



Improving Equity Style Investing using Machine Learning

Tony Guida

Important Information



This material is addressed to professional clients for informative purposes only. It is neither an offer nor an invitation to buy or sell investment products and may not be interpreted as investment advice. It is not intended to be distributed, published or used in a jurisdiction where such distribution, publication or use is forbidden, and is not intended for any person or entity to whom or to which it would be illegal to address such a material. In particular, investment products are not offered for sale in the United States or its territories and possessions, nor to any US person (citizens or residents of the United States of America). The opinions herein do not consider individual clients' circumstances, objectives, or needs. Before entering into any transaction, clients are advised to form their own opinion and consult professional advisors to obtain an independent review of the specific risks incurred (tax, financial etc.). Upon request, RAM AI Group is available to provide more information to clients on risks associated with investments. The information and analysis contained herein are based on sources deemed reliable. However, RAM AI Group does not guarantee their accuracy, correctness or completeness, and it does not accept any liability for any loss or damage resulting from their use. All information and assessments are subject to change without notice. Changes in exchange rates may cause the NAV per share in the investor's base currency to fluctuate. There is no guarantee to get back the full amount invested. Past performances, whether actual or back-tested, are not necessarily indicative of future performance. Without prejudice of the due addressee's own analysis, RAM understands that this communication should be regarded as a minor non-monetary benefit according to MIFID regulations. Clients are invited to base their investment decisions on the most recent prospectus, key investor information document (KIID) and financial reports which contain additional information relating to the investment product. These documents are available free of charge from the SICAV's and Management Company's registered offices, its representative and distributor in Switzerland, RAM Active Investments S.A. and at Macard Stein & Co AG, Paying and Information Agent in Germany; and at RAM Active Investments (Europe) SA – Succursale Milano in Italy. This marketing material has not been approved by any financial Authority, it is confidential and addressed solely to its intended recipient; its total or partial reproduction and distribution are prohibited. Issued in Switzerland by RAM Active Investments S.A. which is authorised and regulated in Switzerland by the Swiss Financial Market Supervisory Authority (FINMA). Issued in the European Union and the EEA by the Management Company RAM Active Investments (Europe) S.A., 51 av. John F. Kennedy L-1855 Luxembourg, Grand Duchy of Luxembourg. The reference to RAM AI Group includes both entities, RAM Active Investments S.A. and RAM Active Investments (Europe) S.A.

Why this topic?



Given the **exponential increase in data availability**, the obvious temptation of any asset manager is to try to infer future returns from the abundance of attributes available at the firm level.

Current computational power allows to “test” almost all types of new characteristics/signals.

Sharing knowledge effect. Cross fertilization between Hard science and finance is increasing (implicitly and explicitly).

A need to innovate. Legacy approach for constructing Style/Factor equity portfolio has been delivering less return than 10 years ago.

What can we expect from ML in Style Investing?



To test **more** characteristics/signals

To leverage on **non-linear** complex patterns, rule based

To **adapt and identify** to trends by re-running models

To **ensemble** more models, wisdom of the crowd

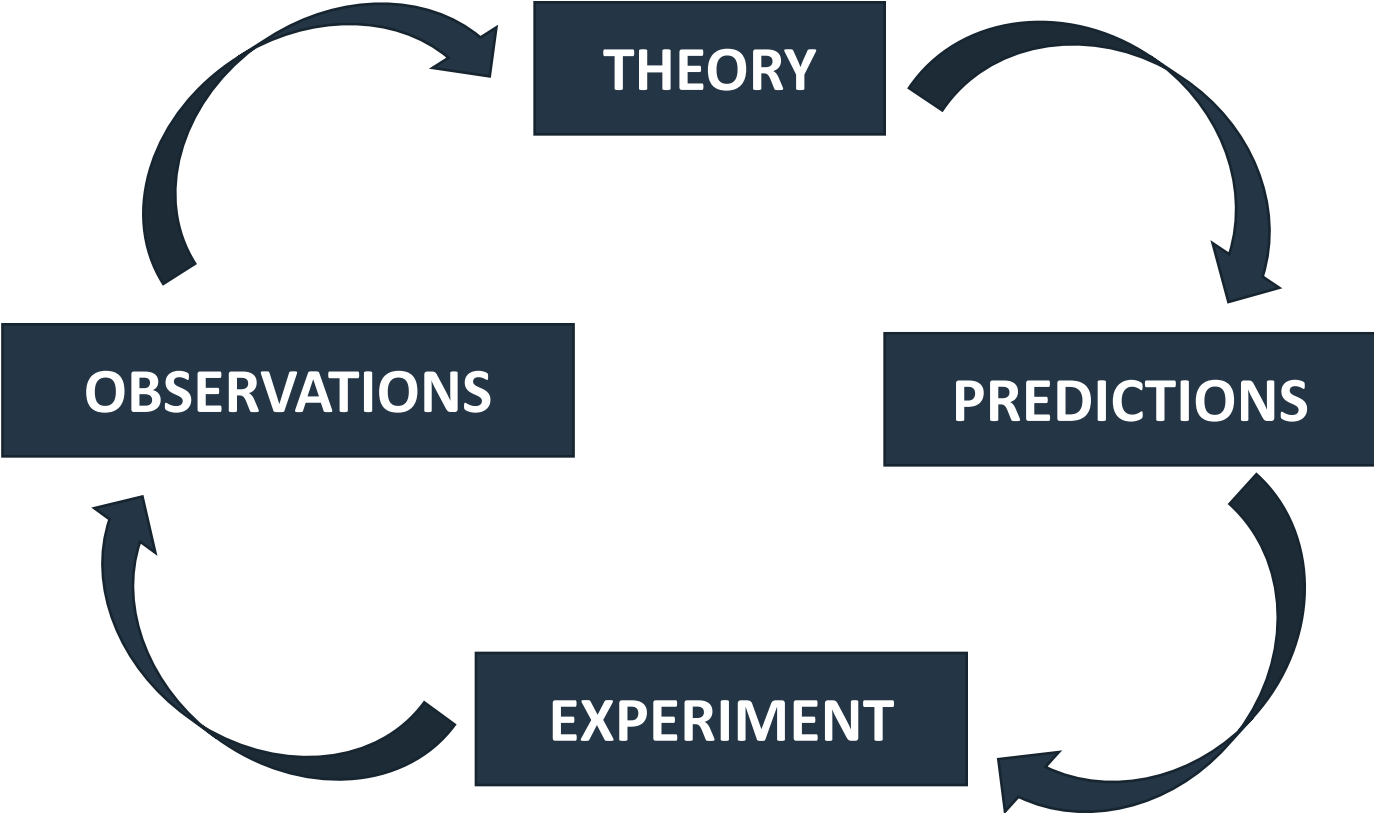
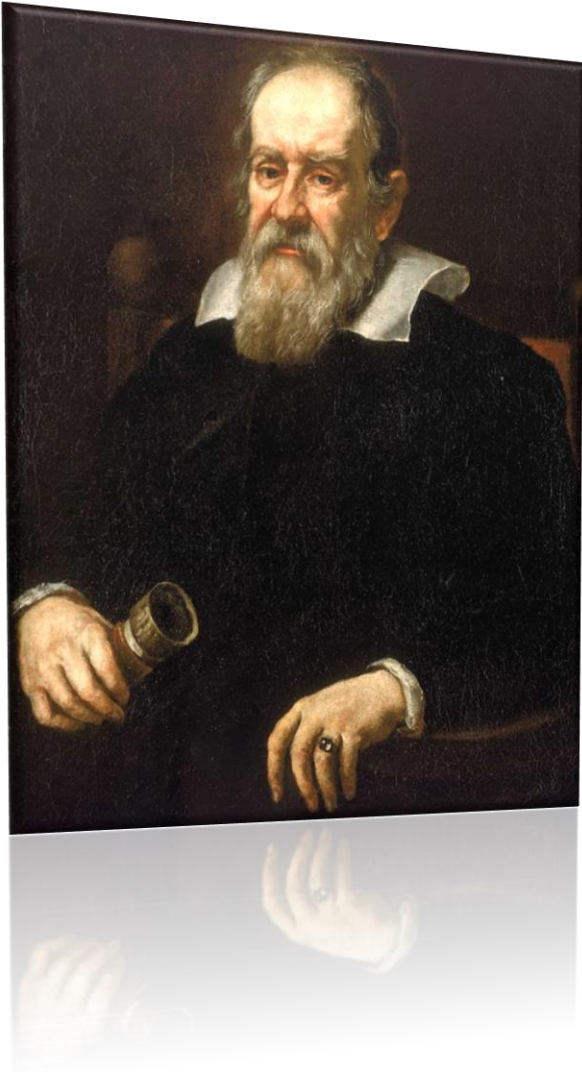
To be **less biased** than trad. dogmatic quant. approach



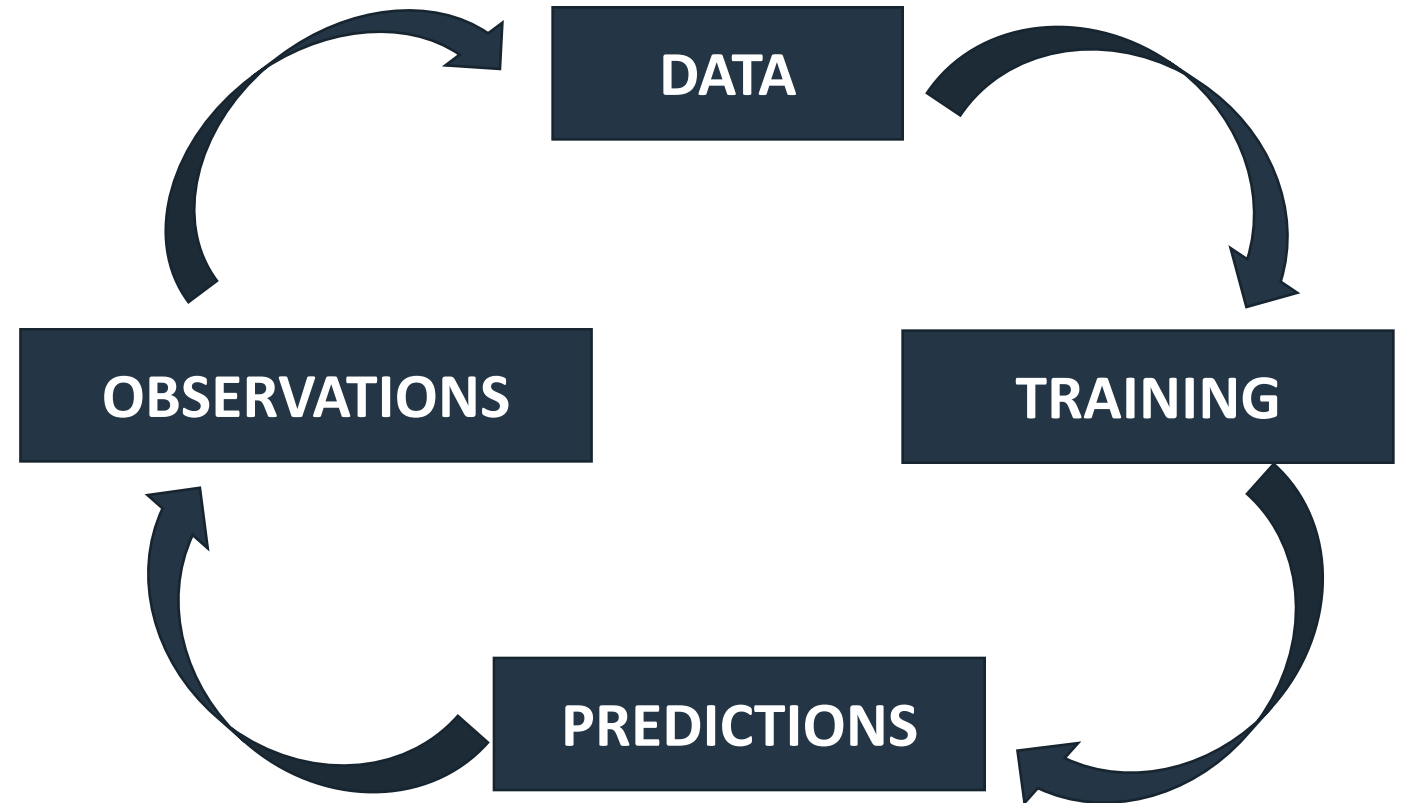
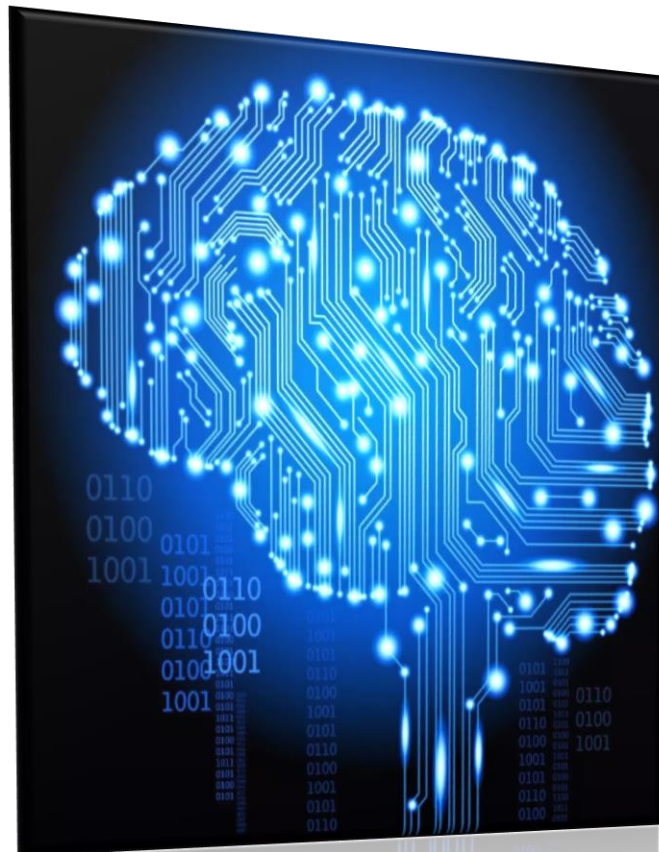
CASE STUDY :

Data “the” key element

A bit of Epistemology



A New Way for Research





CASE STUDY :

Concept and protocol

Protocol for ML



We will predict **1M future** performance

We will use a boosted tree classification ML model

Our Investment universe is composed of **global stocks including EM (~1700)**

Full dataset from Dec-1999 until May-2020. **Style datasets** are sub parts of the full one.

