Journal of Administrative Development, Vol. 2 (1), 2024

UNLOCKING MARKET SECRETS: STATISTICAL FACTOR MODEL IN STOCK RETURN PREDICTION

Shantnu Sood[*] & Dr. Puneet Bhushan[†]

Abstract

This paper delves into the domain of statistical factor models as a unique approach for constructing factor-based investment strategies with the goal of outperforming the stock market. Unlike conventional factor models that explicitly identify and quantify the factors influencing stock returns, the statistical factor model operates with latent and concealed factors, rendering it akin to a 'black box' strategy.The primary allure of the statistical factor model lies in its potential to unearth and utilize hidden factors that may remain unnoticed through traditional analysis. This approach allows investors to potentially tap into novel sources of return, enhancing their portfolio performance. However, this unorthodox characteristic also presents a challenge: the lack of transparency regarding the precise factors driving returns. The study tests the statistical factor approach on the constituents of NSE 500 index from 2013 to 2023 and found that the model comprehensively beats NSE 500 buy and hold strategy.

Keywords: *Machine Learning, PCA, Deep Learning, Quantitative Finance, Unsupervised Learning*

1. Introduction

In the dynamic and ever-evolving landscape of finance and investment, traditional methods for constructing and managing portfolios have encountered challenges and limitations. As financial markets grow

[*] *Research Scholar, Himachal Pradesh University Business School, email – soodshantnu2@gmail.com*
[†] *Assistant Professor, Himachal Pradesh University Business School, email – pbsud.sml@gmail.com,*

increasingly intricate and information becomes more accessible, investors are constantly in search of innovative strategies to achieve superior risk-adjusted returns and protect against unpredictable market conditions. Among these emerging approaches, factor investing has emerged as a prominent and promising paradigm, offering an alternative perspective for portfolio construction.

Factor investing is based on the belief that specific risk factors drive the returns of financial assets, presenting unique opportunities for investors to optimize their portfolios. This research paper delves into the concept of factor investing, rooted in the understanding that factors such as market risk, size, value, momentum, and quality, among others, underlie asset returns.

In a world where conventional portfolio strategies are being challenged by the complexities of the modern financial landscape, factor investing emerges as a compelling and innovative approach worthy of comprehensive investigation. Therefore, this paper aims to provide the back tested results of one such factor investing model called statistical factor model that aims to derive factors from the covariance matrix of the returns of each stock comprising Nifty 500 index for the period of 10 years instead of using traditional factors such as momentum, size, quality etc.

II. Objectives of the Study

- To develop a statistical factor model using PCA on the constituents of NSE 500 index.
- To evaluate the performance of developed model against NSE 500 buy and hold strategy.

Iii. Review of Literature

The journey in factor models commenced with Harry Markowitz's groundbreaking 1952 research on Modern Portfolio Theory. This theory essentially attributed the return of a portfolio to the weighted returns of individual assets, defining portfolio risk as the standard deviation of the individual assets along with their correlations. Post-Modern Portfolio Theory, an extensive body of literature emerged, aiming to formulate a market equilibrium theory for asset prices in the presence of risk. Notable contributions included (Sharpe, 1964) proposition of a linear relationship between expected return and standard deviation of return (systematic risk) for efficient combinations of risky assets. This was depicted through an investment opportunity curve that highlighted the advantages of diversification alongside riskless borrowing and lending.

Subsequently, (Lintner, 1965) sought to establish conditions under which investors would optimally hold stocks (long or short) in their portfolios, even when risk premiums varied. (Fama& Kenneth, 1992), introduced their three-factor model, identifying three stock market risk factors that explained a significant portion of stock returns. These factors encompassed overall market risk (calculated as the beta of the stock), firm size (differentiating between low and high market capitalization stocks), and book-to-market equity (discriminating between low and high book-to-market capitalization stocks). These pioneering studies paved the way for subsequent factor research. It is worth noting that many of these factors relied on fundamental stock data, which can be challenging to acquire and work with due to its irregular release schedule, often limited to annual or quarterly updates in contrast to the more frequently released price data (daily, hourly, or at tick levels).

