Statistical Arbitrage

MS&E 448 Midterm Presentation

Shane Barratt Russell Clarida Mert Esencan Francesco Insulla
Cole Kierszनowski Andrew Perry[1]

Stanford University

May 5, 2020

[1] Authors listed in alphabetical order

Overview

1. Background
2. Data
3. Preliminary Implementation
4. Next Steps

Background

- Statistical arbitrage = short-term trading strategy that bets on mean-reversion of asset baskets (more later)
- The intuition of statistical arbitrage is based on the idea that the difference between what an equities' price is and what it should be is driven by idiosyncratic shocks
- Statistical arbitrage requires 3 steps:
  1. Finding asset baskets
  2. Prediction based on mean-reversion
  3. Portfolio construction

Market–Neutral Investments

- n assets, with prices $p_t \in \mathbb{R}^n_+$ at time period $t = 1, \ldots, T$
- Assume assets are hedged w.r.t. market, *i.e.*, each asset is actually 1 unit of the asset and $-\beta$ units of the market, where β is the correlation of the market returns with the asset returns
- Observation: **Investing (long or short) in any of these assets is market-neutral**

Finding Pairs

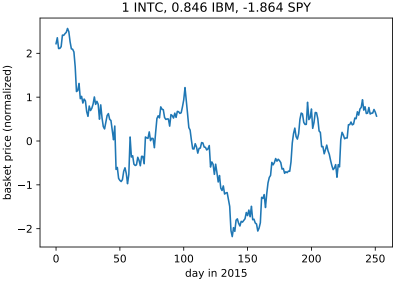
- For each pair of assets, consider the asset basket where we invest in 1 unit of asset i and θ units of asset j
- Choose θ as solution to the problem

$$\text{minimize} \quad \sum_{i=1}^{T}((p_t - \mathbf{E}[p_t])_i + \theta(p_t - \mathbf{E}[p_t])_j)^2,$$

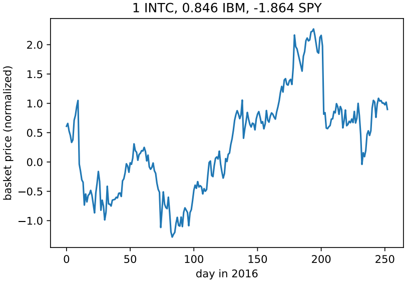
  with variable $\theta \in \mathbb{R}$.
- Results in $n(n-1)/2$ baskets (for S&P 1500, $\approx$ 1.1 million baskets).

Example: Finding Pairs

- Consider INTC hedged (1 INTC, -1.006 SPY) and IBM hedged (1 IBM, -1.014 SPY), in 2015
- Solving problem on previous slide gives $\theta = 0.846$, so final asset basket is (1 INTC, 0.846 IBM, -1.864 SPY)



What Happened in 2016?



Ornstein–Uhlenbeck Process

- We model the residual returns as a mean-reverting processes the Ornstein–Uhlenbeck (OU) process [1]
- The OU process is given by

$$dX_t = \rho(\mu - X_t)dt + \sigma dB_t, \quad \rho, \sigma > 0 \tag{1}$$

  where ρ is the speed of mean reversion, μ is the long run average, σ is the instantaneous volatility, and B_t is a standard Brownian motion
- We allow for the assumption that over a short trading period that ρ, μ, and σ stay constant

AR(1) Process

- To determine the constants ρ, μ, and σ in the OU process we use an auto-regressive process with a lag of one (AR(1) process)
- An AR(1) process is given by

$$X_{t+1} = \lambda + \phi X_t + \epsilon_t \tag{2}$$

- The interpretation of λ, ϕ, and ϵ as it relates to the OU process is

$$\begin{aligned} \lambda &= \mu(1 - e^{-\rho\delta t}) \\ \phi &= e^{-\rho\delta t} \\ \epsilon_t &\sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}(1 - e^{-2\rho\delta t}\right) \end{aligned} \tag{3}$$