Stocks are filtered according absolute and relative metrics for **MCAP** and **ADV**.

Data engineering for training is based on:

- *Training on tails (extreme quantile from Label/fit cross section) training*
- *Outliers removals (from label and features)*
- *Low-coverage instance (row) removal*
- *Low-coverage feature (column) removal*

(~ 200) features, monthly normalised in percentile

We use a rolling window of 5 years- **80% Training 20% Testing**

From the ML output (probability of outperforming) we create a signal and we construct portfolio from top/bottom decile (around 150 stocks in each portfolio).

Implementation ML vs trad. Signal blending

EW portfolios based on ML alpha. Selecting top decile.

Comparing against simple linear average of composite factor MF made of Profitability/Quality, Volatility, Value, Momentum and Size. **TRAD VS MODERN**

Comparing against linear average of the top 15 most important feature. **PROXY FOR NON LINEARITY ADDED VALUE.**

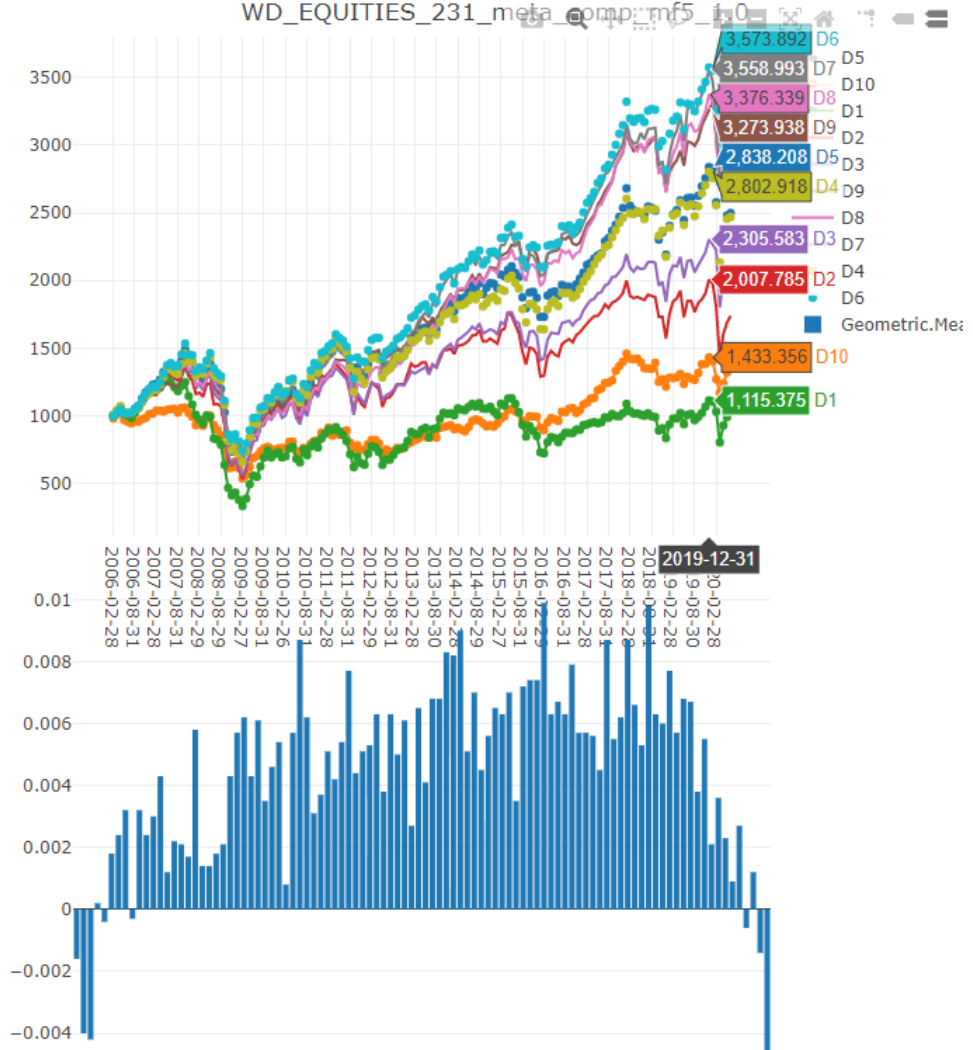
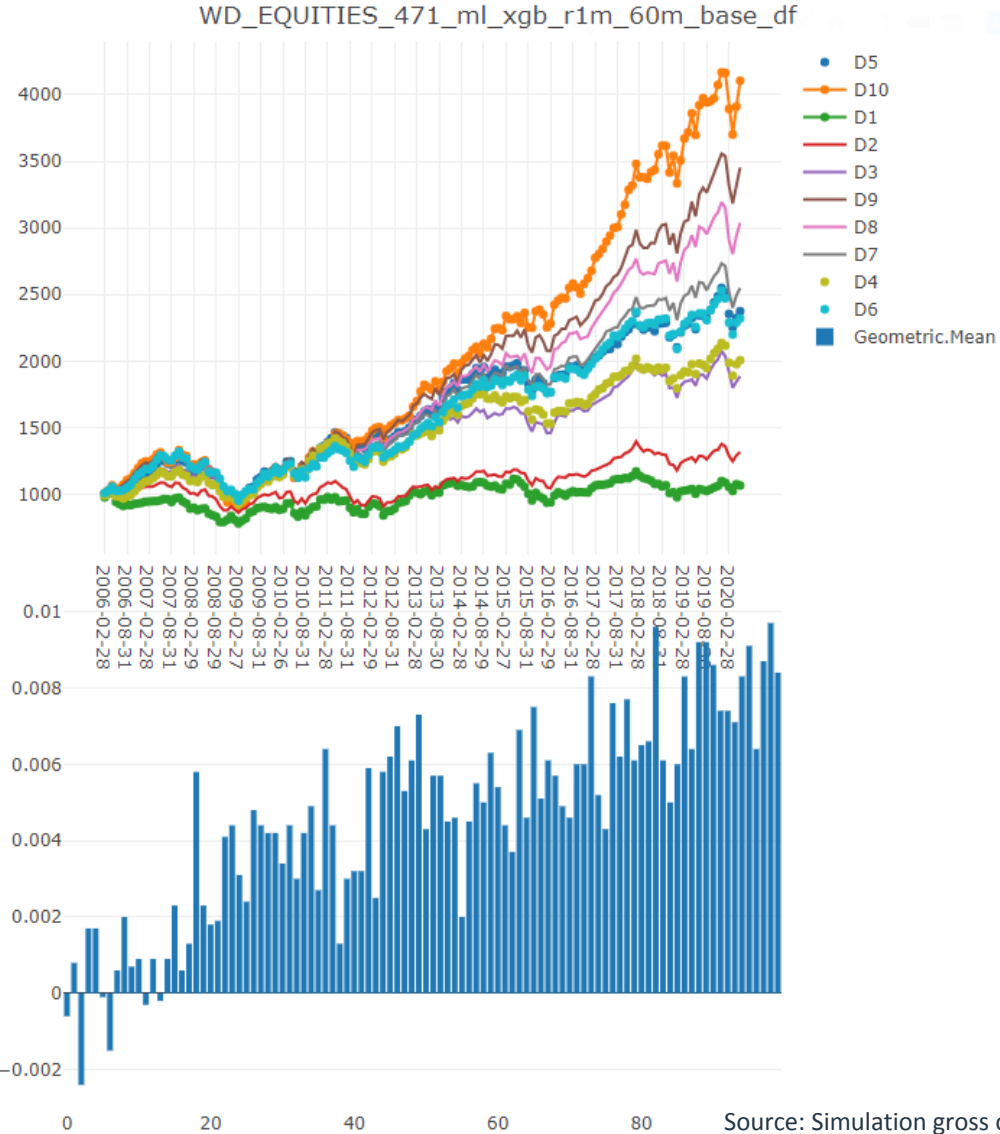
ALL ML models in this case study have the SAME hyperparameters



