Mean-Reverting / Statistical Arbitrage Strategy

Jonathan Tuck Raphael Abbou Vin Sachidananda

Stanford University

May 5, 2020

Co-integrated groupings

- Pairs trading: find a pair of assets S_t, F_t such that:

$$\frac{dS_t}{S_t} = \alpha dt + \beta \frac{dF(t)}{F(t)} + dX_t$$

  where X_t is a mean reverting process.

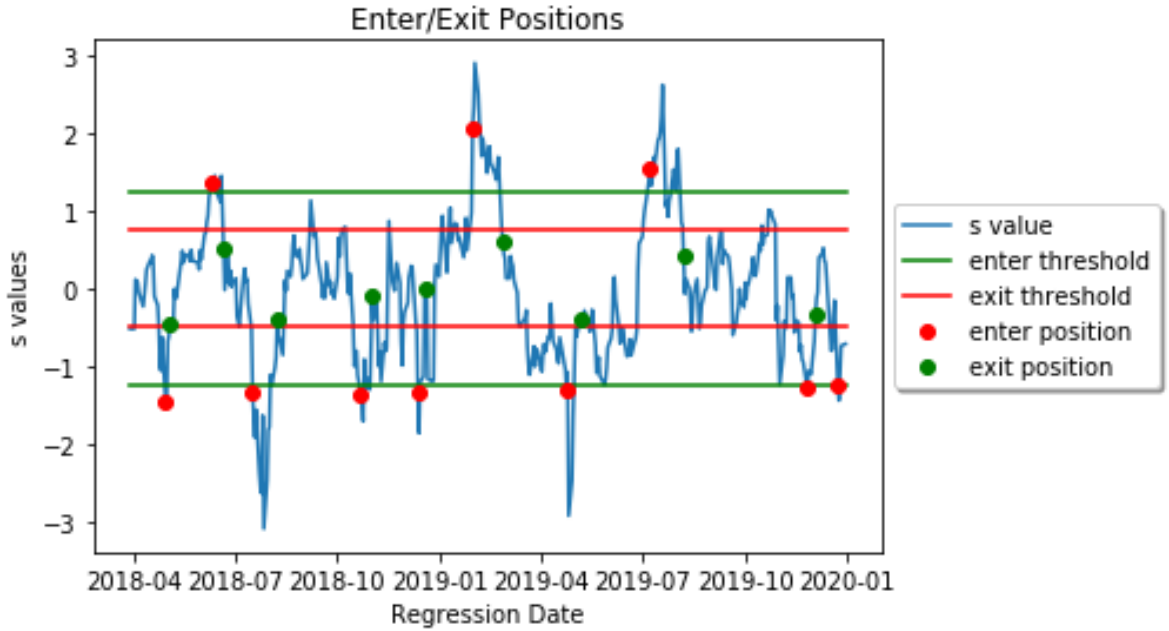
- We want to generalize this strategy to m assets, *i.e.*, find a group of assets $(S^{(i)})_i$ and an allocation $w \in \mathbf{R}^m$ such that

$$\sum_i w_i \frac{dS_t^{(i)}}{S_t^{(i)}} = dX_t$$

- Constraints
  - Max assets with open positions, $\mathbf{card}(w) < n$
  - Max allocation to individual asset, $\|w\|_\infty < k$

Strategy

Large positive (resp. negative) values in our mean reverting process dX_t mean that our basket of stocks is likely to drop and produce negative (resp. positive) returns, and we want to short (resp. long) it.

