

Combining Alpha Signals Using Ensemble Methods for Enhanced Alpha

Siddhant Sarkar¹, Swarnim Shukla², Sourabh Patel³, Rahul Painkra⁴

¹⁻⁴B.E. Final Year Student, Computer Science and Engineering, Government Engineering College, Bilaspur, India.

Abstract - Alpha is the risk-adjusted performance of the portfolio, or a manager's 'value added'. Mathematically, alpha is the incremental difference, given the degree of risk, between the actual results of a manager and its predicted results. A positive alpha implies that a portfolio has achieved returns above the anticipated level, and a negative alpha suggests the underperformed portfolio, despite the assumed level of risk. Alpha is one of five basic risk ratios used in the Theory of Modern Portfolios. The remaining four are beta, R-squared, standard deviation, and the ratio of Sharpe. Many of these statistical metrics assist investors in assessing a fund's risk-reward profile. In this paper we will explore a method for mixing outputs from multiple models of Alpha Factors using a form of ensemble modelling known as stacked generalization. This approach excels in creating models that "generalize" well to unknown future data, making them an excellent choice for the financial domain where it is a major challenge to overfit past data. In markets, there are many minor inefficiencies and trends which can be found and used to gain slight advantage on the competition. These edges are rarely sufficiently large to trade in isolation-transaction costs and overheads can easily exceed the expected profits. But when we can combine several of those tiny edges, the rewards can be great.

Key Words: Quantitative Trading, Alpha Factor, Statistical Risk Model, Portfolio Optimization, Factor Modelling, Alpha Combination, Ensemble Modelling.

1. INTRODUCTION

In order to produce better returns, multiple alpha factors are combined to produce portfolio long / short weights we achieve this by using non overlapping voters classifier with random forest as the base estimator then after training the model on previous 3 years of data of Nifty 500 Universe we generated alpha factors next year's data and then we used a statistical risk model (PCA Based) to optimize the entire portfolio's weights.

1.1 Factor investing

The return of an asset is a function of the financial investment volatility or risk associated with that. For example, an equity investment implies assuming the business risk of a company, and a bond investment implies assuming the risk of default. Alpha factors are designed to derive signals from data to forecast returns of assets over the trading period for a given investment universe. When evaluated, a factor assumes a single value for each asset, but can combine one or more input variables. The Trading

Strategy workflow research phase includes the design, assessment, and combination of alpha factors.

Machine Learning plays a large role in this process because the complexity of factors has increased as investors react to both the signal decay of simpler factors and the much richer data available today.

Alpha factors are transformations of market, fundamental, and alternative data that contain predictive signals. They are designed to capture risks that drive asset returns. One set of factors describes fundamental, economy-wide variables such as growth, inflation, volatility, productivity, and demographic risk. Another set consists of tradeable investment styles such as the market portfolio, value-growth investing, and momentum investing. There are also factors that explain price movements based on the economics or institutional setting of financial markets, or investor behaviour, including known biases of this behaviour. The economic theory behind factors can be rational, where the factors have high returns over the long run to compensate for their low returns during bad times, or behavioural, where factor risk premiums result from the possibly biased, or not entirely rational behaviour of agents that is not arbitrated away.

In this paper, we used the following Alpha factors:

- Momentum 1 Year Factor
- Mean Reversion 5 Day Sector Neutral Smoothed Factor
- Overnight Sentiment Smoothed Factor
- 15/30 Days RSI Of Returns
- Regression Against Time to Compute Beta and Gamma

We used the following universal quant factors:

- Stock Volatility 20d, 120d
- Stock Dollar Volume 20d, 120d
- Sector

We used the following features to capture market-wide regimes:

- High and low volatility 20d, 120d
- High and low dispersion 20d, 120d

We used the following date features that might capture trader/investor behavior due to calendar anomalies:

- is_March (End of Financial Year)
- is_April (Beginning of Financial Year)

- weekday
- month_end (End of Month)
- month_start (Start of Month)
- qtr_end (End of Quarter)
- qtr_start (Start of Quarter)

1.2 Ensemble Method

Decision trees are not only useful for their transparency and interpretability but are also fundamental building blocks for much more powerful ensemble models that combine many individual trees with strategies to randomly vary their design to address the overfitting and high variance problems discussed in the preceding section.