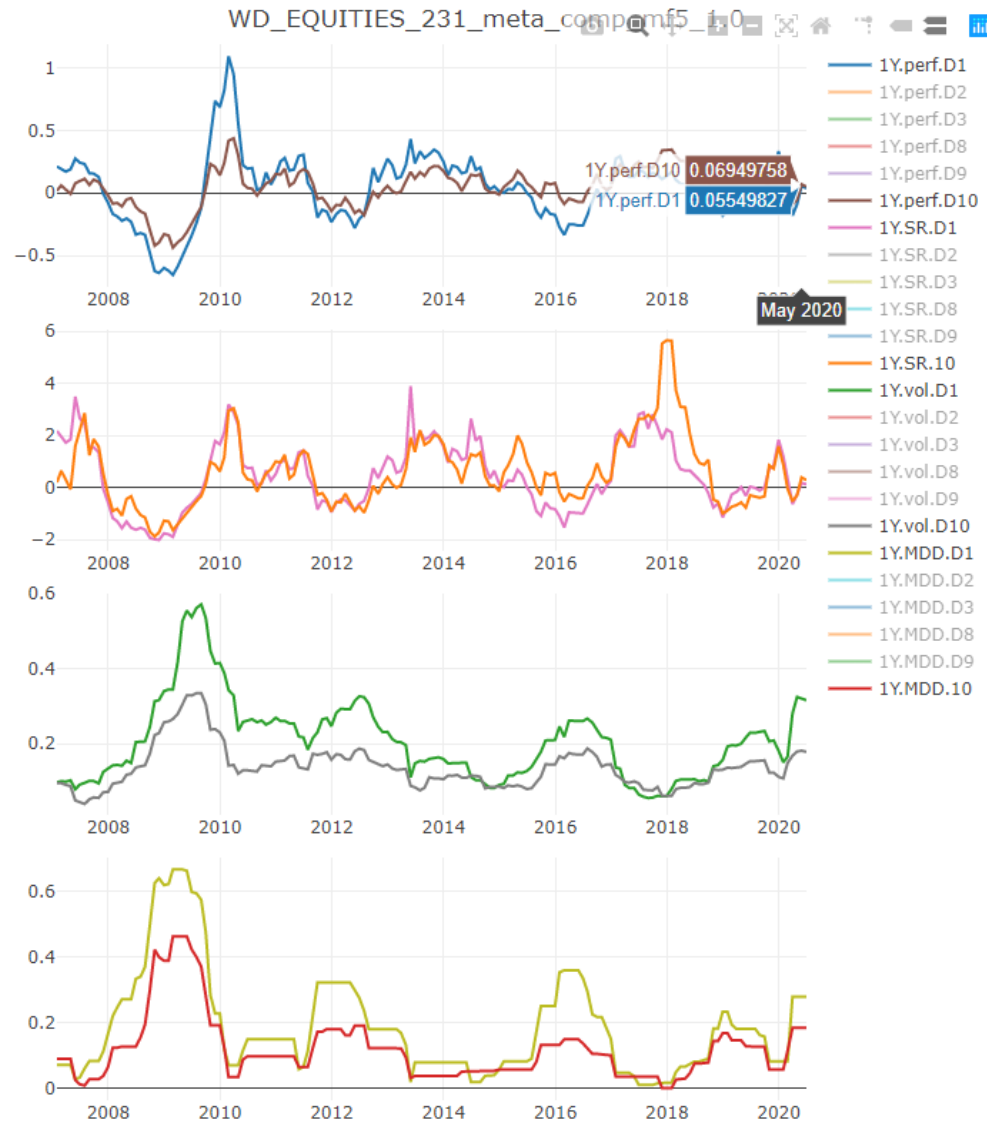
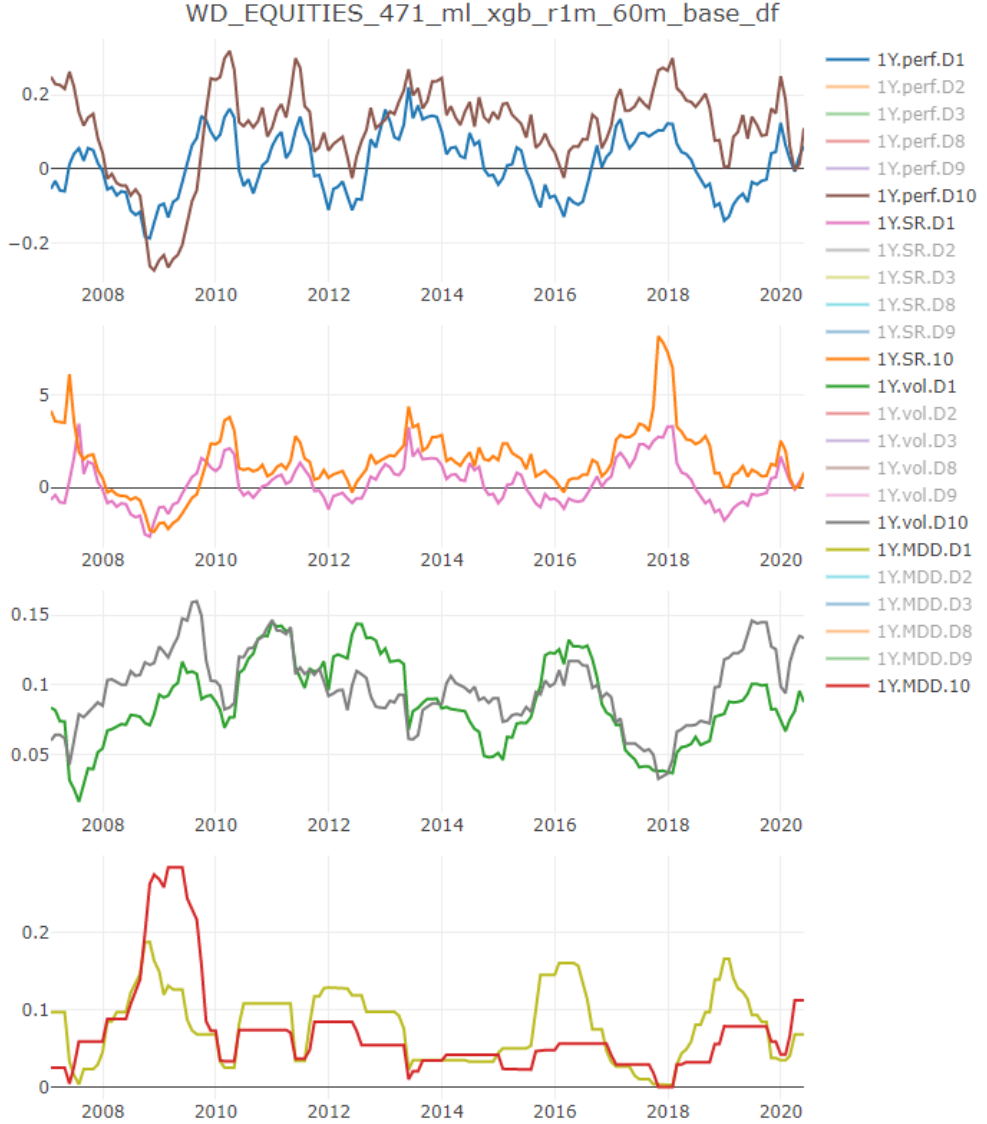
Check #:

ML vs MF

Comparing Monotonicity

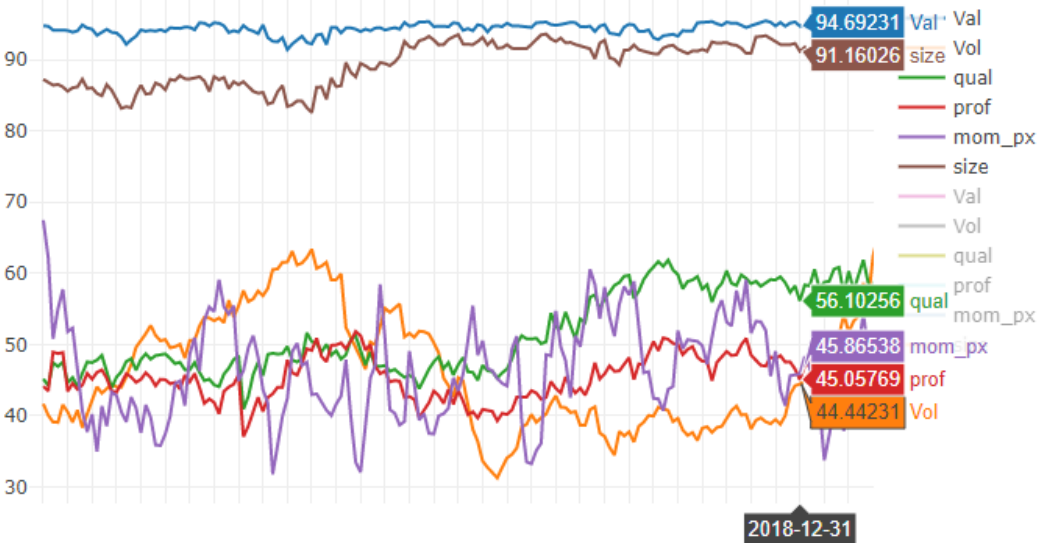
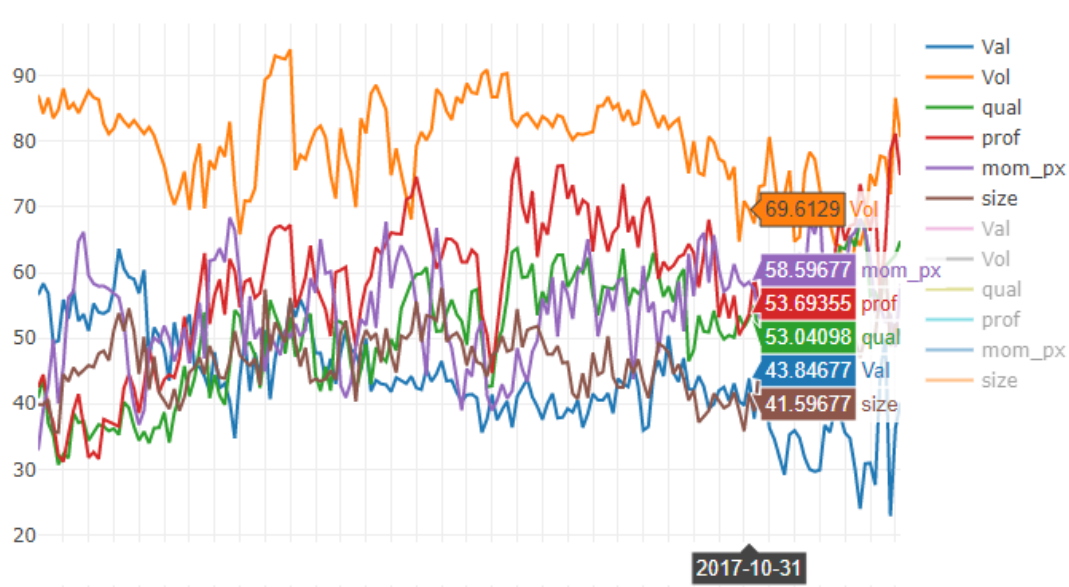


Comparing Performance



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

Comparing Factor Exposure



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

ML vs MF: What can we learn from that?



Basic unconditioned linear signal blended MF signal **show very poor results** (due also to the **complexity of the universe**). Averaged quantile performance shows an **unrewarded extreme tail**. Factors revealed the **basic small cap/Value** cluster.

Factor exposure is more stable but load on **less rewarded features**.

Level of average tail **turnover is comparable** ~25% for MF vs 32% for ML.

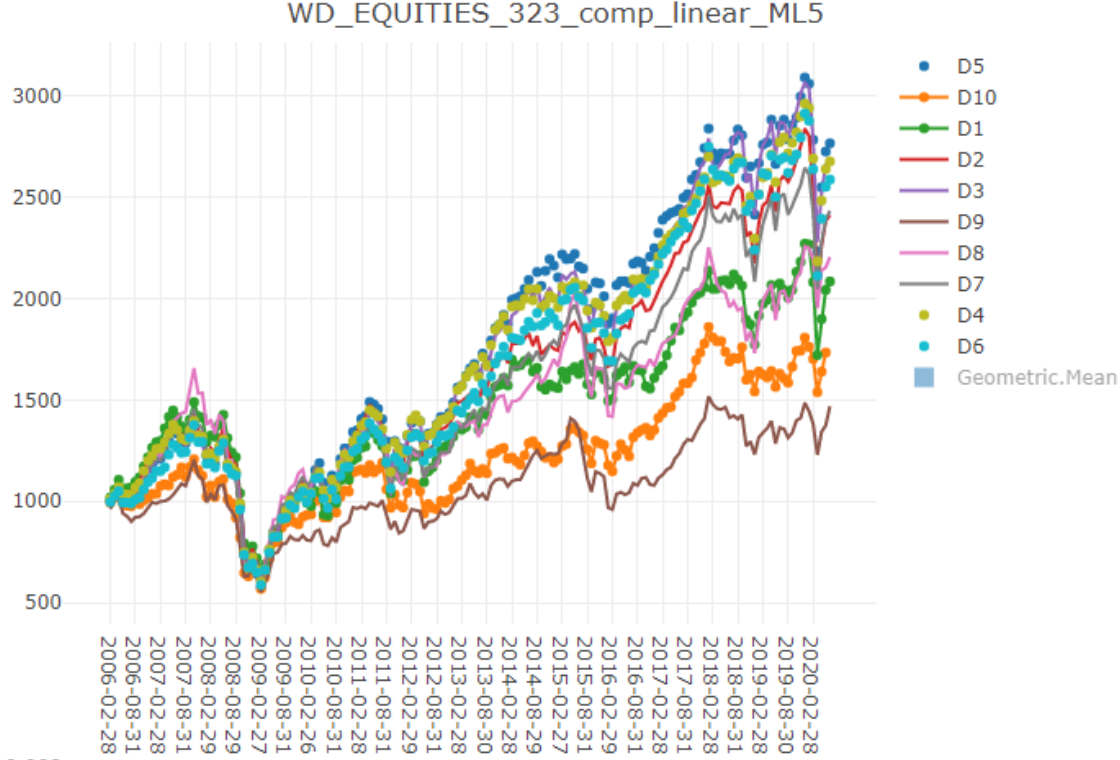
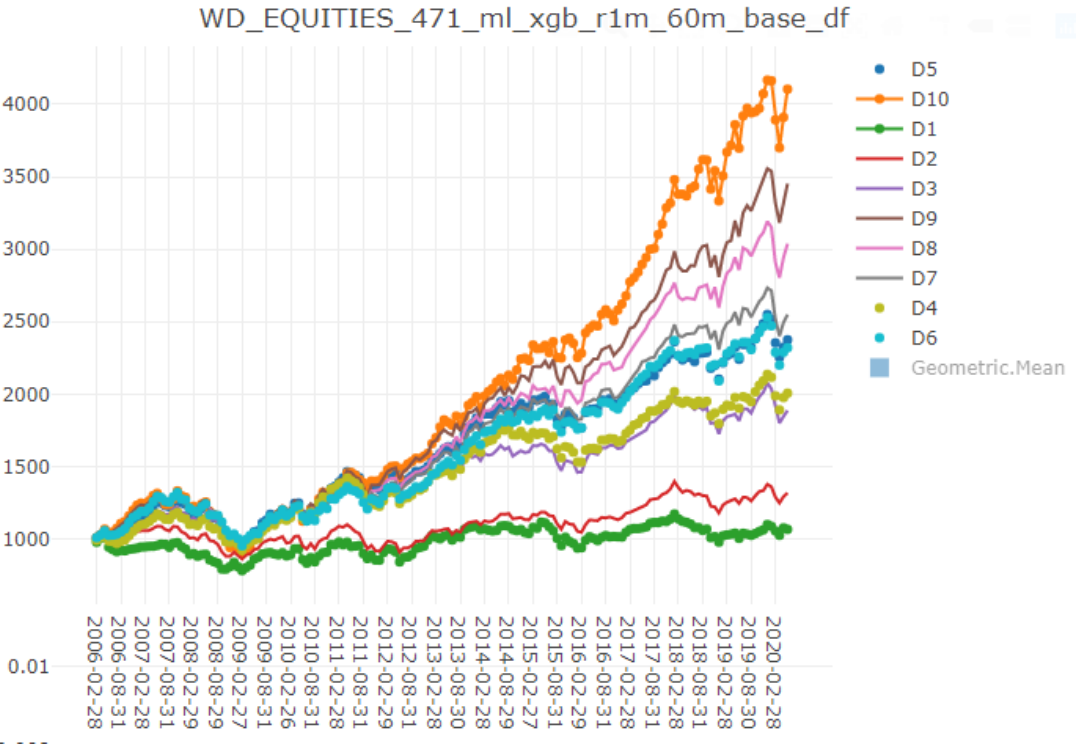
ML base case (on the left) **is superior on a risk/performance/MaxDD** standpoint.



