Aspects of machine learning in asset management:
A managed futures perspective

Tobias Rydén

Lynx Asset Management

September 2020

Important Information

This document has been provided to you for information purposes and does not constitute an offer, solicitation, advice or a recommendation to purchase financial instruments or financial services. The factual information and views expressed herein are those of the author and they do not necessarily represent the views of Lynx Asset Management AB.

Past performance is not necessarily indicative of future results.

Outline

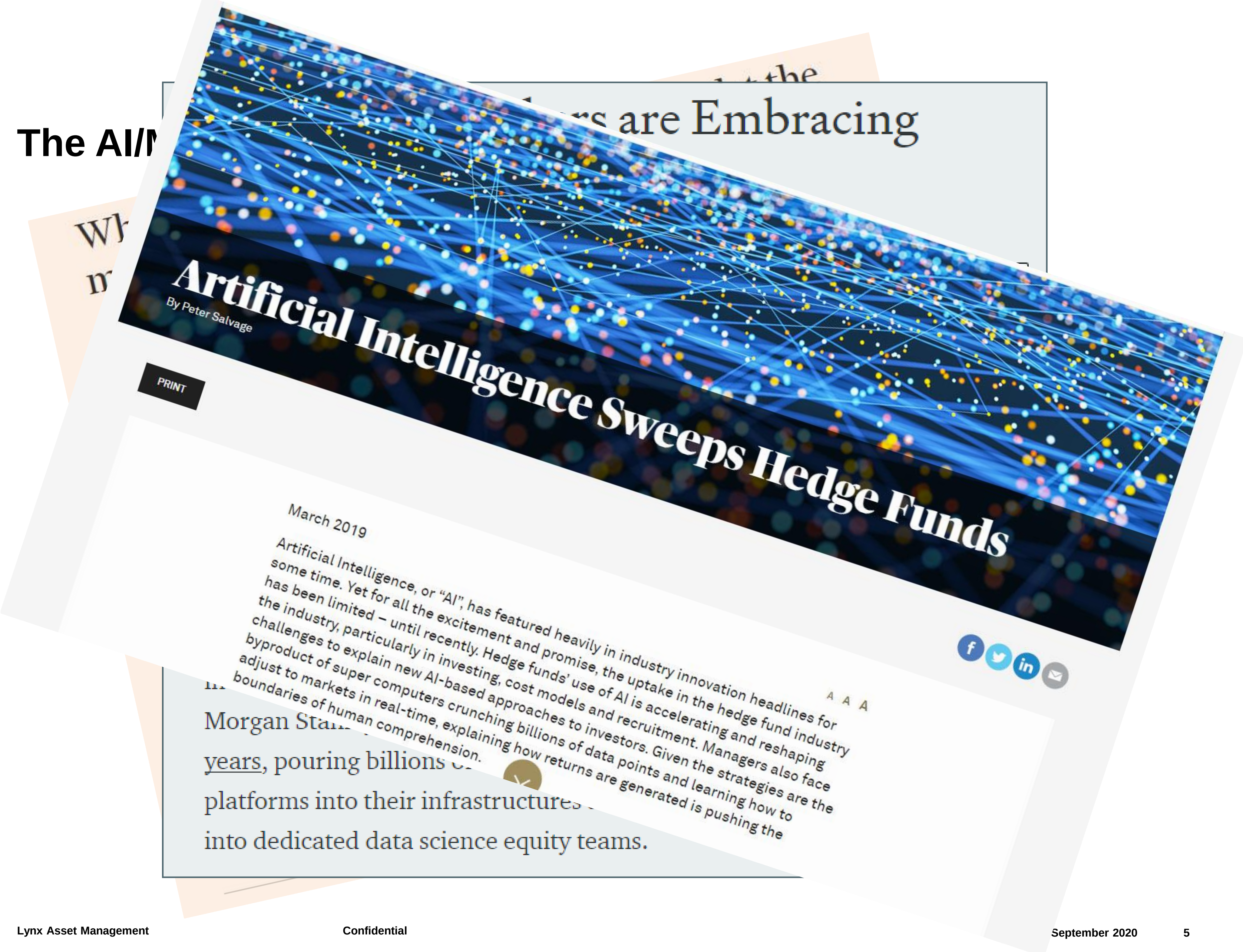
- Industry perspective on AI and ML
- What can we expect from and achieve with ML
- Example: Diversification through ensembling
- Example: An approach to linear filter feature extraction

The AI/ML hype

- **Today, AI and ML are buzz words in as good as any potential area of application**
- **So also in asset management**
- **The terms *big data* (and even *alternative data*), AI and ML are sometimes used almost interchangeably**

- **ML techniques are quantitative, and can be used for any quantitative purposes such as**
  - ✓allocation to asset classes, sectors, …
  - ✓credit risk prediction
  - ✓risk management
  - ✓trading (transactions) of assets
  - ✓market return predictions

