Machine Learning in Finance: practice

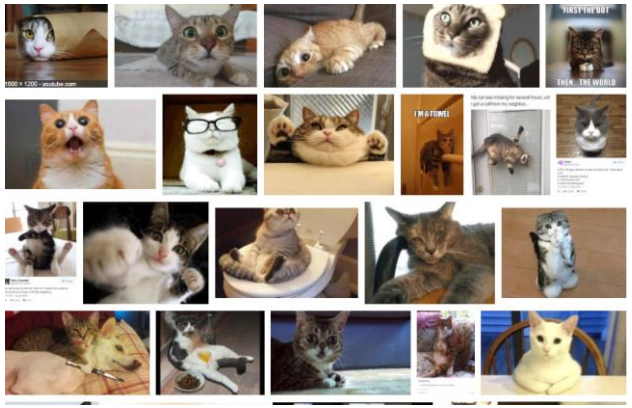
Tony Guida, Senior Investment Manager – Quant PM

What this presentation is NOT about

AlphaGo

#AlphaGo WINS!!!! We landed it on the moon. So proud of the team!! Respect to the amazing Lee Sedol too

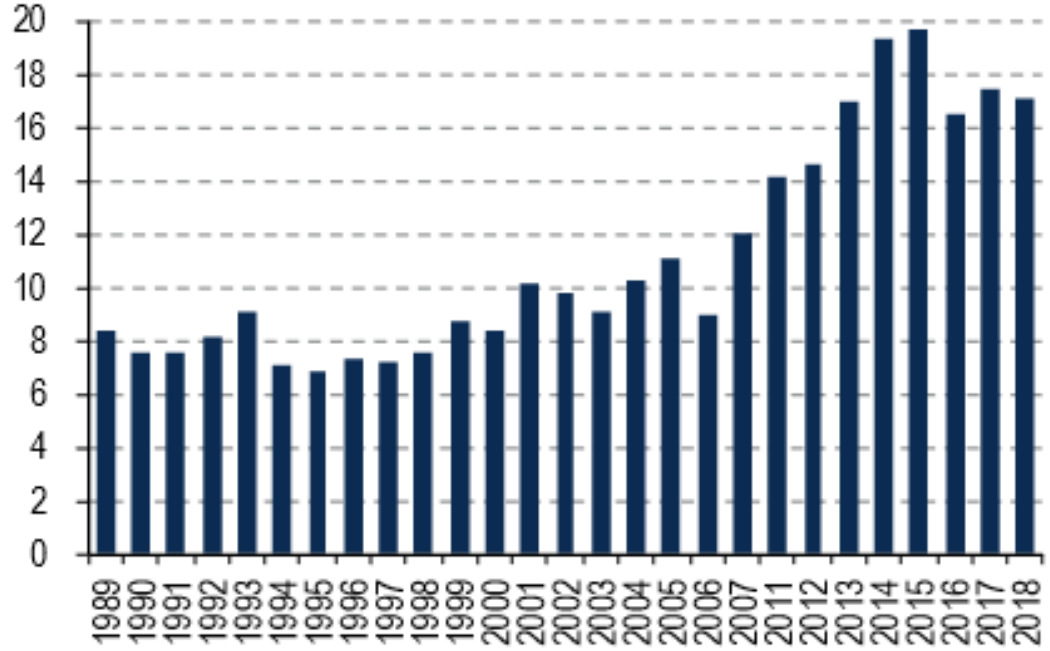
246 222



Part I: Quant Equity Nowadays, Everybody is a Quant

The search for Quant equity innovation…

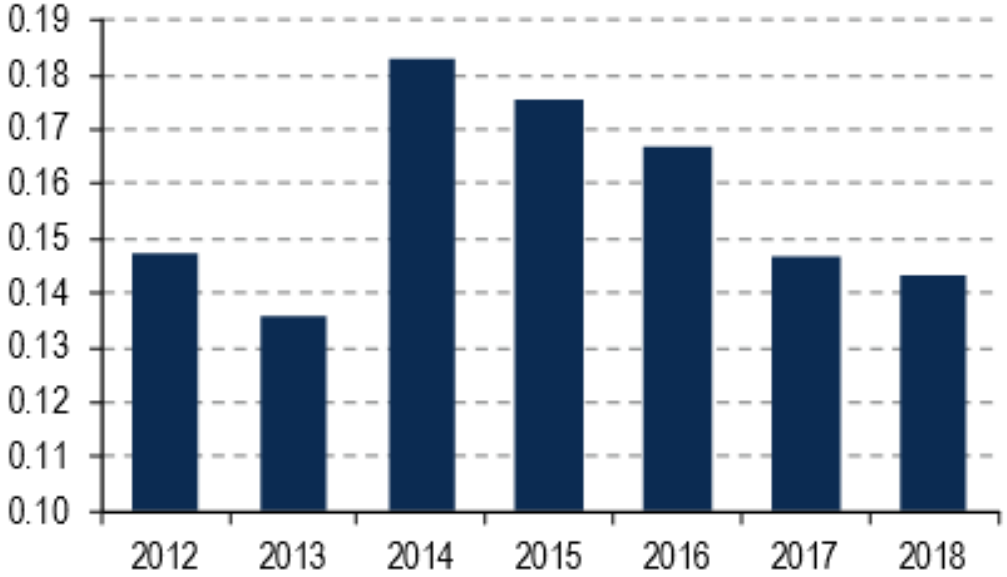
Chart 7: Average number of factors used by respondents, 1989-2018



Note: Does not include data for 2008-10 due to lack of sufficient responses
Source: BofA Merrill Lynch US Equity & US Strategy

Chart 8: Correlations between clients have generally decreased (as clients are using different factors than their peers)

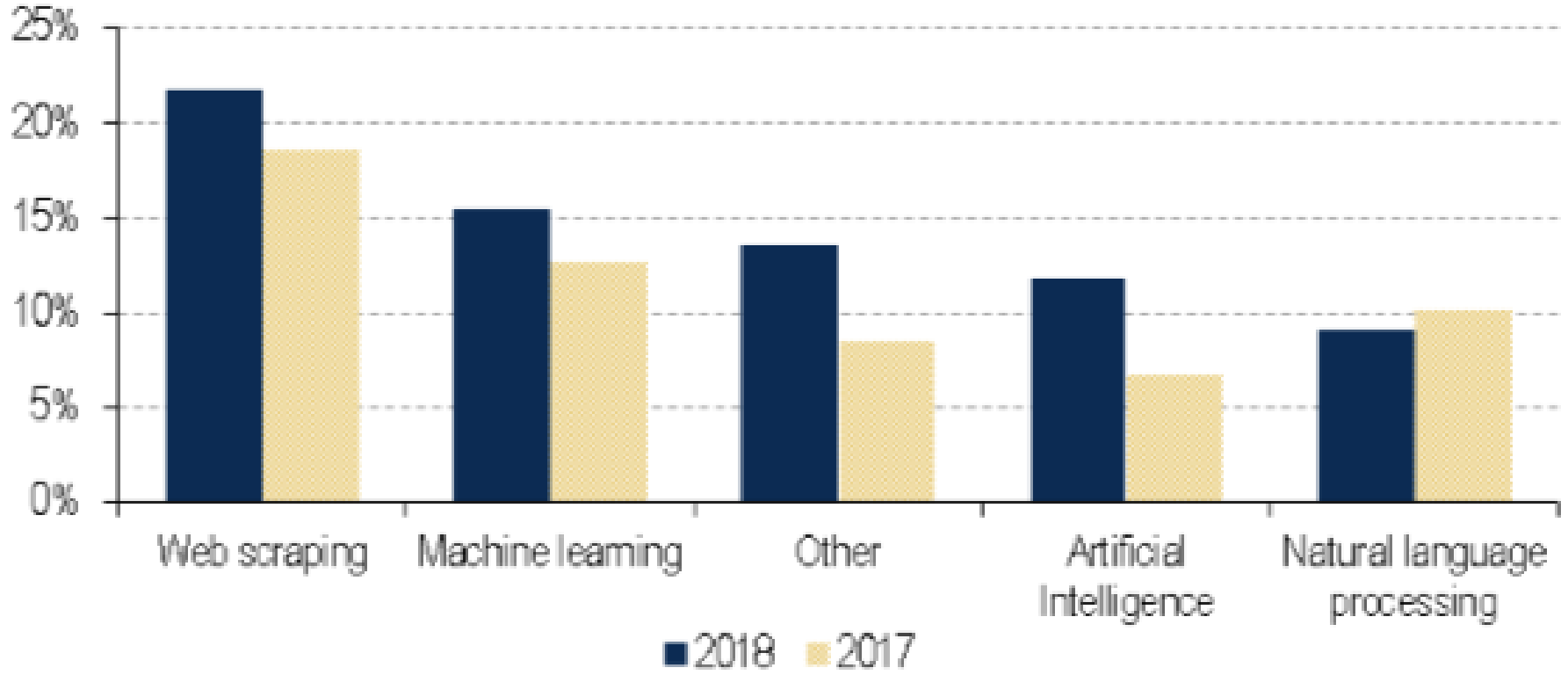
Avg. pair-wise correlations between survey respondents' factor usage by year



Source: BofA Merrill Lynch US Equity & US Strategy

… in Technology

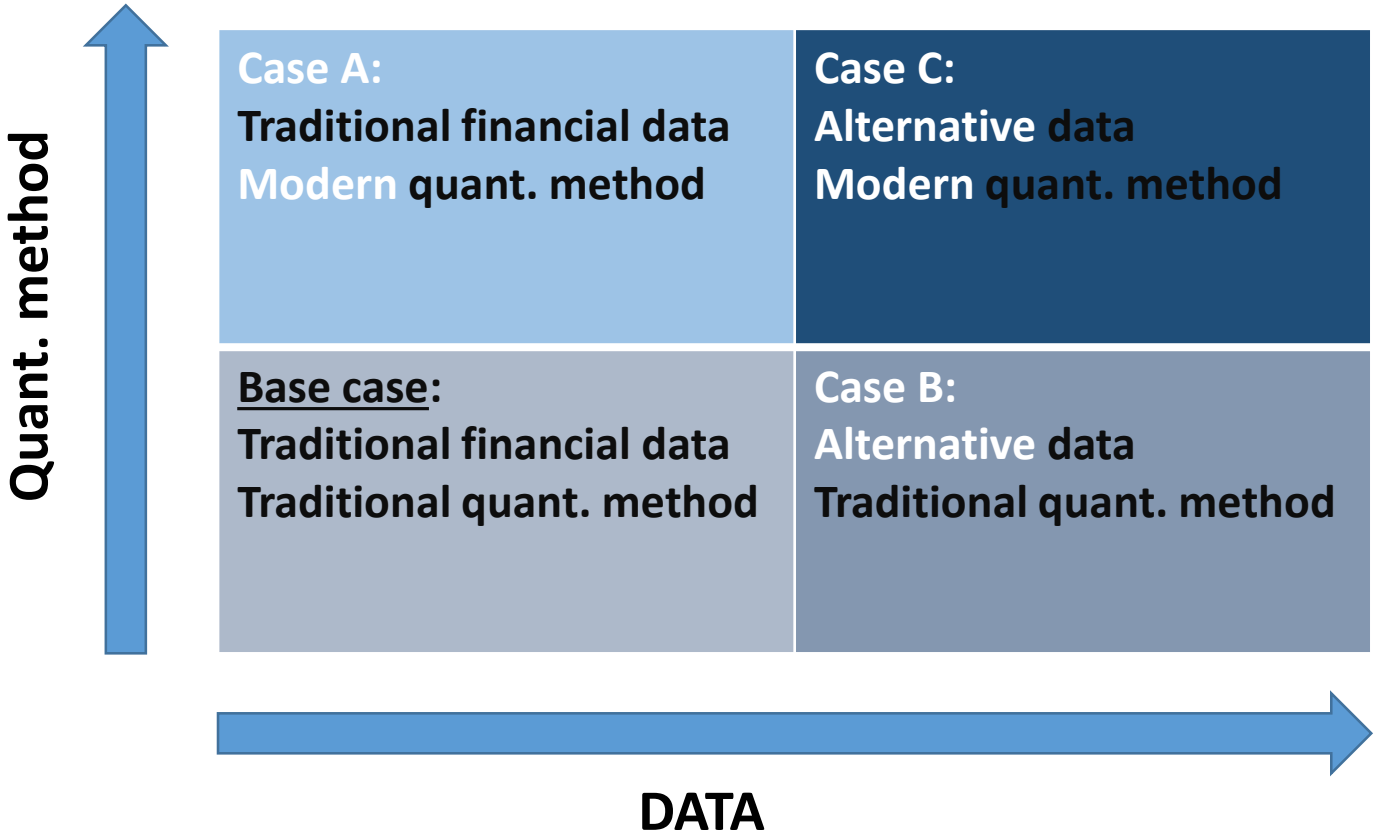
Chart 13: Which of the following quantitative tools do you use in your investment process? (% of 2018 and 2017 survey respondents)



Note: "Natural language processing" answer choice renamed in 2018 survey from "Text-based algorithms" previously
Source: BofA Merrill Lynch US Equity & US Quant Strategy

Quant (R)evolution ?

Quant (R)evolution ?



Usual criticisms against ML in finance

- **Low** signal to noise ratio
- Blackbox, **cannot be interpreted**
- Overfit, **not valid out of sample**
- **Only** for high frequency
- Data mining, prone to **spurious correlation**

Limitations of ML in quantitative investment

- Finance is not a **“plug and play”** for ML
- **Causality** between labels and outcomes
- **Laws** of the **financial ecosystem** change with agents’ interactions

Beyond Machine Learning buzzword in finance… How ?

Intelligence is <u>NOT</u> consciousness

Technology is <u>NOT</u> deterministic

- You have to : have a ML "**Why**"
- You have to : **frame** the question, put financial and economic structure in dataset. **<u>Your</u>** domain knowledge.
- You have to : **understand** and know your **Data**

Modern approaches are made for PREDICTIONS

Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106

Machine Learning: An Applied Econometric Approach

Sendhil Mullainathan and Jann Spiess

Machines are increasingly doing "intelligent" things: Facebook recognizes faces in photos, Siri understands voices, and Google translates websites. The fundamental insight behind these breakthroughs is as much statistical as computational. Machine intelligence became possible once researchers stopped approaching intelligence tasks procedurally and began tackling them empirically. Face recognition algorithms, for example, do not consist of hard-wired rules to scan for certain pixel combinations, based on human understanding of what constitutes a face. Instead, these algorithms use a large dataset of photos labeled as having a face or not to estimate a function $f(x)$ that predicts the presence y of a face from pixels x. This similarity to econometrics raises questions: Are these algorithms merely applying standard techniques to novel and large datasets? If there are fundamentally new empirical tools, how do they fit with what we know? As empirical economists, how can we use them?[1]

Goals of Prediction and Estimation

- Goal of prediction: loss minimization

$$\hat{f} \approx \min_{f \in \mathcal{F}} E_{(y,x)} L(f(x), x)$$

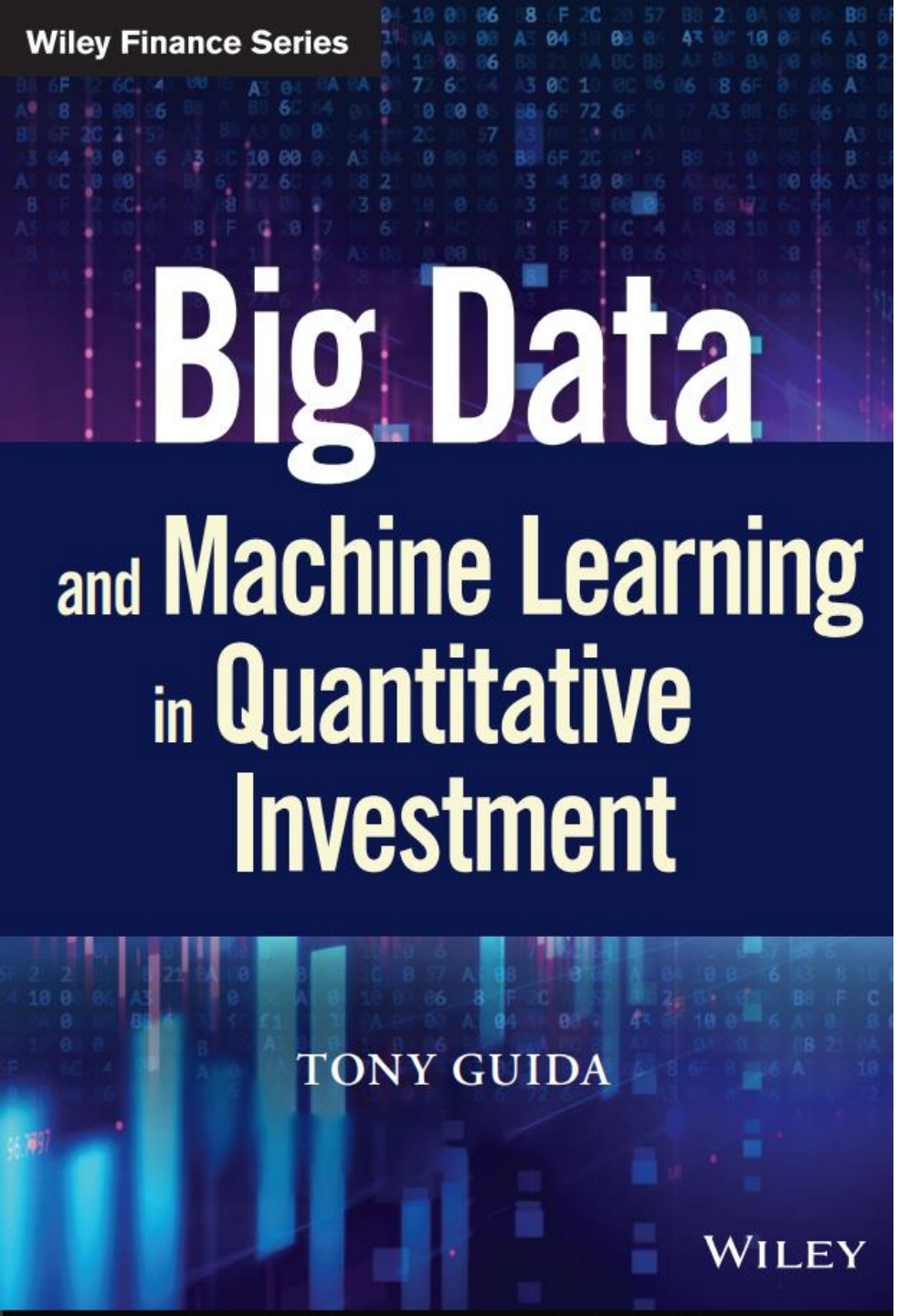
  – Use the data to pick a function that does well on a new data point

- Goal of estimation: unbiasedness

$$E[\hat{f}] = f$$

Source: AFA Lecture: Machine Learning and Prediction in Economics and Finance

PART II: Big Data and Machine Learning, The book.