Check #:

ML vs Linear MIF ML

Comparing Monotonicity

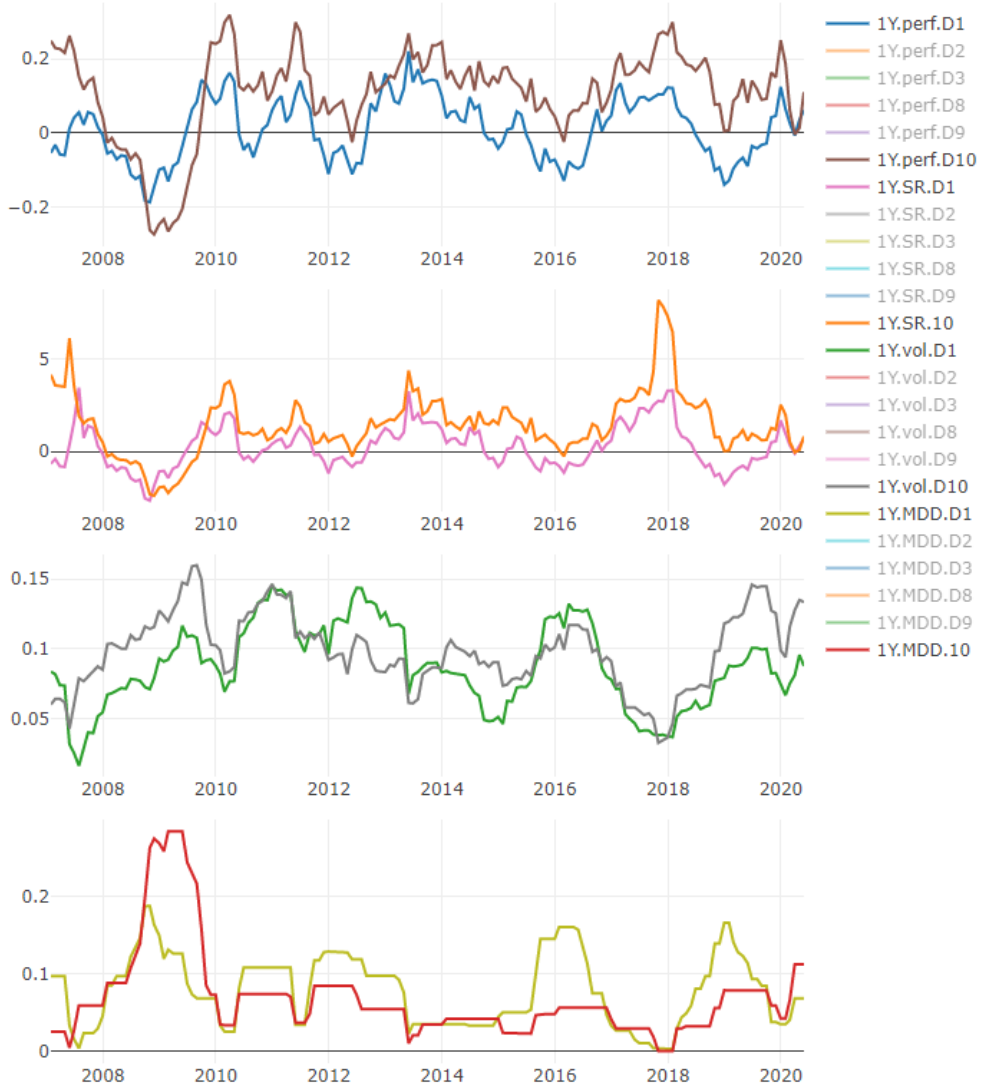


Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

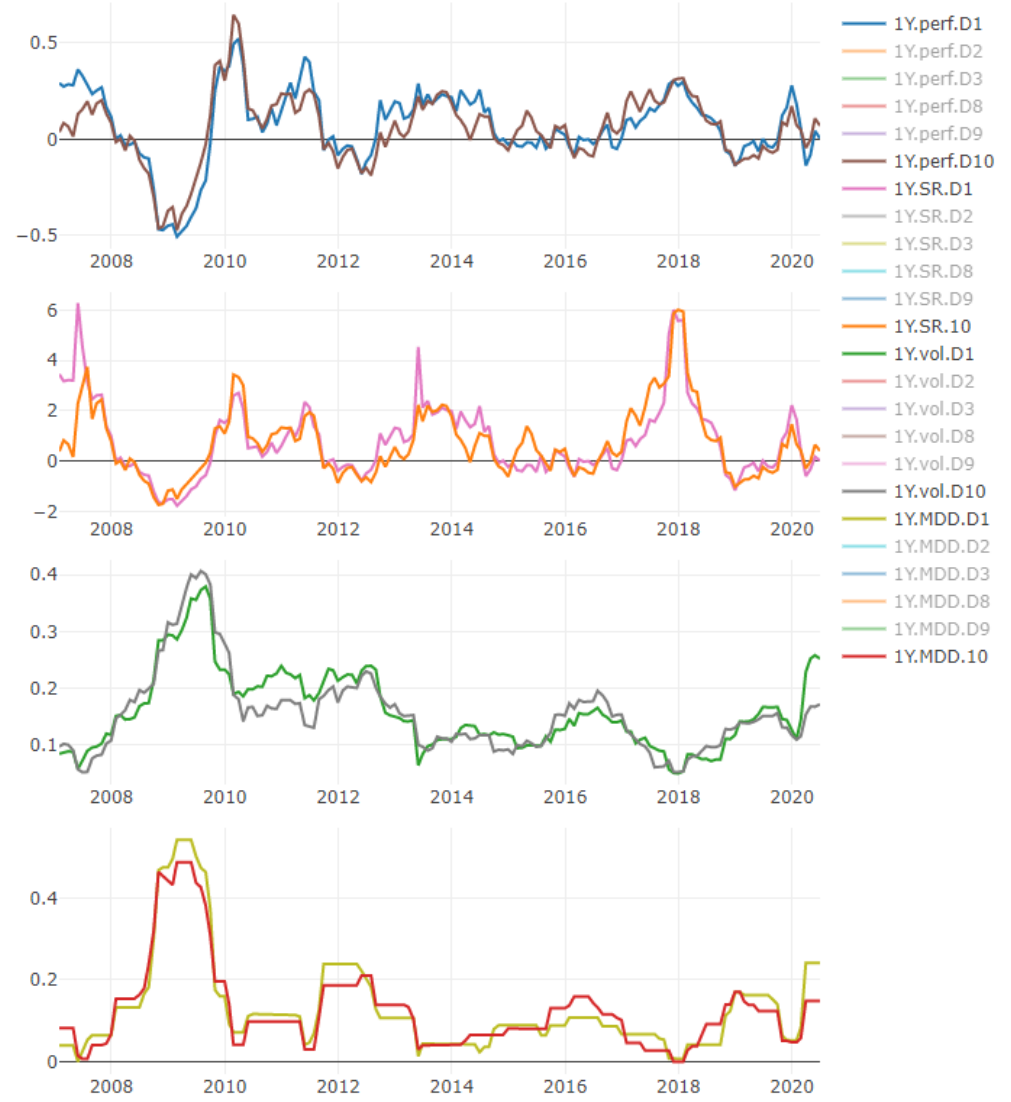
Comparing Performance



WD_EQUITIES_471_ml_xgb_r1m_60m_base_df

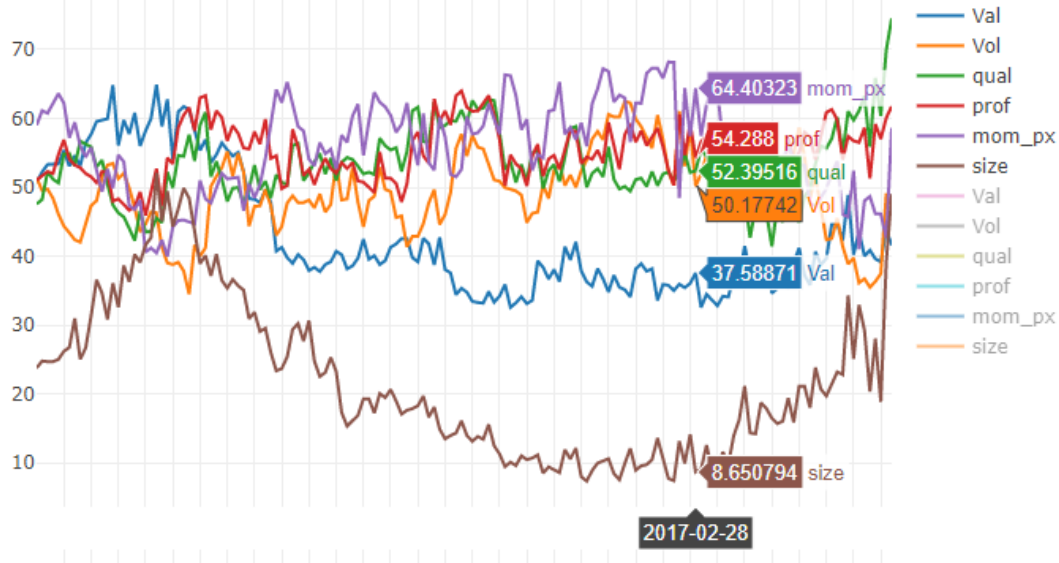
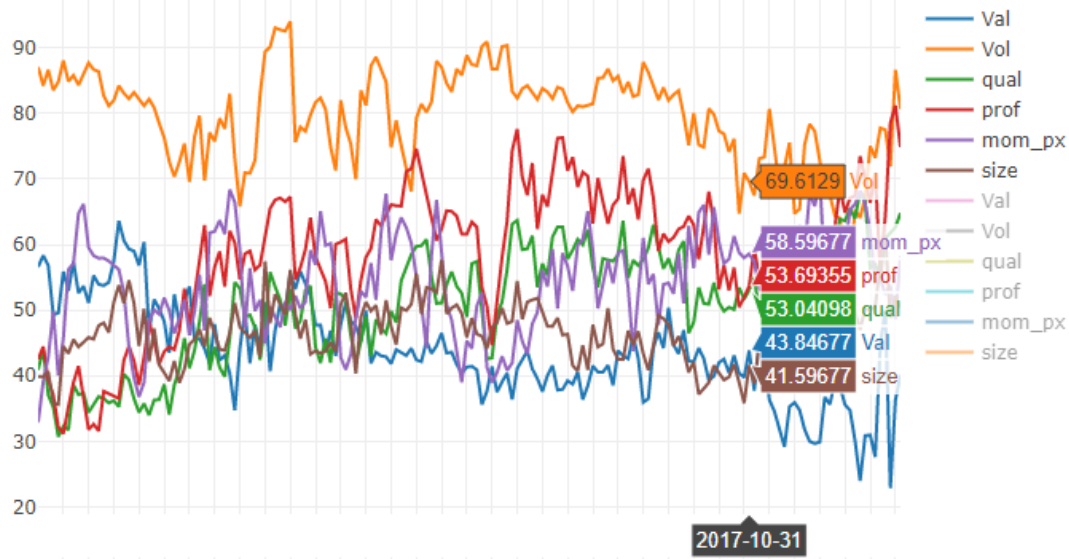


WD_EQUITIES_323_comp_linear_ML5



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

Comparing Factor Exposure



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

ML vs MIF ML: What can we learn from that?

Using ML to time “features” and blend them equally **does not show good results**. Pure noise. Very poor monotonicity.

ML base case (on the left) **is superior on a risk/performance/MaxDD** standpoint.