- We fit an AR(1) to all 1 million baskets and select the ones we are most confident about by using a Dickey - Fuller test

Trading Signal

- Now that we are able to model stocks as an OU process we need a dimensionless trading signal
- We will use the distance that X_t is from the mean μ by the long - run standard deviation $\overline{\sigma} = \frac{\sigma}{\sqrt{2\rho}}$. Giving us the signal

$$s_t = \frac{X_t - \mu}{\overline{\sigma}} \tag{4}$$

- We use the signal as follows: given an $s_t \gg 0$ we short S and buy F, when $s_t > 0$ we exit that position. Conversely, when $s_t \ll 0$ we long S and short F, when $s_t < 0$ we exit that position.

Trading Signal

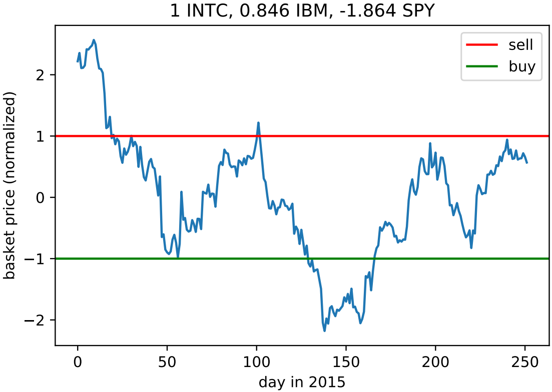


Figure: Trading Signal for previous example.

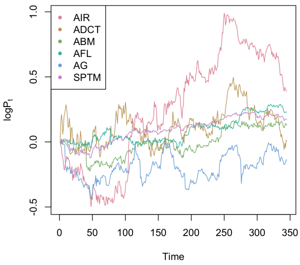
Data

- The data we are using comes from CRSP (Center for Research in Security Prices) accessed through WRDS (Wharton Research Data Services)
- Universe: We decided to use equities that are present in the S&P 1500 every year from 2003 to 2019
- Timescale: Currently we are using daily data to create our statistical arbitrage models, but we will hopefully extend our model to use 30 minute price intervals

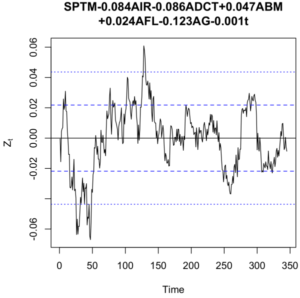
Preliminary Implementation

We constructed a data pipeline that does the following given a selection of stocks:

- Constructs the optimal basket
- Tests the stationarity of the residuals using the Augmented Dickey-Fuller (ADF) test



(a) Log Price



(b) Stationary residuals

Next Steps

- Determine best thresholds for the trading signal
- Extending our model to work with higher frequency data rather than daily data
- Develop various approaches to construct baskets
  - We already have a Greedy Bottom-Up ADF-based implementation
- Produce more residual returns and develop a robust back testing method

Questions

Thank You

References

1. Steven Finch. Ornstein-Uhlenbeck Process by Steven R. Finch. pages 1–14, 2004.
2. Marco Avellaneda et al. Statistical Arbitrage in the U.S. Equities Market, 2009

Appendix A: Portfolio Construction

- Let Σ be the estimated covariance of our predictions
- Basic: Long k baskets with highest μ_i/Σ_{ii}, short k baskets with most negative μ_i/Σ_{ii}

- Sophisticated: Convex Markowitz portfolio optimization

$$\begin{array}{ll} \text{maximize} & \mu^\top x - \frac{\gamma}{2} x^\top \Sigma x \\ \text{subject to} & \mathbf{1}^\top x = 1 \end{array}$$

where $x \in \mathbb{R}^m$ and risk-aversion $\gamma \in \mathbb{R}_+$.

- All portfolios are (by definition) market neutral