

Machine learning in Finance

introduction. why it's not a “normal” ML. mistakes :)



Machine Learning
Milan

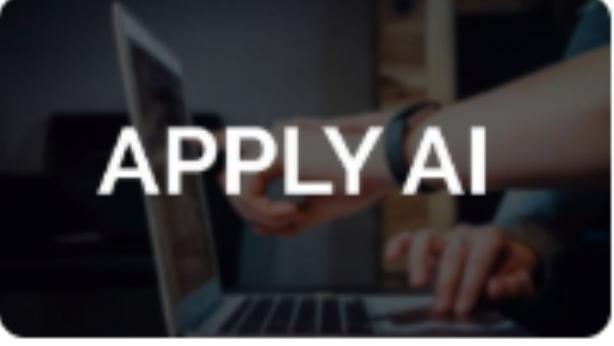
- We are meetup in Milan for academics, hobbyists or professionals interested in Machine Learning solutions and latest development of AI
- We want to build a connected community of AI practitioners in Milan and bring ML Milan at an European level
- We are a no profit group that aims to show hands-on use cases/application
- From 2019 we are partners of

Apply AI @4 and @5 online

SAT, FEB 29, 10:00 AM

ApplyAI@4 - Hands-on workshop to explore ML in Finance - PART 1 -...

Online



Hey all! ApplyAI is back...with a new use workshop on ML in finance. PLEASE NOTE: Due to the current situation in Milan we have decided to not meet together but do anyway the first part of the workshop Online. Alex will do an introduction of the fu...

80 attendees

5

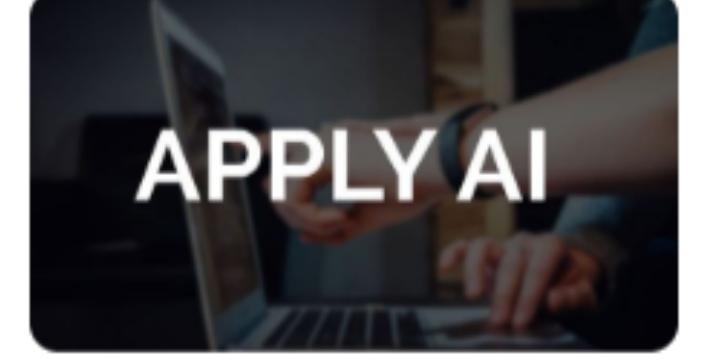
Manage

Going

sab 14 mar, 10:00

ApplyAI@5 - GAN-ime: an use case to turn selfies alive with DL - ONLINE

Online



Hey all! We decided to do another Online event in the meanwhile, with the same logic as the one on 29th of February. Milan will probably be blocked for a while, so we thought it could be a nice idea for all of us to learn in a live environment. ML...

2 partecipanti

Gestisci

Parteciperò

Today: Finance

In 2 weeks: GANs

About Alex Honchar



Practitioner

~7 years developing
data-driven products



Entrepreneur

building a distributed
consulting company



Educator

blogging, workshops,
teaching at universities

About you?

- No experience in finance, but interesting
- SWE with no experience in finance
- Consulting, lots of ML, no finance
- Basic ML, no finance
- Crypto assets, experience with bots, prior ML experience
- NLP, no finance
- Basic quant and ML
- Prior banking and insurance

what this workshop is **not** about

machine learning algorithms
deep stochastic mathematics
making bots for crypto trading
easy ways to hack the market

what this workshop is about

how to doubt everything in your data modeling activities
from now and forever

Plan

- What people in Finance do
- What AI in Finance can do
- Practice. “Classical” ML in Finance.
 - Fixing mistakes
- Why ML in Finance is different from “normal ML”
- Practice. Fixing mistakes again
- Why AI at all, what’s wrong with mathematical modeling?
- Practice. Time series full pipeline rollout
- Practice. ML beyond classification and regression

Today, ~2 hours

Later, ~4-6 hours

People in Finance

	Front Office	Back Office
Buy Side	Asset management at a big bank. Hedge fund (strategies constrained to prospectus) Prop trading (fastest moving) <i>Matlab, Java, Functional Languages.</i> 70-100k+large bonus	Data scraping and maintenance, Execution, Server administration <i>Bash, SQL, SVN, Linux, C++.</i> 90-100k+small bonus
Sell Side	Sales & Trading at a big bank (taking & executing orders, creating derivatives by client reques, execution algos) <i>Excel.</i> 70-80k+medium bonus	Technology, Operations, or Risk Management at a big bank (hard to transition to front office) <i>C++, internal language, hacky code.</i> 90-100k+small bonus

<http://isomorphisms.sdf.org/maxdama.pdf>

Applied directions

Banking

Asset Management

Trading

Retail operations

Representation learning

Price forecasting

P2P operations

Portfolio optimization

Optimal execution

Lending, credit scoring

Risk management

Optimal strategy development

Trading roles

Analysts

Data curators

Feature analysts

Strategists

Backtesters

Deployment team

Portfolio managers

AI in Finance

Supervised learning: regression and classification

Unsupervised learning: clustering and representation learning

Reinforcement learning: optimal strategy, “learning” from the expert

Regression:
earnings prediction
credit loss forecasting
price forecasting

Classification:
rating prediction
default modeling
credit card fraud
anti-money laundering

Clustering:
customer segmentation
stock segmentation

Representation learning:
factor modeling
de-noising
regime change detection

Banking data: clients information, transactions...

Fundamental data: assets, liabilities, sales, earnings...

Market data: price, volume, dividends, open interest...