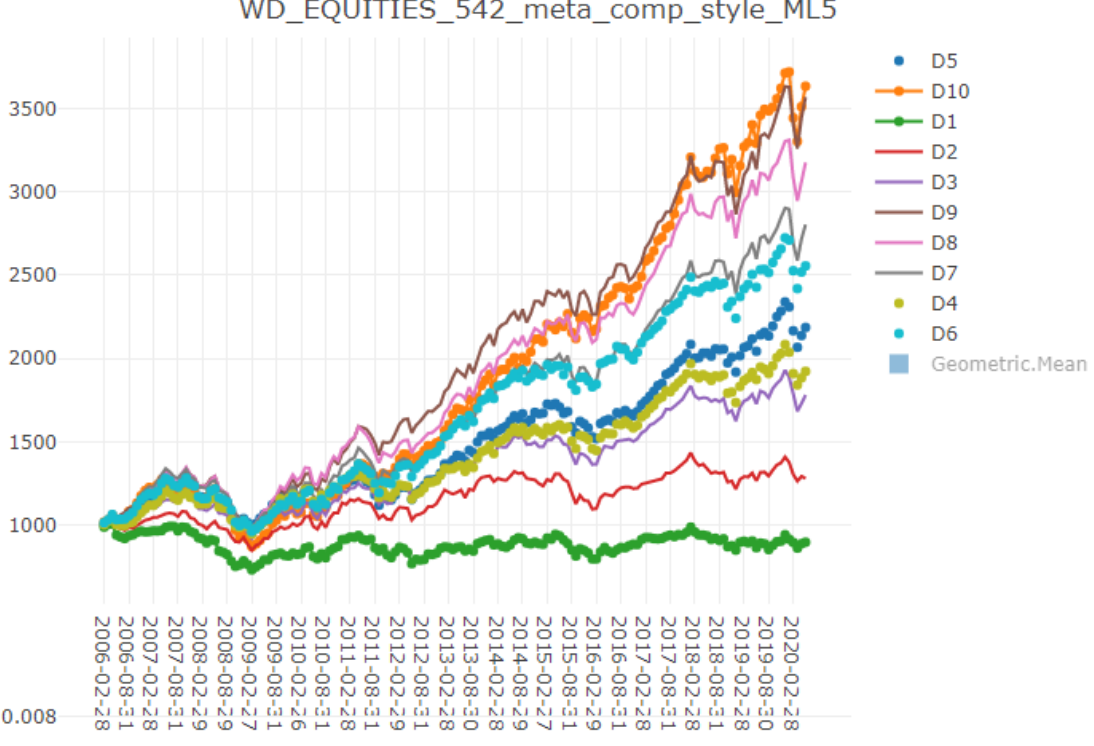
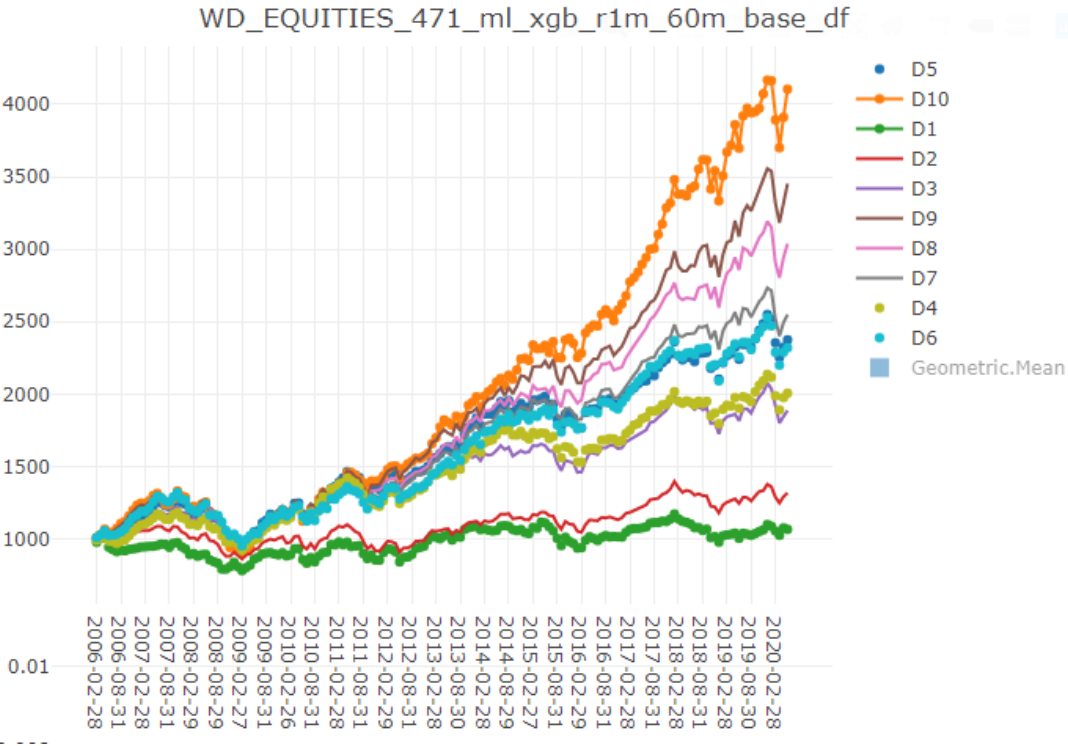
Most Important Feature (MIF) signal **is a good proxy for benchmarking** that your ML model is **using non-linearities and interaction** effect.



Check #:

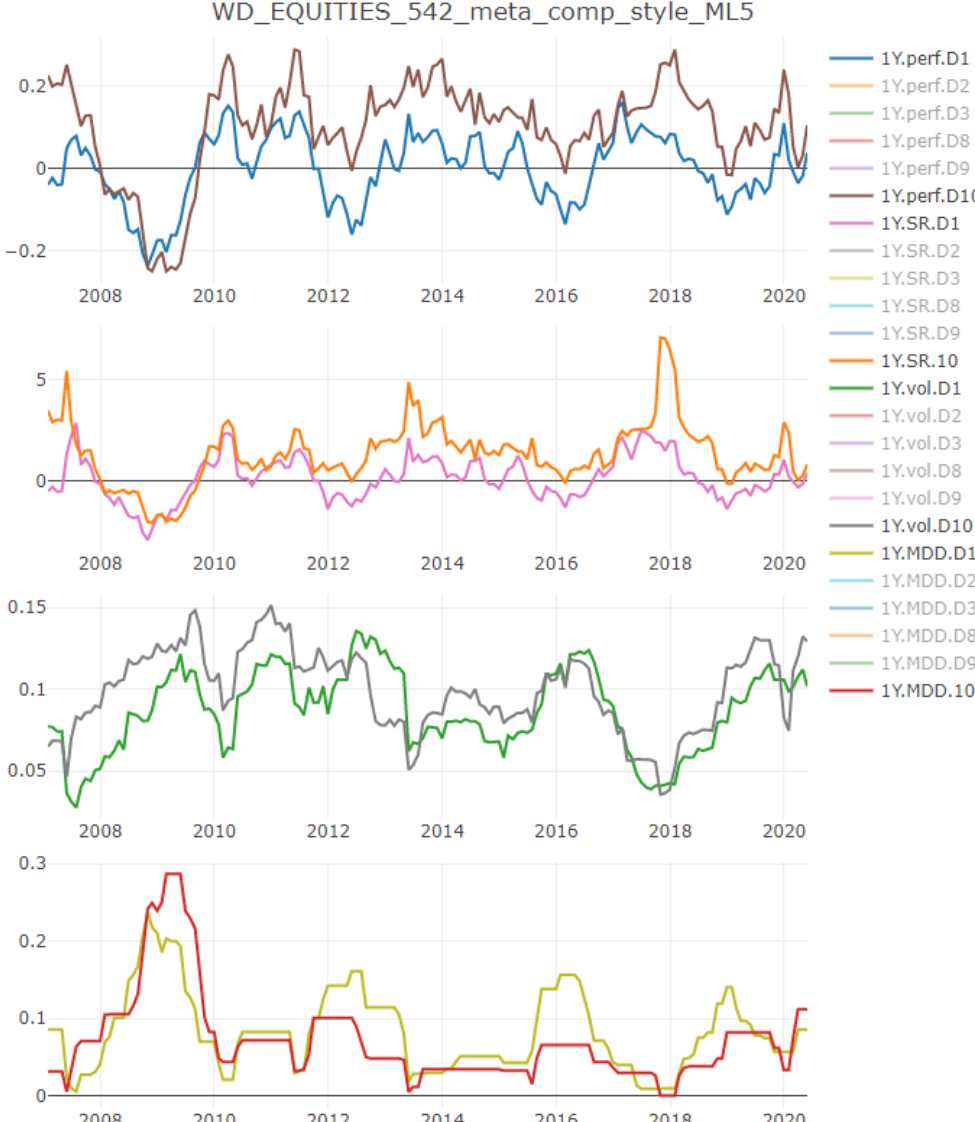
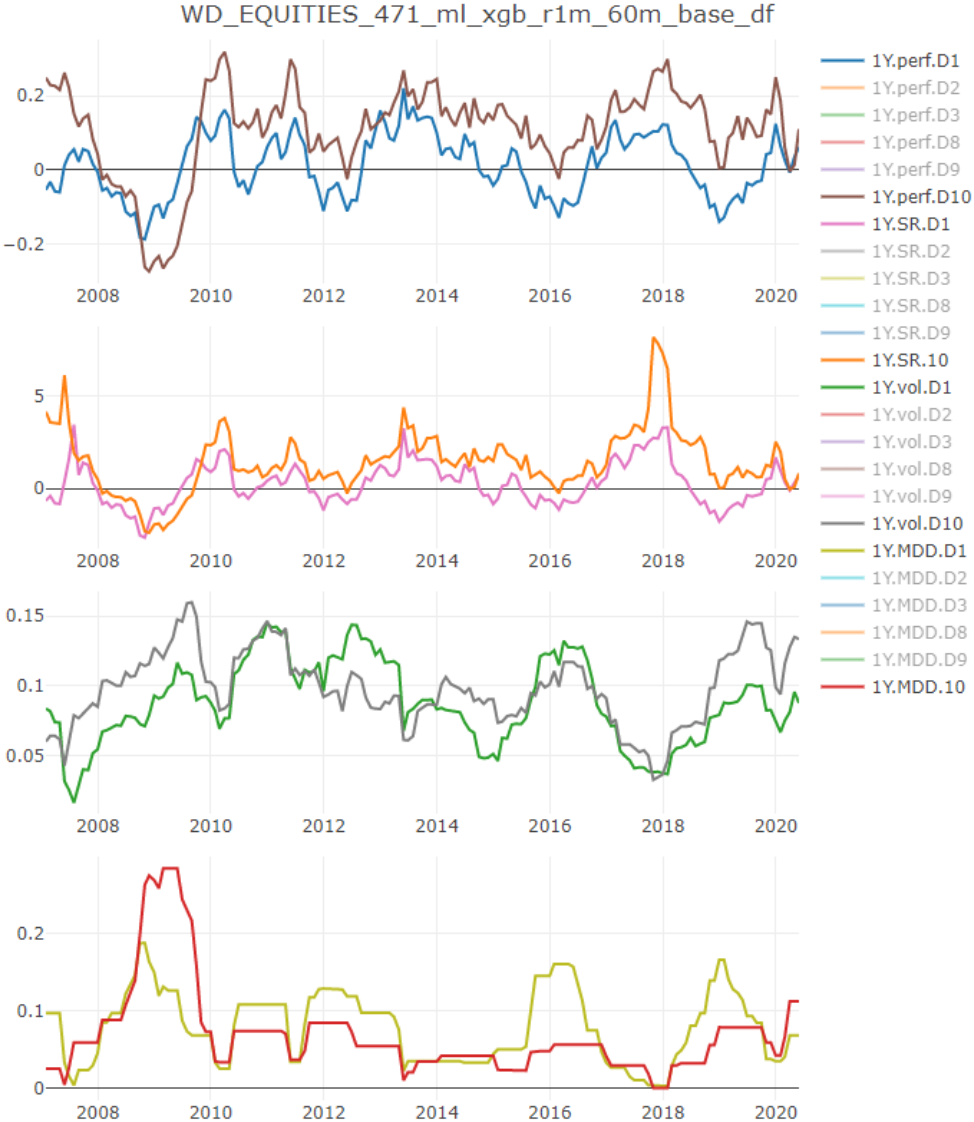
ML vs Style ML