Recognizing the limitations of fundamental data, (Connor &Korajczyk, 1988) explored Principal Component Analysis (PCA) as a tool to uncover latent factors using solely the price data of stocks. Their research highlighted PCA's potential in this context. Moreover, (Bai et al. ,2012) conducted a comparative study, pitting cross-sectional and time series statistical factor models against the Fama-French four-factor model. Their findings revealed that statistical factor models outperformed the Fama-French model, suggesting alternative avenues for factor modeling. PCA, a statistical method, plays a pivotal role in this context by transforming high-dimensional data into a new coordinate system while retaining critical information from the original dataset. It achieves this by identifying principal components, which are orthogonal linear combinations of the original features, ranked by the amount of variance they capture.

(Chatopadhyay et al., 2015) proposed a factor investing model incorporating two factors, Return on Equity (ROE) and Book-to-Price (B/P). Their research demonstrated that this model outperformed a passive investing strategy based on the S&P 500. Meanwhile, (Cueto et al., 2020) examined the coefficient of variation as an explanatory variable for stock returns and found it to be statistically insignificant.

In this paper, a trading strategy based on a PCA-driven statistical factor model has been tested using the constituents of the Nifty 500 index over a ten-year period.

IV. Research Methodology

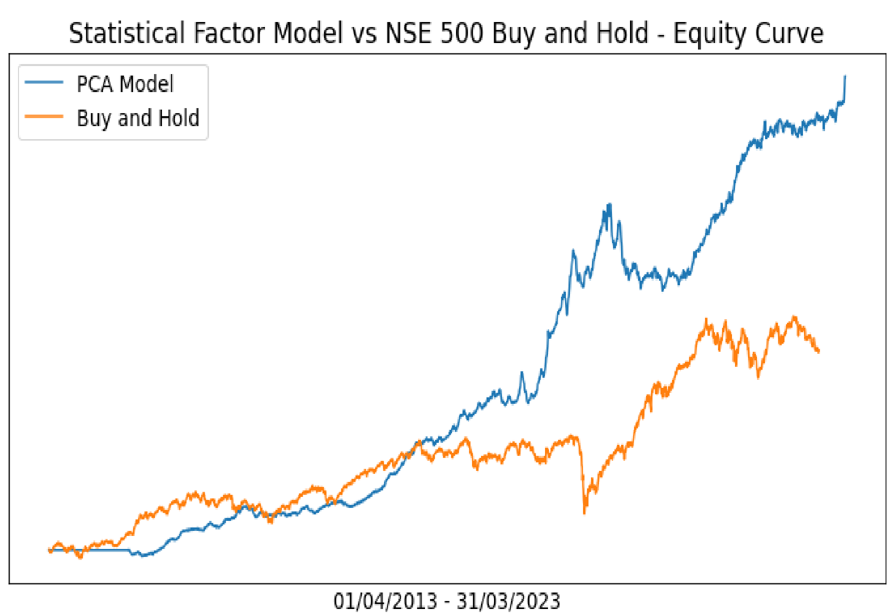
- Data – The study has been conducted on the constituents of NSE 500 index for the period 1/04/2013 to 31/03/2023 using daily adjusted closing prices. There were a total of 956 companies during the entire period.
- This research employs statistical factors (factor loadings) derived from the covariance matrix of stock returns using principal component analysis and sorting the eigenvectors in descending order of eigenvalues to find independent variables to construct a time-series factor model. The decomposition of a stock's return is represented as:

$$return(t, s) - rf = \alpha(t, s) + \beta 1(s) * factor1(t) + \beta 2(s) * factor2(t) + \cdots + \varepsilon(t, s) \tag{1}$$

$$return(t+1, s) - rf = \alpha(t, s) + \beta 1(s) * factor1(t) + \beta 2(s) * factor2(t) + \cdots + \varepsilon(t, s) \tag{2}$$

- **Trading strategy**
  - To create a predictive model, a trading strategy based on the principal components is implemented. For each trading day, the strategy considers the returns of the past n-252 trading days for all 500 stocks that are part of NSE 500 index to calculate the principal components which explain the most amount of variance.
  - These components are then used as independent variables to forecast the next day's returns for all the stocks.
  - Based on these predictions, we take long positions in the top 50 stocks with the highest predicted returns for the following day and take short positions in the bottom 50 stocks with the lowest predicted returns.
  - These positions are held till market's closing. This process is repeated for 2294 days i.e. 2546 (total no. of days during study period – starting 252 days).