Ensemble learning involves combining several machine learning models into a single new model that aims to make better predictions than any individual model. More specifically, an ensemble integrates the predictions of several base estimators trained using one or more given learning algorithms to reduce the generalization error that these models may produce on their own.

In this paper, we learned how to use two new classes of machine learning models for trading: decision trees and random forests. We saw how decision trees learn rules from data that encode non-linear relationships between the input and the output variables. We illustrated how to train a decision tree and use it for prediction for regression and classification problems such as asset returns and price moves. We will also visualize and interpret the rules learned by the model, and tune the model's hyperparameters to optimize the bias-variance trade-off and prevent overfitting.

Then we proceeded to build a long-short trading strategy that uses a Non-Overlapping Voters Ensemble with Random Forest as base estimator to generate profitable signals for Indian equities over the last 4 years. We will source and prepare the stock price data, tune the hyperparameters of a Random Forest model, and backtest trading rules based on the models' signals. The resulting long-short strategy uses machine learning rather than the cointegration relationship to identify and trade baskets of securities whose prices will likely move in opposite directions over a given investment horizon.

1.3 Portfolio Optimization

Machine Learning is about optimizing objective functions. In algorithmic trading, the objectives are the return and the risk of the overall investment portfolio, typically relative to a benchmark (which may be cash, the risk-free interest rate, or an asset price index like the Nifty and Sensex).

Mean-Variance optimization solves for the optimal portfolio weights to minimize volatility for a given expected return, or maximize returns for a given level of volatility. The key requisite input is expected asset returns, standard deviations, and the covariance matrix.

A statistical risk model has its factors constructed using principal component analysis (PCA), which ensures that those factors have the maximum explanatory power by processing asset return time series. This technique dynamically selects factors based on the maximum commonality among asset returns rather than constraining the model on a set of pre-defined factors. To elaborate further, the use of the PCA technique enables the formulation of statistical model factors (principal explanatory component, in this case) by clustering securities in sets in order to maximize asset return correlation within the cluster. At the same time, the clustered securities will have negligible correlations with the rest of the securities' returns, thus enabling the derived factors to capture maximum risk. Mathematically, both fundamental and statistical risk models begin with the same linear factor model of asset returns:

$$R = Bf + u \dots\dots\dots (1)$$

R is a vector of asset returns, B is a matrix of factor exposures, f is a vector of factor returns, and u is a vector of asset-specific, idiosyncratic returns. While R is known, fundamental and statistical risk models approach the solution of the rest of the terms in this equation differently. While fundamental models have the factors and their exposures (B) given, the equation is solved for the factor return, f, using regression; in macroeconomic models, the returns are available while the factor exposures are estimated. However, for statistical risk models, both the matrix of factor exposures, B, and the vector of factor returns, f, are solved for simultaneously to maximize the predictive power of the above equation. In a nutshell, the factors of the statistical model showcase high adaptability of the factors, especially relevant in times when the market is predominantly driven by unexpected factors or extreme events.

2. Results

Information Analysis

	1D
IC Mean	0.058
IC Std.	0.174
Risk-Adjusted IC	0.333
t-stat(IC)	4.659
p-value(IC)	0.000
IC Skew	0.158
IC Kurtosis	-0.436

