Improved Signals

Does Meta-Labeling Add to Signal Efficacy?

Table of Contents

Overview

At Hudson and Thames we believe the scientific method is the best way to approach investment management. We pride ourselves on our ability to productionize machine learning models for our clients and we do this by following strict software engineering protocols. Our models are reproducible and work at scale.

This presentation outlines meta-labeling and shows that it improves performance metrics out-of-sample.

Goal

- Build an open-source quantitative finance research platform.
- Implement core functionality from the book, AFML by Dr. Lopez de Prado.
- Implement key principles that make M/L in finance more effective:
  - Event based data sampling
  - Triple-barrier method
  - Meta-labeling
- Implement trading strategies on this platform to test the efficacy of the above principles
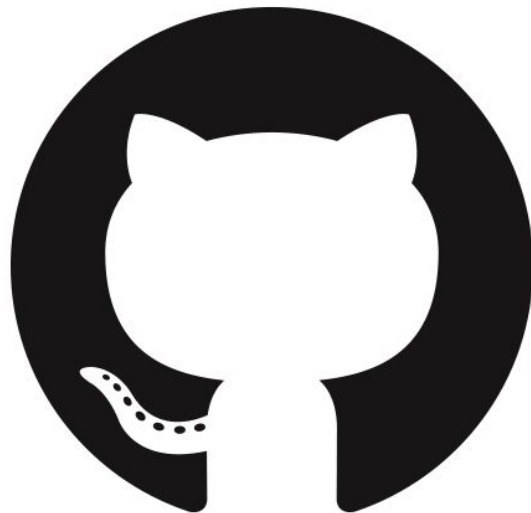
Accomplishments (so far) …

- Created an open-source package called `mlfinlab` that has code for:
    - Event based data sampling – dollar, volume and tick bars
    - Classes that implement:
        - Symmetric CUSUM filter
        - Triple-barrier
        - Meta-labeling
    - Functions that compute Bollinger bands and RSI
    - Utility functions to compute daily returns from intraday positions
- Implemented moving-average trend-following strategy
- Implemented Bollinger band mean-reverting strategy
- Prove that meta-labeling improves performance metrics (out-of-sample).

Marcos López de Prado @lopezdeprado · Mar 14

Ashutosh Singh and @JacquesQuant have published a nice notebook on Meta-Labeling:

github.com/hudson-and-tha…

I recommend it to anyone interested in boosting the Sharpe ratio of their investment strategies.
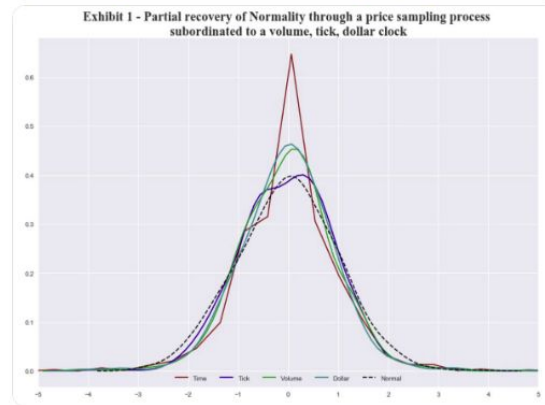
Great job, and thanks for sharing!

Meta Labeling
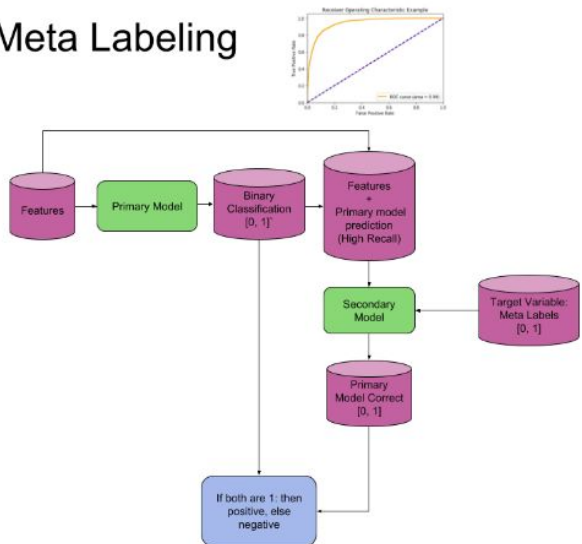
💬 2   🔁 31   ❤️ 118   ✉️

Corey Hoffstein ✔️ @choffstein · Mar 15

3/ Suddenly, "returns" became a lot more normal and stable looking.