The AI/M



Artificial Intelligence Sweeps Hedge Funds

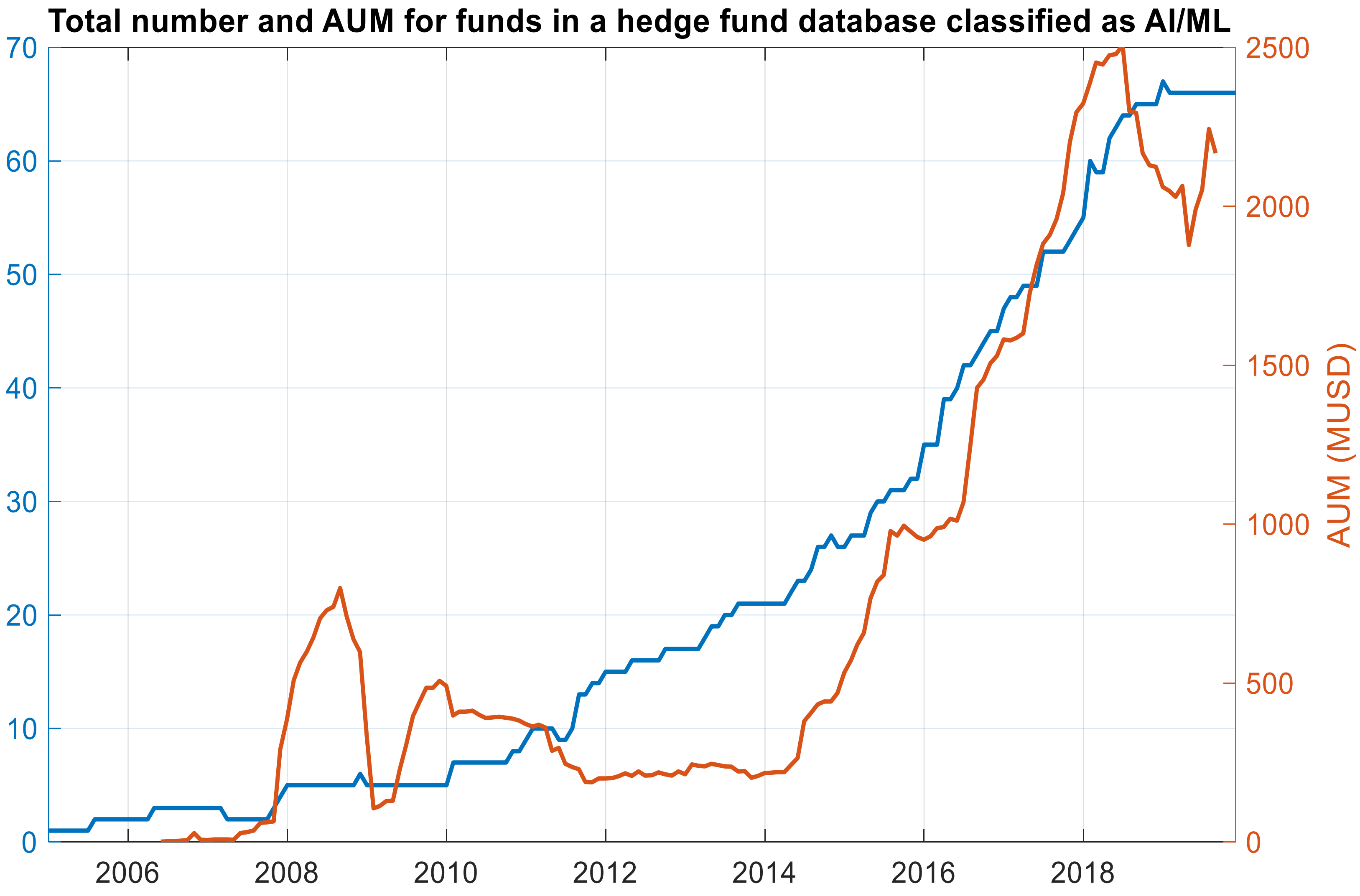
By Peter Salvage

PRINT

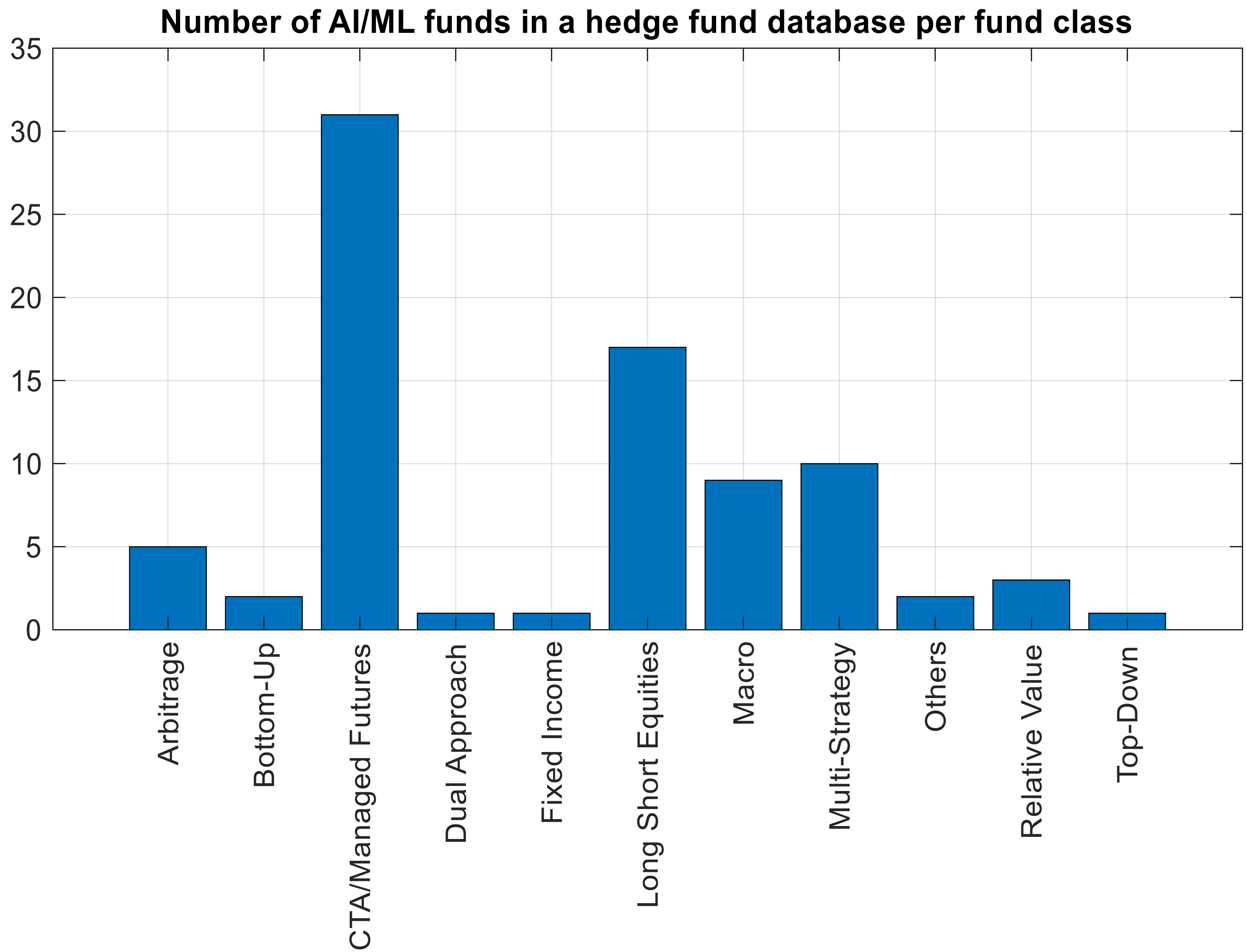
March 2019

Artificial Intelligence, or "AI", has featured heavily in industry innovation headlines for some time. Yet for all the excitement and promise, the uptake in the hedge fund industry has been limited – until recently. Hedge funds' use of AI is accelerating and reshaping the industry, particularly in investing, cost models and recruitment. Managers also face challenges to explain new AI-based approaches to investors. Given the strategies are the byproduct of super computers crunching billions of data points and learning how to adjust to markets in real-time, explaining how returns are generated is pushing the boundaries of human comprehension.