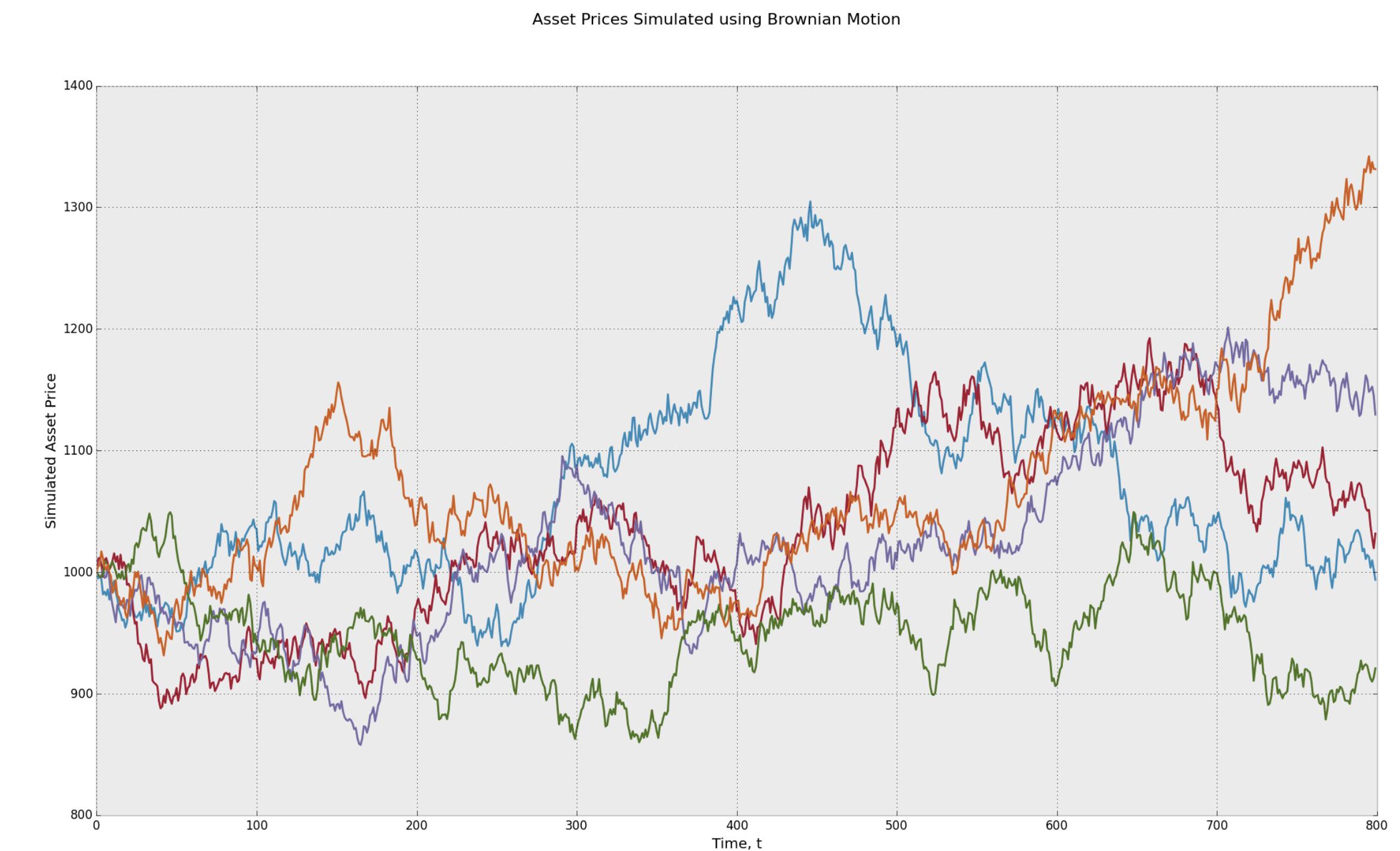
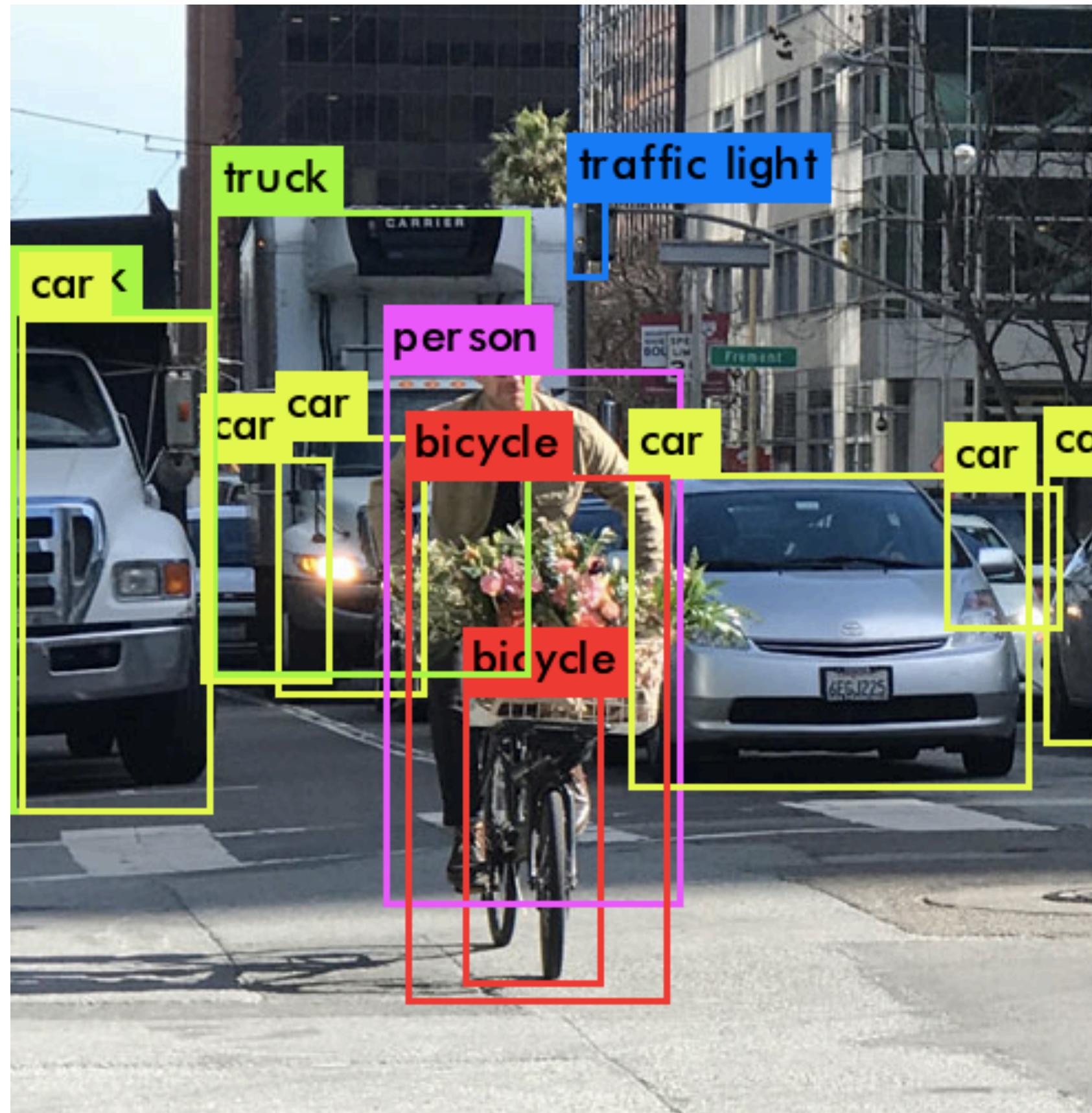
Analytics data: recommendations, credit rankings, new sentiment...

Alternative data: CCTV, satellite images, Twitter...