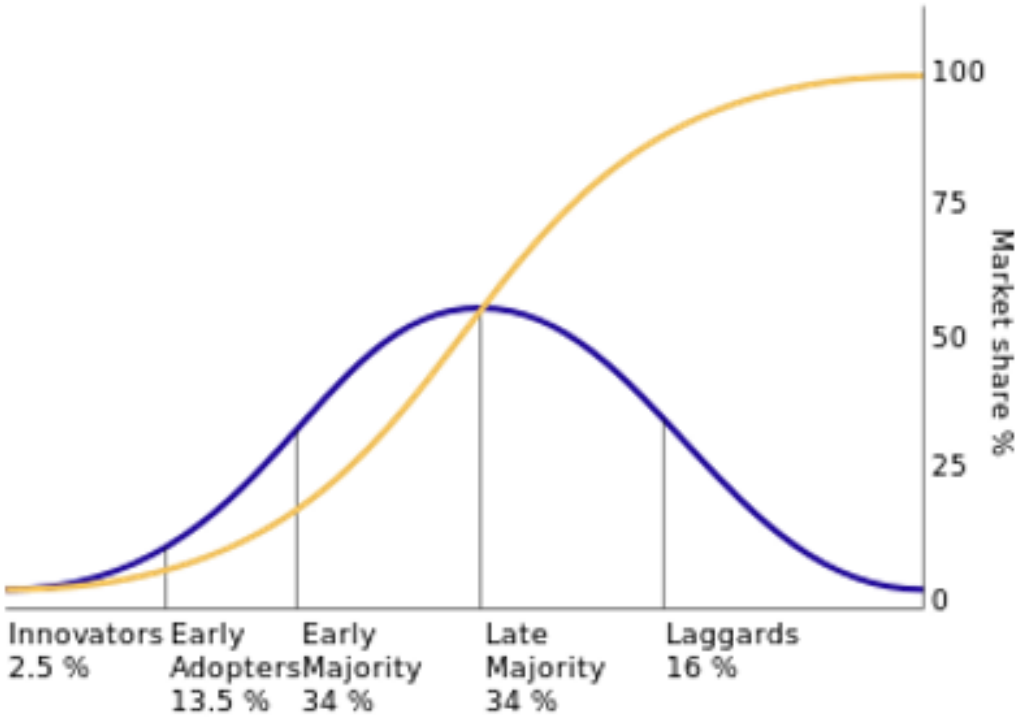
A book by practitioners for practitioners

Why doing a book ?

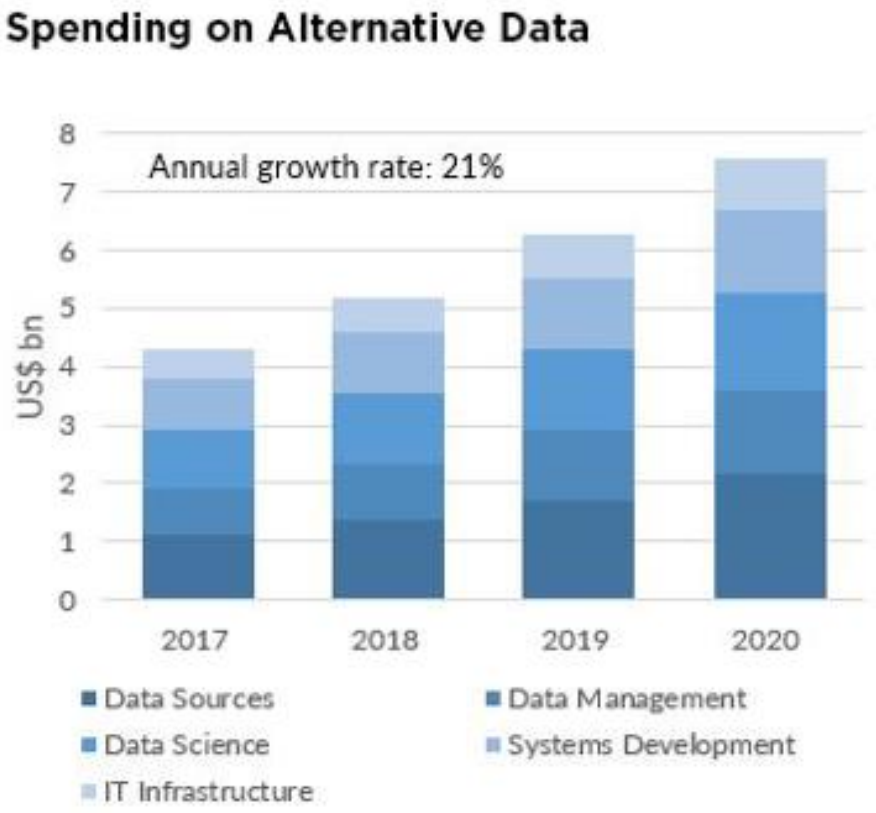
- The way Alternative data and ML in finance have been promoted **was wrong**.
- **Wisdom** of the crowd approach.
- **Lack of content** for real application.
- I Believe in the concept of **“quant community”**
- **A need for elaborated debate**

Key takeaways from the book

…"The end of the beginning for alternative data"



Source: Rogers Everett - Based on Rogers, E. (1962) Diffusion of innovations[1]



Source: Opimas Analysis[4]

Source: R. Lipus, Neudata. Chapter 2: Taming Big Data

Key takeaways from the book

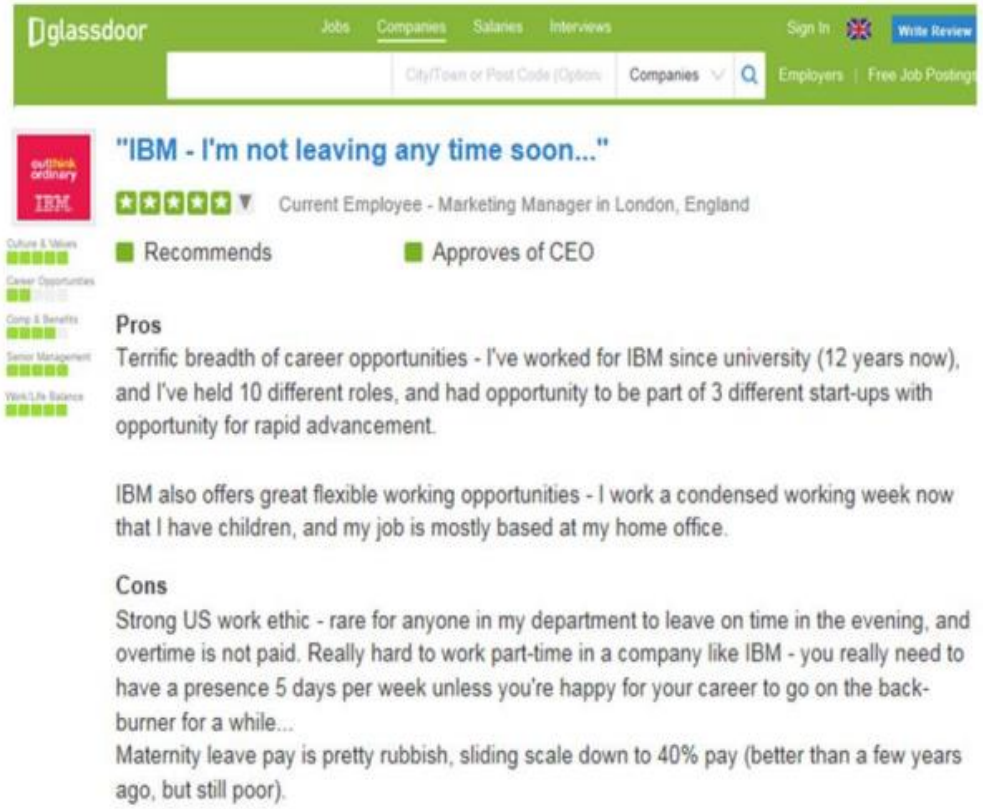
Chart: Applications of Big Data and Artificial Intelligence in Alternative Investment Strategies

Increasing volume, variety, velocity, and veracity of data	No AI	Advanced Data Processing	AI Research & Trading
Big Data	**Analyst Army** Analysis of voluminous, real-time, noisy data sets using traditional tools such as multi-variate regression. • **MCD:** Live collection and analysis of restaurant parking lot information across the globe. • Potentially feasible but time and cost prohibitive as the manual work required would entail an army of analysts.	**Master Data Scientists** Identification of novel, preferably proprietary, data sets using cutting edge techniques to identify alpha signals across multiple sources. • **MCD:** Use of machine vision to count cars in satellite images of parking lots, whilst adjusting for local weather, seasonality, and local holidays/events. • The trading of commodity markets by using real time geo data regarding ports/harbours/roads combined with meta data (such as identity and capacity of vessels).	**Holy Grail or Apocalypse** Autonomous data collection, analysis of unstructured data and independent strategy development. • **MCD:** The machine collects information from promising sources and develops new trading strategies wholly independently. • Humans spend their time on the beach or under the control of their robot masters (!).
Unstructured Data	**Data Collectors** Large structured proprietary data sets analysed using traditional methods to identify potential alpha signals. • **MCD:** Social media sentiment from Google Trends used to analyse mentions of healthy lifestyles vs. convenience and the impact on the firm. • Systematic collection of bid/ask and flow information from multiple banks on FX pairs for signal generation (quoting real time spot and forwards on a range of $ sizes).	**Competitive Data Scientists** Use of established data science techniques on public or proprietary structured and unstructured data sets. • **MCD:** Natural Language Processing (NLP) techniques used to analyse competitive positioning in other countries (press releases and texts in different languages). • Use of NLP to analyse if media is positive or negative on a group of stocks in a region/sector/country.	**Strategy Development & Execution** Machine identifies promising structured and unstructured data sets. Development and execution of trading strategies is conducted with very limited human intervention. • **MCD:** The machine identifies superior data sources and is able to evaluate and process them independently. • The machine aspires to assess other machines/models and decide which to allocate to. Models are further advanced to reconstruct data sets.
Structured Data	**Traditional Alternatives** Use of traditional analytical tools to build sophisticated models incorporating publicly available structured data. Known in the industry as Systematic, Algo, Quant, and Discretionary Traders. • **MCD:** Sentiment analysis from exchanges, macro data from regions McDonalds operates in, sub-sector fundamental analysis ($ consumer spends on junk food vs healthy lifestyles) executed in a systematic or discretionary way. • Typical approaches include Momentum, Trend Following, OTC (Over The Counter) Trend Following, and Volatility Strategies.	**Advanced Trading** More sophisticated analytical techniques to process existing data, enabling faster reaction times, a larger research universe and extracting information that single factor strategies miss. • **MCD:** Combination of technical and fundamental information and its utilisation in the context of other stocks. Ability to structure and carve out nuances in financial statement analysis. • Use of Machine Learning to guide positioning in the markets traded (i.e. examine how the market reacted to features historically, and take real time readings of the current feature).	**Autonomous Trading** The machine develops and executes trading strategies with little to no human intervention based on preselected structured data sets. Systems are very advanced and able to act qualitatively and think in terms of what-if-scenarios. • **MCD:** The trading system takes long or short positions in the stock autonomously (advanced algorithms identify market regime changes). • The system is advanced enough to adjust the portfolio composition, net /gross exposures, take profits, cut losing positions and maintain high conviction trades.

Increasing use of advanced data processing techniques

Source: E. Sirotyup, CS. Chapter 3: State of machine learning applications in investment management

Key takeaways from the book



	(1)	(2)	(3)
lagged dependent	-0.012	-0.015	-0.012
	(-0.358)	(-0.423)	(-0.351)
Forecast dispersion	-2.700	-2.806	-2.581
	(-3.196)	(-3.318)	(-2.916)
OVERALL	0.067	0.053	0.079
	(0.761)	(0.505)	(0.755)
GOAL		1.770	4.477
		(2.536)	(3.751)
TONE		0.054	1.714
		(2.071)	(1.796)
High_expectations			14.180
			(2.892)
Analyst revisions	15.130	14.730	18.050
	(4.749)	(4.622)	(5.173)
Log(Market Equity)	0.000	0.000	0.000
	(-1.078)	(-1.021)	(-1.552)
Log(Book/Market)	-0.006	-0.018	-0.053
	(-0.096)	(-0.294)	(-0.857)
Pmom	0.716	0.738	0.774
	(7.411)	(7.612)	(8.007)
KLD			0.055
			(1.904)
BC			-0.974
			(-1.699)

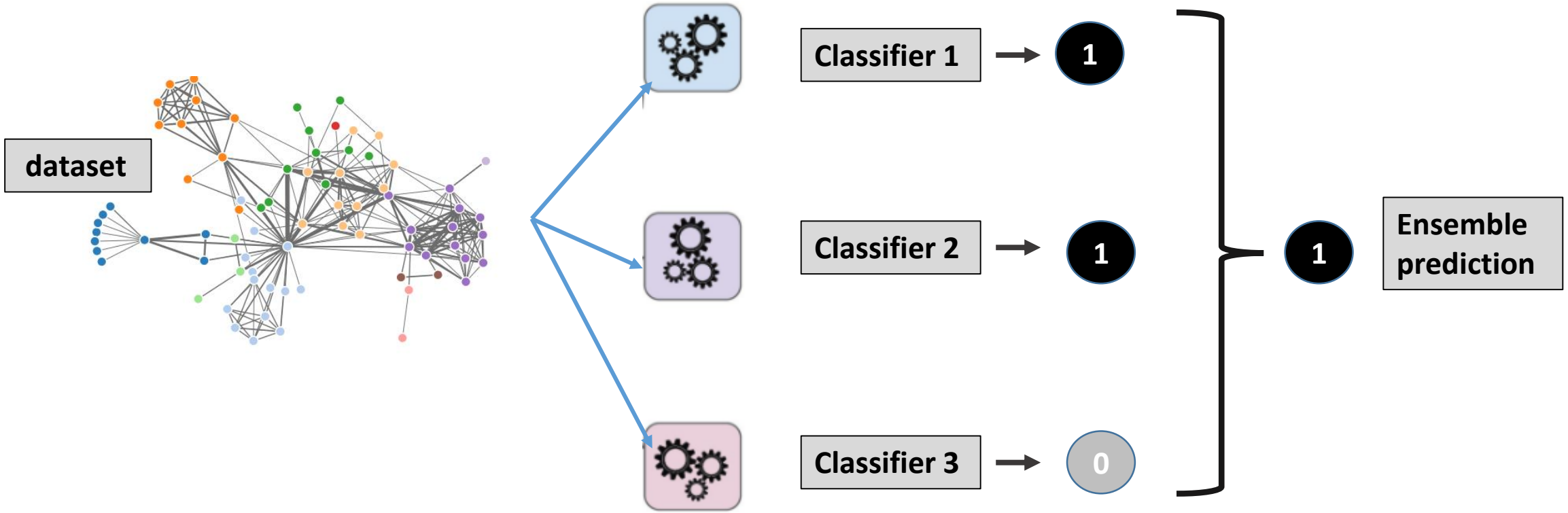
Source: A. Moniz, DB. Chapter 8: **A social media analysis of corporate culture**