Comparing Monotonicity



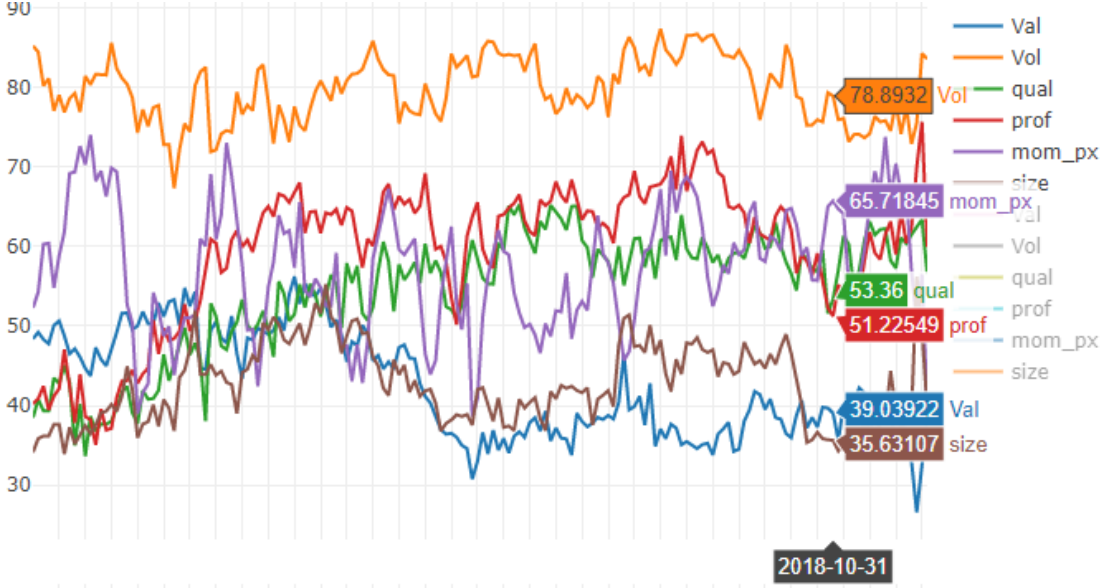
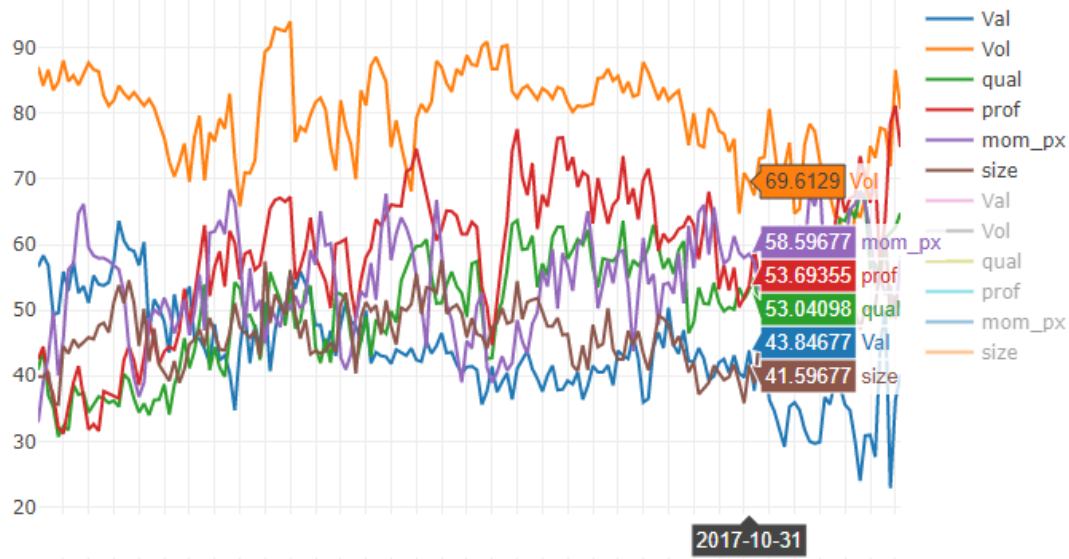
Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

Comparing Performance



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

Comparing Factor Exposure



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset.

ML vs Style ML: What can we learn from that?

To construct Style ML, we slice the core dataset according to the 5 style's families of signals. We then run the same ML model for each style dataset. We finally **average the 5 scores into one blended score**.

Results are very similar with still a slight advance on the ML base case side.

Style ML is a good proxy for checking that **one style/subset of the dataset is not driving all the modelling**.

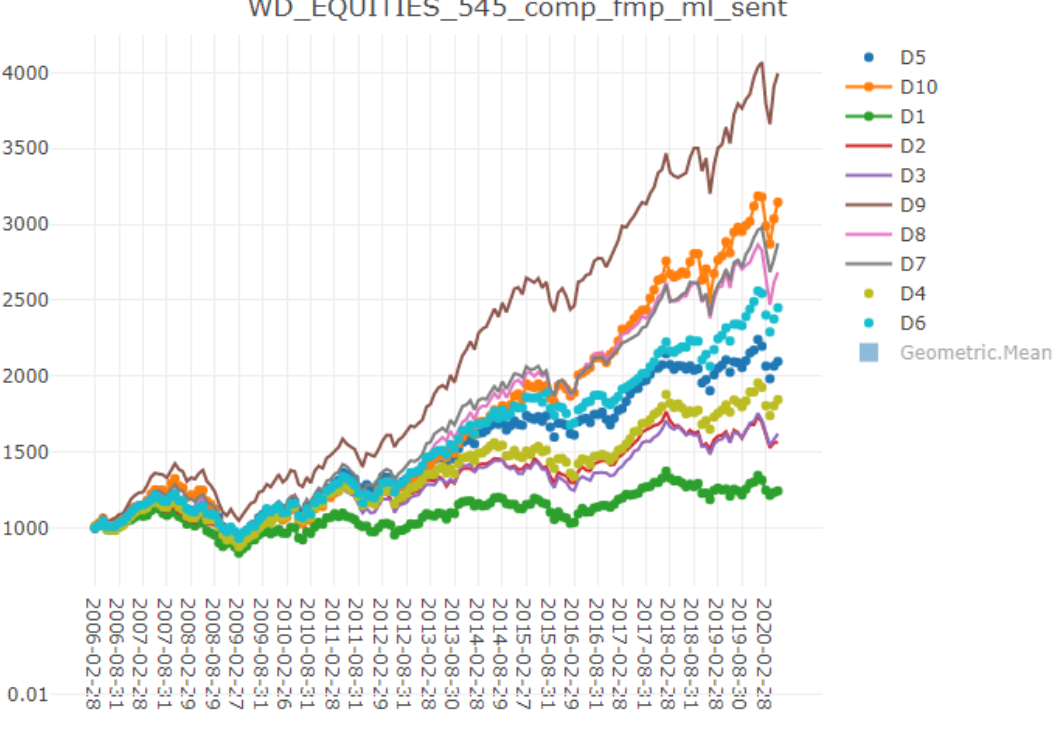
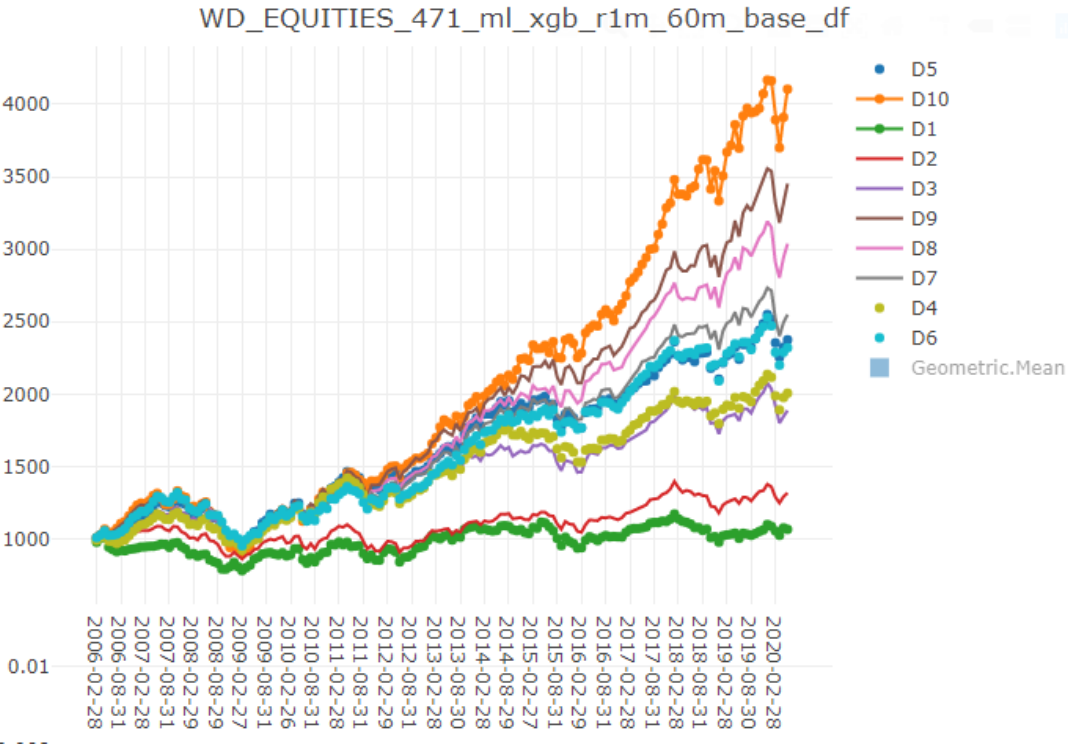


Check #:

ML vs NLP Sentiment

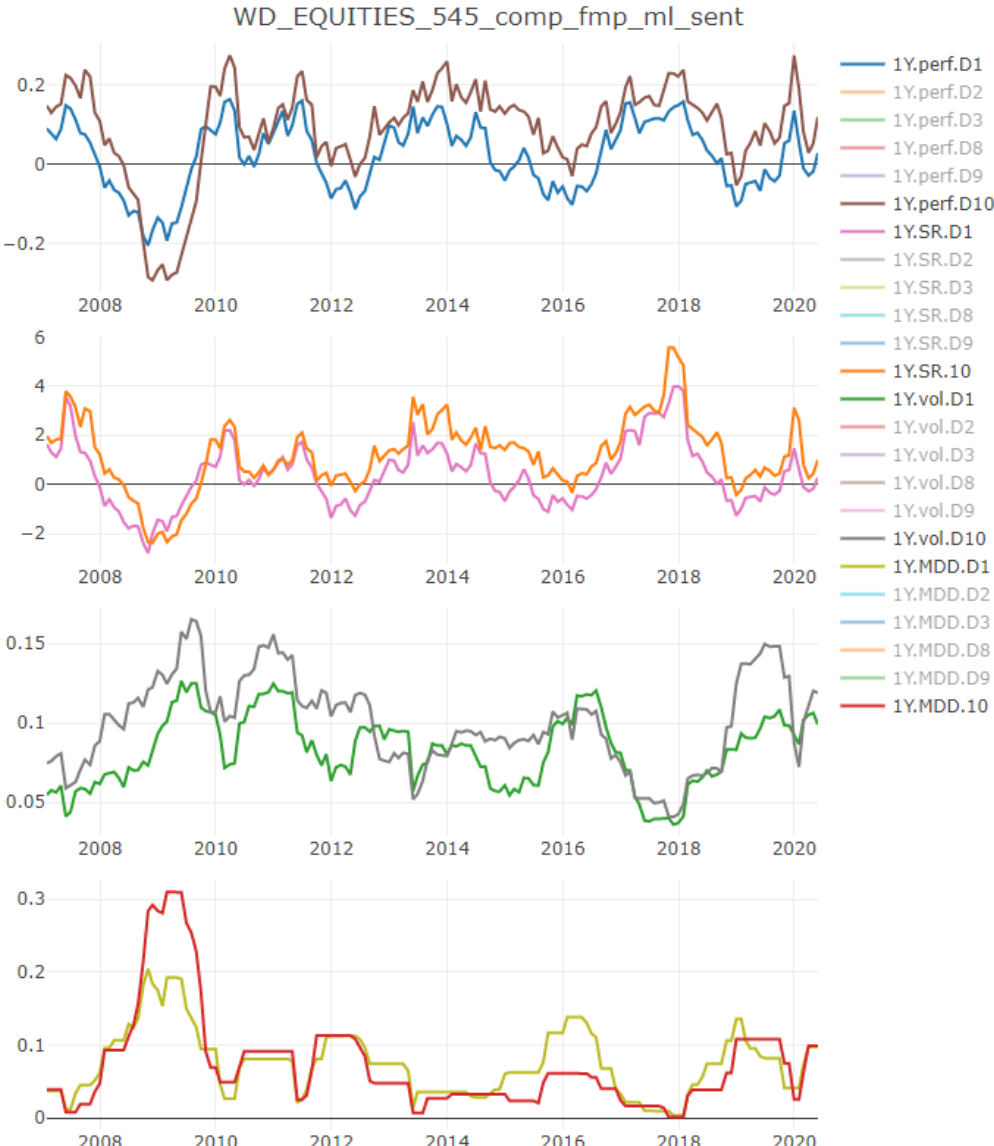
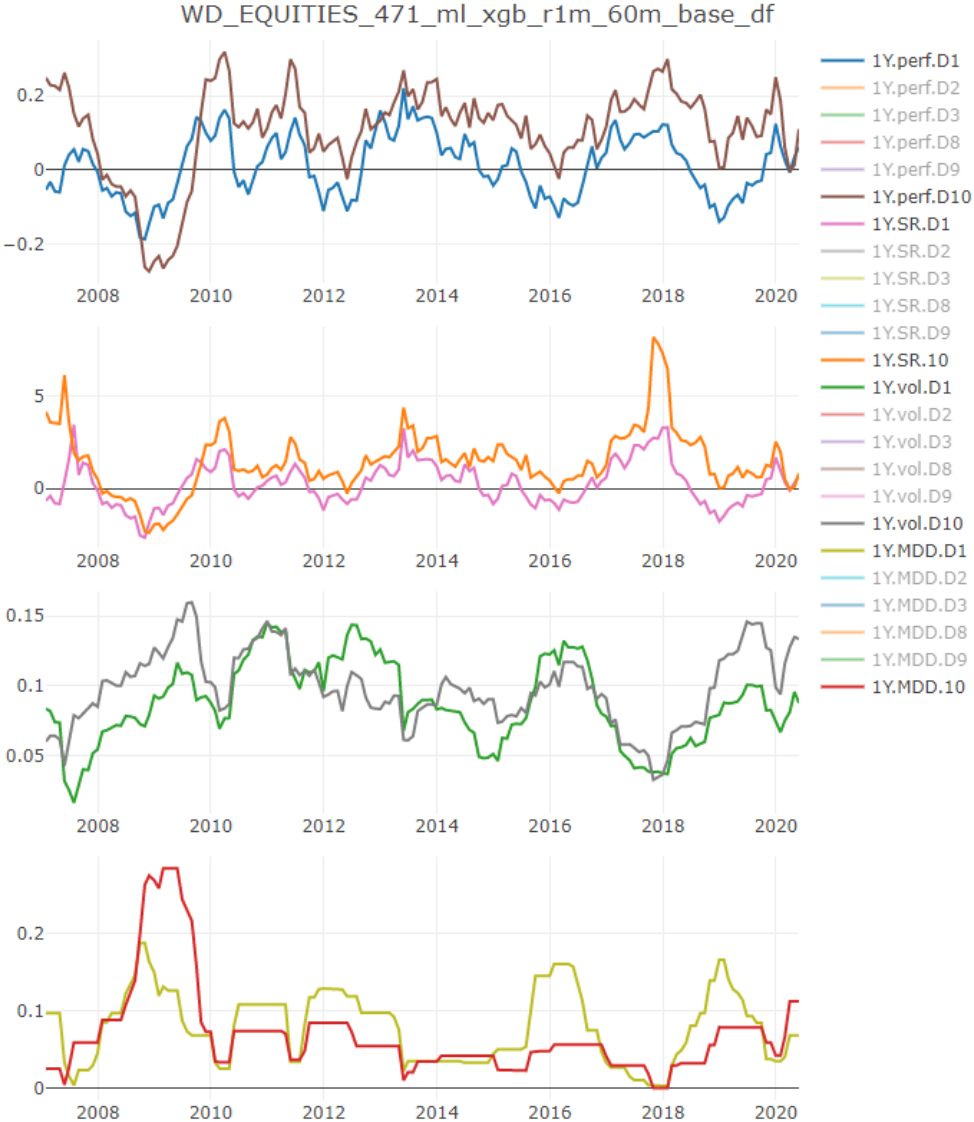
Dataset