Growth of AI/ML hedge funds



Types of AI/ML hedge funds



The other point of view…

Not everybody is ready to jump onto the bandwagon…

- **Many investors are skeptical, or adverse to, investment strategies that are, in some sense, non-explainable**

ML in market forecasting: challenges

Data availability
in particular for financial time series, macro data

Data quality
(including revisions)

Very low signal-to-noise ratio

Machine learning: "traditional" applications vs. finance

Image classification: which picture is an orange, a bicycle, and a monkey respectively?

Simple task for humans – and for a machine too;
accuracies well above 90%
Such tasks are about *automation*

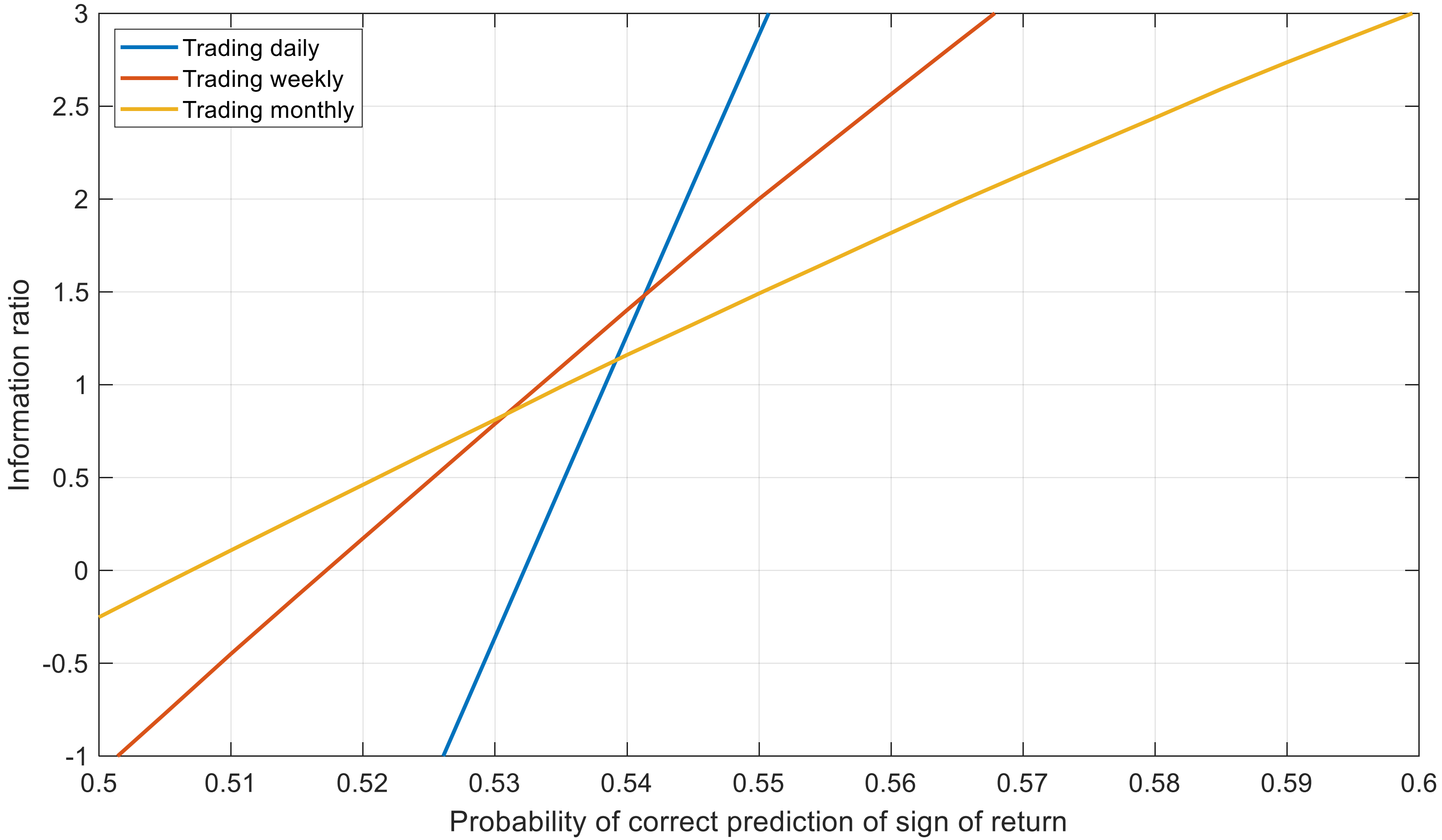
Machine learning: application to financial prediction

Six months of prices (and volumes) for S&P 500; will price go up or down during the next day, week, or month?

Difficult task for humans – and for a machine too; achievable accuracies are barely above 50%

Machine learning: application to financial prediction

The good news is: we *do not need to* achieve accuracies much above 50% !