PART III: Training Equity Models

Concepts and definitions

Introduction and definition: Ensemble Learning

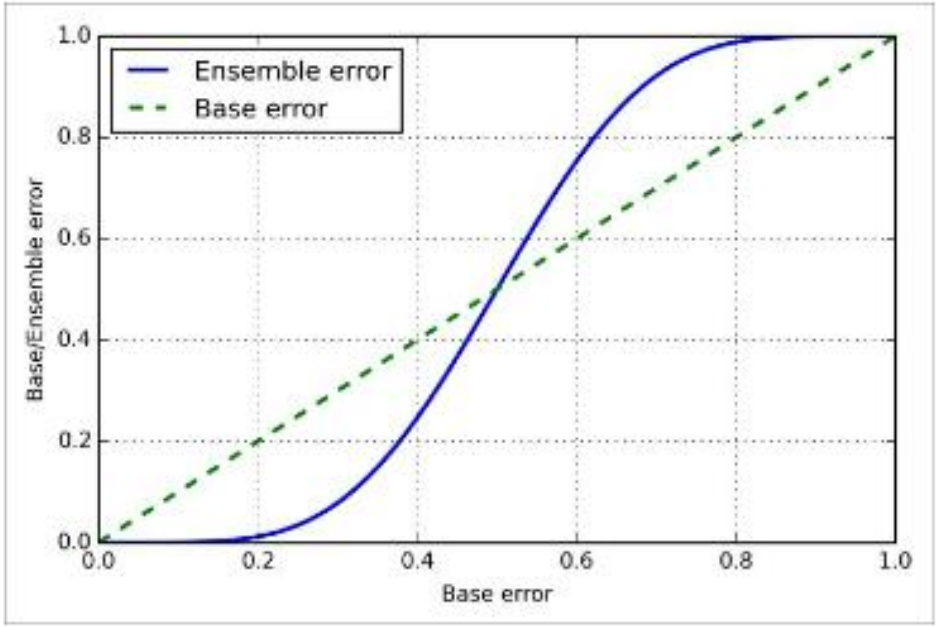
"**Ensemble Learning purpose** is to combine the predictions of several classifiers built with a given learning algorithm in order to improve bias and variance over a single estimator.



Wisdom of the crowd: ensemble classifier error

Simple example

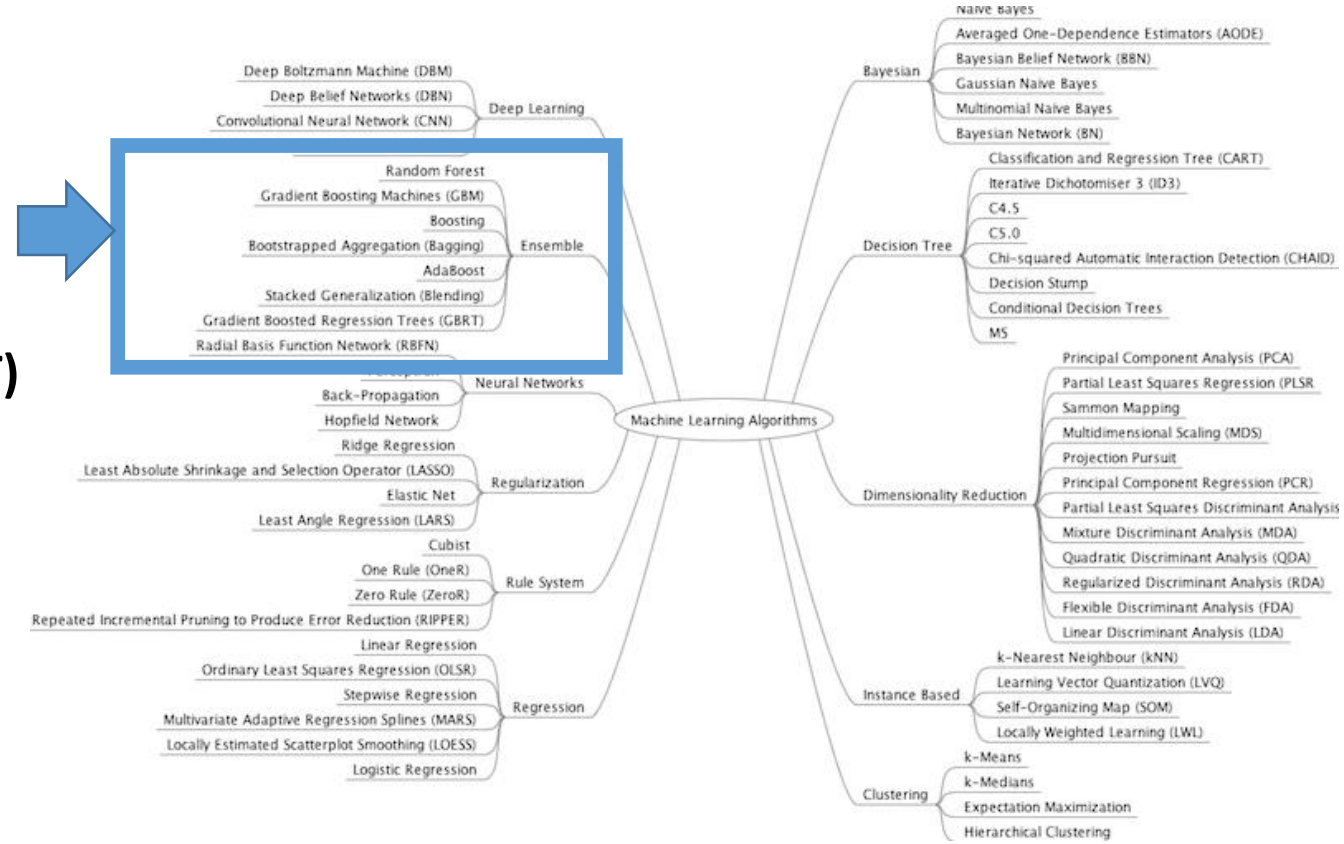
- Assuming independent classifiers
- Classifier has an error rate $\varepsilon < 0.5$
  - Ensemble prediction better than random guess
- If $\varepsilon > 0.5$ for each classifier, ensemble wrong prediction will increase



Raschka, Sebastian. Python Machine Learning (p. 202). Packt Publishing.

Wisdom of the crowd :ensemble learning family

- **Boosting**
- **Bootstrapped Aggregation (Bagging)**
- **AdaBoost**
- **Stacked Generalization (blending)**
- **Gradient Boosted Regression Trees (GBRT)**
- **Random Forest**



https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

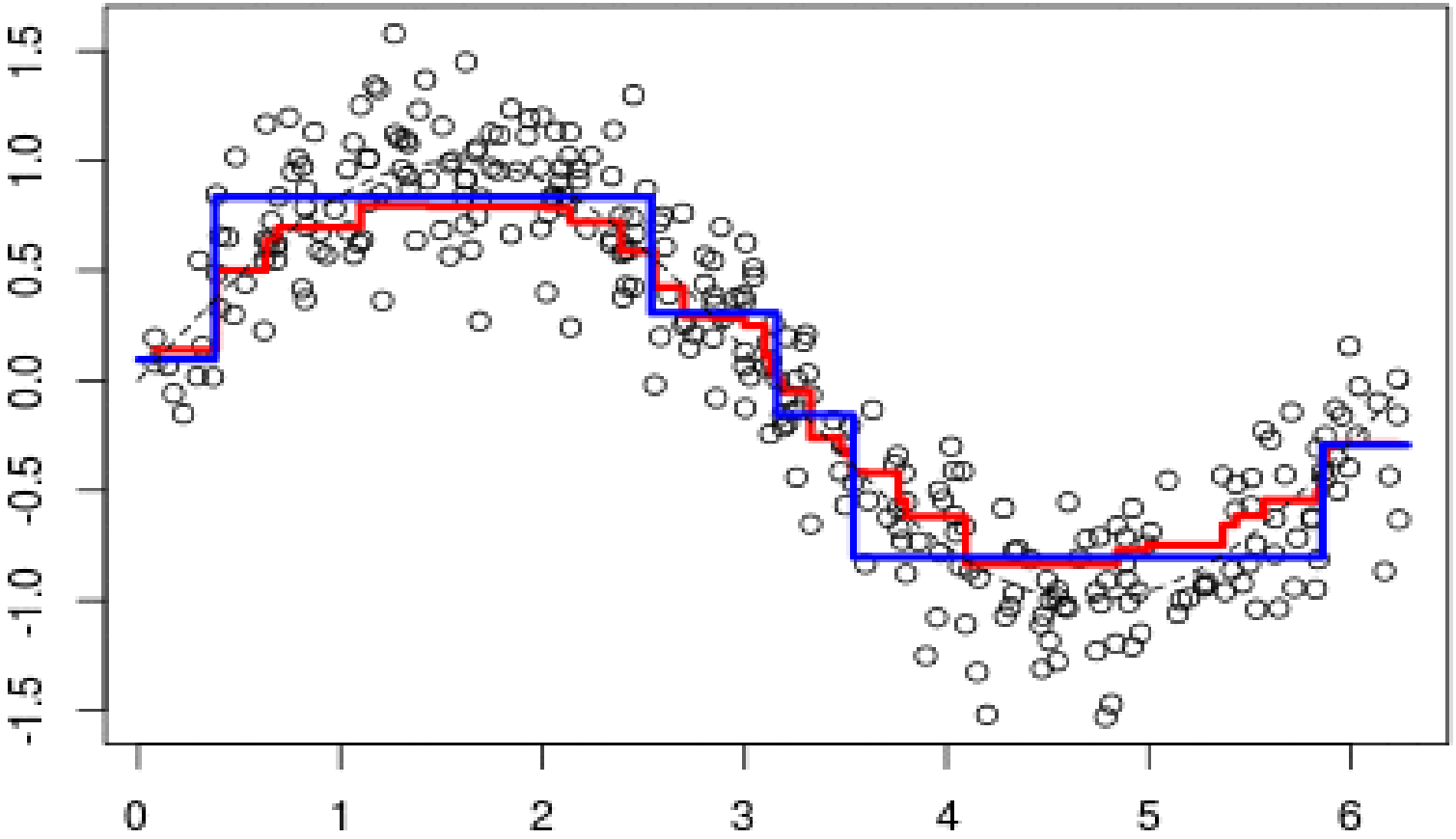
Wisdom of the crowd :ensemble learning family

Bagging: short for bootstrap aggregating is a way to decrease the variance of your prediction by generating bootstrapped samples. Each bootstrap sample is then used to fit a classifier.

Boosting: multi-step approach, where one first uses subsets of the original data to produce a series of averagely performing models and then "boosts" their performance by combining them together. Every new subsets contains the elements that were (likely to be) misclassified by previous models.

Stacking is a way of combining multiple models as inputs, that introduces the concept of a meta learner.

Bias vs Variance tradeoff



Training Equity Models: Naive example

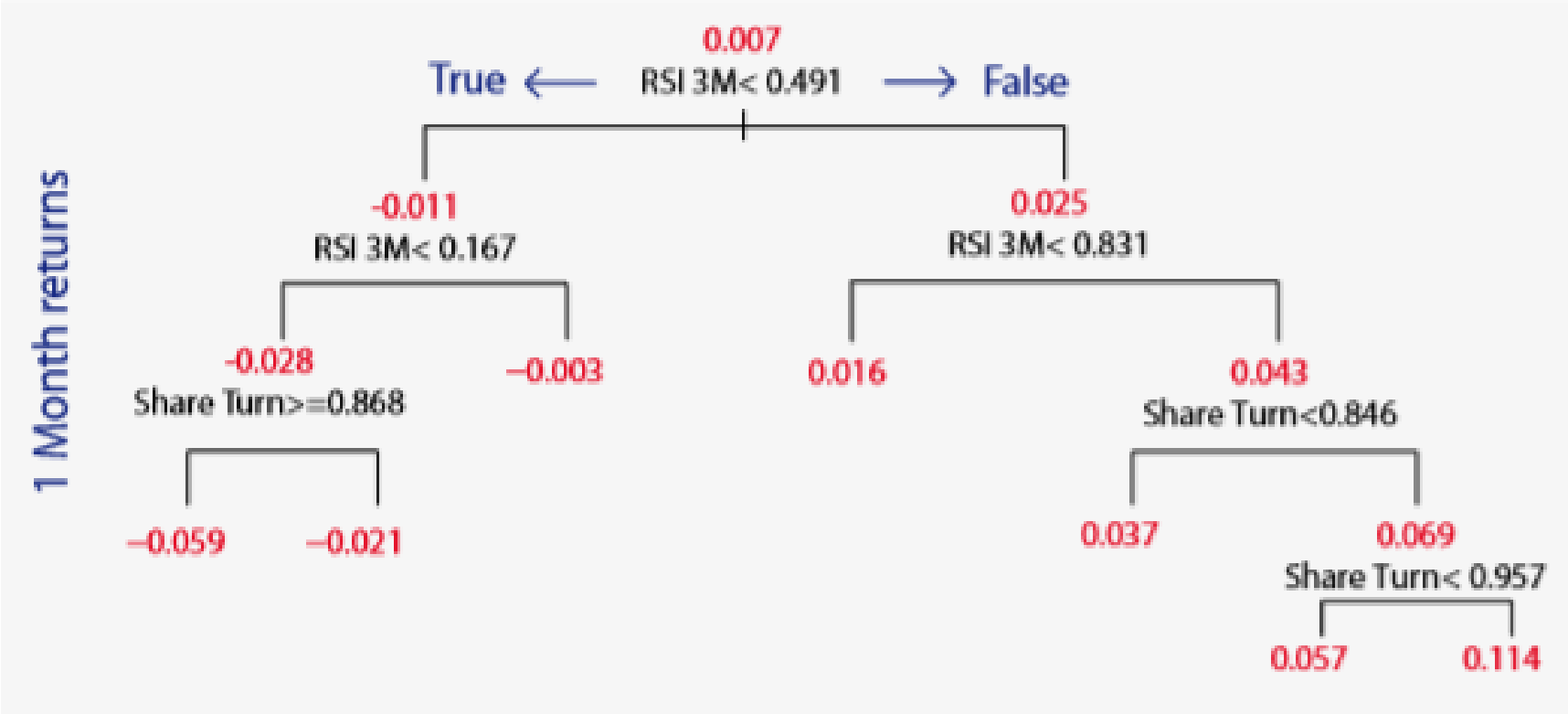
Naive approach: Objective, data and protocol

- **1M future absolute** performance
- Using simple regression trees
- Investment universe is **US stocks (~500)**
- **Full** dataset from Dec-2000 until Dec-2016
- **(~ 30) features**, monthly normalised.
- No pre-processing except winsorizing