(Image source: github.com/Jackal08/finan…; credit @JacquesQuant)

Exhibit 1 - Partial recovery of Normality through a price sampling process subordinated to a volume, tick, dollar clock

💬 4   🔁 2   ❤️ 22   ✉️

[–] **killsecurity**   6 points 12 days ago

Many thanks

permalink   embed   save   report   give award   reply

[–] **beegreen**   4 points 12 days ago

Oh so wet rn

permalink   embed   save   report   give award   reply

[–] **AloneThanks**   1 point 12 days ago

This is fantastic, thank you!

permalink   embed   save   report   give award   reply

[–] **JeffreyChl**   1 point 12 days ago

This field lacks open source contribution by its nature but people like you change the world and this industry step by step. Kudos!

permalink   embed   save   report   give award   reply

📖 **mlfinlab**

Package based on the work of Dr Marcos Lopez de Prado regarding his research with respect to Advances in Financial Machine Learning
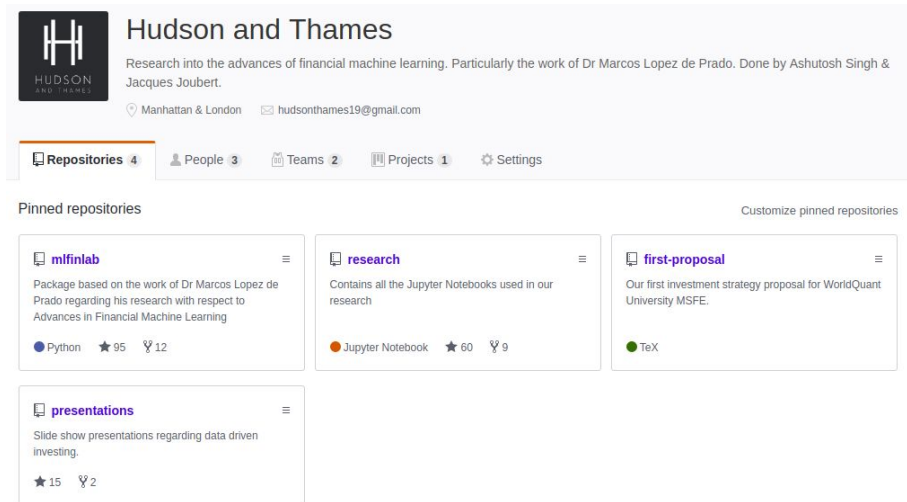
🔵 Python   ⭐ 95   ⑂ 12

📖 **research**

Contains all the Jupyter Notebooks used in our research

🟠 Jupyter Notebook   ⭐ 59   ⑂ 9

Project Principles

- Open-source (Github)
- Agile (Agile Board)
- Continuous Integration (Travis-CI)
- 100% code coverage and Unit-Test
- Research environment (Jupyter Lab)
- Development environment (Pycharm)
- Latex for published research
- Markdown for readme and community documentation

Hudson and Thames

Research into the advances of financial machine learning. Particularly the work of Dr Marcos Lopez de Prado. Done by Ashutosh Singh & Jacques Joubert.

⌖ Manhattan & London   ✉ hudsonthames19@gmail.com

| Repositories 4 | People 3 | Teams 2 | Projects 1 | Settings |

Pinned repositories   Customize pinned repositories

mlfinlab

Package based on the work of Dr Marcos Lopez de Prado regarding his research with respect to Advances in Financial Machine Learning

● Python   ★ 95   ⑂ 12

research

Contains all the Jupyter Notebooks used in our research

● Jupyter Notebook   ★ 60   ⑂ 9

first-proposal

Our first investment strategy proposal for WorldQuant University MSFE.

● TeX

presentations

Slide show presentations regarding data driven investing.