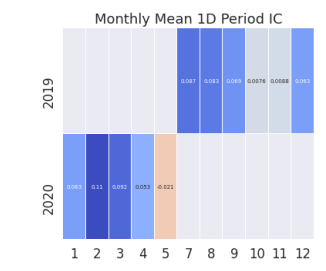
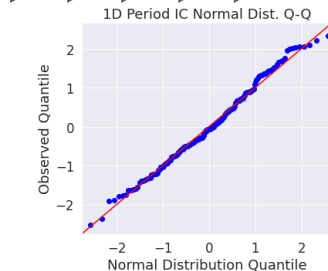
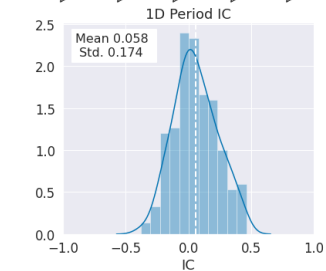
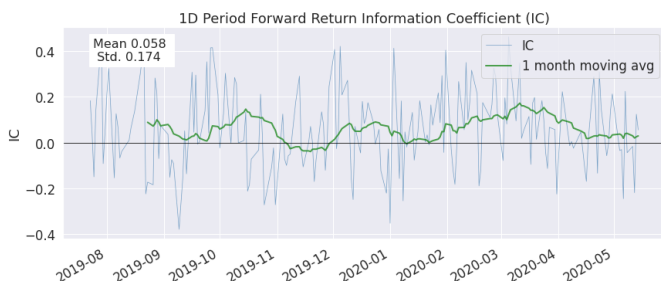
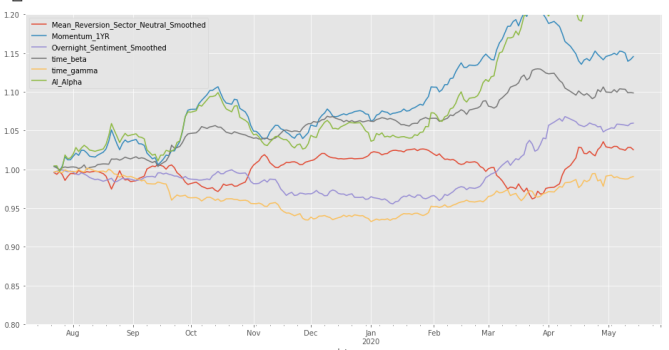
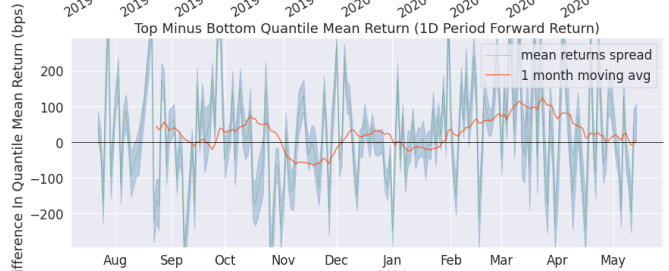
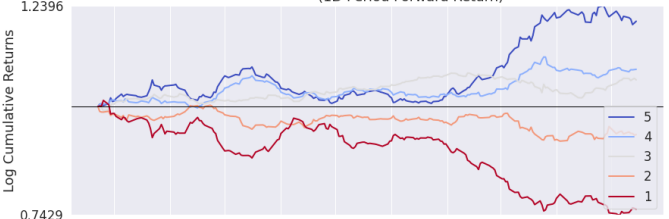
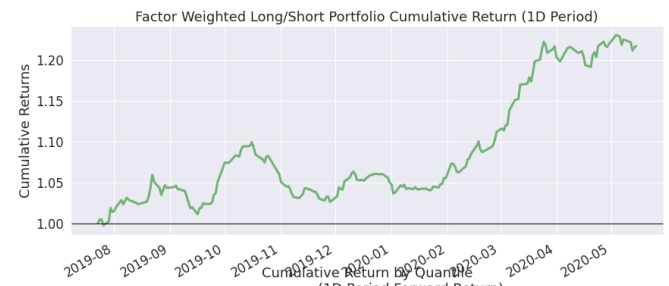
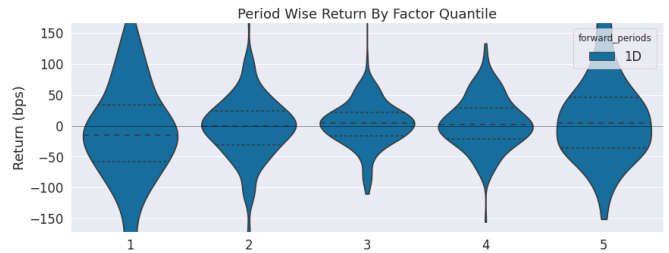
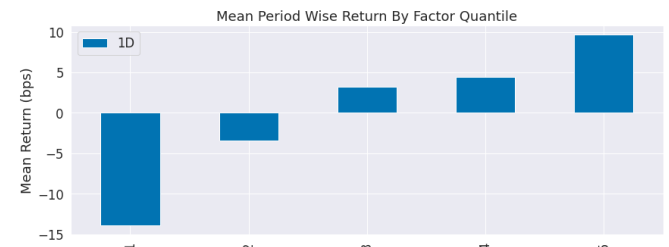
Returns Analysis

	1D
Ann. alpha	0.251
beta	-0.127
Mean Period Wise Return Top Quantile (bps)	9.638
Mean Period Wise Return Bottom Quantile (bps)	-13.871
Mean Period Wise Spread (bps)	23.510

Sharpe Ratios

Mean_Reversion_Sector_Neutral_Smoothed	0.55
Momentum_1YR	2.16
Overnight_Sentiment_Smoothed	1.49
time_beta	2.56
time_gamma	-0.21
AI_Alpha	2.76
dtype: float64	

Despite significant difference between other Alpha factors our AI Alpha is able to generate a good return as compared to other Alphas. The Mean Period Wise Returns are evenly distributed across the Factor Quantile. Also, the Sharpe Ratio of our AI Alpha is Better than other Alpha Factors from which it is made off.