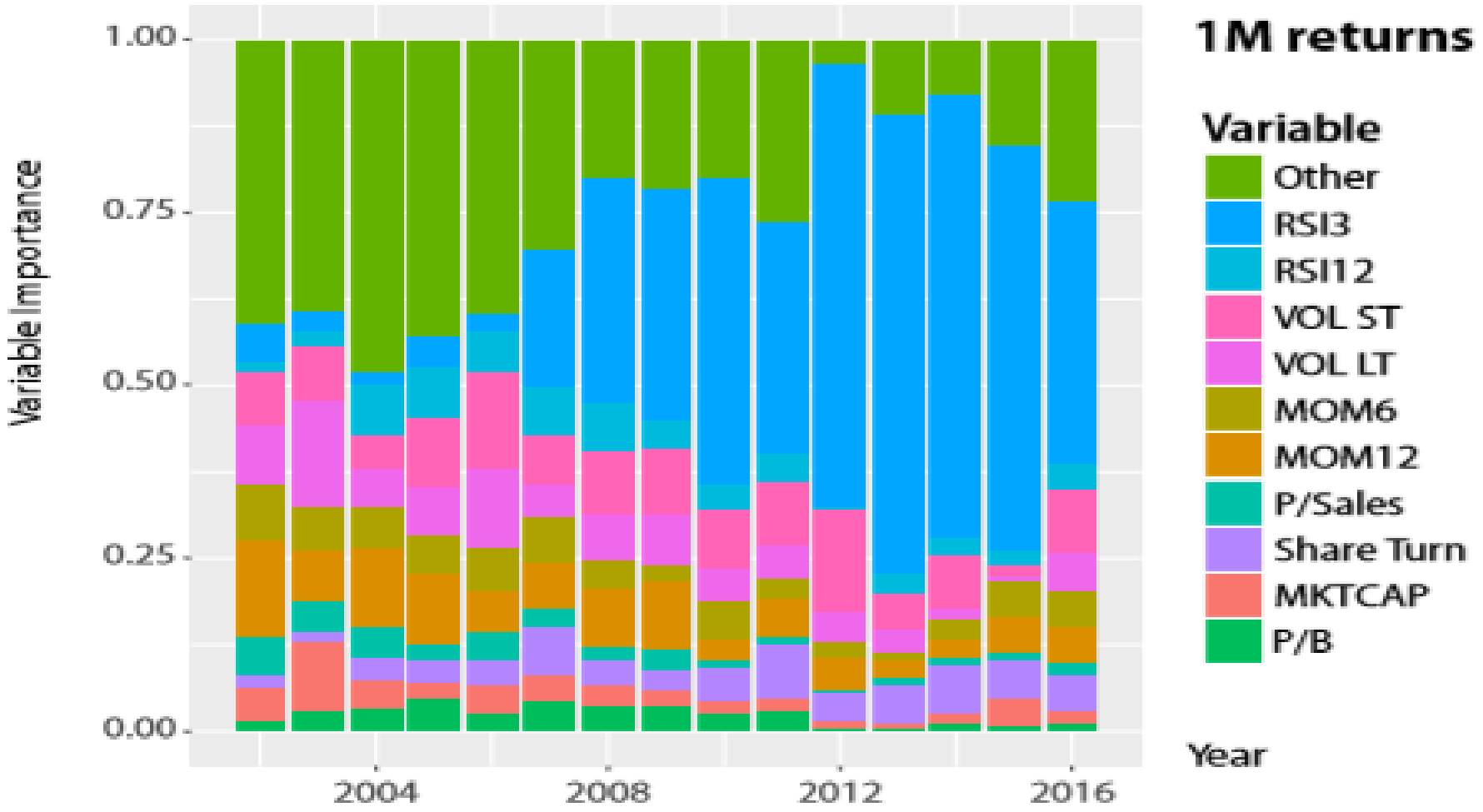
Creating the dataset

K	Short Name	Description
		Fundamental / Accounting
1	BuyBackYld	Buyback Yield
2	Cash Conversion	Cash to EBIT ratio
3	Debt/Equity	Debt on Equity
4	DivYld	Dividend Yield
5	EV/EBITDA	Enterprise Value to EBITDA ratio
6	FCFYld	Free Cash Flow Yield using FCF per Share
7	GrProfitGrowth	Gross Profits Growth (12 months)
8	Net Debt	Net Debt
9	NetDbtYld	Net Debt Yield
10	Net Prt Margin	Net Profit Margin
11	Op Prt Margin	Operating Profit Margin
12	P/B	Price to Book ratio
13	P/E	Price-Earnings ratio
14	P/Sales	Price to Sales ratio
15	PyoutRatio	Payout Ratio (Divided Per Share / Earnings Per Share)
16	ROC	Return On Capital
17	ROE	Return On Equity
18	ShareHldrYld	Total Shareholer Yield
		Momentum
19	MOM6	Momentum 6-1 months
20	MOM12	Momentum 12-1 months
21	RSI 3	RSI over 3 months
22	RSI 12	Relative Strength Index over 12 months using weekly returns
		Volatility
23	Vol LT	Ex-post volatility over 156 weeks
24	Vol ST	Ex-post volatility over 52 weeks
		Liquidity
25	ADV20	Average Daily Volume over 20 days
26	ADV60	Average Daily Volume over 60 days
27	ADV12M	Average Daily Volume over 12 months
28	BidAsk	Bid-Ask Spread
29	MKTCAP	Market Capitalization
30	Share Turn	Share turnover ratio

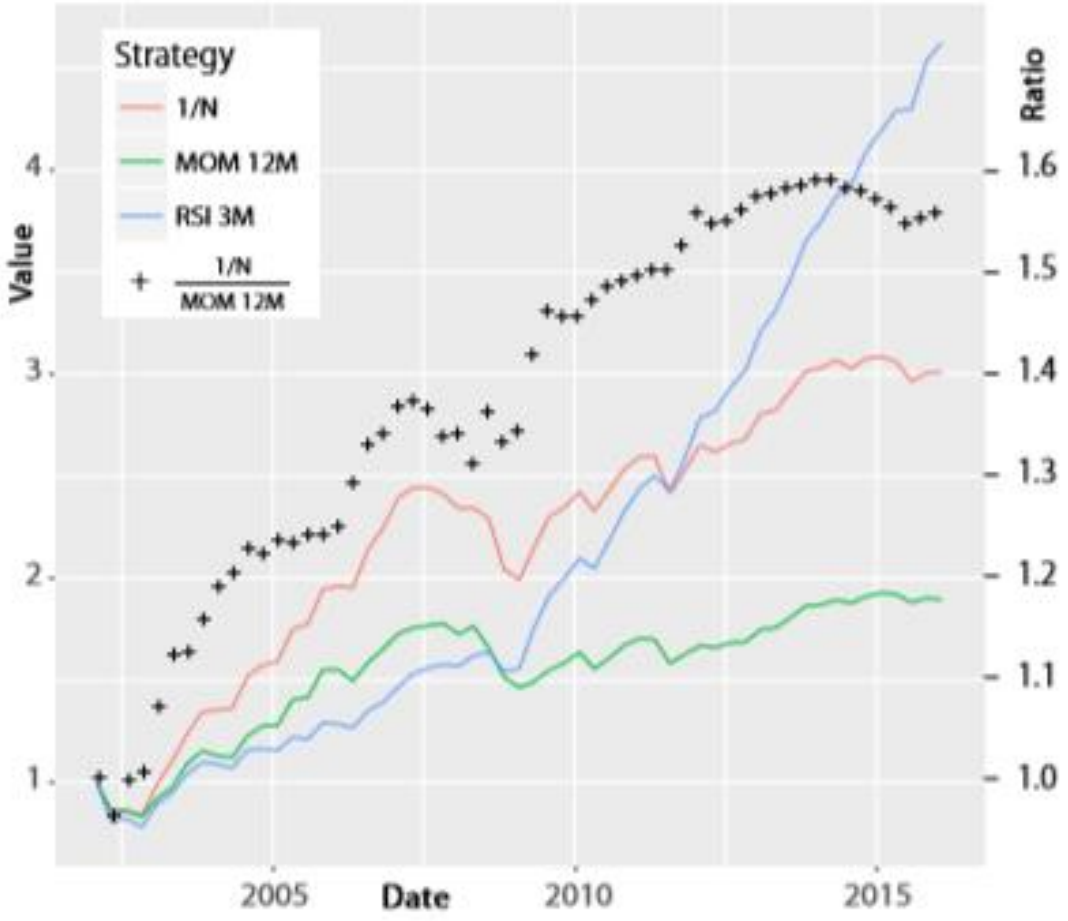
Result for the model full sample



Variable importance for the model



Implementing the model



Training Equity Models: Using a robust approach

eXtreme Gradient Boosting

XGBoost is an open-source software library which provides the gradient boosting framework. It became popular in the computer science community after being used by most of winners in kaggle competition

kaggle is a platform for predictive modelling and analytics competitions in which statisticians and data miners compete to produce the best models for predicting and describing the datasets uploaded by companies and users.

eXtreme Gradient Boosting : introduction to the model behind

General objective of tree ensemble for K trees

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

$$Obj = \boxed{\sum_{i=1}^{n} l(y_i, \hat{y}_i)} + \boxed{\sum_{k=1}^{K} \Omega(f_k)}$$

Training on loss **Complexity of the trees**

$$\begin{aligned}
\hat{y}_i^{(0)} &= 0 \\
\hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\
\hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\
&\dots \\
\hat{y}_i^{(t)} &= \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)
\end{aligned}$$

Additive training

http://xgboost.readthedocs.io/en/latest/model.html#

Why using eXtreme Gradient Boosting ?

- **Easier** to detect future **outperforming** stocks using non-linearity and **variables interactions**
- Wisdom of the **crowd**
- **Model complexity controlled** through **regularization parameters**
- Computation speed, **parallelized**
- Interpretability through **variable importance**

XGBoost applied to Quant Equity

Objective, data and protocol

- We will try to predict **12M future** performance relative to their sectors.
- Investment universe is **US stocks (~3000)**
- **Full** dataset from Dec-2000 until Dec-2017
- **(~ 200) features**, monthly normalised in percentrank.
- Dataset pre-processed, outliers removed, focussing on training on the tails of the distribution **(top and bottom 20%)**
- Split the dataset between **Training** (80%) and **Testing** (20%)
- We will use **k-fold cross validation to find hyperparameters**
- Rolling window of **24 months**

Creating the dataset

Features 1 >>> ~ 200

Instances >>> ~ 400 000

Results | Messages

	DataDate	dateReturnPerf	SecID	Fact2	Fact3	Fact4	Fact7	Fact8	Fact9	Fact10	Fact11	Fact12	Fact13	Fact14	Fact16	Fact17	Fact18	Fact20	Fact21	Fact22	Fact23	Fact25	Fact26	Fact27
1	2010-12-31	2011-12-31	1195515867	19	70	94	94	13	99	15	34	55	65	42	32	28	21	27	31	49	38	53	53	76
2	2014-08-31	2015-08-31	570191681	44	NULL	83	48	11	7	19	72	72	28	33	32	31	21	30	22	31	51	13	14	49
3	2012-01-31	2013-01-31	290849864	55	66	98	15	3	100	7	2	2	58	43	14	38	25	19	22	11	11	10	11	65
4	2012-02-29	2013-02-28	324622763	69	69	99	27	4	100	7	2	1	65	42	14	39	16	18	21	6	6	13	15	62
5	2012-08-31	2013-08-31	1528063821	73	70	99	61	4	99	14	2	2	47	41	19	38	16	25	26	43	45	83	84	50
6	2012-03-31	2013-03-31	1850474528	61	67	99	27	4	100	11	2	1	60											
7	2012-11-30	2013-11-30	2021474168	75	74	99	60	4	NULL	13	2	10	47											
8	2017-01-31	2018-01-31	408403206	42	31	61	68	10	2	26	94	92	14											
9	2013-01-31	2014-01-31	1593812596	65	73	99	53	4	NULL	18	3	10	47											
10	2016-05-31	2017-05-31	352206094	NULL	NULL	NULL	NULL	NULL	74	NULL	77	77	33											
11	2009-03-31	2010-03-31	1251799406	55	65	49	30	10	100	11	3	1	83											
12	2007-02-28	2008-02-29	1092089186	36	69	21	97	8	90	10	74	63	33											
13	2016-09-30	2017-09-30	1171513832	30	32	60	67	10	100	27	89	83	13											
14	2003-07-31	2004-07-31	1668572152	48	30	70	35	8	91	30	3	1	41											
15	2010-08-31	2011-08-31	1317881264	20	4	87	96	15	99	10	29	56	62											
16	2008-02-29	2009-02-28	31085621	46	81	19	24	7	63	5	77	62	51											
17	2009-04-30	2010-04-30	512957258	52	62	50	32	10	100	9	3	1	84											
18	2012-04-30	2013-04-30	2143460900	67	68	99	27	4	100	10	2	1	60											
19	2016-10-31	2017-10-31	1415589206	20	NULL	NULL	64	11	58	35	79	73	15											
20	2011-01-31	2012-01-31	486144046	21	70	95	95	13	99	12	35	55	65											
21	2013-03-31	2014-03-31	156902714	54	64	99	67	1	100	19	5	35	36											
22	2009-08-31	2010-08-31	290508352	29	3	53	94	18	100	19	1	1	71											
23	2010-09-30	2011-09-30	1527687499	20	4	87	96	15	99	12	29	56	62											
24	2016-06-30	2017-06-30	2006847672	NULL	NULL	NULL	NULL	NULL	80	NULL	75	74	23	NULL	41	56	37	30	26	74	64	87	82	50
25	2007-12-31	2008-12-31	156190601	47	84	12	83	7	100	8	90	77	53	72	57	60	62	64	72	31	30	18	15	29
26	2014-07-31	2015-07-31	1573508350	48	NULL	89	84	12	6	19	71	71	31	32	31	32	25	31	23	35	61	51	61	53
27	2007-06-30	2008-06-30	1049253878	38	70	26	97	6	96	10	81	70	32	18	67	23	79	75	75	34	42	71	71	30
28	2007-03-31	2008-03-31	400280762	36	70	21	97	8	90	11	74	62	32	17	68	53	81	81	74	34	51	71	77	30
29	2016-08-31	2017-08-31	1510213818	32	30	60	70	10	100	24	89	83	14	23	44	58	29	36	28	83	56	94	88	51
30	2011-02-28	2012-02-29	231369713	19	65	95	58	15	100	11	23	45	63	48	31	34	26	24	31	57	59	60	62	78
31	2008-04-30	2009-04-30	474472098	45	81	20	24	6	65	6	78	62	51	77	54	34	53	59	67	14	10	38	26	26

Some features examples

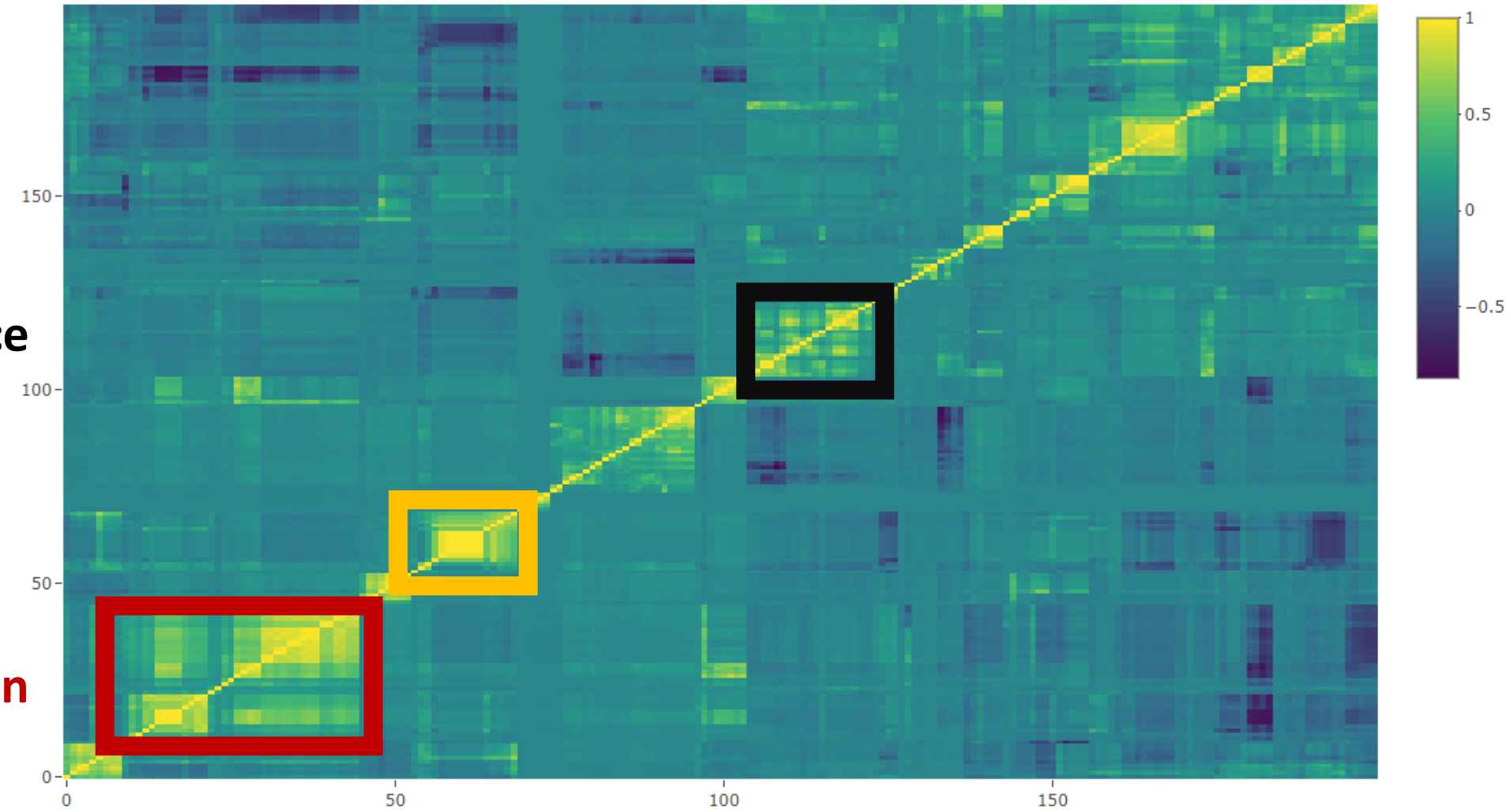
- *Fundamental trailing*
- *Estimates*
- *Price based*
- *Volume based*
- *Composites*
- *Regime variables*