Practice time

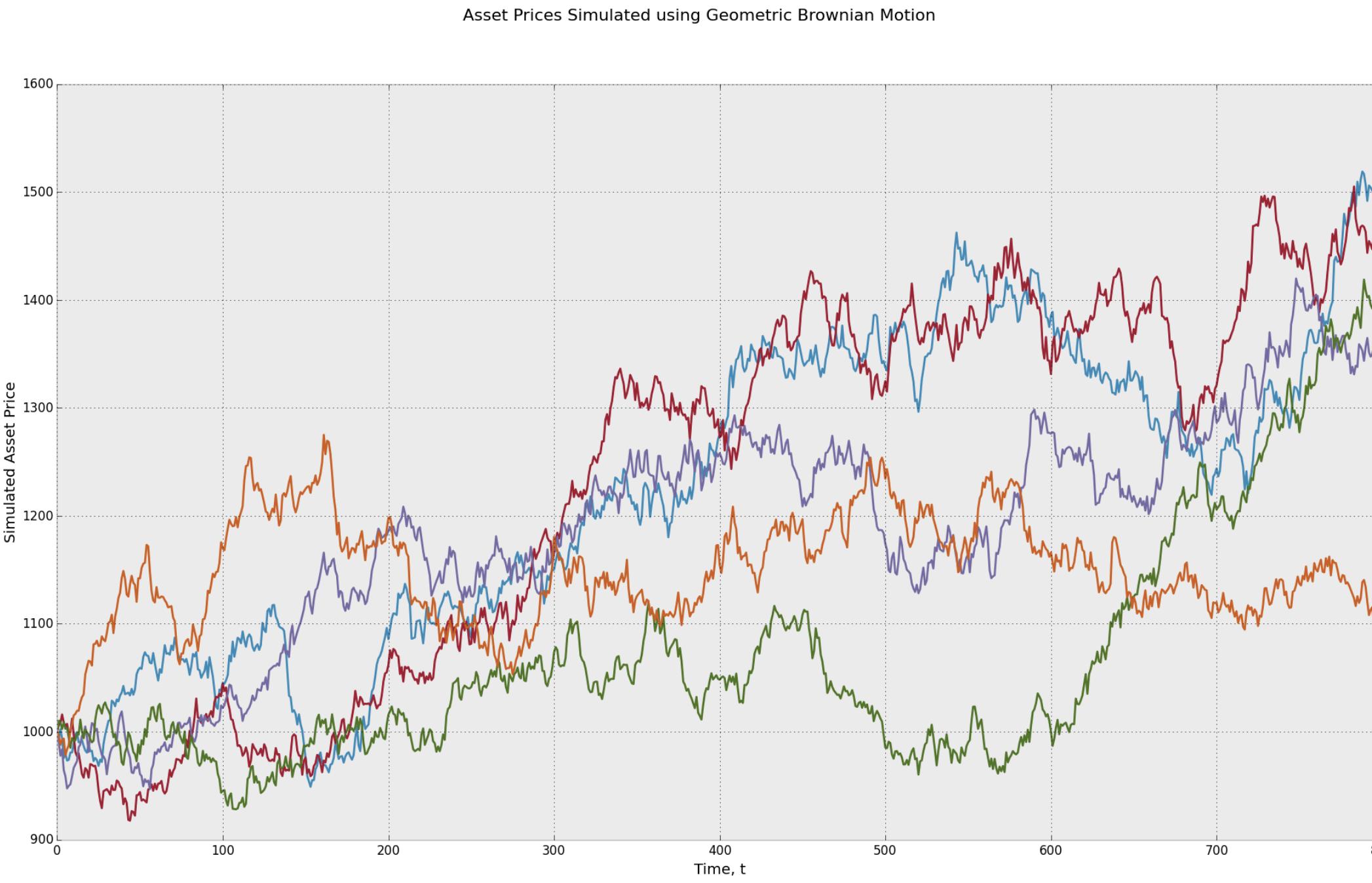
Finance ML vs “Normal” ML

Treating it as a regular ML exercise



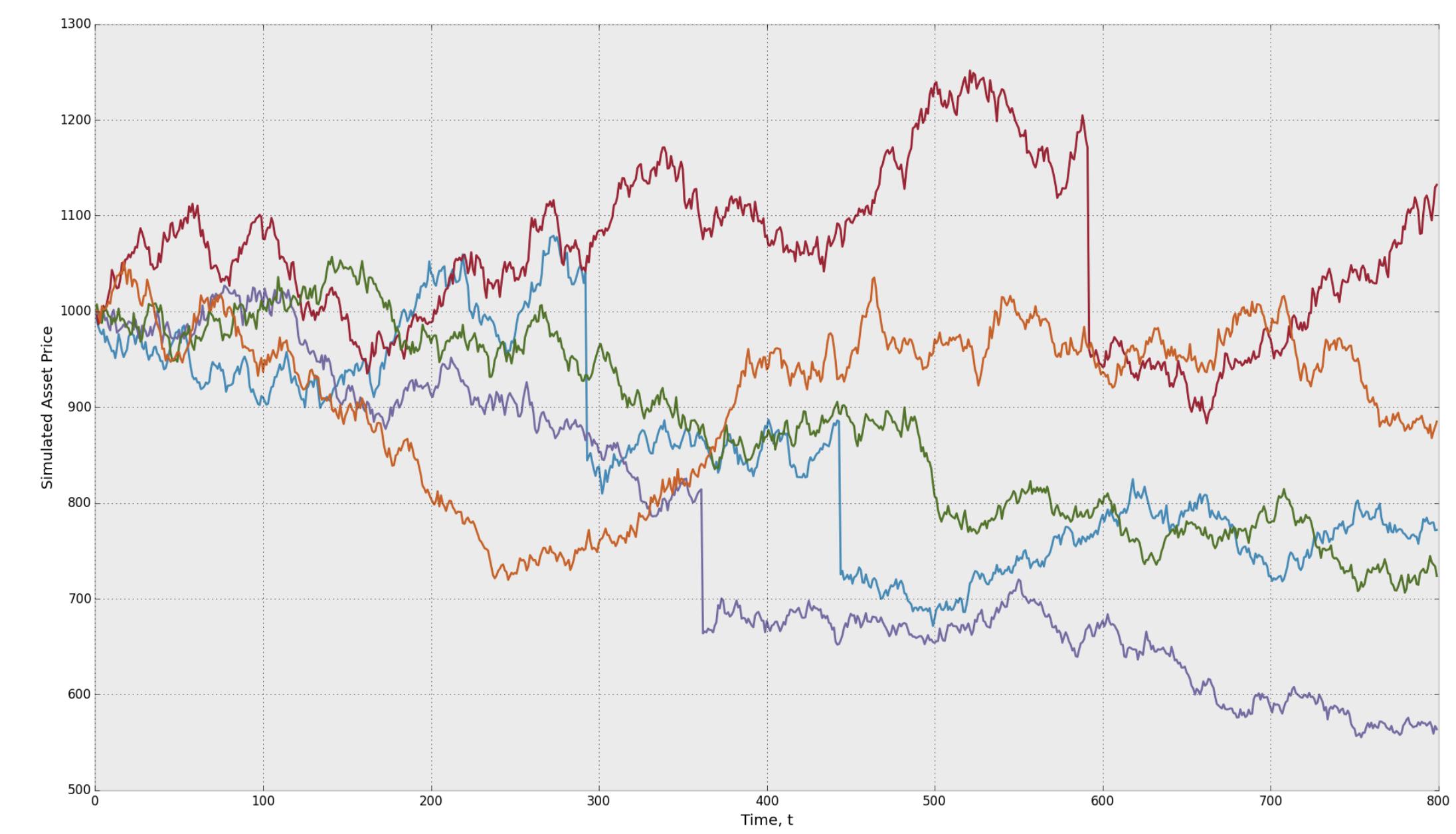
Treating it as a regular ML exercise

$$dS_t = \mu S_t dt + \sigma dS_t W_t$$



$$dS_t = \mu S_t dt + \sigma S_t dW_t + dJ_t$$