Trading 90 futures contracts
[26 equity indices,
18 bonds,
29 commodities,
17 currencies]

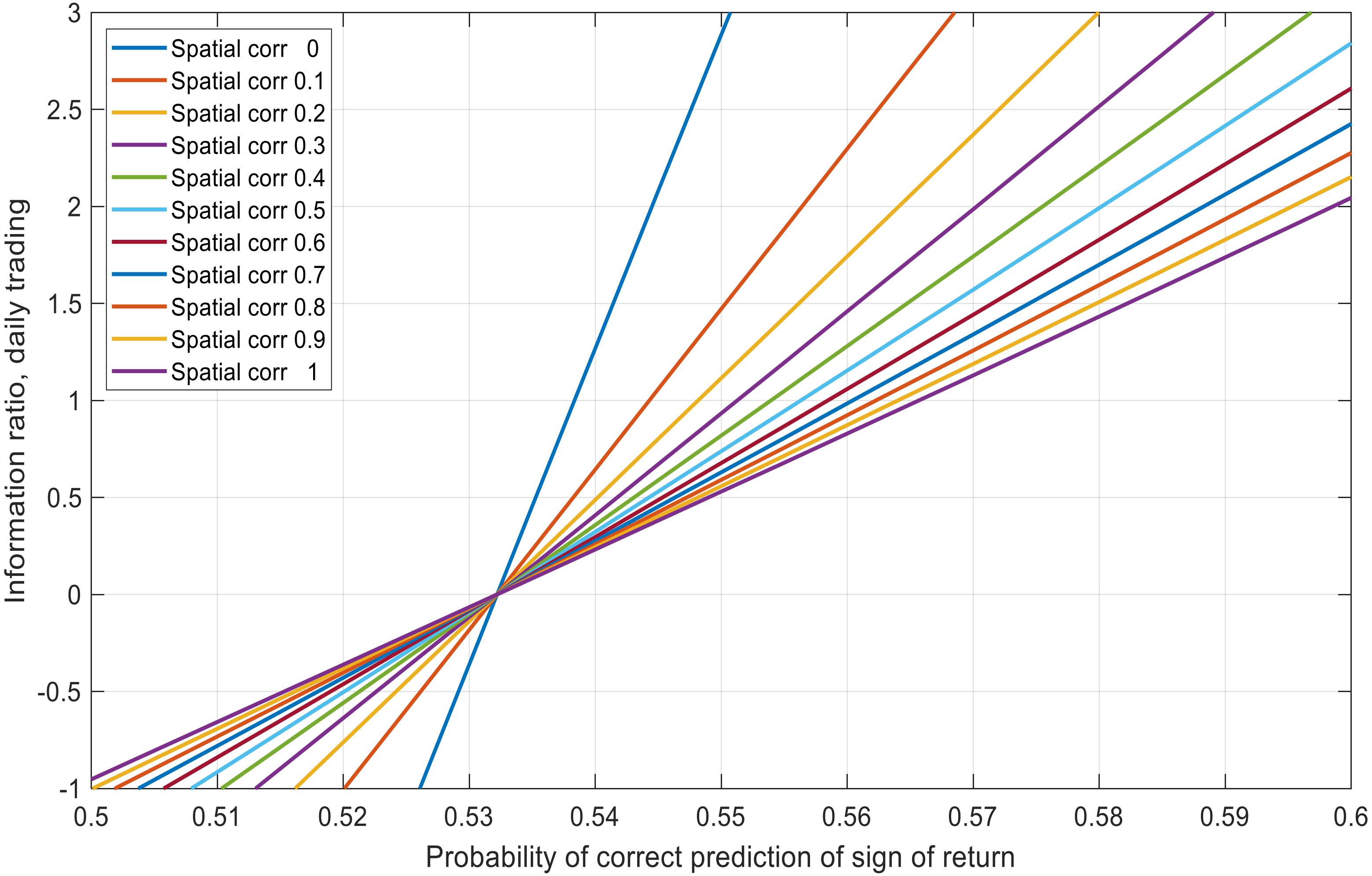
Liquidity-weighted

Realistic trading costs

No temporal correlation. (*Big*) caveat though: plot assumes *no* spatial correlation btw predictions

Machine learning: application to financial prediction

However, correlations can be a headache



Trading 90 futures contracts daily

Liquidity-weighted

Realistic trading costs

From simple strategy to ensemble of prediction models

Starting point: Simple EMA 60 / EMA 20 cross-over model (binary positions)

-- Indicators are EMA's with centres of mass at 60 and 20 days, of normalised returns

-- Positions are +1 (-1) if short EMA is above (below) long EMA

-- Trading 95 futures
-- No trading costs
-- No liquidity concerns

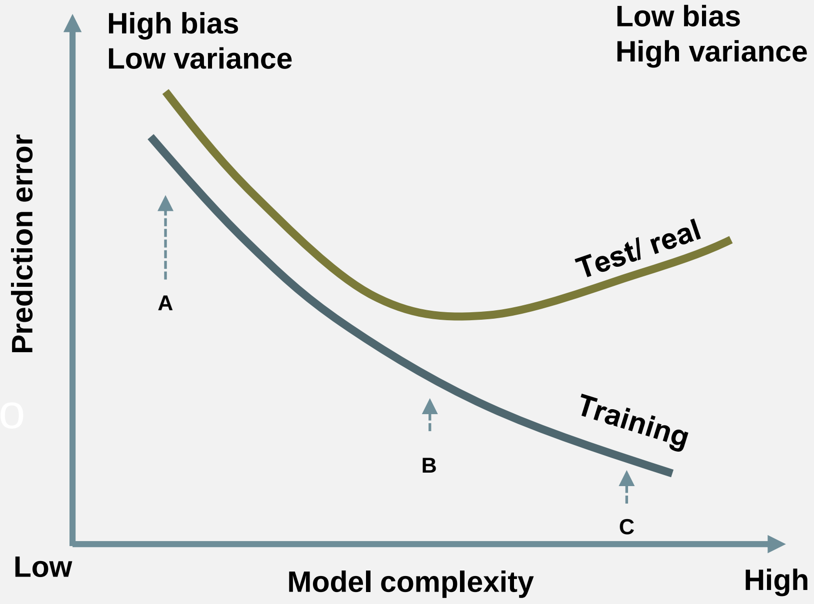
-- IR = 0.36 (ratio btw mean and std dev of daily returns, annualised)

[*EMA: *Exponential moving average*]