Features discovery : « brute force » approach

Risk Price based

Volume

Valuation



Hyperparameters used for tuning

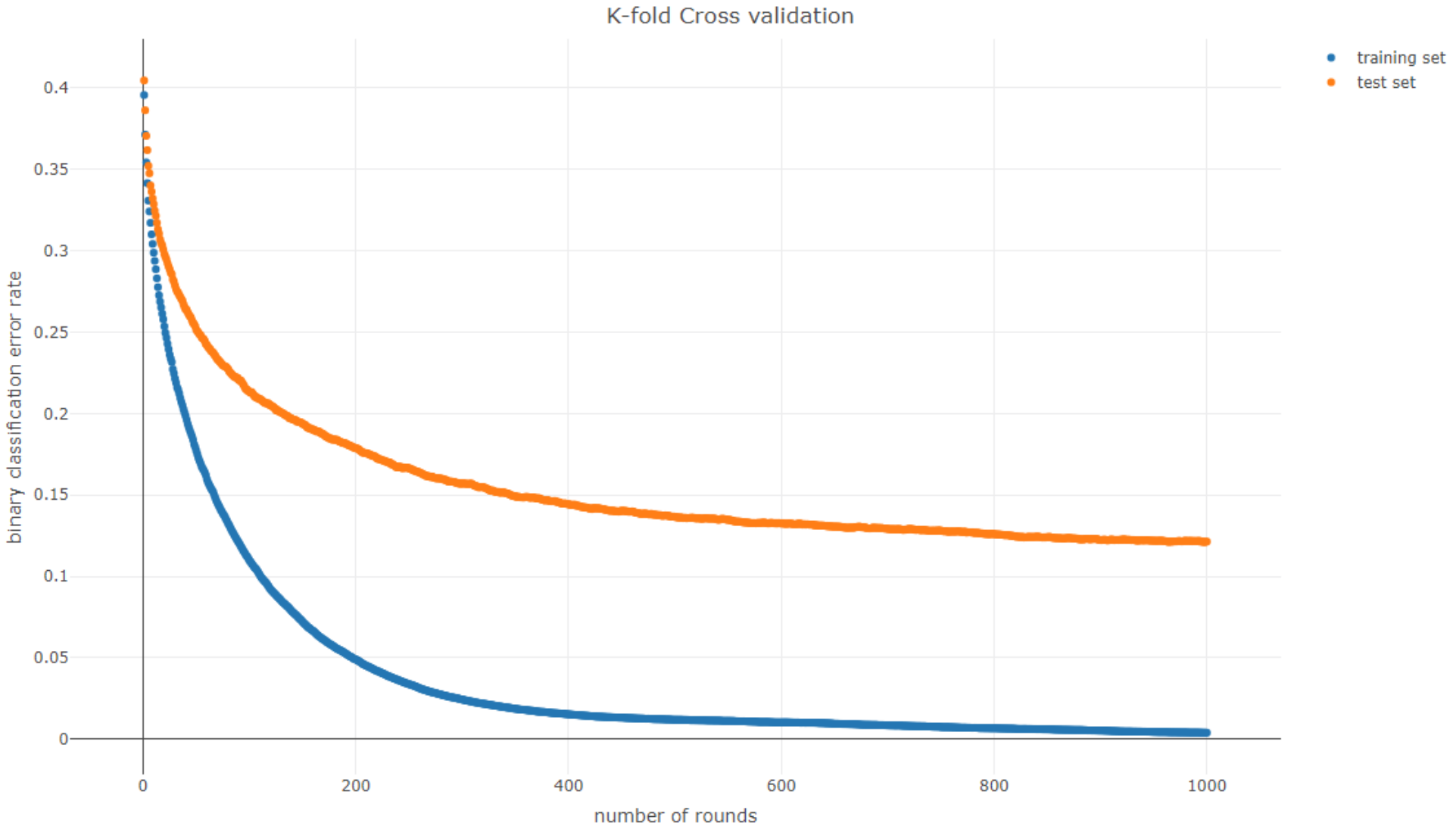
learning_rate : step size shrinkage used in update to prevents overfitting. After each boosting step, we can directly get the weights of new features. and eta actually shrinks the feature weights to make the boosting process more conservative.

min_split_loss : minimum loss reduction required to make a further partition on a leaf node of the tree. The larger, the more conservative the algorithm will be.

max_depth : maximum depth of a tree, increase this value will make the model more complex / likely to be overfitting. 0 indicates no limit, limit is required for depth-wise grow policy.

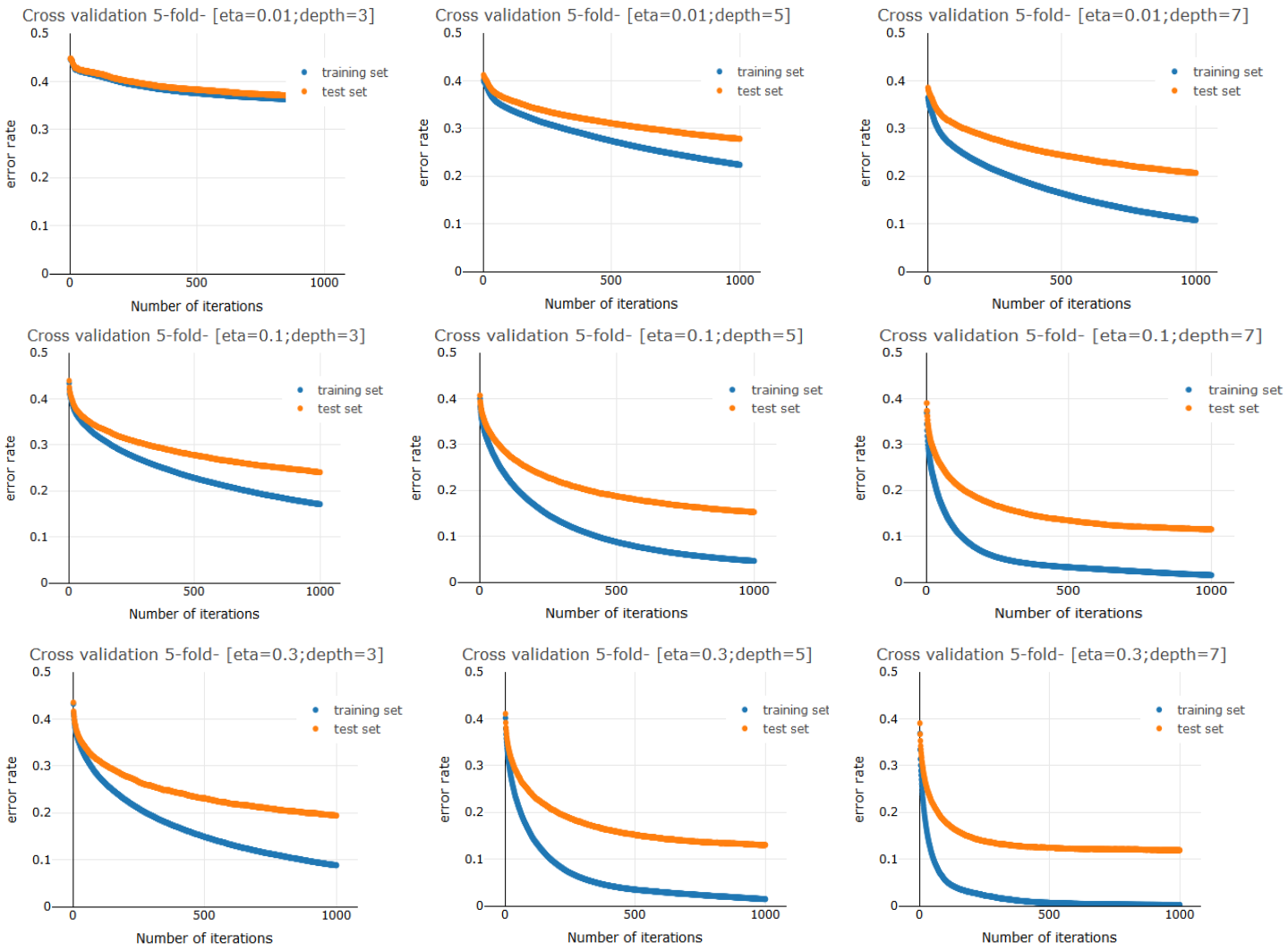
scale_pos_weight : Control the balance of positive and negative weights, useful for unbalanced classes.

Cross validation: choosing the hyperparameters



Max_depth=5, eta=0.1 gamma=0.1, scale_pos_weight =sum_pos/sum_neg

Cross validation: choosing the hyperparameters



type I type II errors

	OUTPERFORMED	UNDERPERFORMED
OUTERPERFORM	**True Positive:** Stock WAS classified as outperforming sector and DID outperformed	**False negative:** Stock was NOT classified as outperforming sector and DID outperformed
UNDERPERF.	**False Positive:** Stock WAS classified as outperforming sector and did NOT outperformed	**True Negative:** Stock was NOT classified as outperforming sector and did not outerperformed

- **Left side** of the matrix shows **Actual**
- **Top side** shows **predicted**

Class error 29%

Class error 25%

Error Rate 27%

Beyond type I type II errors

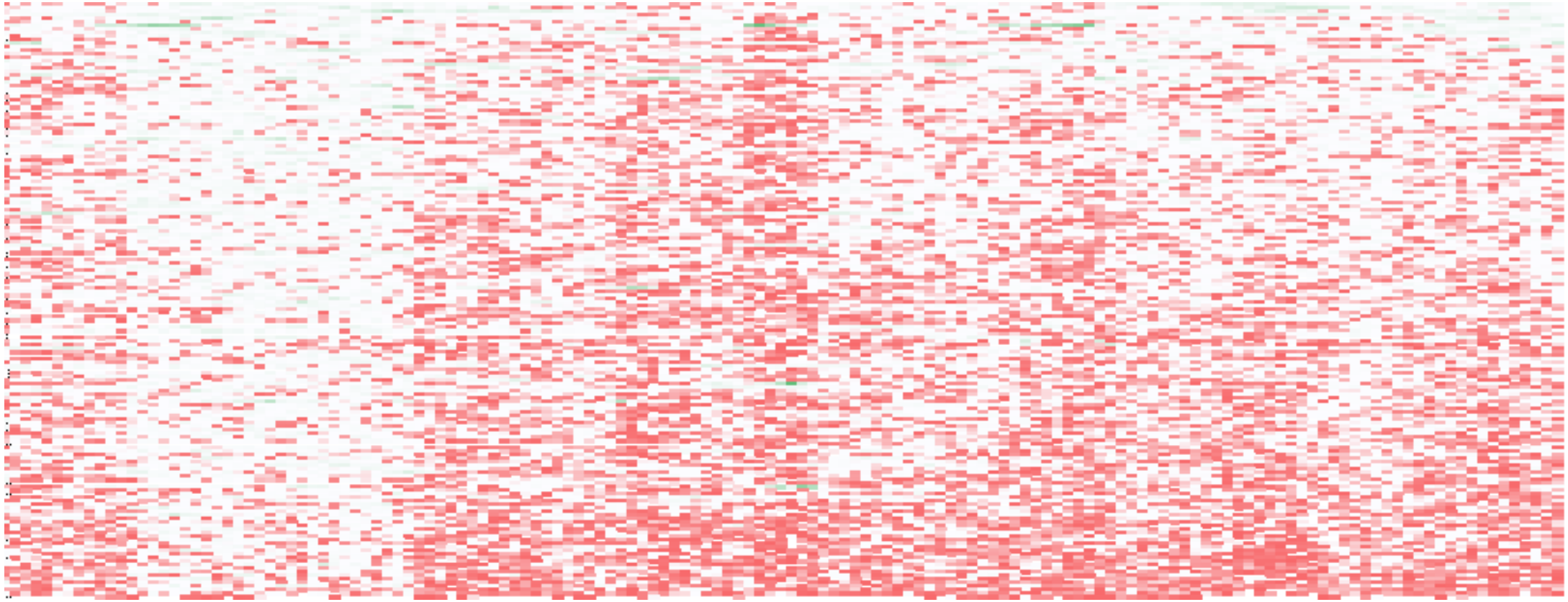
- **Fp** : false positive. Stock predicted to outperform and that did not outperform out of sample.
- **Fn** : false negative. Stock predicted to underperform that outperform out of sample.
- **Tp**: true positive. Stock predicted to outperform which outperform out of sample.
- **Tn**: true negative. Stock predicted to underperform which underperform out of sample.

Precision: **Tp / (Tp + Fp)**
Precision could be defined as a rate of successful prediction for sector neutral outperforming stocks.

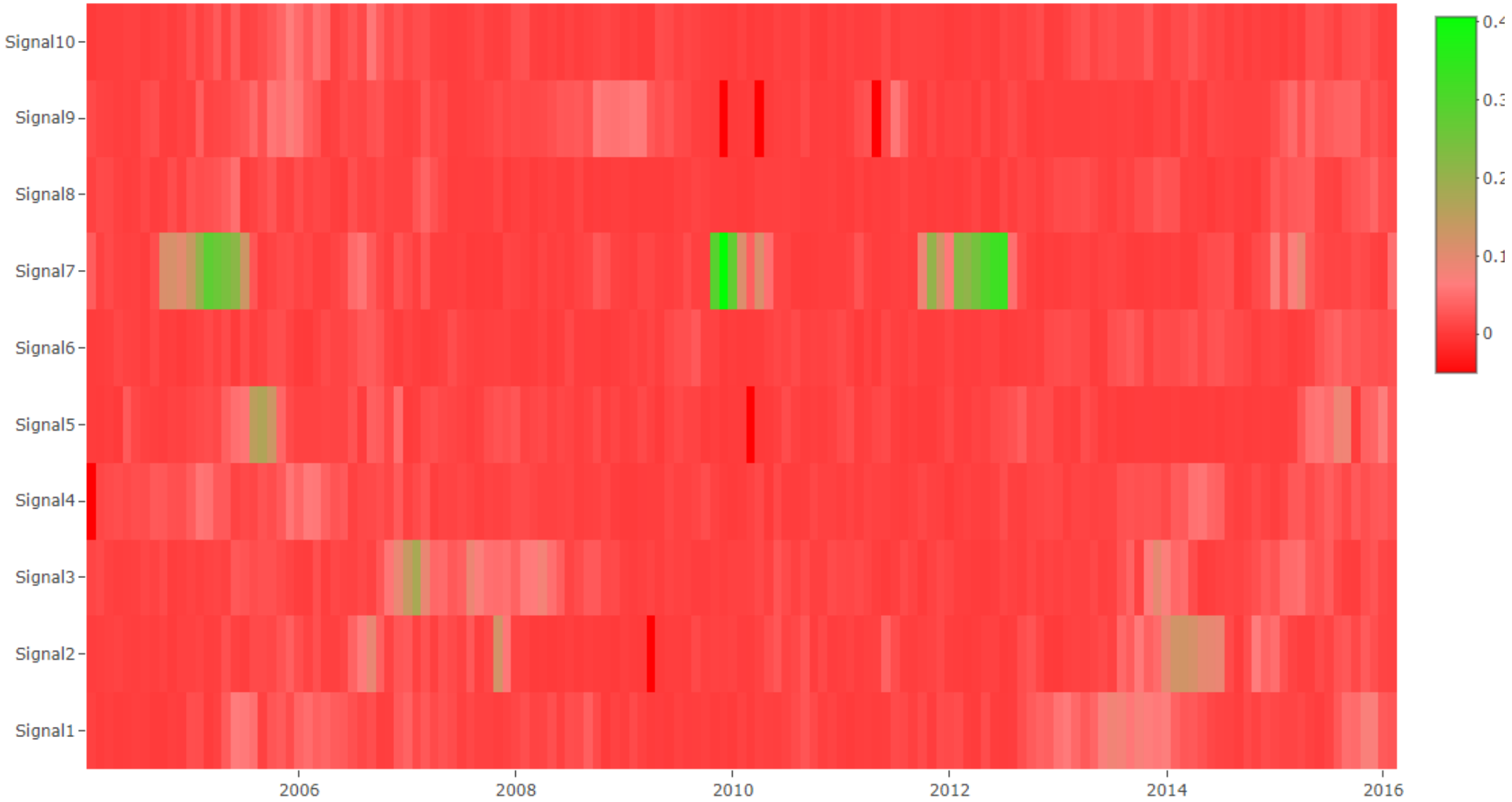
Recall: **Tp / (Tp + Fn)**
Recall could be defined as a true rate, since we include the instances that have been wrongly classified in negative.

