Executive Summary: Quantitative Portfolio Management For Statistical Arbitrage

Introduction

The future of relative stock prices may be glimpsed in certain datasets. We utilise a suite of algorithms and research processes to harness and refine information from these datasets to predict relative price movements. Our approach thrives on scale, which enhances performance and builds a protective moat against competitors. Furthermore, a fully automated pipeline ensures costly human capital can focus on innovation and refinement.

Dataset Acquisition

To uncover potential predictors of stock prices, we first seek datasets with comprehensive coverage and those with extensive historical records. We run our search algorithms on samples of datasets to assess the likelihood and quality of alpha discovery. Due to their inherent complexity and the significant costs associated with both acquisition and processing, we exert careful discretion in the procurement of each dataset.

Search Algorithms

Upon dataset acquisition, our search algorithms, ranging from formula-based to machine learning models are employed. These algorithms derive mathematical models or "alphas" from the datasets. Alphas are essentially weak predictors relying on a couple of data fields to forecast relative future prices. Search algorithms are only allowed to utilise 80% of available historical data to find good alphas. Good alphas meet criteria such as working in large universe sizes, low exposure to known risk factors, high sharpe ratios, low correlation to production alphas, robust to different universe sizes and low turnover. Good search algorithms have high yield, produce highly diverse alphas with high out-sample performance and low complexity, and have a low correlation to other production search algorithms.

Validation Processes

Given the risk of overfitting, our alphas undergo stringent checks to increase the probability of robustness. A designated hard validation period, spanning 20% of our backtesting history, is kept separate from the search algorithms. Before good alphas are productionized, they have to pass validation processes. There are multiple validation processes, as different processes imprint different strengths and weaknesses on alpha production. It is thus possible for an alpha to pass several validation processes but it only has to pass one to be added to the production pool. A simple example of a validation process checks if an alpha sets a new high in the validation period and also raises the sharpe of an equal weighted combialpha of all production alphas.

Hierarchical Ensembles

By aggregating multiple alphas, we construct more powerful predictors known as "combialphas." These ensembles, which can emerge from clusters of alphas in a dataset or shared statistical behaviours, help reduce transaction costs by crossing alpha signals and improve predictability by aggregating uncorrelated signals. Combialphas reflect predictions from an ensemble of alphas that represent specific viewpoints or "views." Strategies are ensembles of alphas or combialphas that focus on specific objectives (e.g. max sharpe, min vol, hedge drawdown) while adhering to real-world trading limitations. Each strategy is equipped with a trading algorithm that determines its trading pattern to get to its optimal portfolio. Given the no free lunch theorem, there's no universally optimal method for combialpha and strategy construction, hence we have to continually innovate towards hierarchical ensemble structures that excel in given objectives while uniquely contributing the production combialpha and strategy pools. Our final portfolio consists of capital allocation towards these strategies, seeking to maximise our risks and returns given their empirical performance.

Conclusion

Our approach with statistical arbitrage allows our performance to increase with scale. As we expand in alphas, datasets and search algorithms, our scale acts as a moat against competitors. Scale also lets us use signals more effectively and have new search algorithms reduce the costs associated with subsequent datasets. Lastly, our comprehensive automation allows our team to focus on continuous innovation and this creates a distinct competitive advantage.