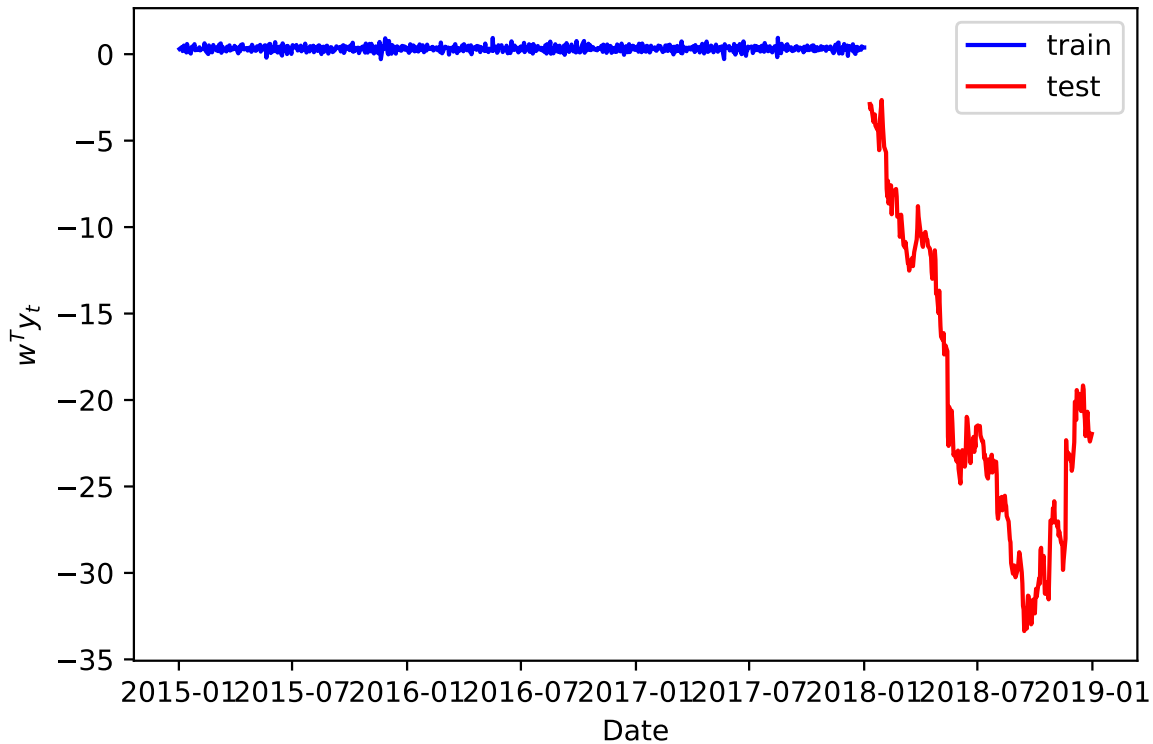


Strategy Roadmap

- Buy when below mean, sell when above mean
- Measure with z-score $z = \frac{s_t - \mu}{\sigma}$,
  - s_t is linear combination of assets
  - μ, σ rolling window samples
  - Window size decided by out-of-sample validation
- Bounds on opening long/short, exiting long/short: $S_{\mathrm{o,l}}, S_{\mathrm{o,s}}, S_{\mathrm{e,l}}, S_{\mathrm{e,s}}$
  - Hyper-parameters
  - Not necessarily symmetric about $z = 0$
- Bet sizing
- Diverse set of strategies, market-neutral

Problem

Naive method badly overfits (perfect in train, completely unusable in test)



Data

- Universe of assets:
  - S&P 500
  - 50 largest cap companies in the US
  - Indices / ETFs
- Exploring tick sizes of 15 minutes, 1 hour, 1 day:
  - Hypothesis: higher frequency $\implies$ more opportunities to enter/exit
  - Sub-selection of stocks to be made in order to reduce overfitting
  - Hyper-parameter to be tuned

Stock groupings

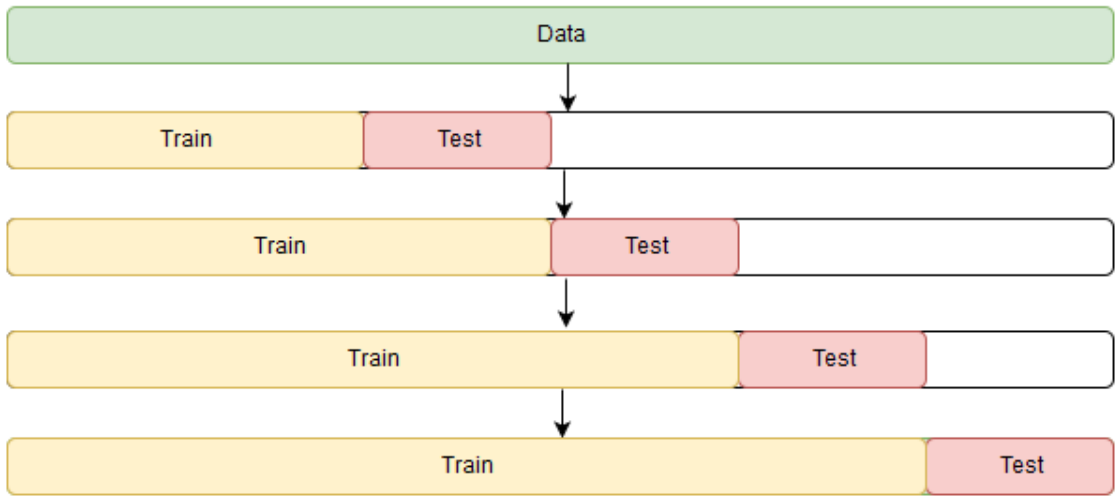
- Groups, not necessarily pairs
- Look at sectors, unsupervised learning methods
- Sparse optimization methods
  - Rounding/Polishing of Regularized (ℓ_1, ℓ_2) solutions
  - Mixed Integer Formulations
- Goal: produce *multiple* methods of finding these groups
  - Hopefully uncorrelated

Criteria

- Drawdown
- Sharpe Ratio
- Overall return
- Rolling portfolio beta

Validation

In order to avoid over-fitting problems, and as we want to take into account the non-stationarity of our data, we develop the following validation scheme to test our model:



Timeline

- Functional pipeline for testing strategies
- Pre-selection of co-integrated pairs
- Solidify strategies
- Hyper-parameter selection