Accuracy: **(Tp + Tf) / (Tp+Tn+Fp+Fn)**
This is the accuracy level used in the cross validation part.

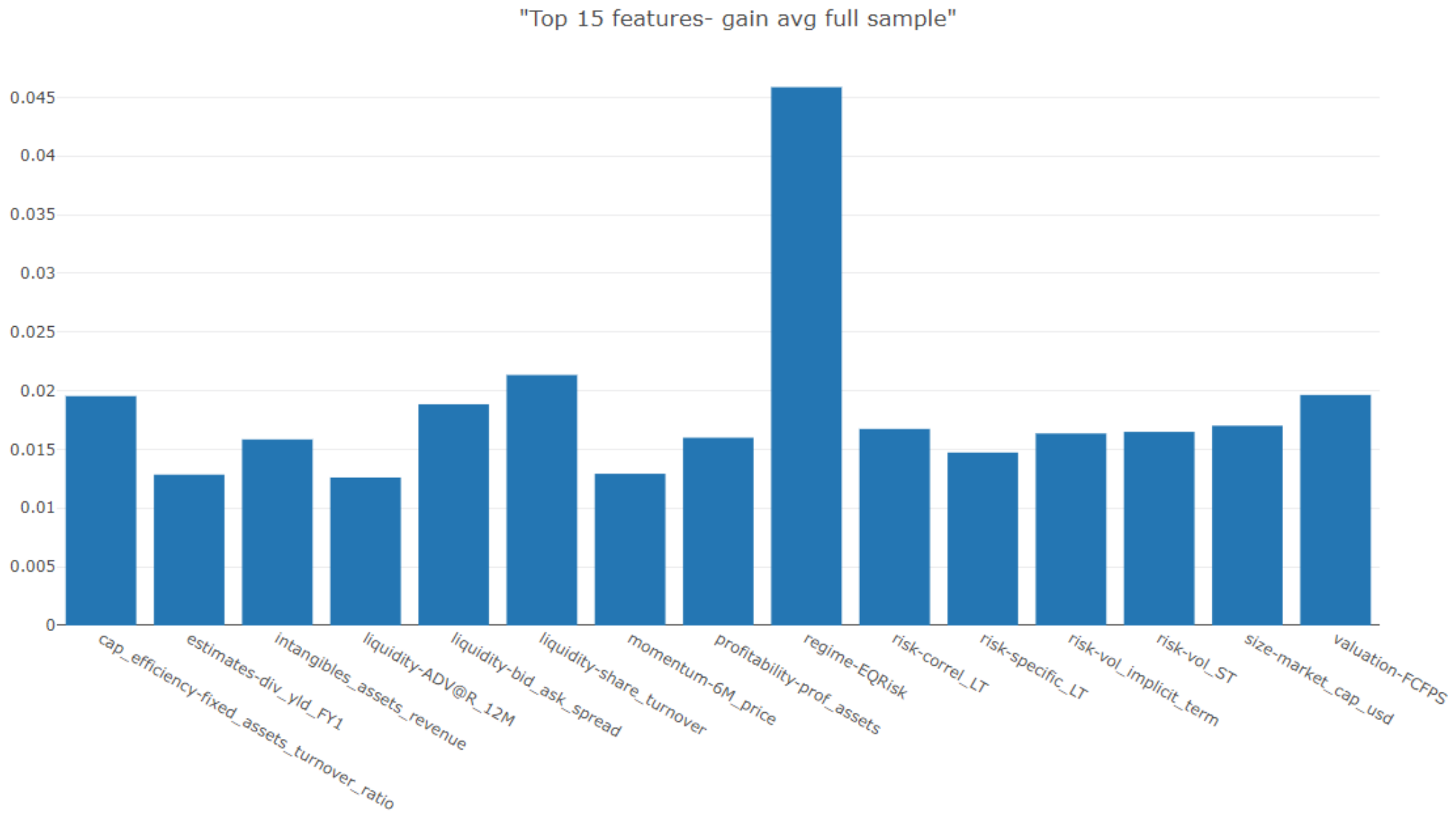
Variable importance rolling model: 100 signals left



Variable importance rolling model : regime based signal

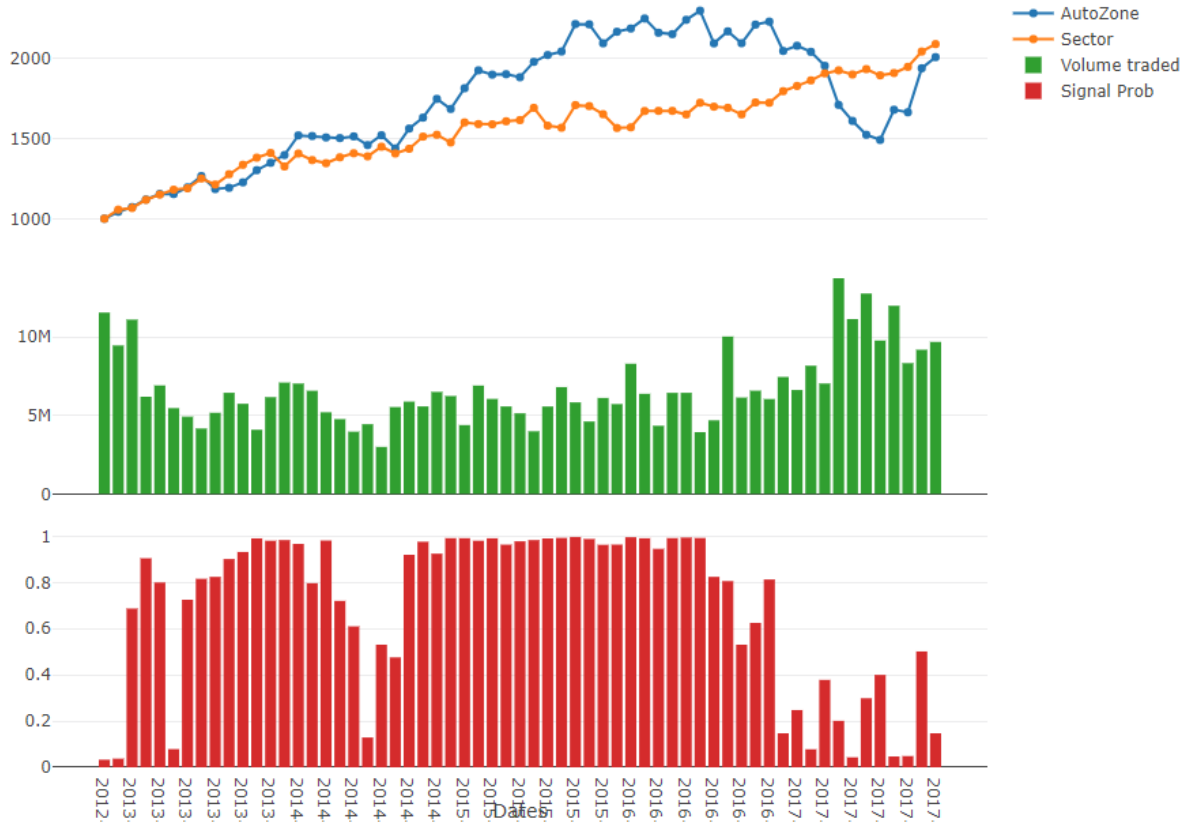


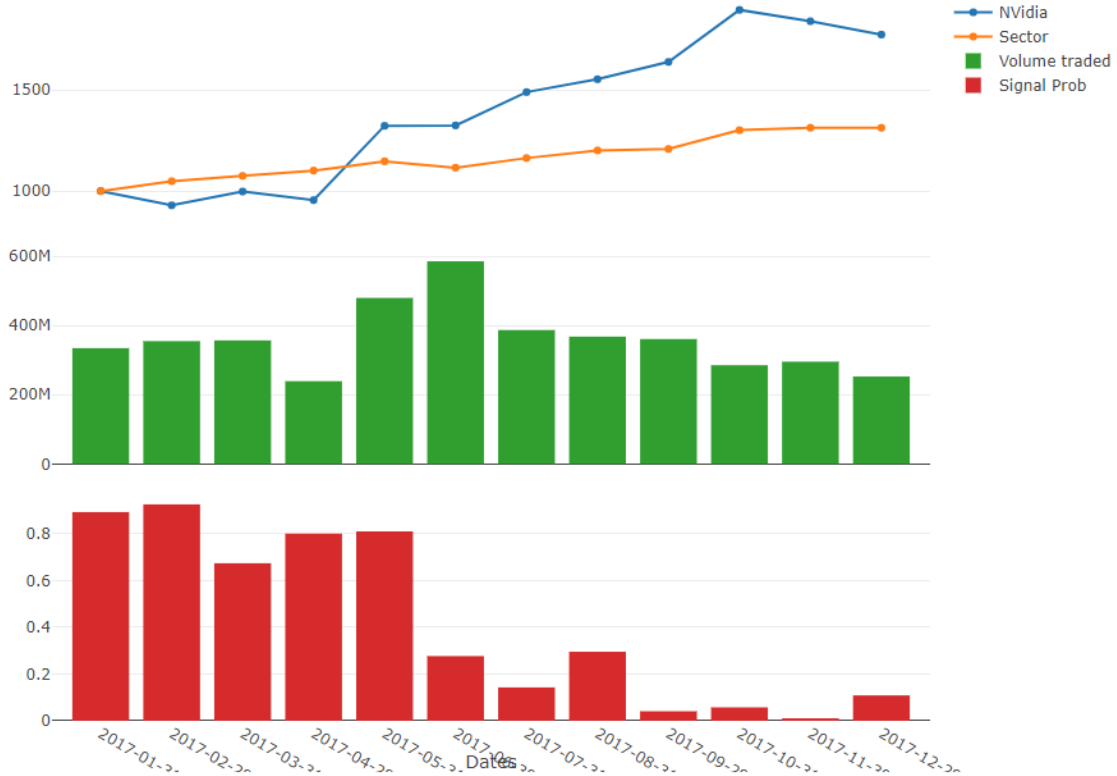
Variable importance rolling model



Implementing in a Quant Process

Trading single stock using the ML signal?

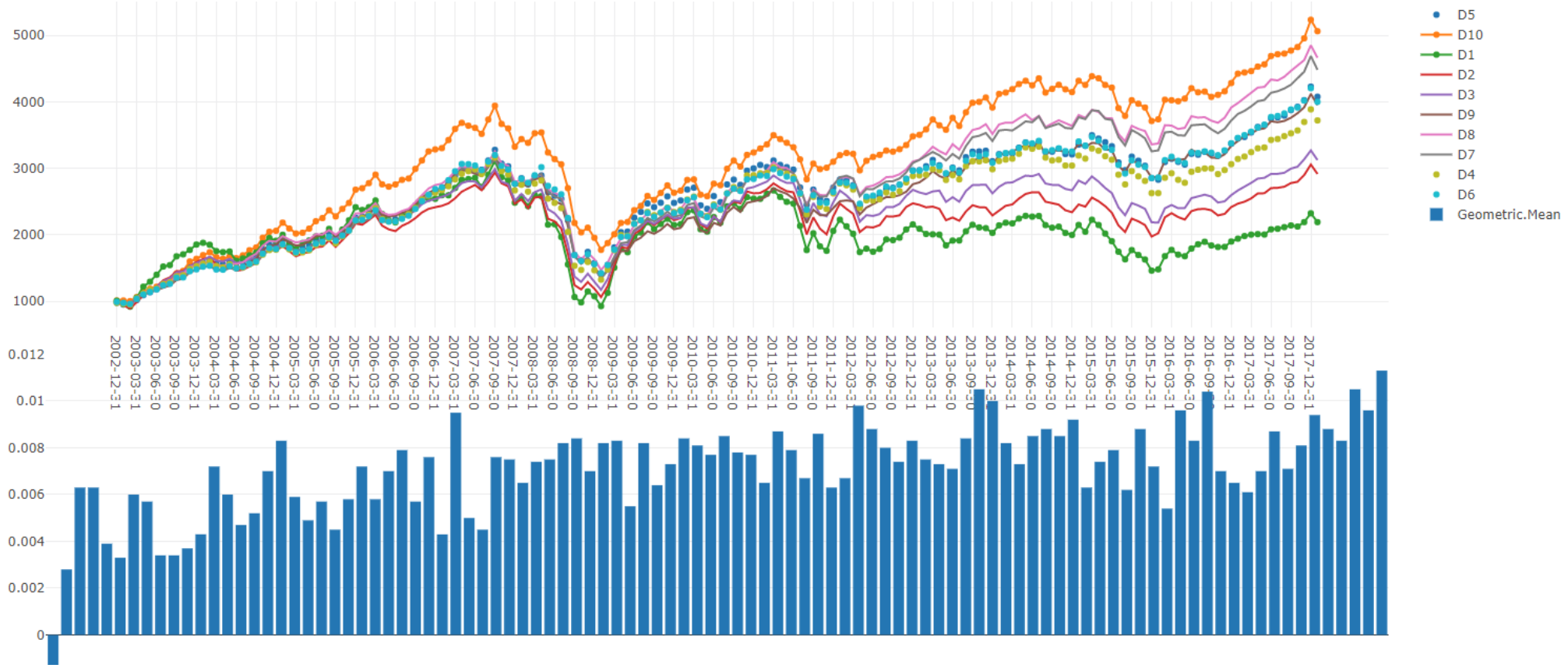




Implementation : ML vs traditional signal blending approach

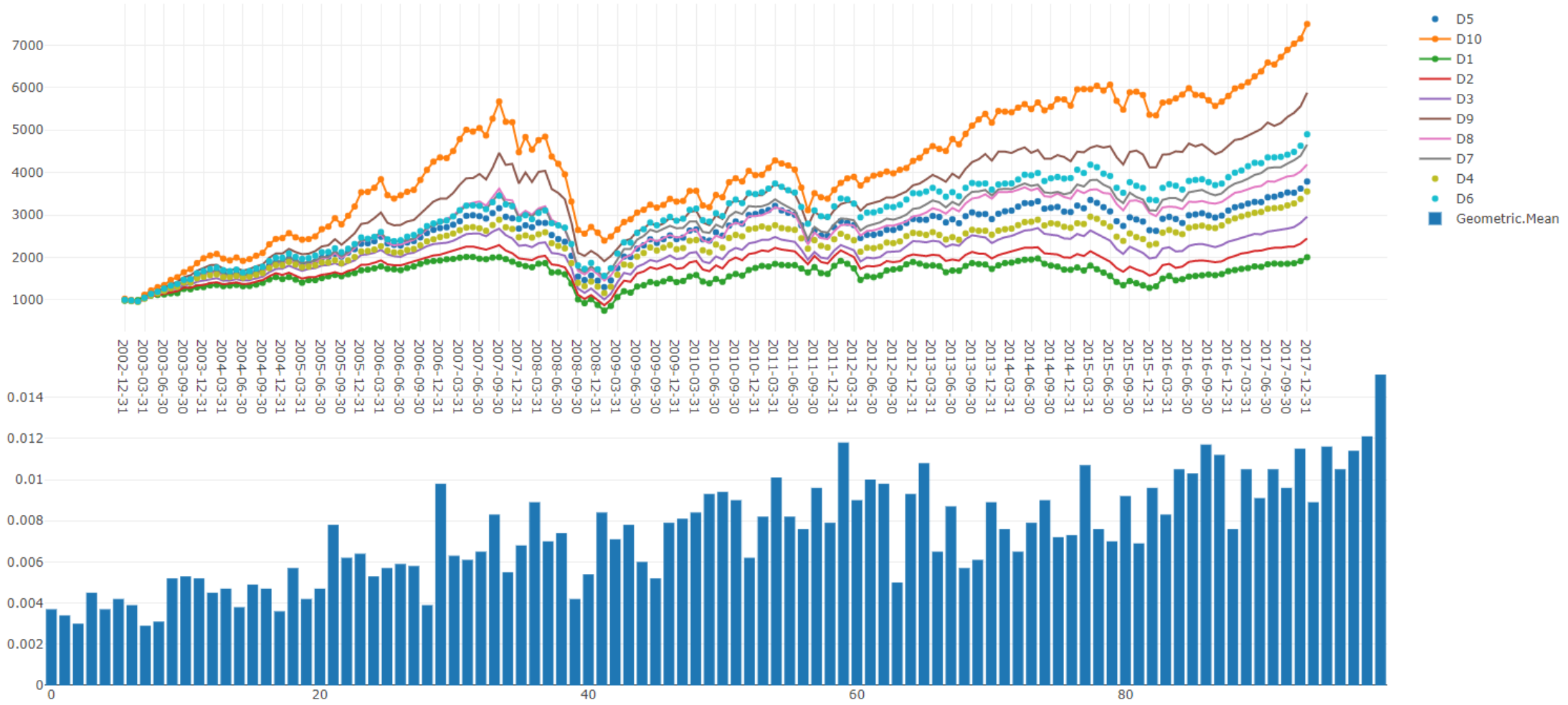
- **EW portfolio** based on the top decile of the ML signal
- Comparing against a simple linear average of composite factor signal **(Mfactor)** made of
    - **Quality** : ROE, Debt/Equity
    - **Volatility**: 3Y and 1Y (weekly returns)
    - **Value**: PE, PB, EV/EBITDA
    - **Momentum:12-1** returns
    - **Size** : Market cap