★ 15   ⑂ 2

Methodology

- Use tick-data to create event based sampling
- Improve data qualities like normality (Jarque-Bera, acf tests)
- Detect events via symmetric CUSUM
- Use triple-barrier (profit-taking, stop-loss and time) to detect exit points
- Use meta-labeling to refine Primary Model
- Use supervised non-linear M/L to train and improve signal efficacy
- Compute OOS performance

Event-based Sampling
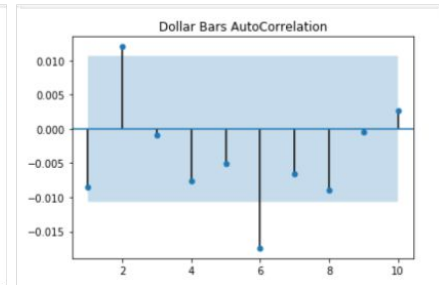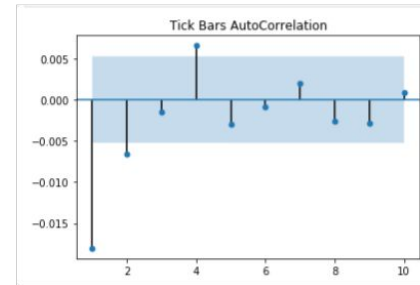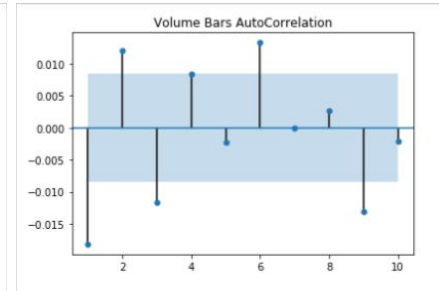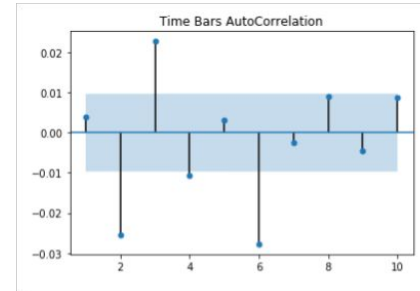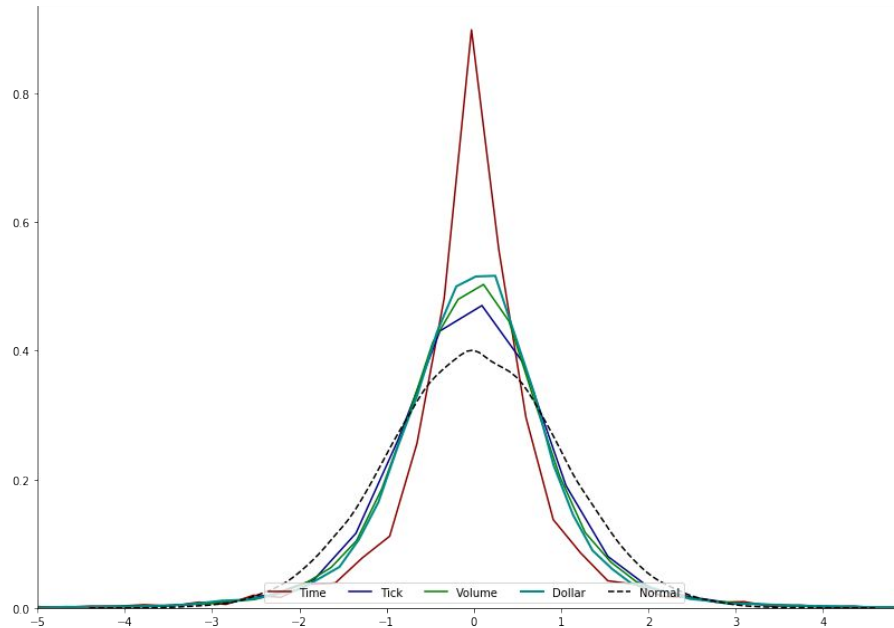
Filtering Data: Symmetric CUSUM

