1016/j.jfineco.2008.12.005.

Bingjie Cai, “Research on MSCI Barra CNE5 Model and Stock Selection,” www.atlantis-press.com (Atlantis Press, April 29, 2022), https://doi.org/10.2991/aebmr.k.220405.252.

Cai, Bingjie. "Research on MSCI Barra CNE5 Model and Stock Selection." *Atlantis Press*, (April 29, 2022). https://doi.org/10.2991/aebmr.k.220405.252.