From simple strategy to ensemble of prediction models

Interlude: Mitigating the risk of overfitting

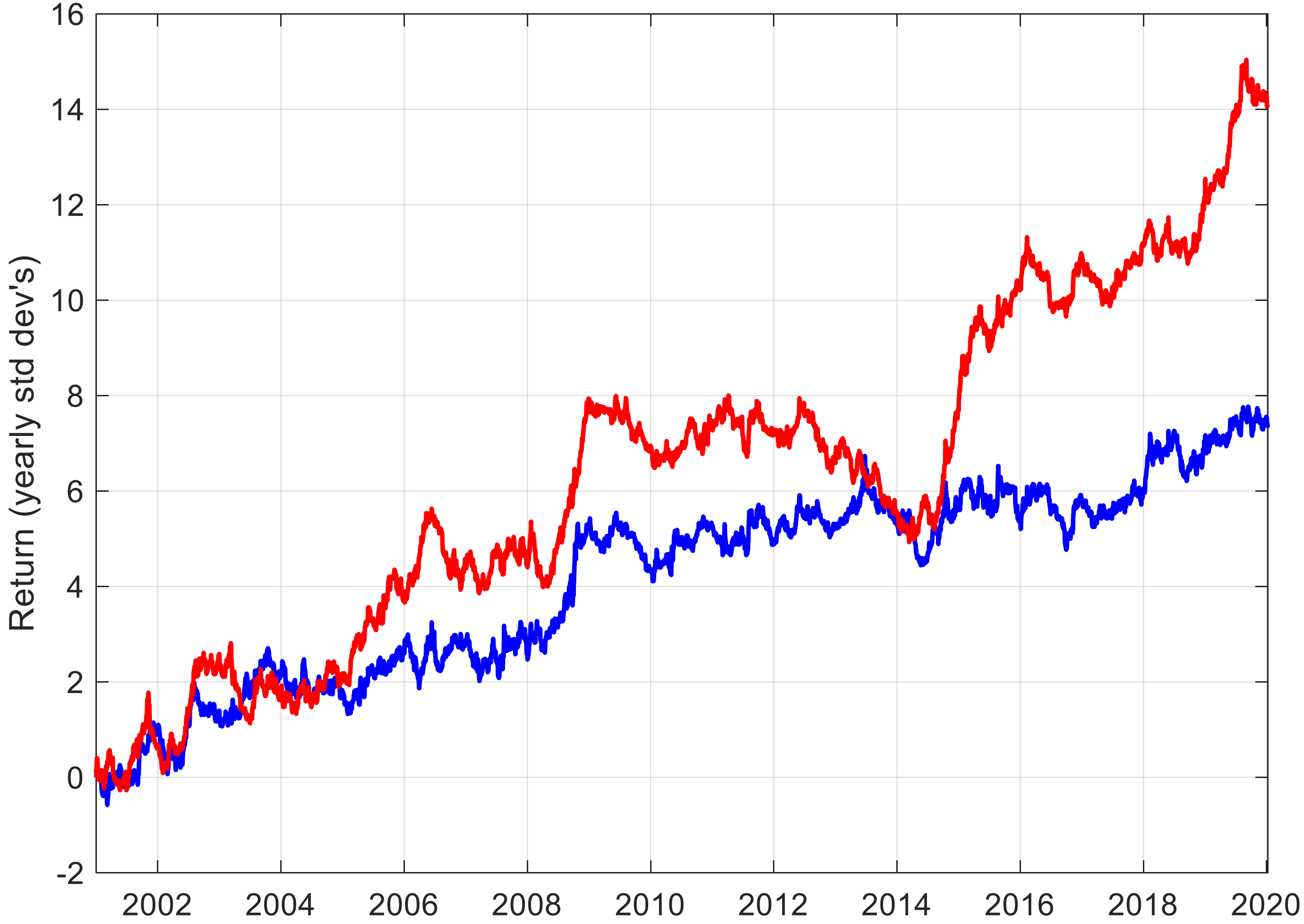
- **Future Selection**
- **Selection of model and complexity**
- **Regularisation**
- **Ensembles**
- **Walk-forward**



Note that ensembles attempt to *decrease variance*, while *maintaining means*.
This is entirely similar to the concept of *diversification* in asset allocation.
Let's see how this works out in the present example!

From simple strategy to ensemble of prediction models

Next attempt: Same EMA indicators, ANN with one hidden layer of 40 nodes, 50% drop-out, MSE predicton error loss for next-day returns, trained once