Sample features by the CUSUM filter

www.larryhua.com
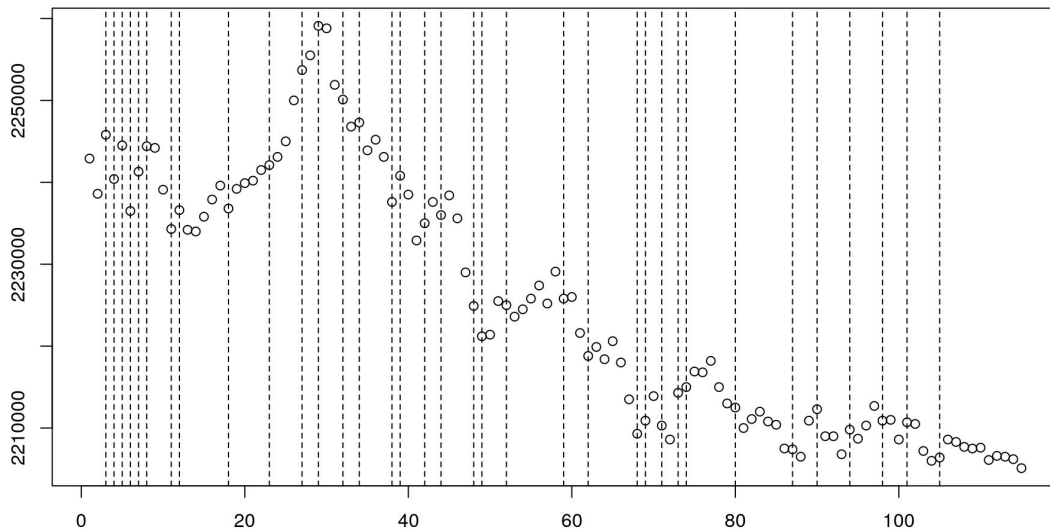
$$S_t^+ = \max\left\{0, S_{t-1}^+ + y_t - \mathrm{E}_{t-1}\left[y_t\right]\right\}, \quad S_0^+ = 0$$

$$S_t^- = \min\left\{0, S_{t-1}^- + y_t - \mathrm{E}_{t-1}\left[y_t\right]\right\}, \quad S_0^- = 0$$

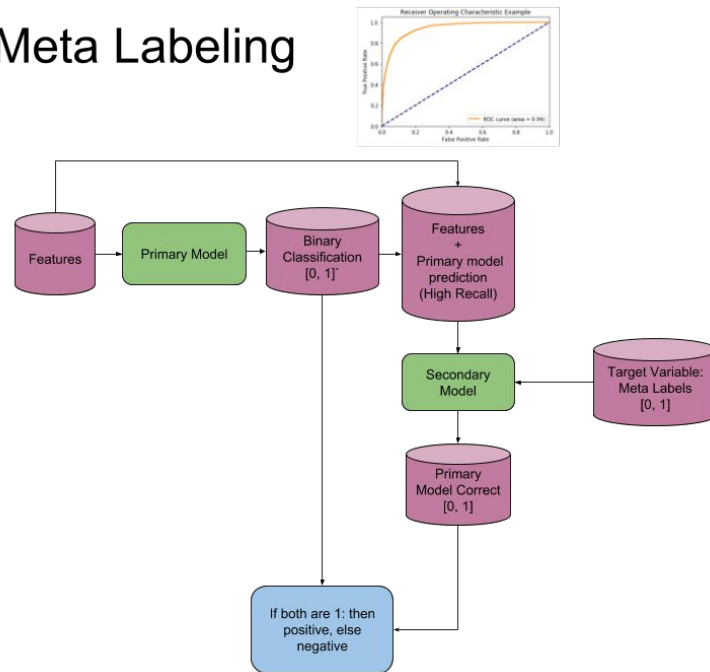$$S_t = \max\left\{S_t^+, -S_t^-\right\}$$