$$dJ_t = S_t d\left(\sum_{i=0}^{N_t} (Y_i - 1) \right)$$

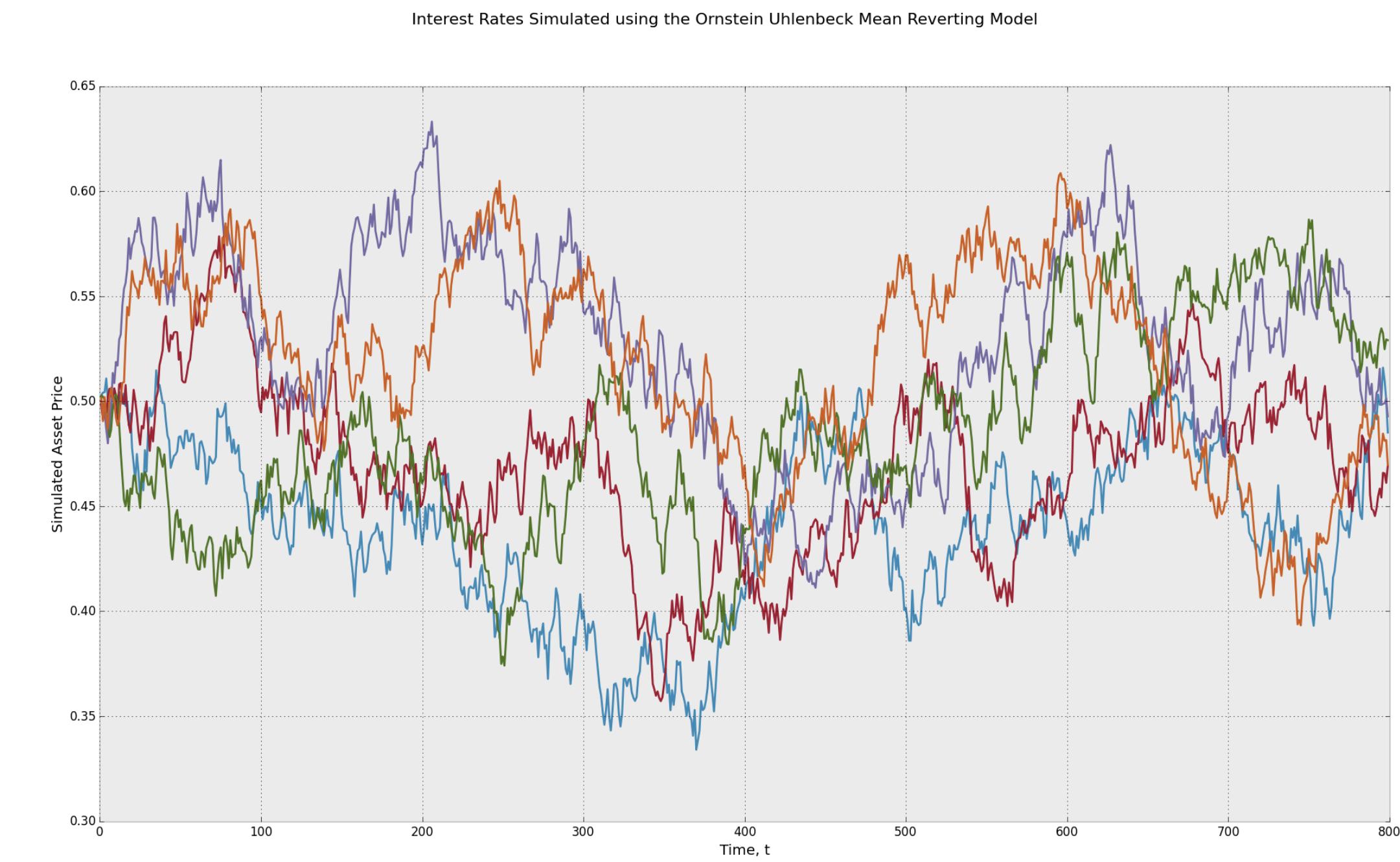
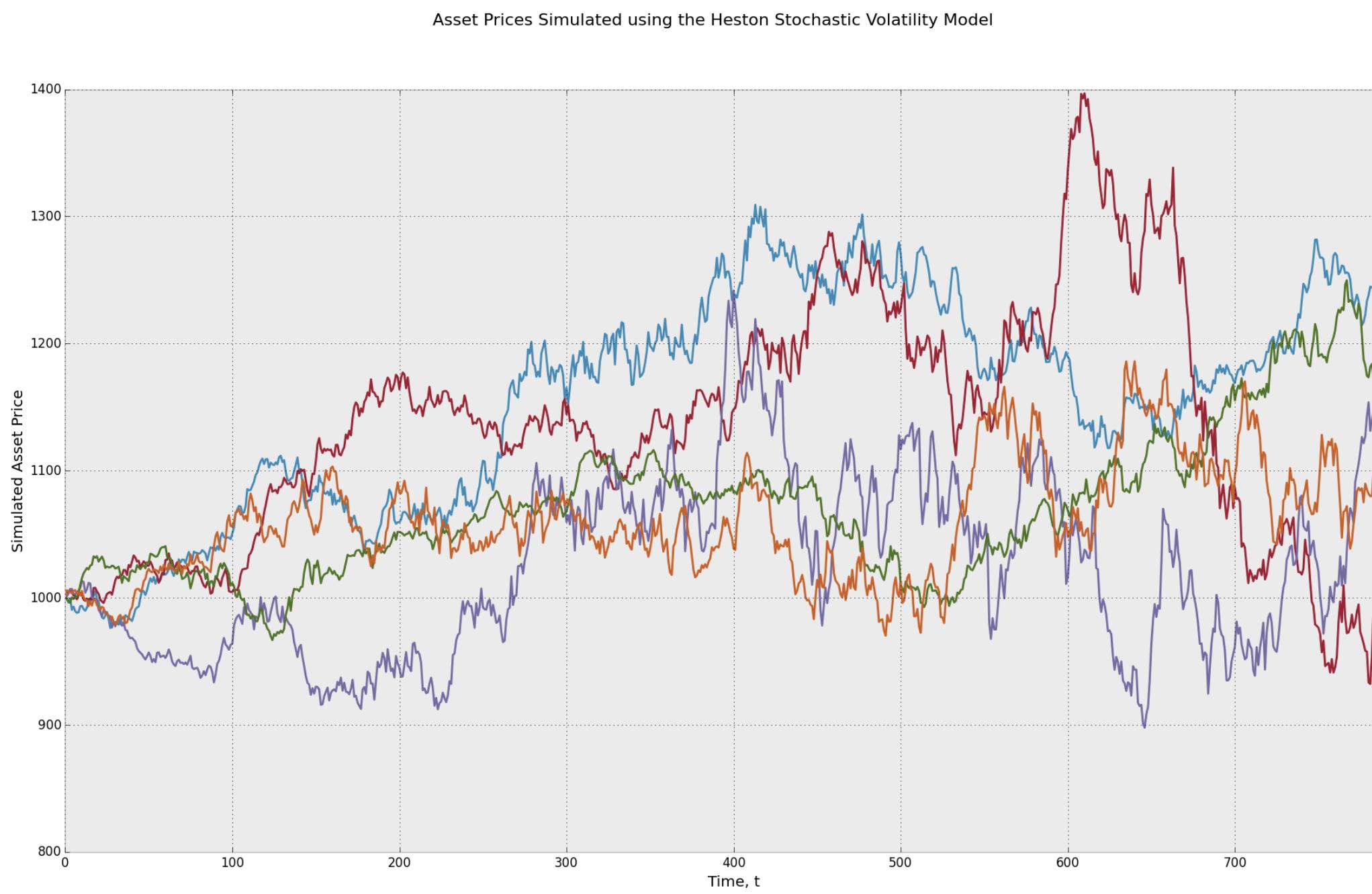


Treating it as a regular ML exercise

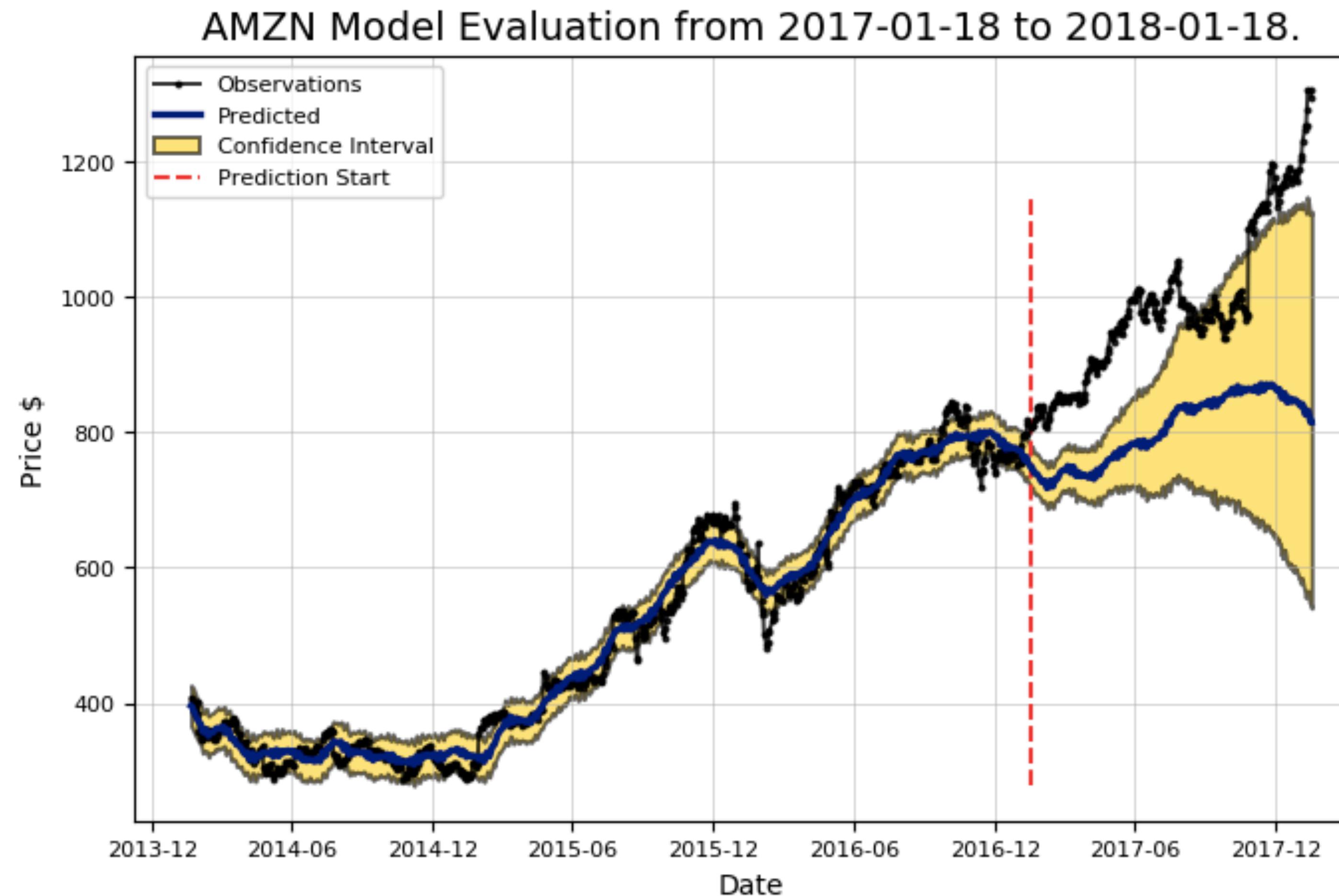
$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^S \text{ where}$$