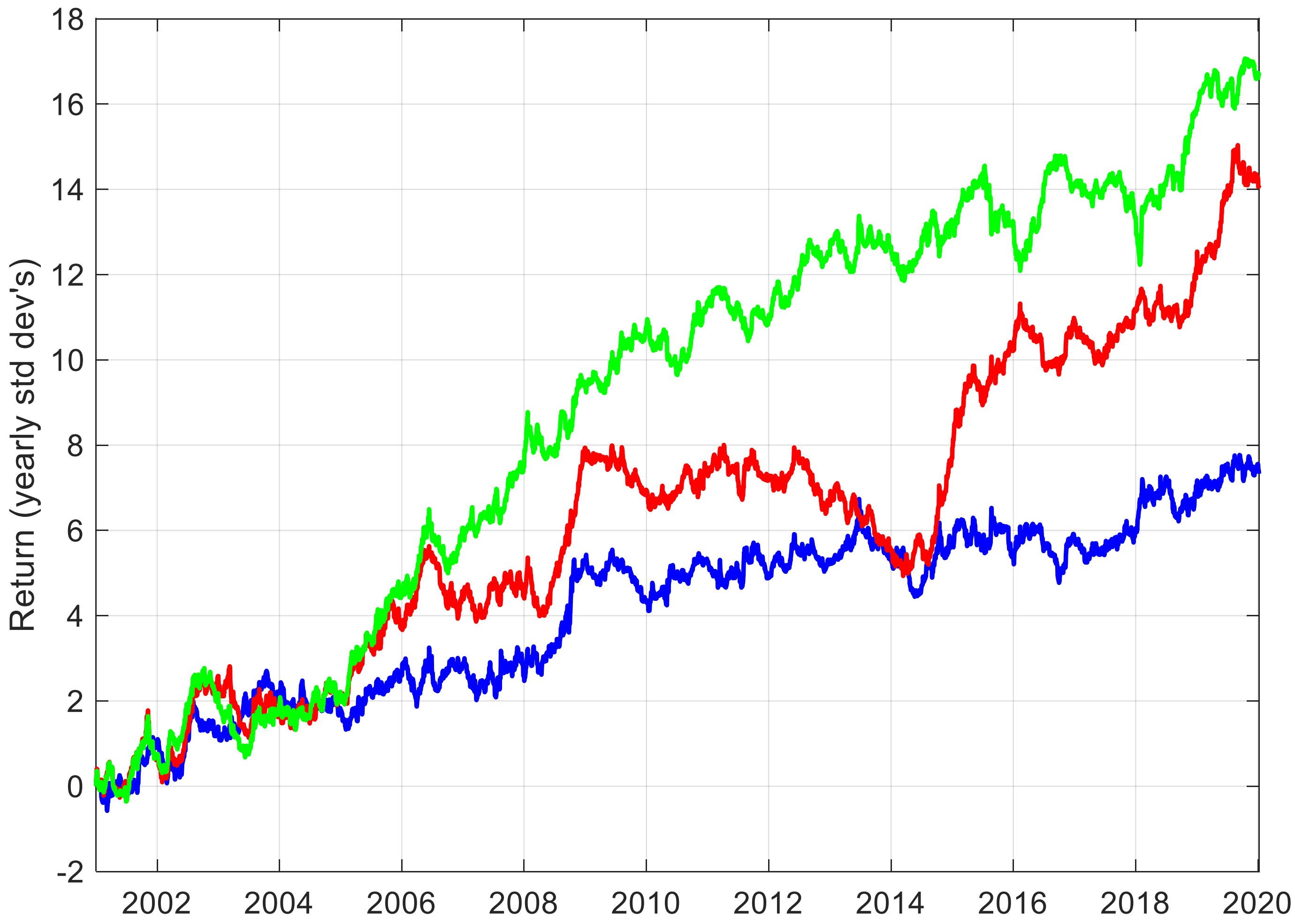


-- Same EMA indicators
-- One hidden layer of 40 nodes
-- Sigmoid activations
-- 50% drop out
-- Loss function is MSE prediction error for next-day returns
-- Trained once, 2000-12-31

-- IR = 0.76

From simple strategy to ensemble of prediction models

Next attempt: Same EMA indicators, ANN with one hidden layer of 40 nodes, 50% drop-out, MSE predicton error loss for next-day returns, retrained once per year



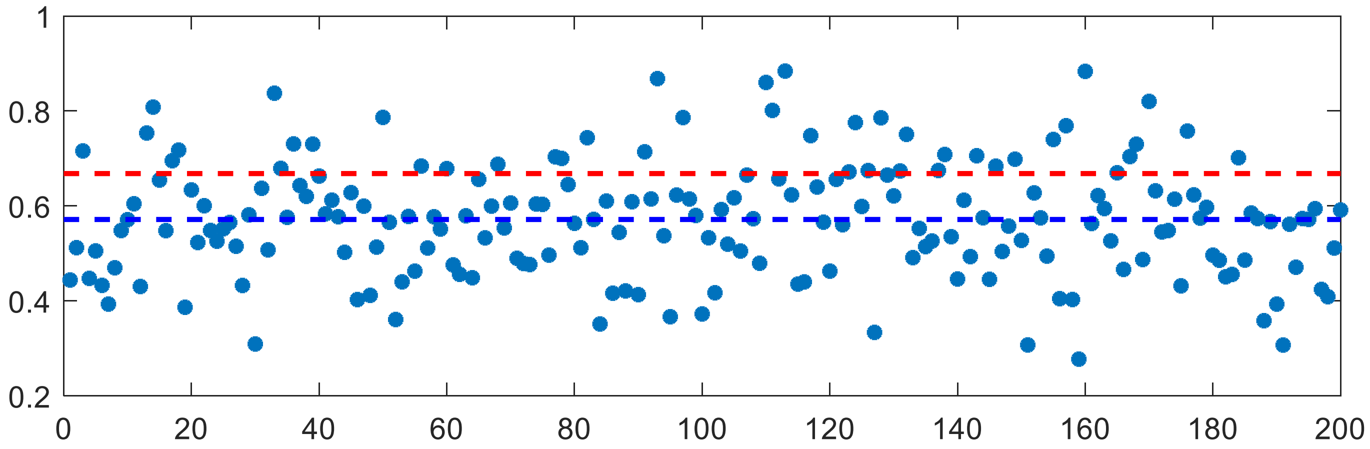
-- Same EMA indicators
-- One hidden layer of 40 nodes
-- Sigmoid activations
-- 50% drop out
-- Loss function is MSE prediction error for next-day returns
-- Retrained once per year

-- IR = 0.77

ANNs are sometimes criticised for their complexity, black-box character, massive amounts of parameters, non-convexity, …
Can we use this to our advantage?

From simple strategy to ensemble of prediction models

Further attempt: Same EMA indicators, ensemble of ANNs diversified over architecture