Comparing Monotonicity



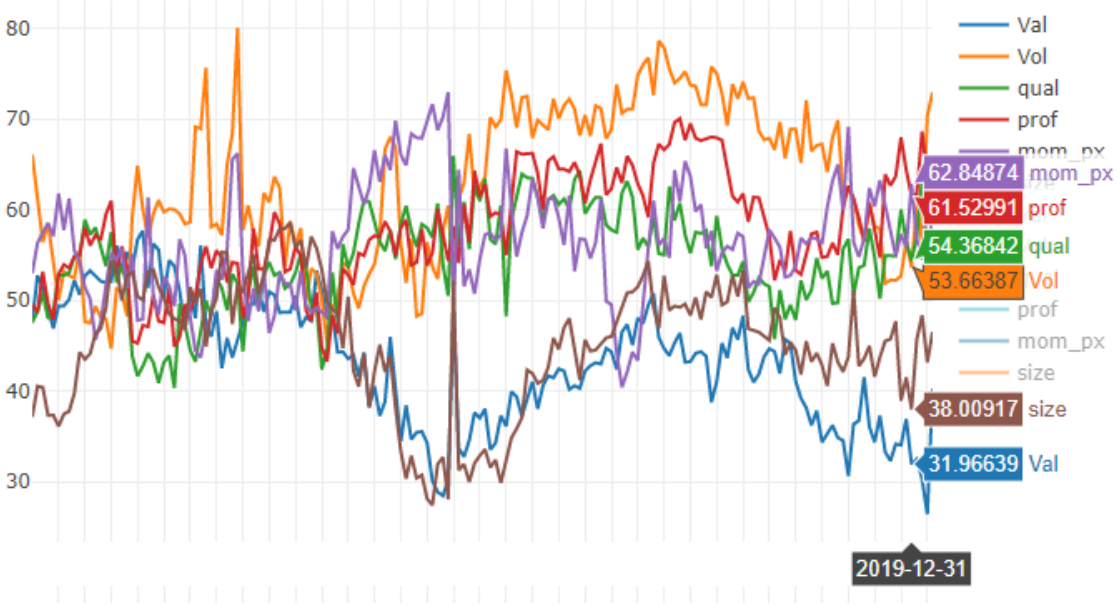
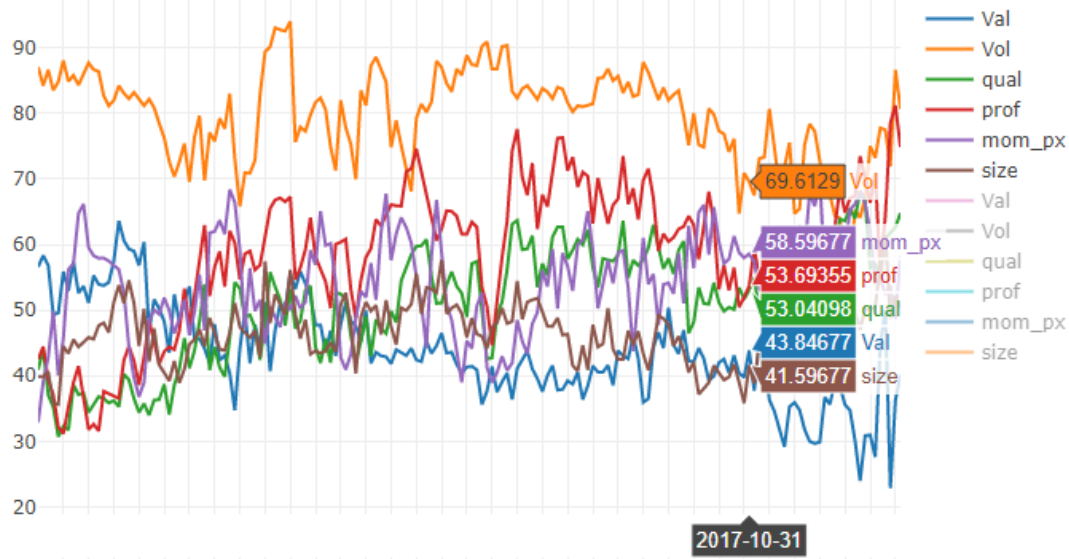
Source: Simulation gross of TC and fees expressed in USD. RAM, Factset., Alt data NEWs NLP 3rd party provider.

Comparing Performance



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset., Alt data NEWs NLP 3rd party provider.

Comparing Factor Exposure



Source: Simulation gross of TC and fees expressed in USD. RAM, Factset., Alt data NEWs NLP 3rd party provider.

ML vs NLP ML: What can we learn from that?

NLP dataset is created using **3rd party NLP/sentiment vendors** and creating ~50 signals based on different sub-topics and checking **relevance, novelty** and **volume of news**.

Global News **NLP dataset with sub-topics** like earnings, stock prices etc.. is a different “**plane**” that incorporates the Human/judgement on text and numbers. Intuitively it should be a **good diversifier**.

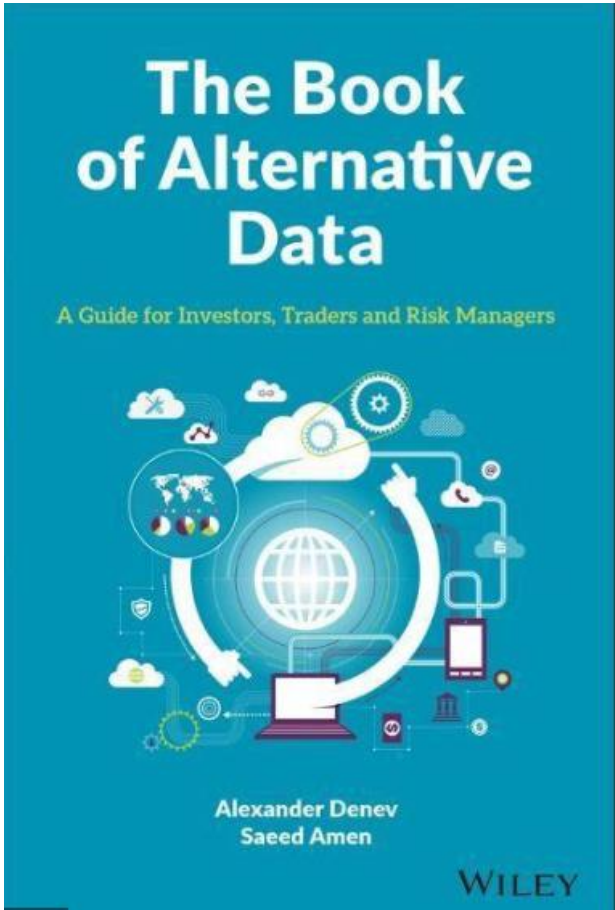
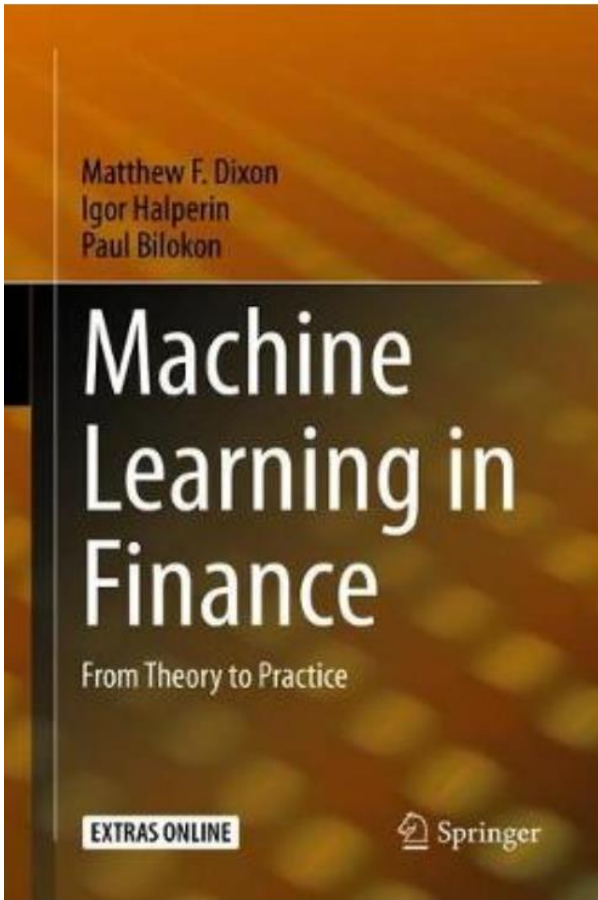
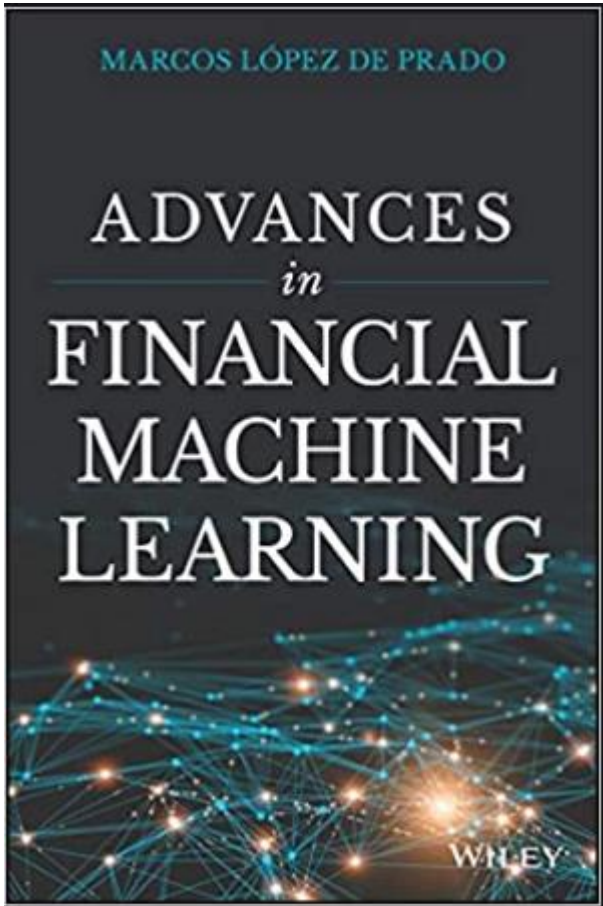
Excess return time series **correlation is 0.5**, which is **low for equities**.