Triple-Barrier
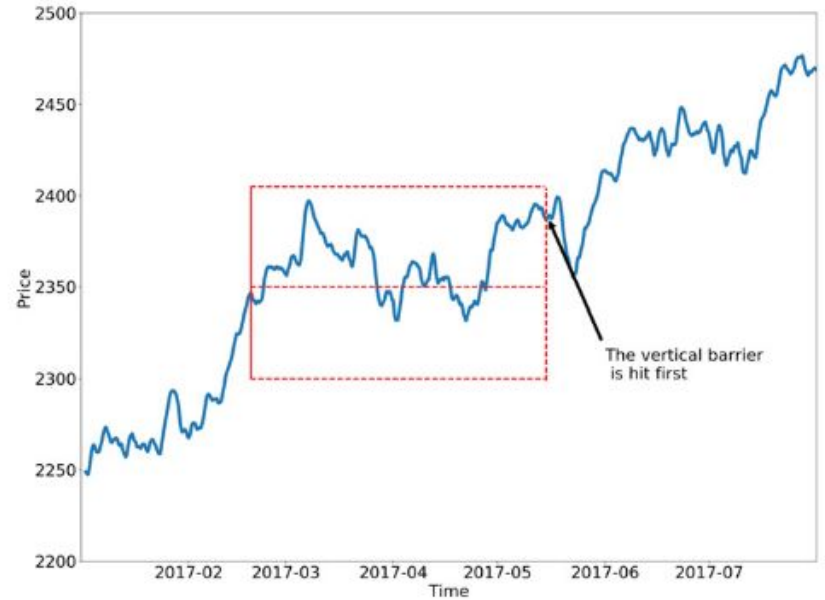
Meta-Labels

To trade or not to trade!
- Takes the side from the primary model (long or short).
- Train a M/L to determine if we should trade on the signal or not.
  - Train Random Forest
  - Use Cross-validation and Grid Search to find the optimal hyperparameters.

Meta Labeling

Model Performance: Mean Reverting

	precision	recall	f1-score	support
0	0.00	0.00	0.00	749
1	0.17	1.00	0.29	151
micro avg	0.17	0.17	0.17	900
macro avg	0.08	0.50	0.14	900
weighted avg	0.03	0.17	0.05	900

Confusion Matrix
```
[[  0 749]
 [  0 151]]
```

Accuracy
0.16777777777777778

	precision	recall	f1-score	support
0	0.85	0.68	0.75	749
1	0.20	0.41	0.27	151
micro avg	0.63	0.63	0.63	900
macro avg	0.53	0.54	0.51	900
weighted avg	0.74	0.63	0.67	900

Confusion Matrix
```
[[506 243]
 [ 89  62]]
```

Accuracy
0.6311111111111111
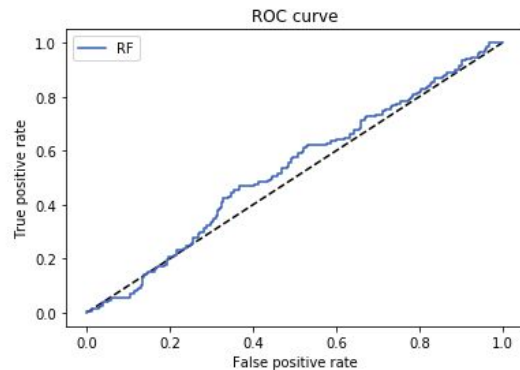
Strategy Performance: Mean Reverting

	Primary Model	Meta Model
Annual return	17.7%	35.3%
Cumulative returns	19.7%	39.6%
Annual volatility	95.0%	56.7%
Sharpe ratio	0.65	0.82
Calmar ratio	0.29	0.96
Max drawdown	-61.9%	-36.8%
Daily value at risk	-11.7%	-7.0%

Cumulative returns

Strategy Performance: Mean Reverting
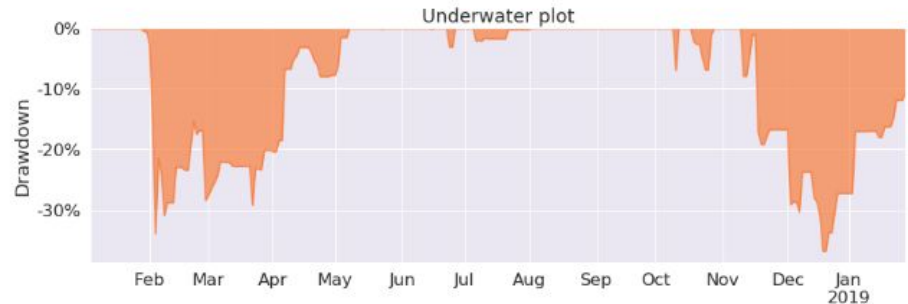
Model Performance: Trend Following

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1088
1	0.48	1.00	0.65	1010
micro avg	0.48	0.48	0.48	2098
macro avg	0.24	0.50	0.32	2098
weighted avg	0.23	0.48	0.31	2098