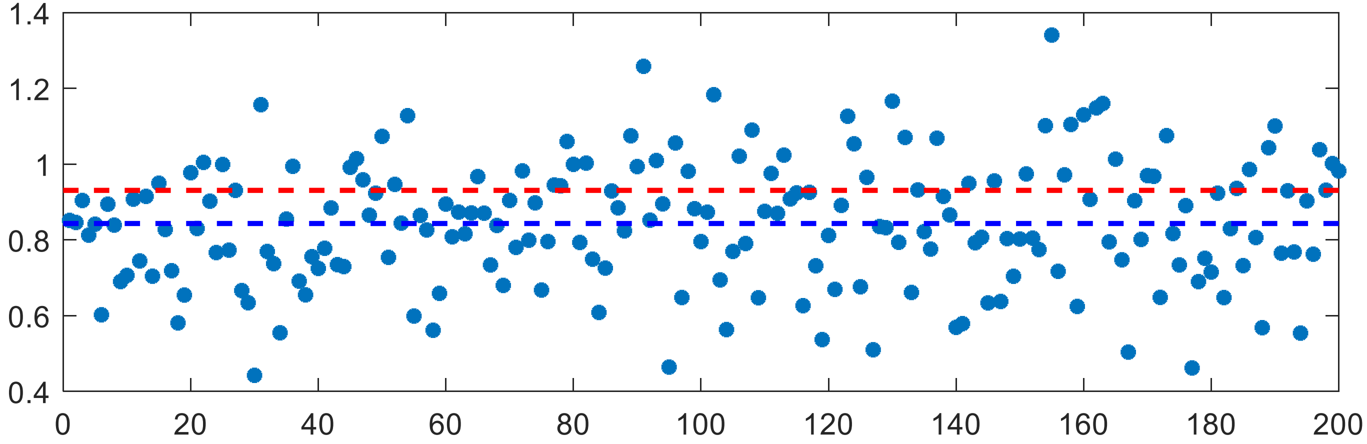


-- Same EMA indicators
-- 50% drop out
-- 0, 1 or 2 hidden layers
-- 10-40 nodes per layer
-- Sigmoid or leaky ReLU activations
-- Loss function is MSE prediction error for next-day returns
-- Retrained once per year
-- Ensemble of 200 instances

-- Average correlation btw instances is 0.71
-- IR = 0.67 (ensemble)

From simple strategy to ensemble of prediction models

Even further attempt: Same EMA indicators, ensemble of ANNs diversified over architecture and markets



-- Same EMA indicators
-- 50% drop out
-- 0, 1 or 2 hidden layers
-- 10-40 nodes per layer
-- Sigmoid or leaky ReLU activations
-- 50% of markets are selected as input, and 50% as output
-- Loss function is MSE prediction error for next-day returns
-- Retrained once per year
-- Ensemble of 200 instances

-- Average correlation btw instances is 0.59
-- IR = 0.93 (ensemble)

An approach to linear feature construction

Smooth prediction filters

- Let $y_k \in \mathrm{R}^m$ be a vector we wish to predict, i.e. a returns per market over some period of time
- Let (x_t) with $x_t \in \mathrm{R}^n$ be a vector time series we have available for the task
- Take the approach $\widehat{y}_{kj} = \sum_i \sum_{s=0}^{d} \beta_{sij}\, x_{t-s,i}$
- The $\beta_{\cdot ij}$ are linear filters; we want them to be smooth, in some suitable sense
- Apply penalties $\lambda \left\| D\beta_{\cdot ij} \right\|_2^2$; for instance, D computes 2nd differences
- x_k is stacked vector $(x_{k-d:k,1}, \dots, x_{k-d:k,n})$ in R^{dn}, B is $dn \times m$ matrix of prediction coefficients
- Then $\widehat{y}_k = B^\top x_k$
- Loss function is

$$\sum_{k=1}^{N} \frac{1}{2} \|y_k - B^\top x_k\|_2^2 = \sum_{k=1}^{N} \mathrm{trace}\left(\frac{1}{2} y_k y_k^\top - y_k x_k^\top B + \frac{1}{2} B^\top x_k x_k^\top B\right) = \mathrm{trace}\left(\frac{1}{2} S_{yy} - S_{yx} B + \frac{1}{2} B^\top S_{xx} B\right)$$

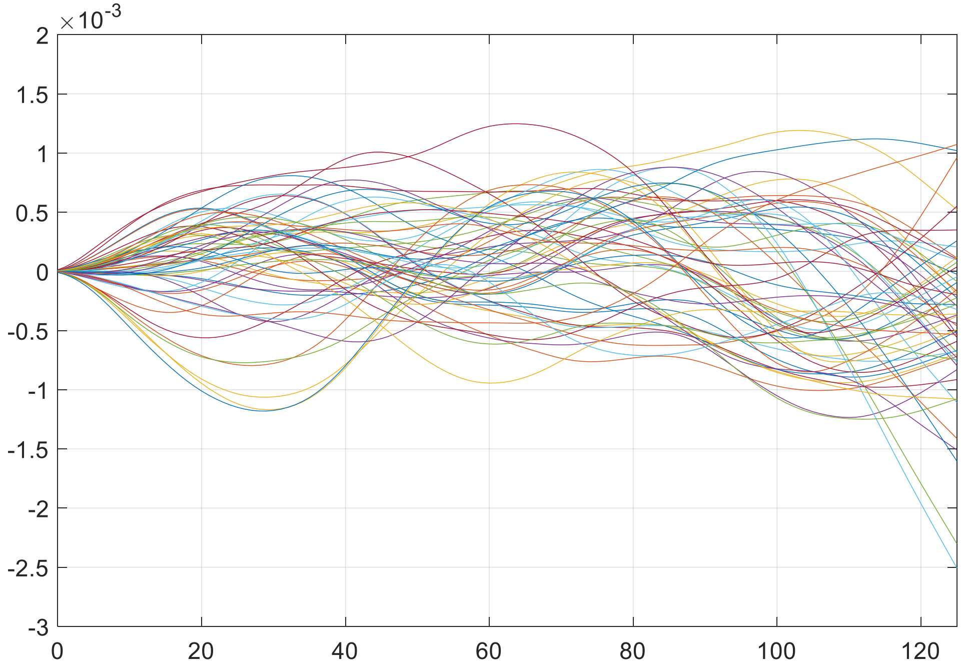
- Total objective (less a constant) becomes $\mathrm{trace}\left(-S_{yx} B + \frac{1}{2} B^\top Q B\right) = -\langle B, S_{xy} \rangle + \frac{1}{2} \left\| Q^{\frac{1}{2}} B \right\|_{\mathrm{F}}^2$
- A change of variables yields objective $-\langle \widetilde{B}, \widetilde{S}_{xy} \rangle + \frac{1}{2} \|\widetilde{B}\|_{\mathrm{F}}^2$
- Can replace $\|\cdot\|_{\mathrm{F}}$ by norm that encourages low rank (nuclear norm or k-support norm)