$$dv_t = a(b - v_t)dt + \sigma \sqrt{v_t} dW_t^v \text{ (Cox Ingersoll Ross),}$$

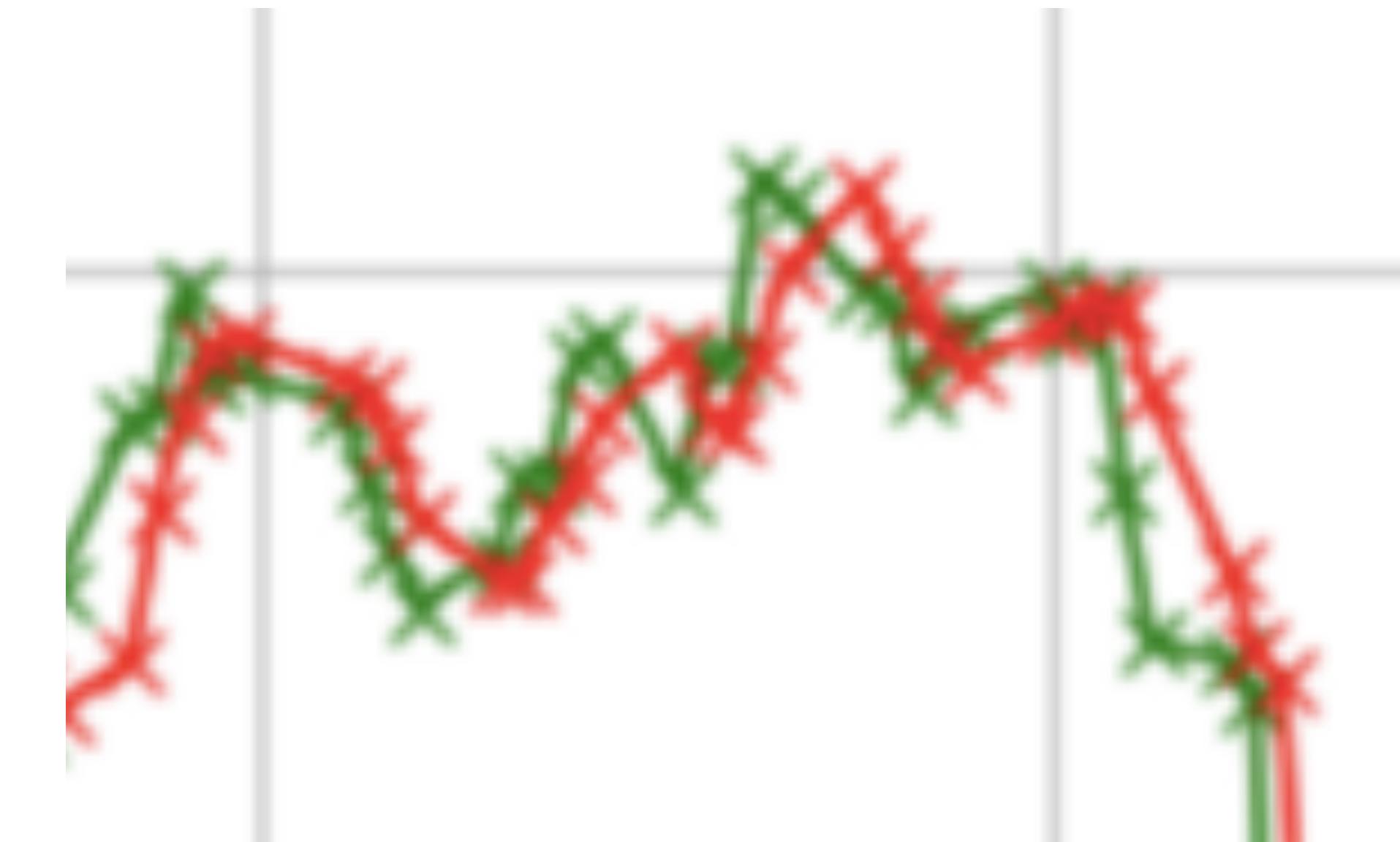
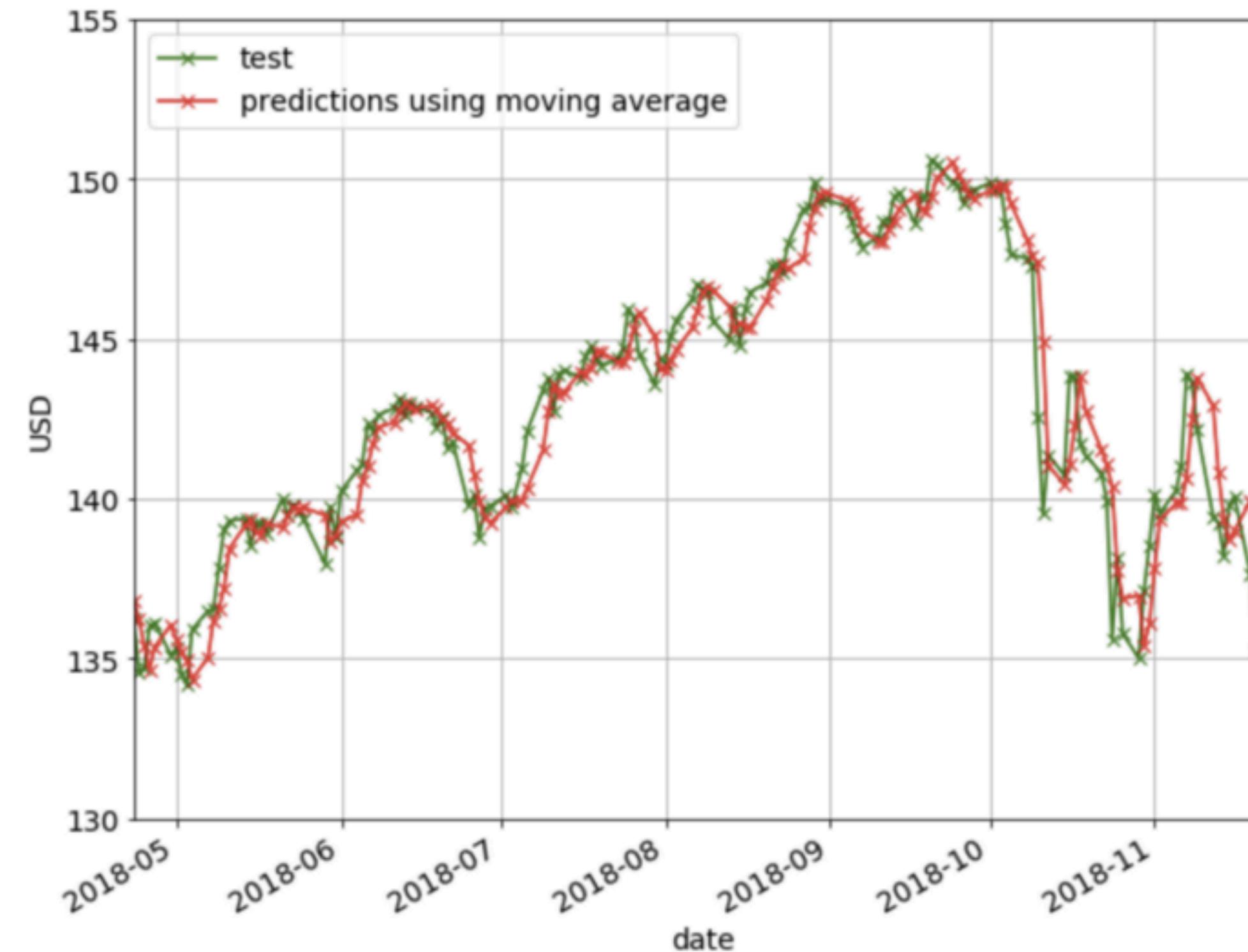
$$dr_t = a(b - r_t)dt + \sigma \sqrt{r_t} dW_t$$