Confusion Matrix
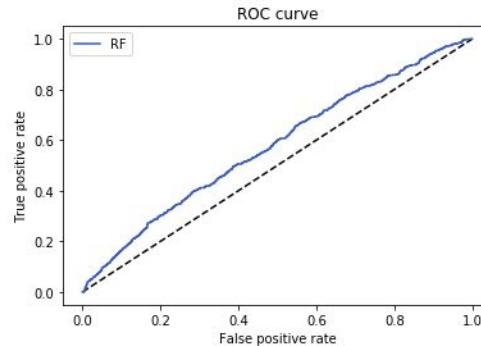[[0 1088]
 [0 1010]]

Accuracy
0.48141086749285034

	precision	recall	f1-score	support
0	0.56	0.65	0.60	1088
1	0.54	0.45	0.49	1010
micro avg	0.55	0.55	0.55	2098
macro avg	0.55	0.55	0.55	2098
weighted avg	0.55	0.55	0.55	2098

Confusion Matrix
[[709 379]
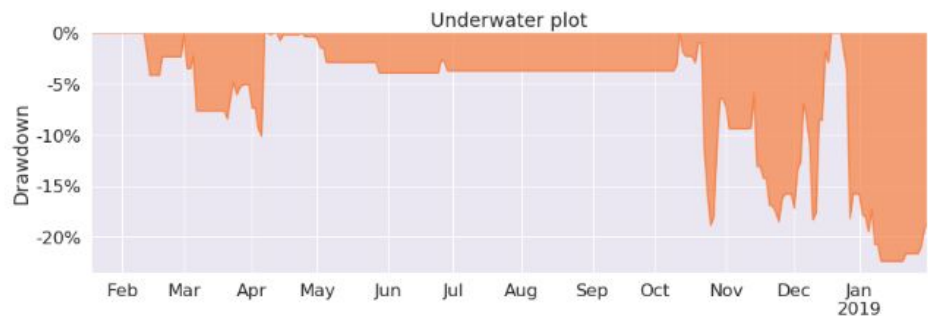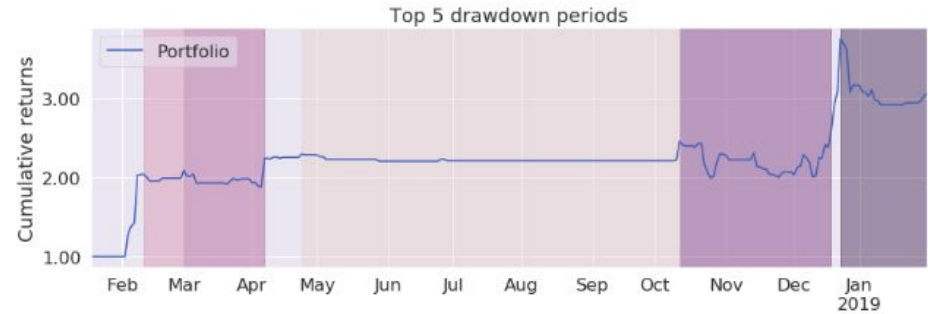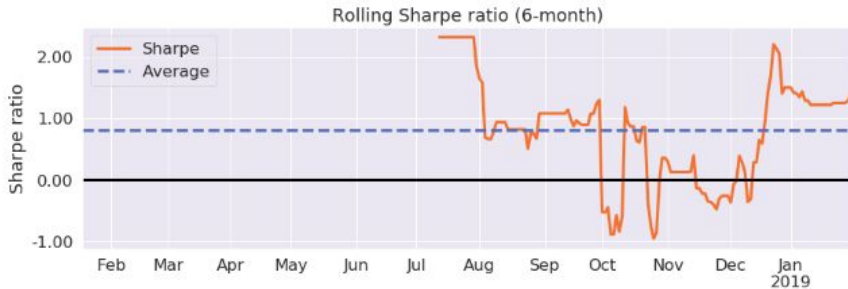 [558 452]]