An approach to linear feature extraction

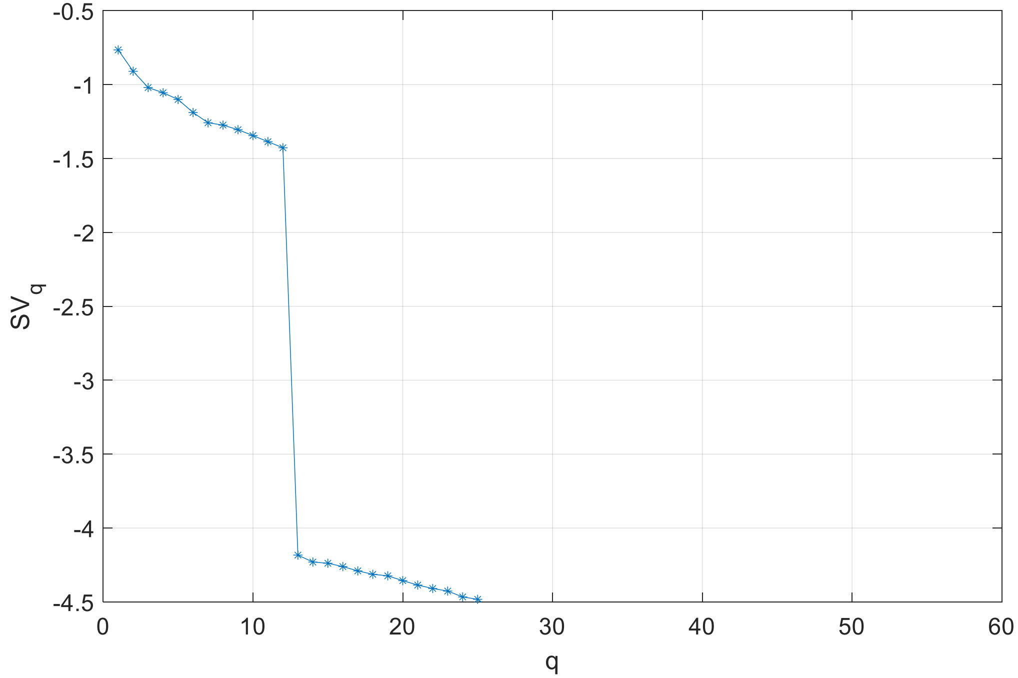
Smooth prediction filters

Solve matrix k-support norm problem $\min_{\tilde{B}} \|\tilde{B}\|_{(k)*}$ given $\langle \tilde{B}, \tilde{S}_{xy} \rangle = 1$ (note: *no* S_{xx} in Q!)

Filters for *one* market



10-log singular values for B ($k = 12$)

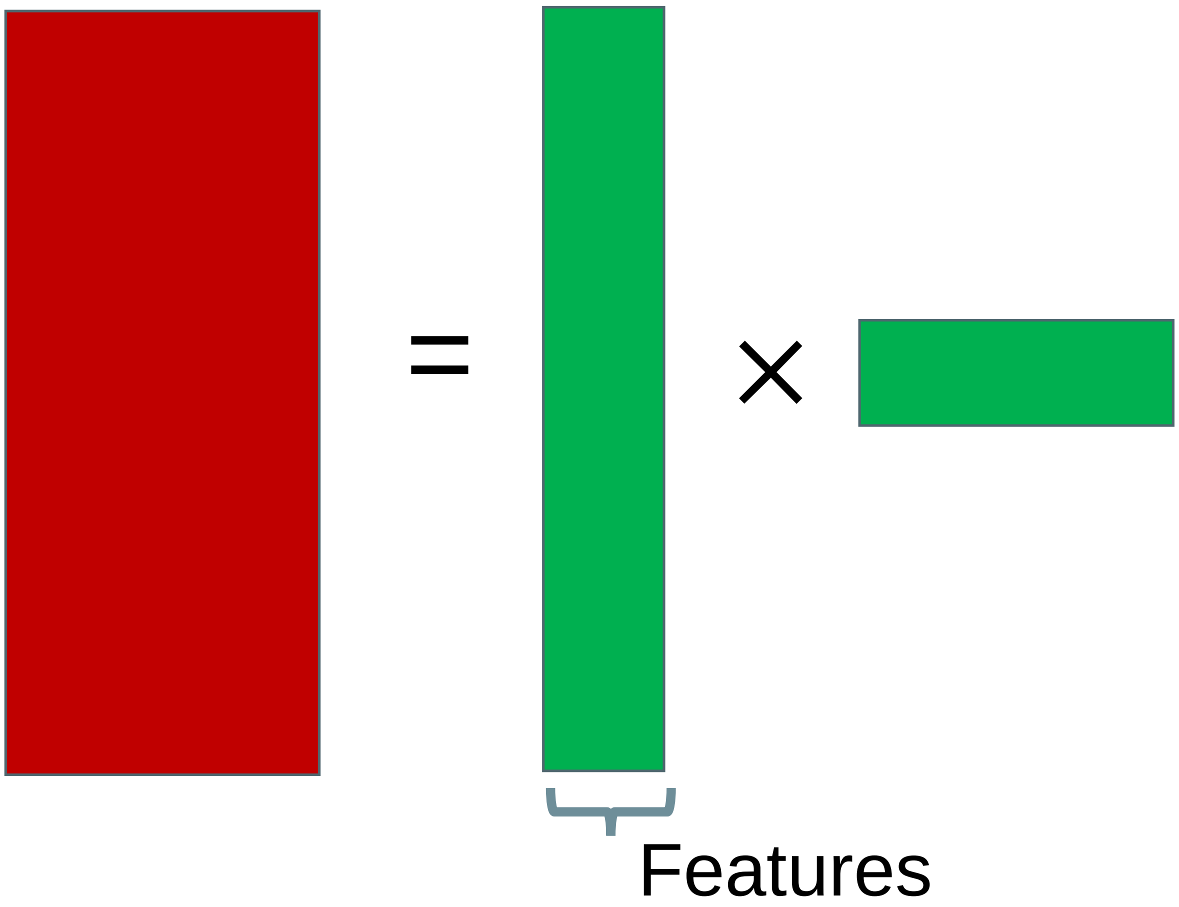


Note 1: k-support norm is similar to 1-norm (cf. Lasso), but encourages more non-zero coeffs.
Note 2: if B is rank defect, we can write $B = B_1 B_2$ with B_1 having k columns (cf. linear ANN)

An approach to linear feature extraction

Smooth prediction filters

If B is rank defect, we can write $B = B_1 B_2$ with B_1 having k columns (cf. linear ANN)



That’s all – thanks for listening!!