Forecasting the price in the future



Forecasting the price in the future



Research via backtesting

SNIPPET 8.1 MARCOS' FIRST LAW OF BACKTESTING—IGNORE AT YOUR OWN PERIL

“Backtesting is not a research tool. Feature importance is.”

—Marcos López de Prado

Advances in Financial Machine Learning (2018)

SNIPPET 11.1 MARCOS' SECOND LAW OF BACKTESTING

“Backtesting while researching is like drinking and driving.

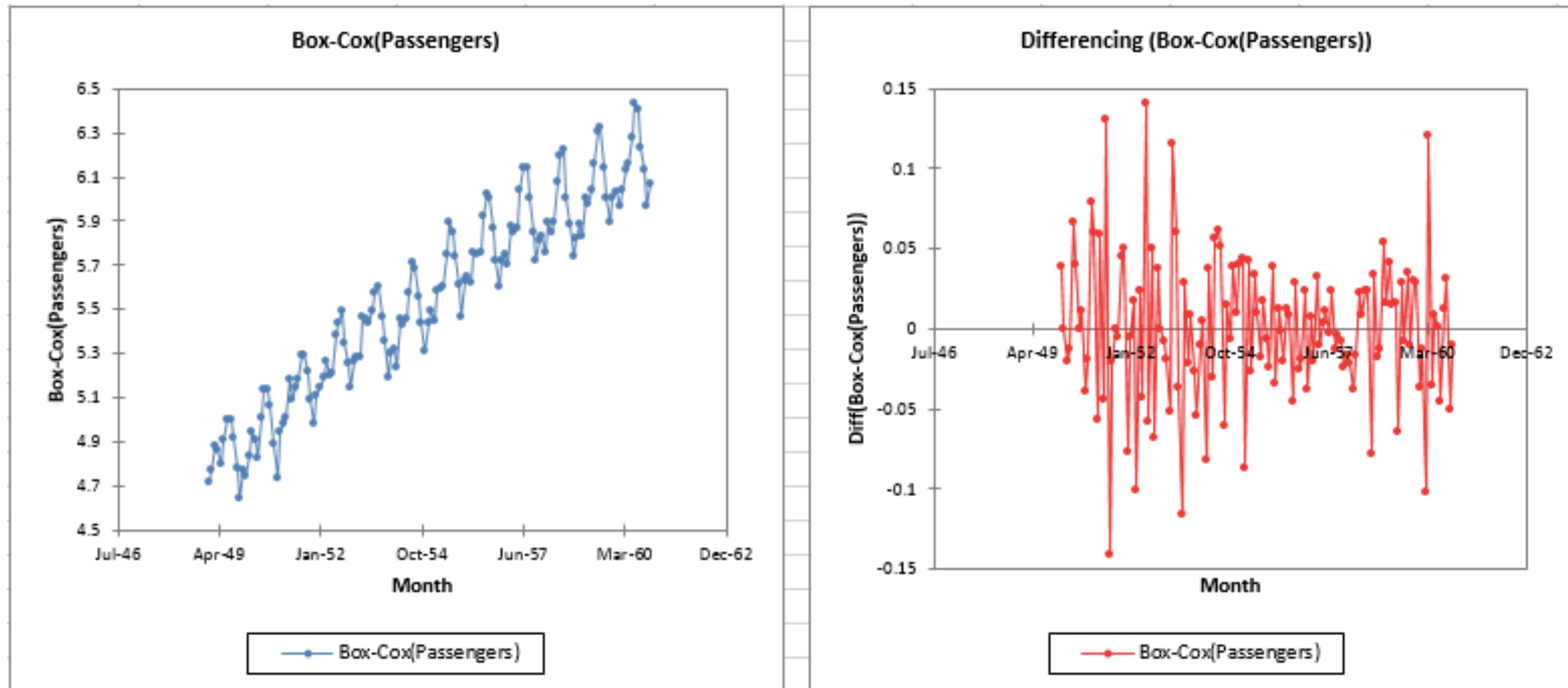
Do not research under the influence of a backtest.”

—Marcos López de Prado

Advances in Financial Machine Learning (2018)

Practice time

Treating it as a regular ML exercise



Treating it as a regular ML exercise

Difference operator [\[edit \]](#)

Main article: [Finite difference](#)

In time series analysis, the first difference operator : ∇

$$\nabla X_t = X_t - X_{t-1}$$

$$\nabla X_t = (1 - L)X_t .$$

Similarly, the second difference operator works as follows:

$$\nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1}$$

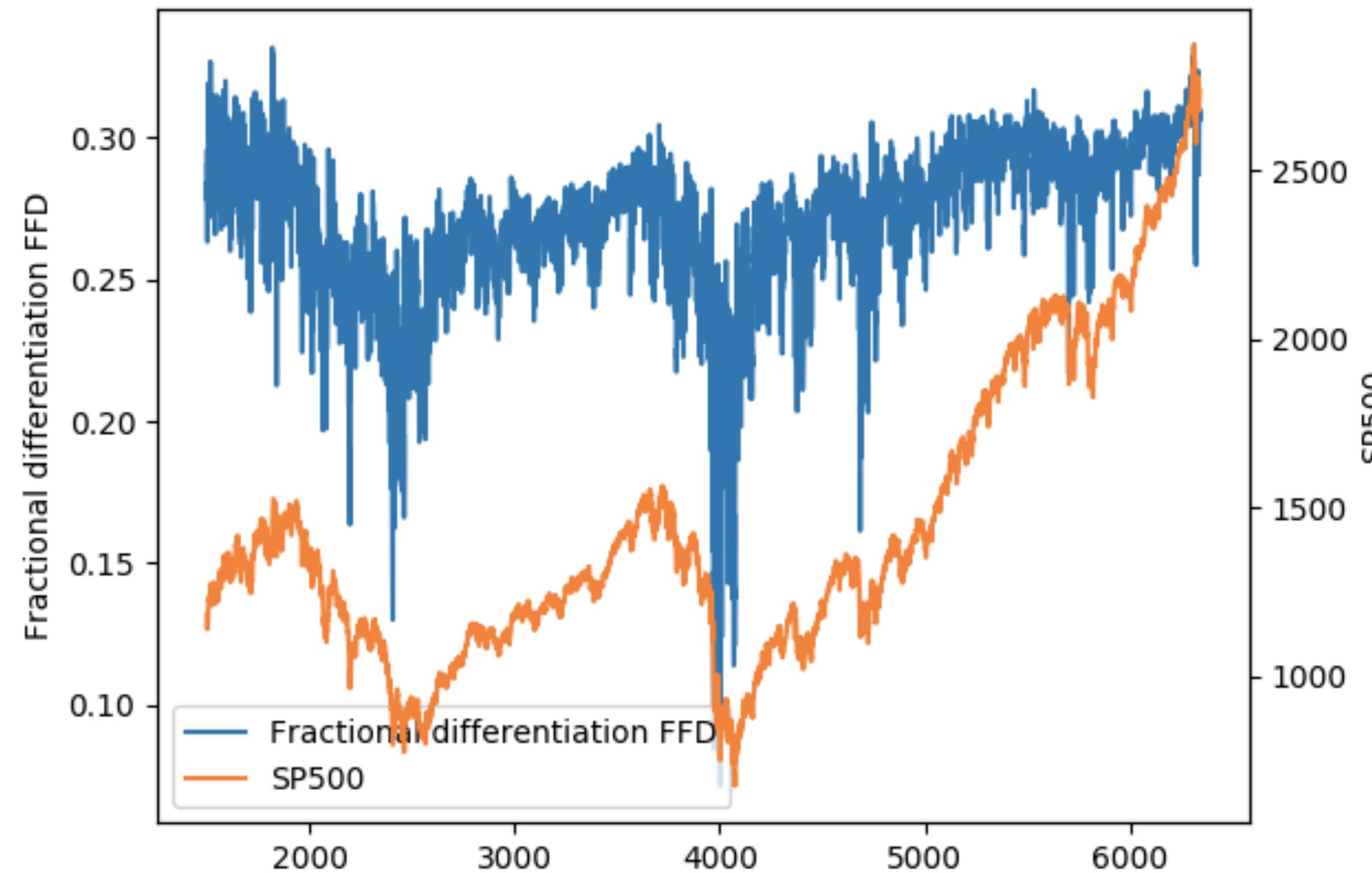
$$\nabla^2 X_t = (1 - L)\nabla X_t$$

$$\nabla^2 X_t = (1 - L)(1 - L)X_t$$

$$\nabla^2 X_t = (1 - L)^2 X_t .$$

The above approach generalises to the i -th difference operator $\nabla^i X_t = (1 - L)^i X_t .$

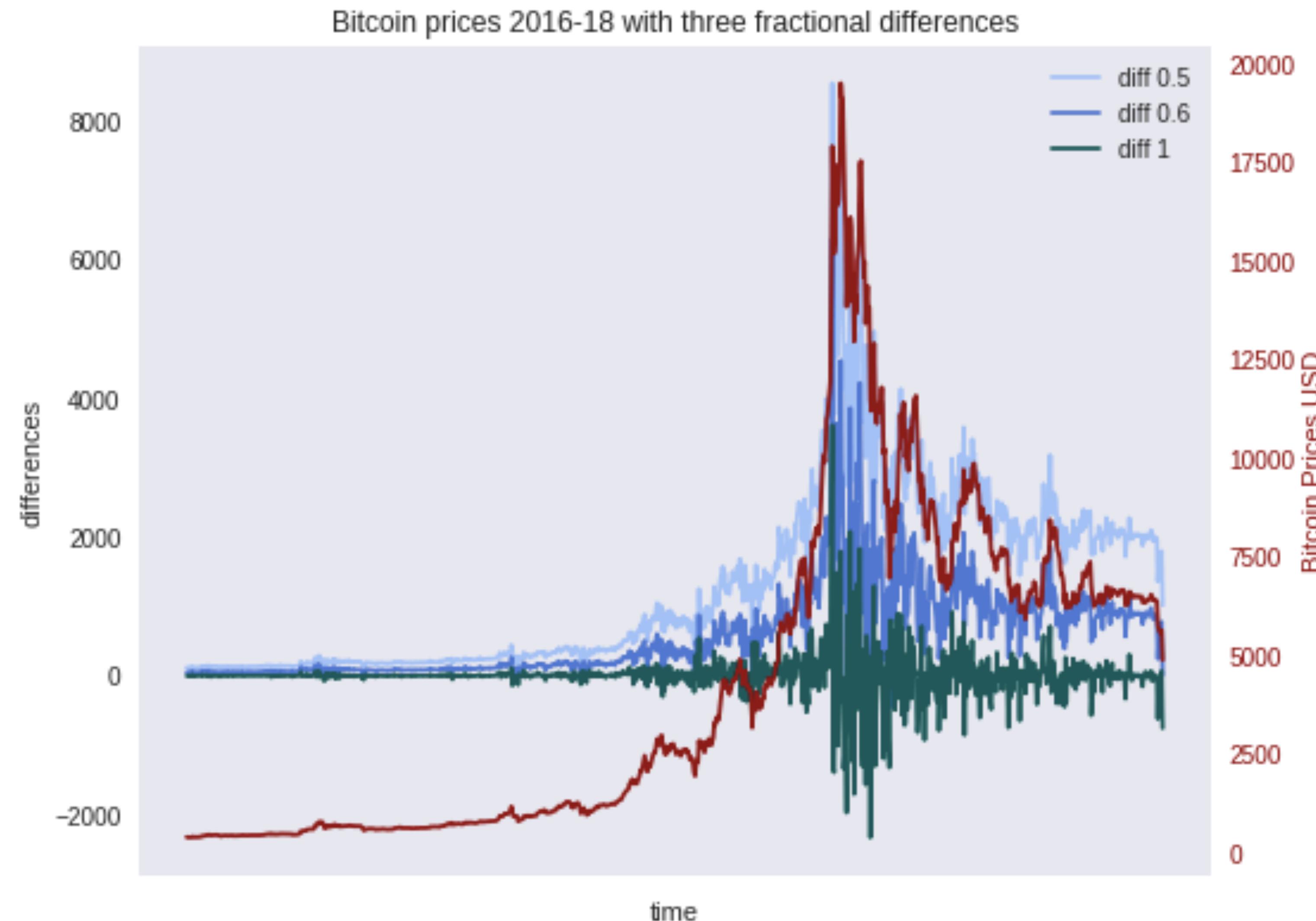
Treating it as a regular ML exercise



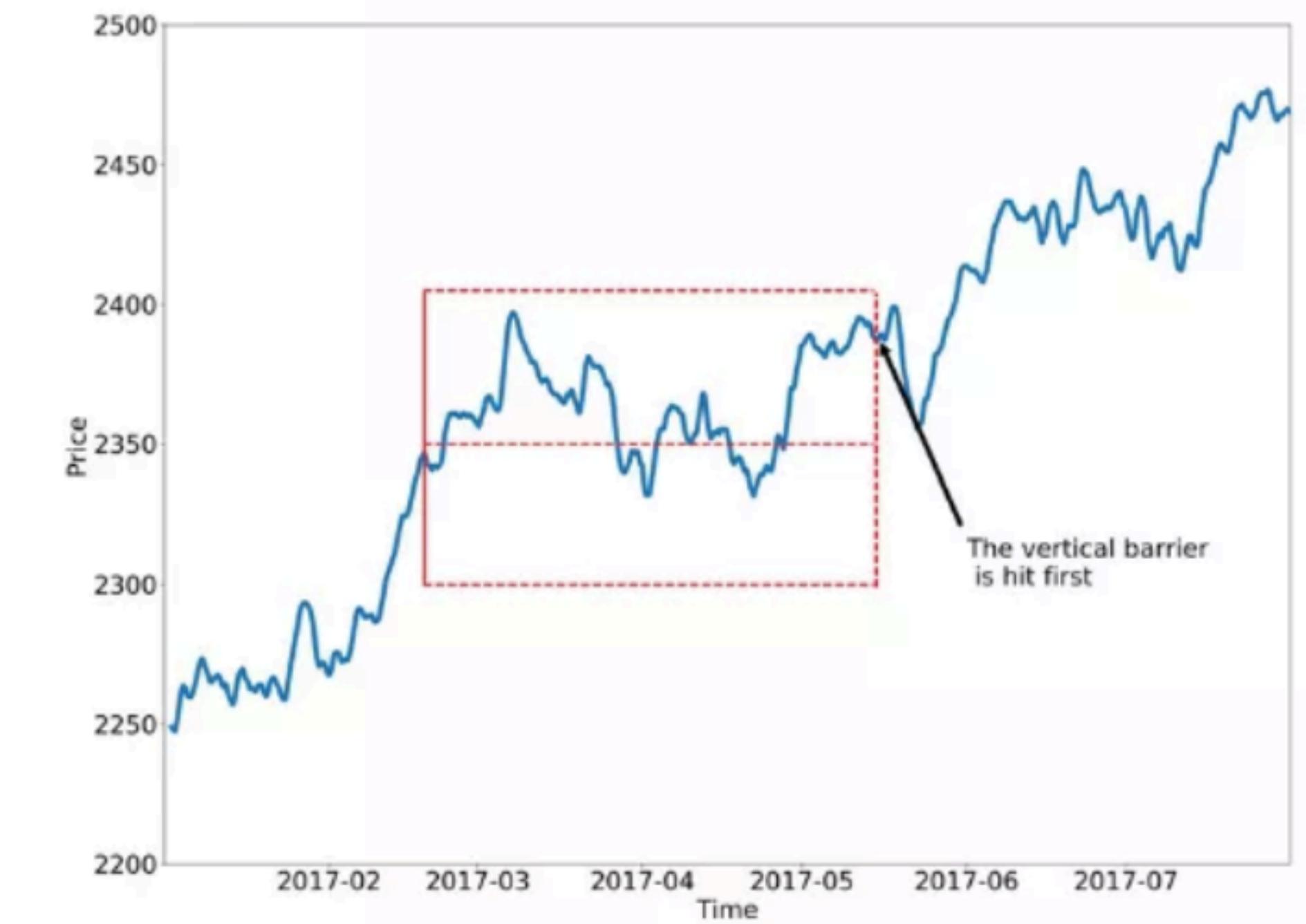
$$(\mathbf{I} - \mathbf{B})\mathbf{X}_t = \mathbf{X}_t - \mathbf{B}\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1}$$

