

# Quantitative Portfolio Management For Statistical Arbitrage

## Introduction

Certain datasets provide a glimpse of the future of relative stock returns. We utilise a suite of algorithms and research processes to refine and harness 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.

## Datasets

We seek datasets with comprehensive coverage and those with extensive histories. We run search algorithms on samples to assess the likelihood of discovering good models. 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. Once we have acquired a dataset, our goal is to extract all useful information in a dataset, by transforming each dataset into multiple new datasets that contain information orthogonal to the original dataset.

## Search Algorithms

Our search algorithms, ranging from formula-based to machine learning models are employed on our datasets. These algorithms derive models from the datasets. Each model is a weak predictor relying on a few data fields to forecast relative future returns. Search algorithms are only allowed to utilise 80% of available historical data to find good models. Good models meet criterias such as working in large universe sizes, low exposure to known risk factors, high sharpe ratios, low correlation to production models, robust to universe sizes, low turnovers and high capacities. Good search algorithms have high yields, produce highly diverse models with high out-of-sample performance and low complexity, and have a low correlation to other search algorithms.

## Validation Process

Given the risk of overfitting, our models 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 models are productionized, they have to pass the validation process. The purpose of a validation process is to validate that the model is able to generalise into an out-of-sample period, and that it value-adds to the production pool of models. An example of a validation process checks if a model has an OS/IS sharpe greater than some threshold, sets a new high in the validation period and also raises the sharpe of an equal weighted ensemble of all production models in the validation period.

## Hierarchical Ensembles

By aggregating multiple models, we construct more powerful predictors, which we will term 'ensembles'. These ensembles help reduce transaction costs by crossing model signals and improve predictability by aggregating uncorrelated model signals. They also reflect predictions that represent specific viewpoints. Strategies are ensembles of ensembles 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 ensemble and strategy construction, hence we have to continually innovate towards ensemble structures that excel in given objectives while uniquely contributing to the production pools. Our portfolio consists of capital allocation to these strategies, seeking to maximise our objectives given their empirical performance.

## Conclusion

Our approach with statistical arbitrage allows our performance to increase with scale. As we expand in models, 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.