Multi-factor signal analysis: Decile performance



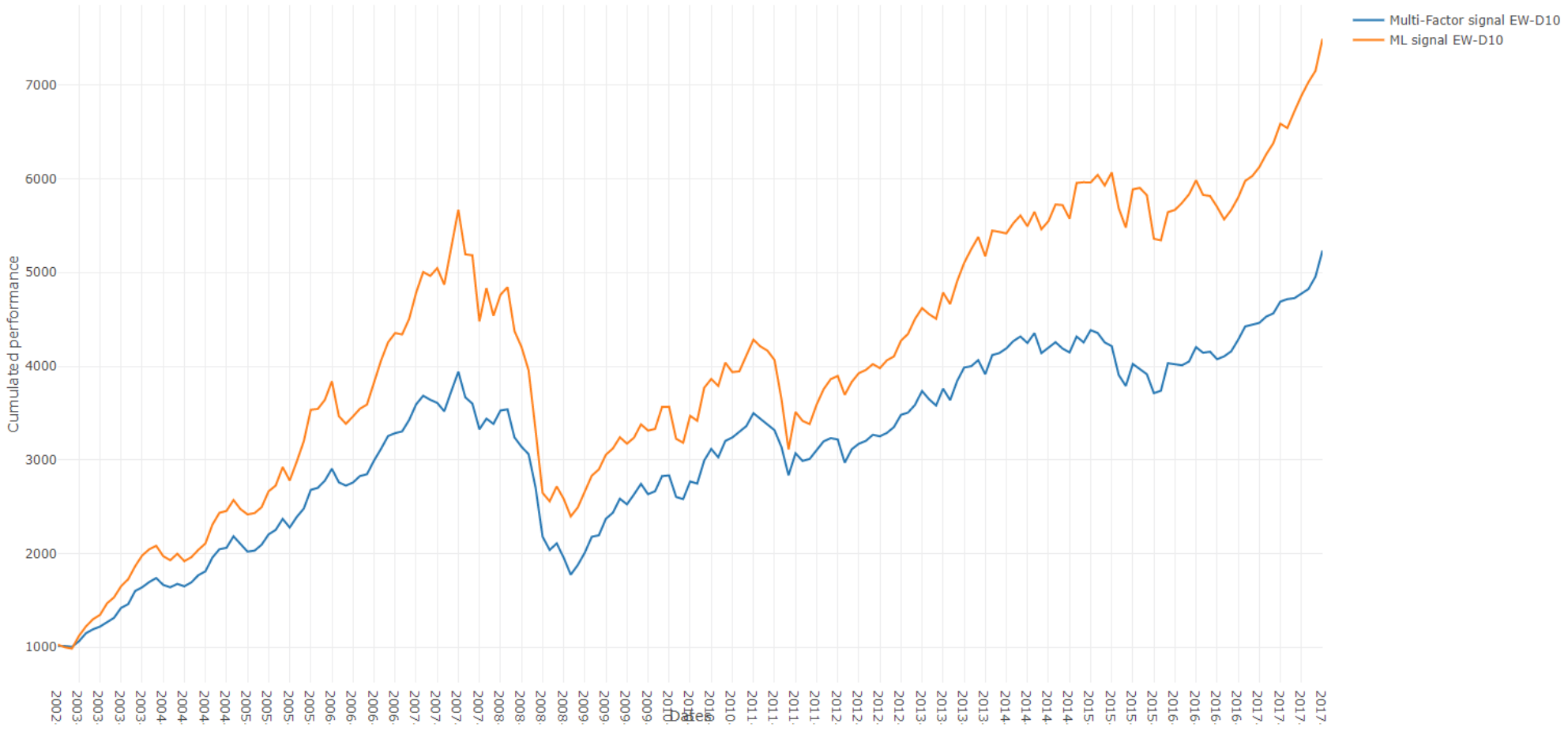
D10 portfolio: tstats =2.69, return p.a.=11.8%, ann. Vol=14.8%, Turn p.a. (2-ways)=123%, TC=62bps p.a.

ML signal analysis: Decile performance



D10 Portfolio: tstats =2.99, return p.a.=16.9%, ann. Vol=16.8%, Turn p.a. (2-ways)=192%, TC=85bps p.a.

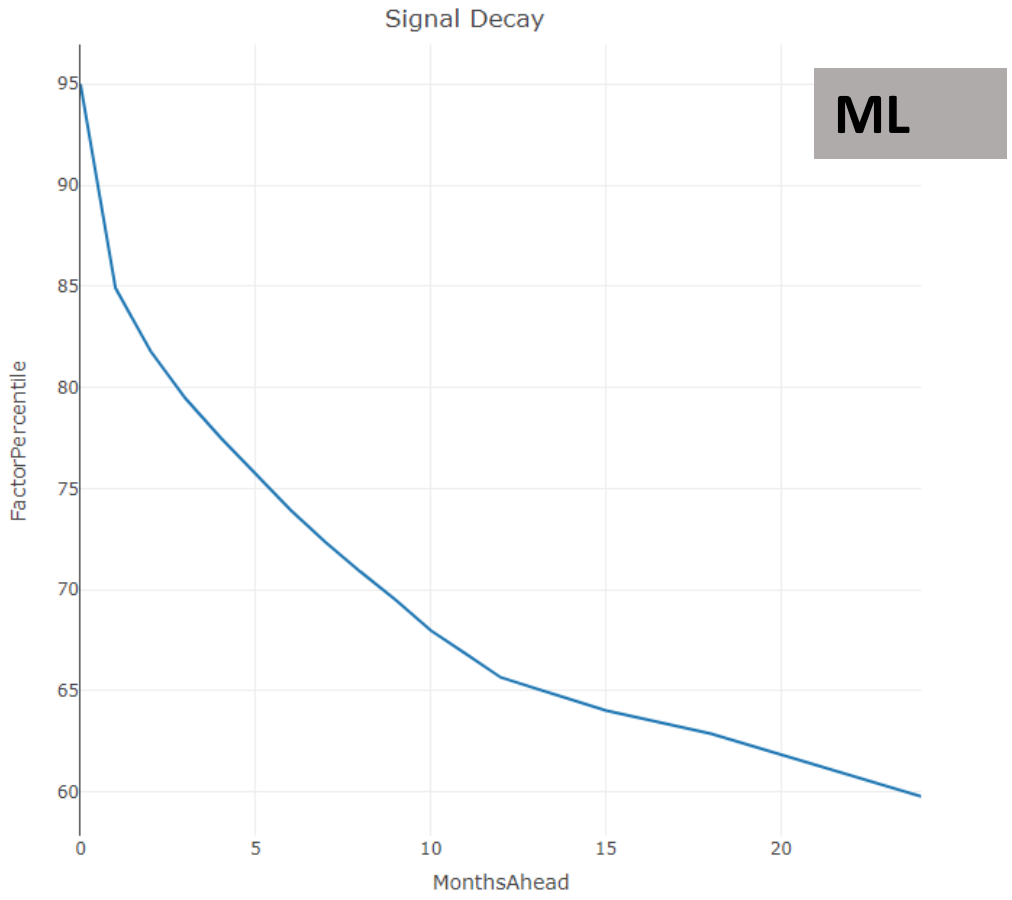
Portfolios' performance comparison

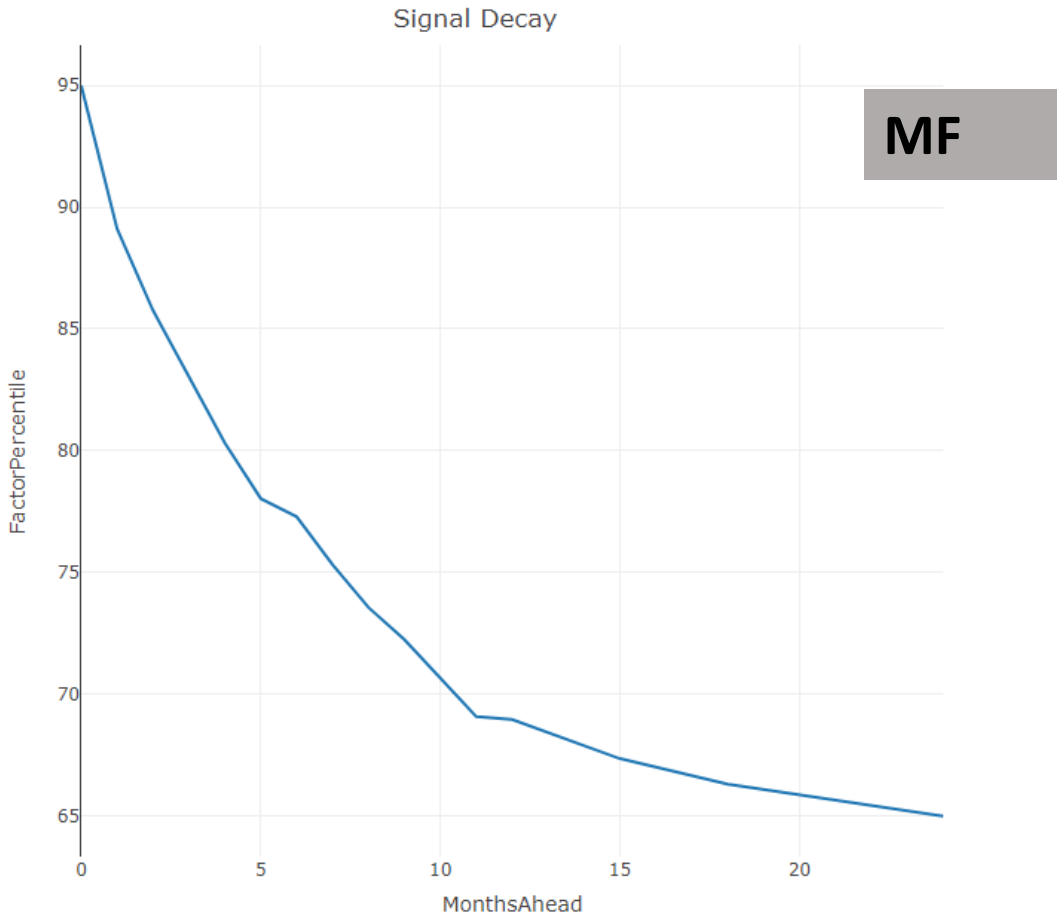


Portfolio ML: tstats =2.99, return p.a.=16.9%, ann. Vol=16.8%, Turnover p.a. (2-ways)=192%

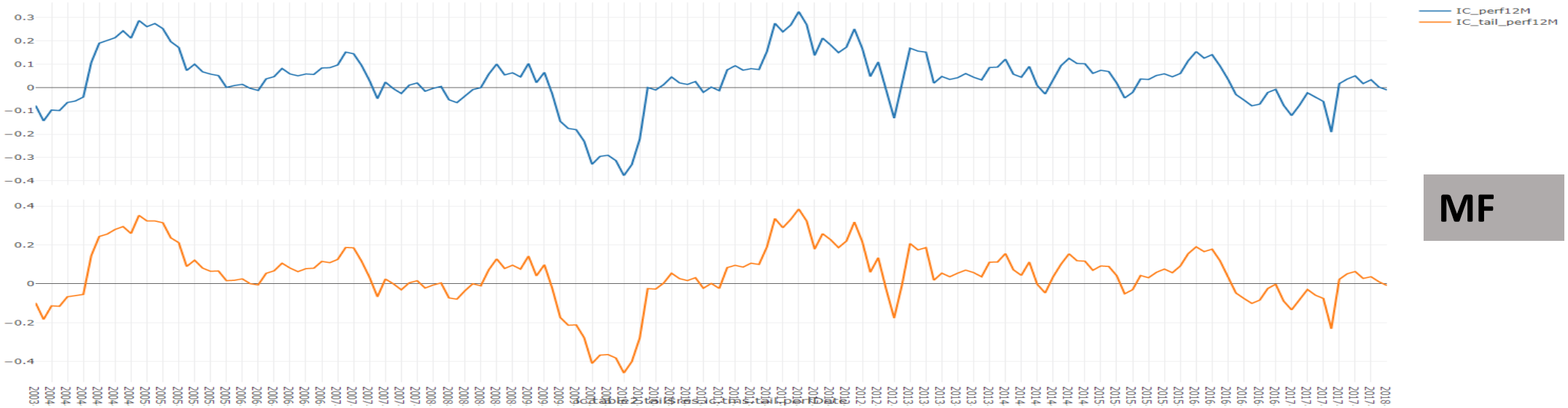
Portfolio MF: tstats =2.69, return p.a.=11.8%, ann. Vol=14.8%, Turnover p.a. (2-ways)=123%