Accuracy
0.5533841754051477

Strategy Performance: Trend Following

|	Primary Model	Meta Model
Annual return	310.2%	182.2%
Cumulative returns	356.2%	205.2%
Annual volatility	121.0%	68.1%
Sharpe ratio	1.66	1.82
Calmar ratio	5.03	8.15
Max drawdown	-61.7%	-22.4%
Daily value at risk	-14.4%	-8.1%

Cumulative returns

# Strategy Performance: Trend Following

Next Steps

- Add position sizing and risk management to the strategies.
- Build unit-tests in each of the library functions.
- Build a feature zoo.
- Use new features and better models for trend-following and mean reverting.
- Test the framework on Euro-STOXX
- Write a paper

MARCOS LÓPEZ DE PRADO

ADVANCES
in
FINANCIAL
MACHINE
LEARNING

WILEY
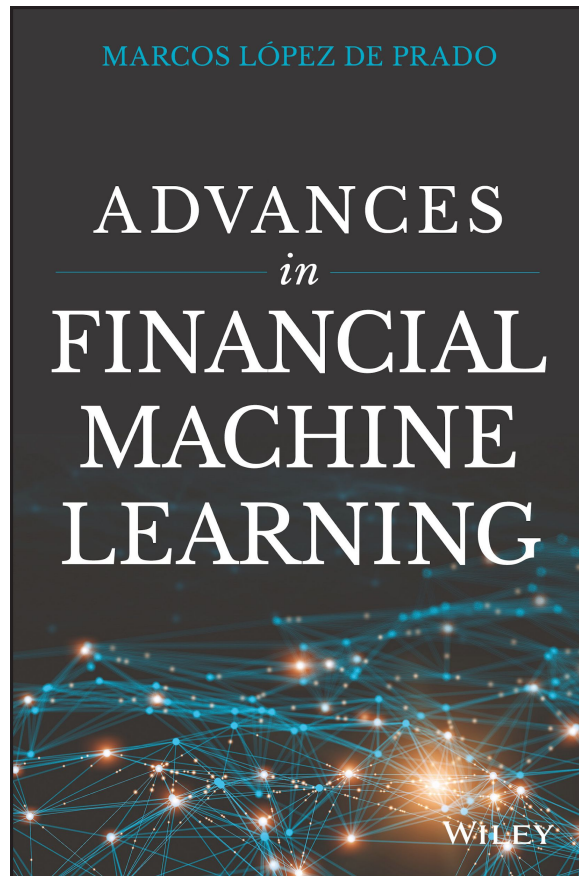
Conclusion

- Our goal was to start an open-source research platform like envisaged by Dr. Lopez de Prado ("meta-strategies") ... we appear to be on the right track given the interest in the community.
- Able to validate that a combination of filtering, triple-barrier and meta-labeling leads to better performing strategies.
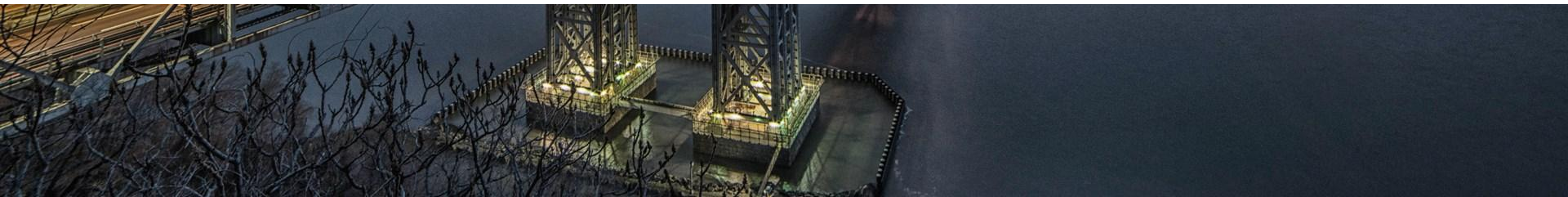
Co-Founder

Ashutosh Singh, CFA

Experienced Executive with a demonstrated history of working in the financial services industry. Strong business development professional skilled in Equity Research, Hedge Funds, Asset Management, Fixed Income, and Derivatives.

HUDSON
AND THAMES

Co-Founder

Jacques Joubert

Machine learning consultant with buy side hedge fund experience. Skilled in machine learning, quantitative finance, systematic investing, software engineering.

Thank You