3. CONCLUSION

Despite using some commonly used alpha factors we were able to get good performance on the testing set i.e. the portfolio's return was nearly 20% of the capital used to trade and we got a good Sharpe ratio of 2.76 also the mean period wise spread is much higher which is a trait of a good Alpha Factor. Despite when the Indian equity markets were falling during the test period this portfolio was able to generate significant returns.

ACKNOWLEDGEMENT

We wish to thank all the people whose assistance was a milestone in the completion of this project. I wish to express my sincere appreciation to my supervisor, Professor Vishnu Verma, who has the substance of a genius: he convincingly guided and encouraged us to be professional and do the right thing even when the road got tough. Without his persistent help, the goal of this project would not have been realized. The smooth functioning and every resource at our disposal are the blessings from our Head of Department Professor Sourabh Yadav (Department of Computer Science & Engineering). Our Head of Institute Dr. B S Chawla Sir (Principal, Government Engineering College, Bilaspur) has provided us the much-needed motivation for this research to be fruitful and effective.

REFERENCES

- [1] [McKinsey on Artificial Intelligence](<https://www.mckinsey.com/featured-insights/artificial-intelligence>)
- [2] [Big Data and AI Strategies](<http://valuesimplex.com/articles/JPM.pdf>), Kolanovic, M. and Krishnamachari, R., JP Morgan, May 2017
- [3] [Quantifying Trading Behaviour in Financial Markets Using Google Trends](<https://www.nature.com/articles/srep01684>), Preis, Moat and Stanley, Nature, 2013
- [4] Dissecting Anomalies by Eugene Fama and Ken French (2008)
- [5] Explaining Stock Returns: A Literature Review by James L. Davis (2001)
- [6] Market Efficiency, Long-Term Returns, and Behavioral Finance by Eugene Fama (1997)
- [7] The Efficient Market Hypothesis and It's Critics by Burton Malkiel (2003)
- [8] The New Palgrave Dictionary of Economics (2008) by Steven Durlauf and Lawrence Blume, 2nd ed.
- [9] Anomalies and Market Efficiency by G. William Schwert²⁵ (Ch. 15 in Handbook of the- "Economics of Finance", by Constantinides, Harris, and Stulz, 2003)
- [10] Investor Psychology and Asset Pricing, by David Hirshleifer (2001)
- [11] The Statistics of Sharpe Ratios, Andrew Lo, Financial Analysts Journal, 2002

- [12] Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk by Richard Grinold and Ronald Kahn, 1999
- [13] How to Use Security Analysis to Improve Portfolio Selection, Jack L Treynor and Fischer Black, Journal of Business, 1973
- [14] Portfolio Constraints and the Fundamental Law of Active Management, Clarke et al 2002
- [15] The Capital Asset Pricing Model: Theory and Evidence, Eugene F. Fama and Kenneth R. French, Journal of Economic Perspectives, 2004

BIOGRAPHIES



Mr. Siddhant Sarkar perusing Bachelor of Engineering with major in Computer Science and Engineering from Government Engineering College Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekananda Technical University, Bilai (C.G) in 2020. Good with Quantitative Finance, Deep Learning and Data Science.



Mr. Swarnim Shukla perusing Bachelor of Engineering with major in Computer Science and Engineering from Government Engineering College Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekananda Technical University, Bilai (C.G) in 2020. Good With Machine Learning And Python programming.



Mr. Sourabh Patel perusing Bachelor of Engineering with major in Computer Science and Engineering from Government Engineering College Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekananda Technical University, Bilai (C.G) in 2020. Good with C++ and Python programming.



Mr. Rahul Painkra perusing Bachelor of Engineering with major in Computer Science and Engineering from Government Engineering College Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekananda Technical University, Bilai (C.G) in 2020. Good with C++ programming.