Analysing signal's decay

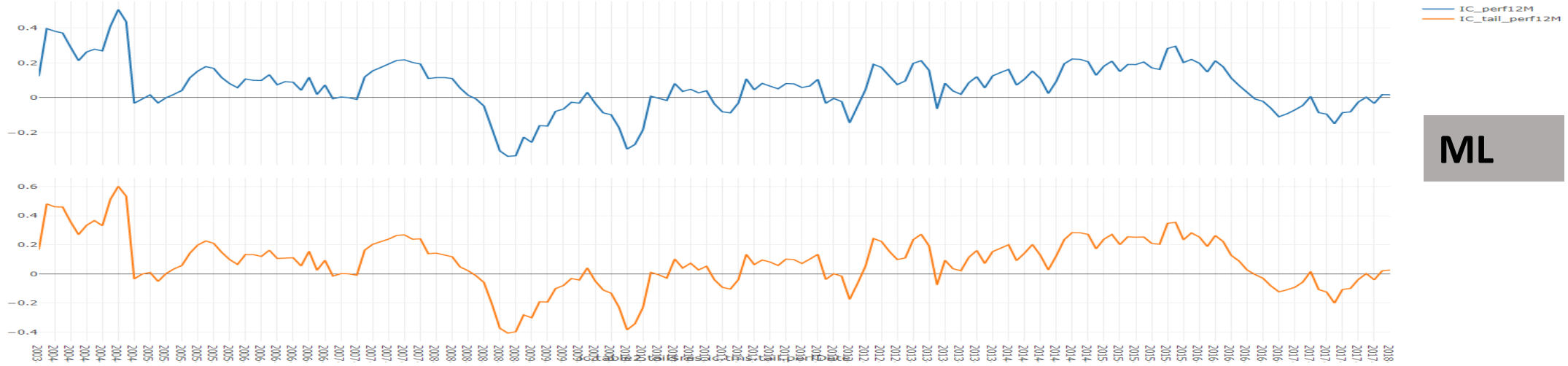




Information coefficient

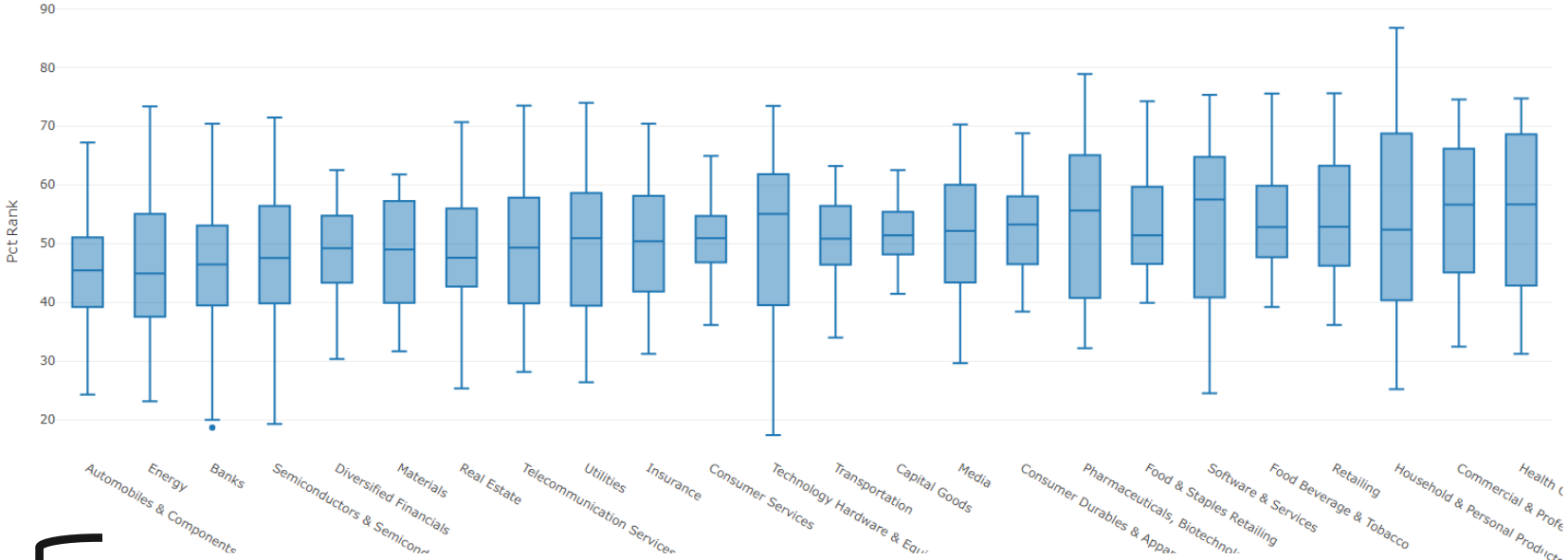


MF

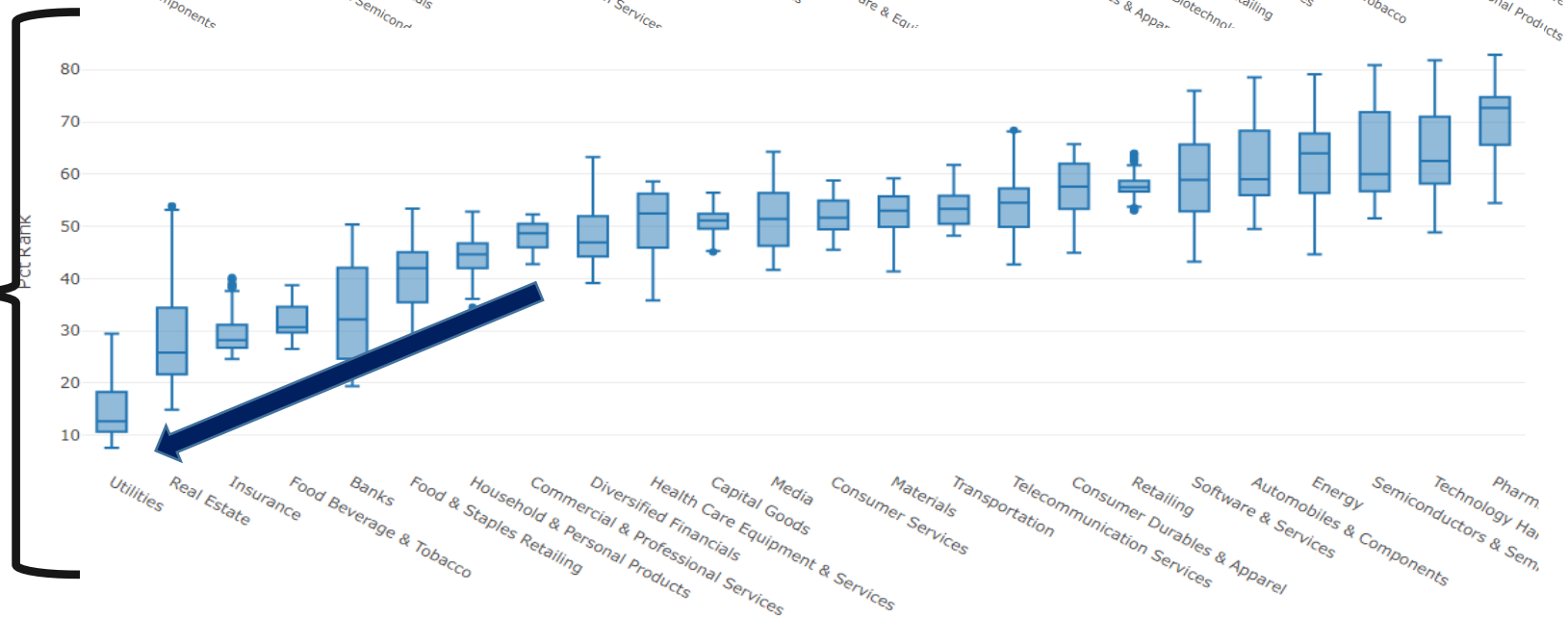


ML

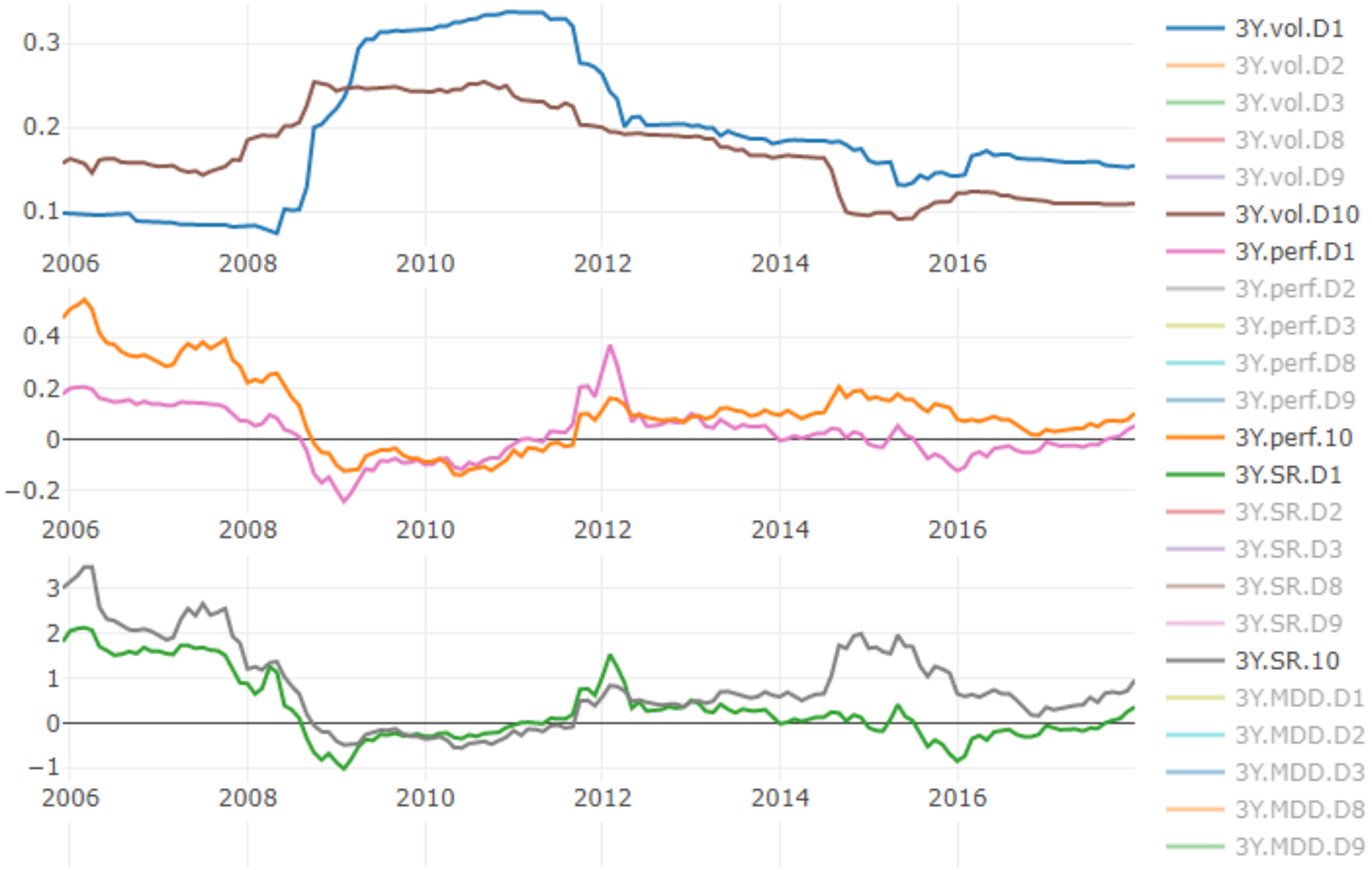
Comparing potential sector bias



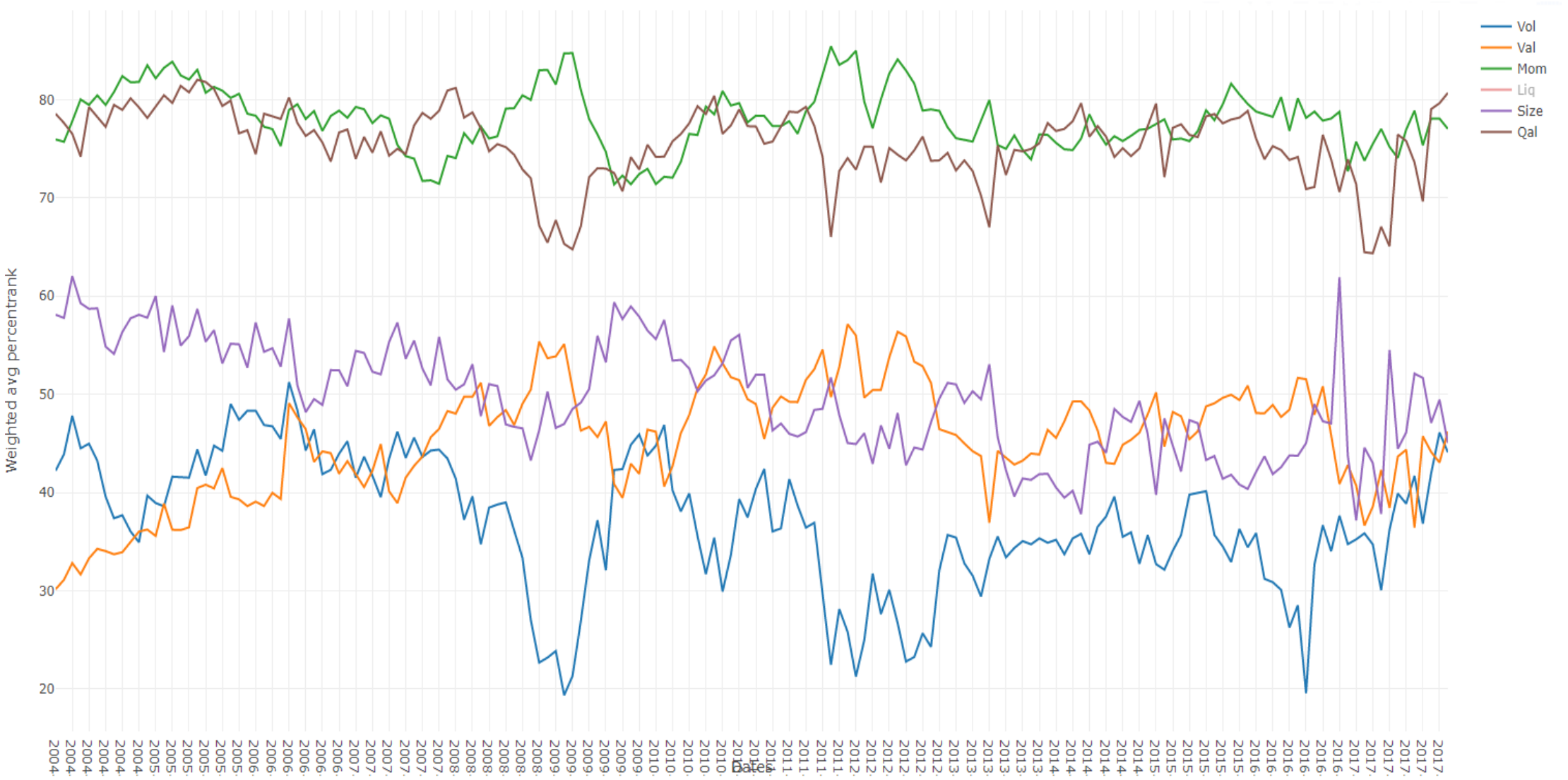
Example for a Low Vol signal



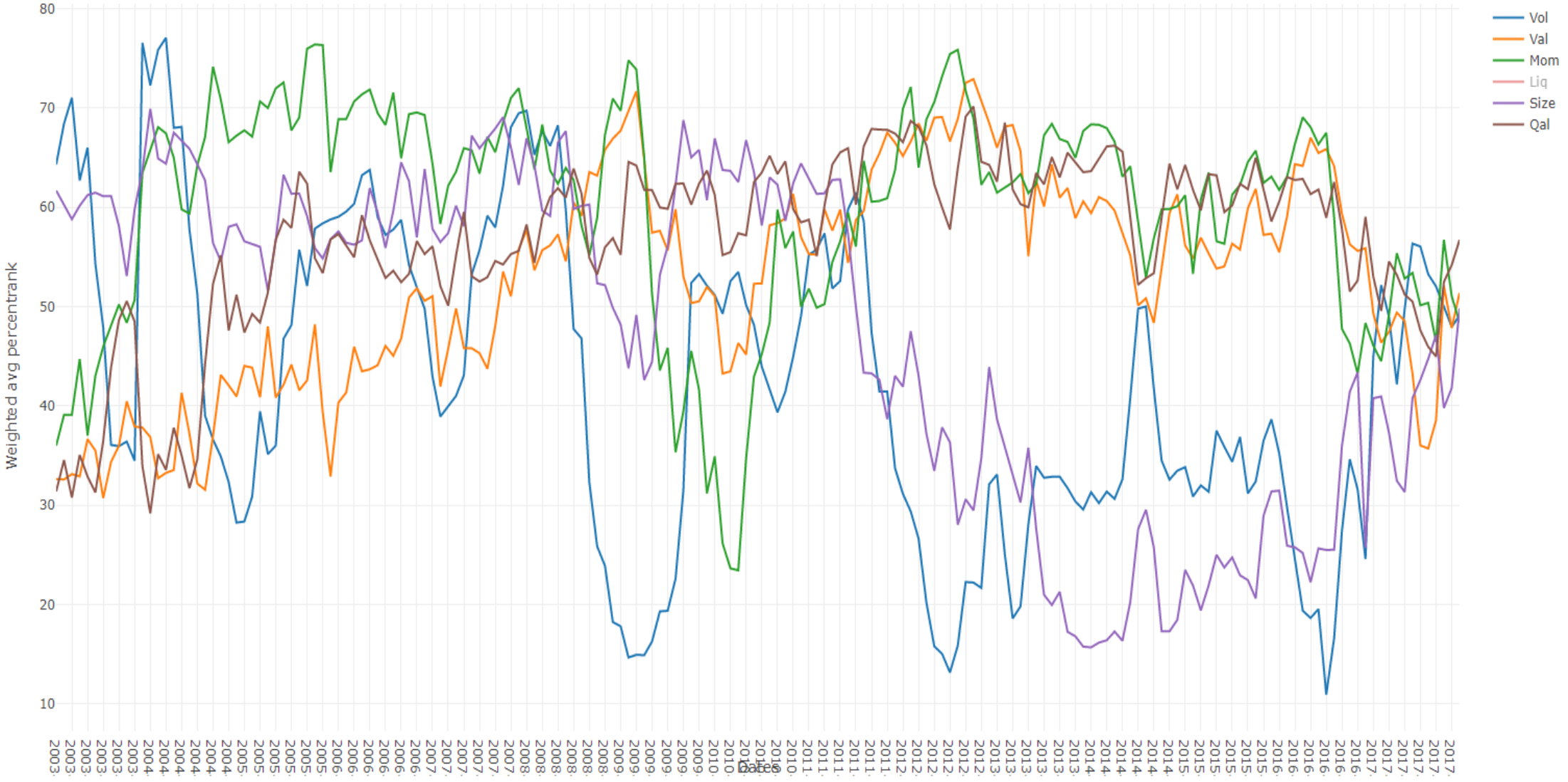
3Y Rolling analytics for EW portfolios based on ML signal



Holdings-based factor exposure: multi factor based portfolio (D10)



Holdings-based factor exposure: ML based portfolio (D10)



Conclusion

- **ML in finance could not be applied as a pure copy paste from computer science.**

- **Yet, when properly addressed, ML algorithms could be extremely useful in <u>any</u> Quant process.**

- **By applying features engineering on core financial data <u>and</u> using more « orthogonal » datasets (NLP based, alternative data..) ML based modelling has a huge potential for improvement.**

- **Embracing modern approaches like ML and alternative/big data will be soon a matter of "survival" for innovation.**

Disclaimer & source

The views expressed in this document are the author's and do not necessarily reflect those of the organizations he is affiliated with.
No investment decision or particular course of action is recommended by this presentation from the author

Source: QuanDL, Alphavantage, R & Rstudio, Python & Spyder, XGBOOST,