V. Results and Discussion



Performance Metrics	Statistical Factor Model (NSE 500)	Buy and Hold (NSE 500)
Average Return (Annualized)	18.43%	13.39%
Standard Deviation (Annualized)	7.92%	16.86%
Sharpe Ratio	2.32	0.79

The present study has successfully implemented an intraday trading statistical factor model. This innovative model leverages the power of Principal Component Analysis (PCA), a sophisticated statistical procedure that uses an orthogonal transformation. This transformation converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, known as principal components. This transformation is particularly effective in dealing with high-dimensional data, reducing dimensionality while retaining the most important information.

In our study, this transformation was applied to the constituents of the NSE 500 index. The NSE 500 is a broad-market equity benchmark index that represents about 96% of the total market capitalization in India. It includes companies from all major sectors of the economy, providing a comprehensive and diversified portfolio for investment.

The implementation of this model has yielded impressive results. The average return achieved was 18.43%, a figure that is noteworthy when compared to the average return of 13.39% achieved by the traditional NSE 500 buy and hold index investing strategy. This significant increase in returns demonstrates the effectiveness of our model in capturing the underlying patterns in the market and translating them into profitable trades. It underscores the potential of PCA-based models in enhancing investment returns in the stock market.

Furthermore, our statistical factor model exhibited a lower annualized standard deviation of 7.92% compared to the 16.86% in the buy and hold strategy. Standard deviation is a widely used measure of the amount of variation or dispersion in a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range. In the context of investment, a lower standard deviation is synonymous with lower volatility, and hence, lower risk. Therefore, the reduced standard deviation in our approach indicates a reduced risk exposure, making it a more stable and reliable investment strategy.

Moreover, the Sharpe ratio of the statistical factor model is 2.32, in comparison to 0.79 of the buy and hold strategy. The Sharpe ratio is a measure used to understand the return of an investment compared to its risk. The ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk. A higher Sharpe ratio indicates better risk-adjusted performance. Therefore, the higher Sharpe ratio of our model further attests to its superior performance in terms of both returns and risk management.

VI. Conclusion

The statistical factor model indeed presents an intriguing approach to construct a factor-based investment strategy aimed at surpassing the performance of the stock market. However, it distinguishes itself from other factor models by its inherent characteristic of dealing with latent and concealed factors. Unlike traditional factor models that explicitly identify and define the factors influencing stock returns, the statistical factor model operates in a somewhat enigmatic manner. In essence, it resembles a 'black box' strategy, wherein the inner workings of the model remain veiled from a comprehensive understanding.

This opaqueness in the functioning of the model raises intriguing questions and concerns. While it has the potential to deliver robust returns, the lack of transparency regarding the specific factors driving those returns poses a significant challenge for investors. Typically, factor-based investing thrives on the ability to discern and exploit the fundamental drivers of asset performance. In contrast, the statistical factor model seems to shroud these underlying factors in mystery, making it difficult for investors to make informed decisions or adjustments based on a clear understanding of what is happening in the background. This black box aspect of the statistical factor model may be both strength and a weakness, depending on one's perspective.

References

1. Chattopadhyay, A., Lyle, M. R., & Wang, C. C. (2015). Accounting data, market values, and the cross section of expected returns worldwide. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2613366
2. Connor, G., &Korajczyk, R. A. (1988). Risk and return in an equilibrium APT. Journal of Financial Economics, 21(2), 255-289. https://doi.org/10.1016/0304-405x(88)90062-1
3. Cueto, J. M., Grané, A., & Cascos, I. (2020). Models for expected returns with statistical factors. Journal of Risk and Financial Management, 13(12), 314. https://doi.org/10.3390/jrfm13120314
4. Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics, 22(1), 3-25. https://doi.org/10.1016/0304-405x(88)90020-7
5. Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2), 427. https://doi.org/10.2307/2329112
6. Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics, 47(1), 13. https://doi.org/10.2307/1924119
7. Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425. https://doi.org/10.2307/2977928