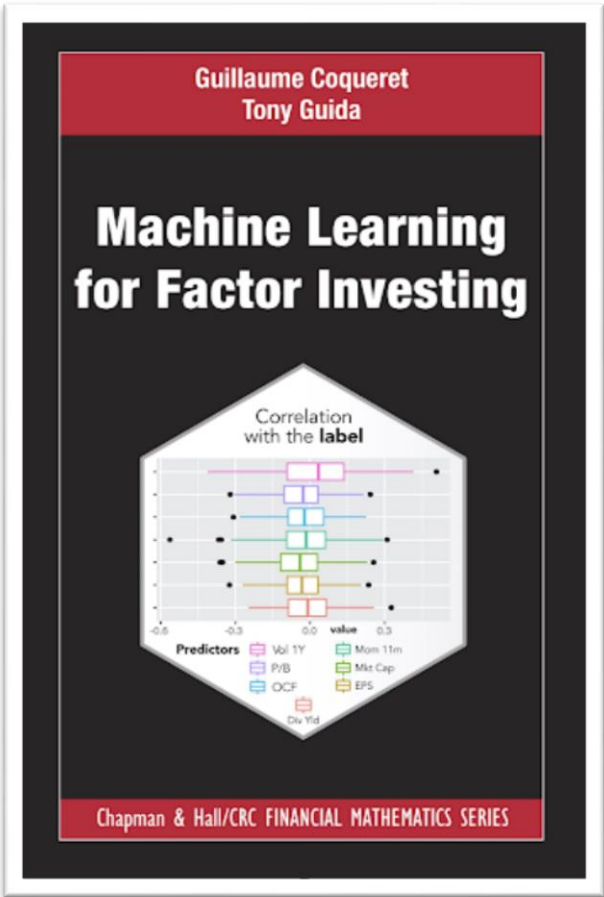
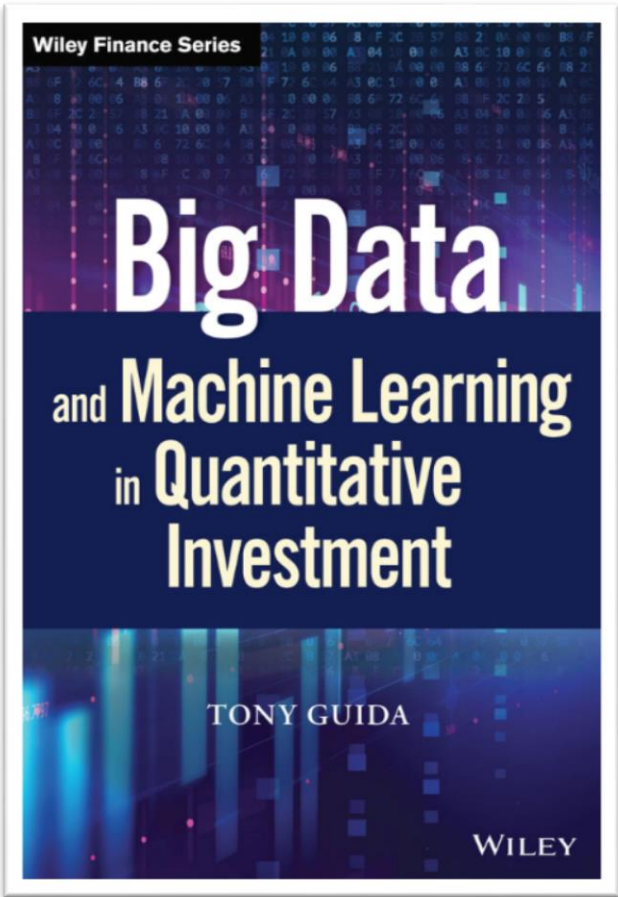
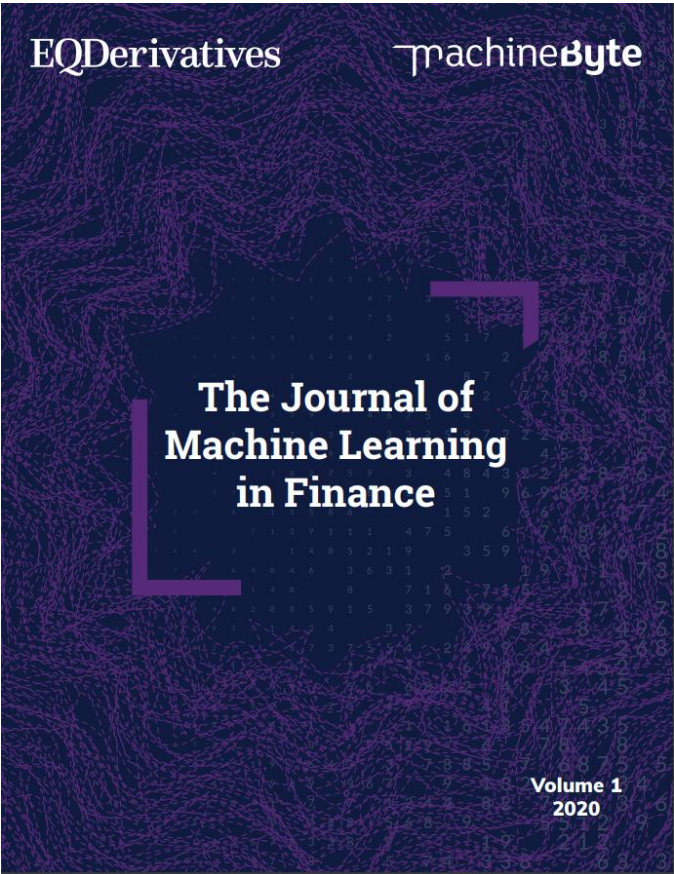
Knowledge is Power:

Some ML/alt. data books

ML books from the “Community”



My contributions to the “community”



What this presentation was NOT about



Machine Learning model debate: **Deep Learning vs RestOfML, TensorFlow vs PyTorch, Python vs R, Classification vs Regression** etc..

What type of **dataset to use or not.**

A lecture on “**ML how to**” for hyperparameters.

A full portfolio construction/optimisation **with investment’s constraints + risk management, trading implementation constraints..**

Conclusion



Machine learning is not new but a “**new**” way for doing **research** today.

ML used with traditional data proved to add a **non-linear rolling** component to alpha prediction.

ML used in isolation for **NLP/sentiment** could be a **good stand alone** and **complement for a multi-factor** strategy in Equity.

Matter of survival to be capable on **onboarding, analysing** and **implementing** ML and Big/alt data in the investment toolbox.



ram

ACTIVE INVESTMENTS

Peer reviewed published paper on tails training



We'd like to understand how you use our websites in order to improve them

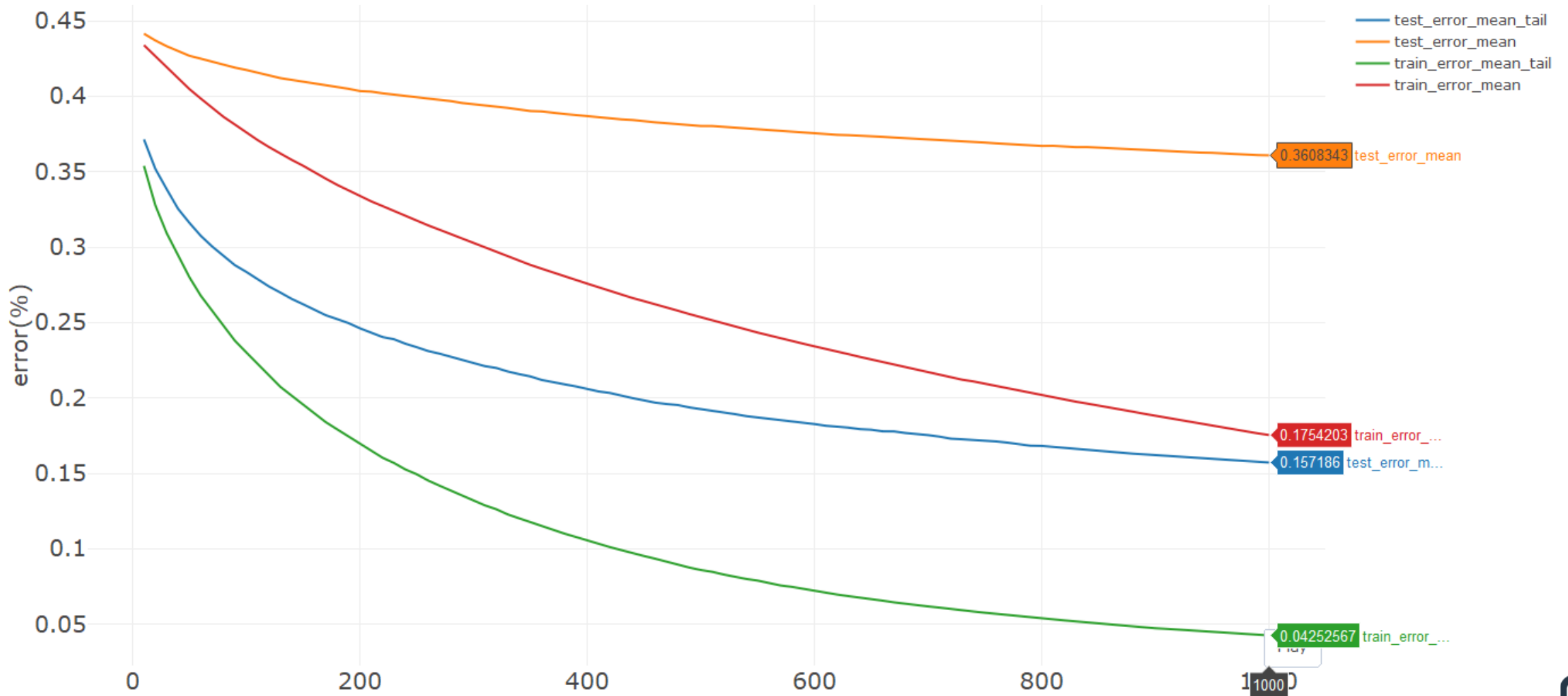
Original Research | Published: 20 February 2020

Training trees on tails with applications to portfolio choice

[Guillaume Coqueret](#)  & [Tony Guida](#)

[Annals of Operations Research](#) **288**, 181–221(2020) | [Cite this article](#)

What tails training does on accuracy



Source: Hypothetical exercise based on a different yet similar datasets (World including EM)

Important Information



This material is addressed to professional clients for informative purposes only. It is neither an offer nor an invitation to buy or sell investment products and may not be interpreted as investment advice. It is not intended to be distributed, published or used in a jurisdiction where such distribution, publication or use is forbidden, and is not intended for any person or entity to whom or to which it would be illegal to address such a material. In particular, investment products are not offered for sale in the United States or its territories and possessions, nor to any US person (citizens or residents of the United States of America). The opinions herein do not consider individual clients' circumstances, objectives, or needs. Before entering into any transaction, clients are advised to form their own opinion and consult professional advisors to obtain an independent review of the specific risks incurred (tax, financial etc.). Upon request, RAM AI Group is available to provide more information to clients on risks associated with investments. The information and analysis contained herein are based on sources deemed reliable. However, RAM AI Group does not guarantee their accuracy, correctness or completeness, and it does not accept any liability for any loss or damage resulting from their use. All information and assessments are subject to change without notice. Changes in exchange rates may cause the NAV per share in the investor's base currency to fluctuate. There is no guarantee to get back the full amount invested. Past performances, whether actual or back-tested, are not necessarily indicative of future performance. Without prejudice of the due addressee's own analysis, RAM understands that this communication should be regarded as a minor non-monetary benefit according to MIFID regulations. Clients are invited to base their investment decisions on the most recent prospectus, key investor information document (KIID) and financial reports which contain additional information relating to the investment product. These documents are available free of charge from the SICAV's and Management Company's registered offices, its representative and distributor in Switzerland, RAM Active Investments S.A. and at Macard Stein & Co AG, Paying and Information Agent in Germany; and at RAM Active Investments (Europe) SA – Succursale Milano in Italy. This marketing material has not been approved by any financial Authority, it is confidential and addressed solely to its intended recipient; its total or partial reproduction and distribution are prohibited. Issued in Switzerland by RAM Active Investments S.A. which is authorised and regulated in Switzerland by the Swiss Financial Market Supervisory Authority (FINMA). Issued in the European Union and the EEA by the Management Company RAM Active Investments (Europe) S.A., 51 av. John F. Kennedy L-1855 Luxembourg, Grand Duchy of Luxembourg. The reference to RAM AI Group includes both entities, RAM Active Investments S.A. and RAM Active Investments (Europe) S.A.