$$\begin{aligned}(1 - B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\&= \sum_{k=0}^{\infty} \frac{\prod_{a=0}^{k-1} (d-a)}{k!} (-B)^k \\&= 1 - dB + \frac{d(d-1)}{2!} B^2 - \dots.\end{aligned}$$

Treating it as a regular ML exercise



Forecasting the price in the future



Backtesting is not a research tool

SNIPPET 8.1 MARCOS' FIRST LAW OF BACKTESTING—IGNORE AT YOUR OWN PERIL

“Backtesting is not a research tool. Feature importance is.”

—Marcos López de Prado

Advances in Financial Machine Learning (2018)

SNIPPET 11.1 MARCOS' SECOND LAW OF BACKTESTING

“Backtesting while researching is like drinking and driving.

Do not research under the influence of a backtest.”

—Marcos López de Prado

Advances in Financial Machine Learning (2018)

Practice time

EDA: correlations won't work

Sampling: all this time we were violating the IID hypothesis

Validation: all this time we even didn't try cross-validation

Feature Importance: all this time there were 5 random features in the dataset

Backtesting: all this time we were testing on one random sample from big universe

Probabilistic interpretation: what to do with that Random Forest?

Hyper-parameters: ... there are some specialties too :)

Homework

Inputs:

obviously, input data is still not perfect, fix the variables, maybe create new ones

Modeling:

why bagging of linear models works better than boosting of trees?

Feature Importance:

what are the redundant variables? how to detect them?

Literature analysis:

read 5+ papers on ML in stock price prediction, find all the mistakes ;)

GOAL:

to maximize the trust to the ML pipeline itself

HINT:

backtesting is not a research tool!

ML research directions

Simulations: scenario generations for the strategy tests

Investment management: beyond classical optimization, goal-based investments

Factor decomposition: reconstructing real market structure

Clustering: regime switching detection

Alternative data analysis: adding external variables to the game

Materials

 Quantopian Lectures

Lectures

Lecture 1	Introduction to Research	A simple tutorial.
Lecture 2	Introduction to Python	Some basic tools.
Lecture 3	Introduction to NumPy	How to use NumPy.
Lecture 4	Introduction to pandas	An introduction to pandas.
Lecture 5	Plotting Data	A brief primer.
Lecture 6	Means	Measures of centrality.
Lecture 7	Variance	Measures of dispersion.

 QuantStart

Successful Algorithmic Trading

A step-by-step guide to developing systematic trading strategies using the Python programming language



Michael L. Halls-Moore, PhD.

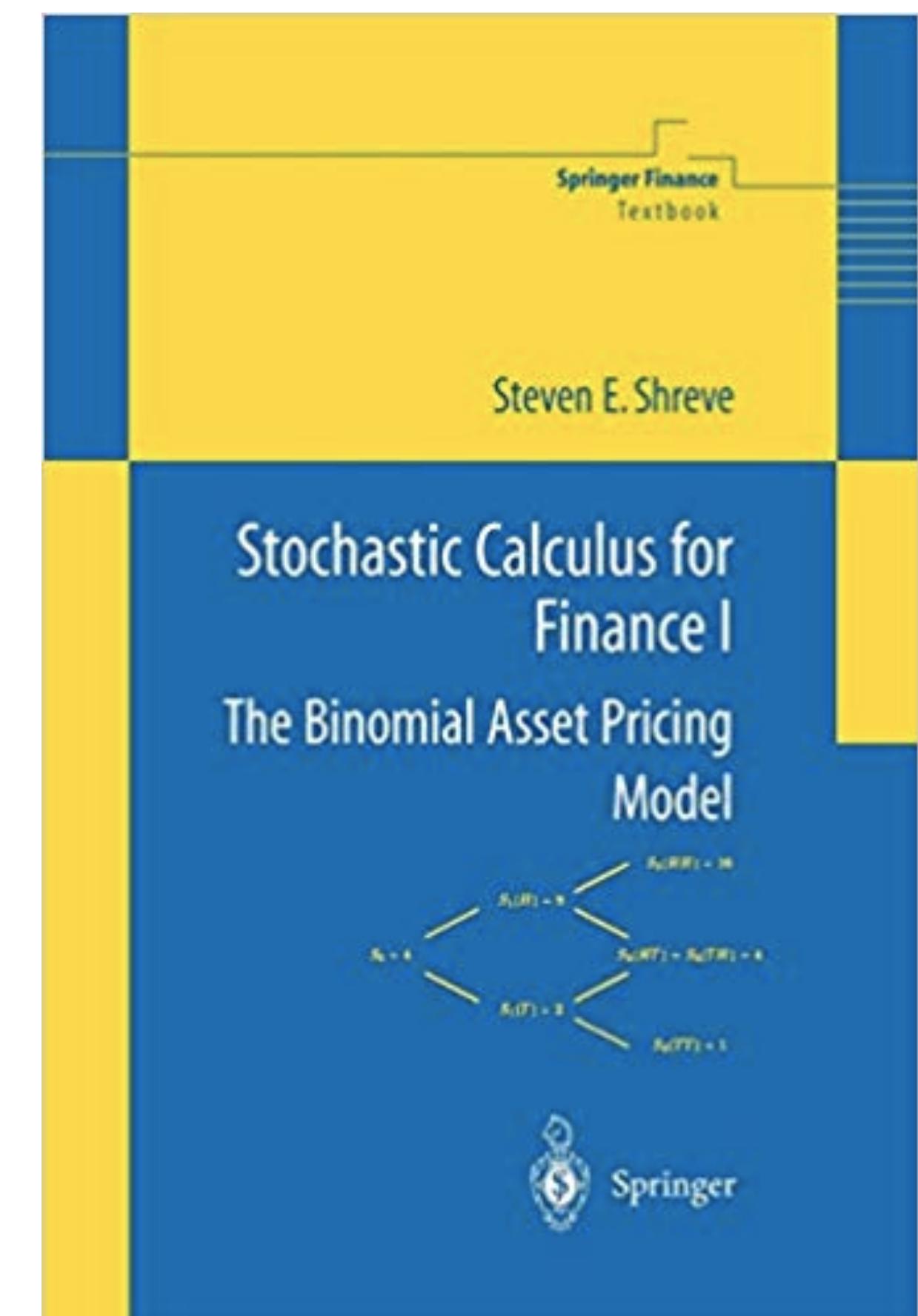
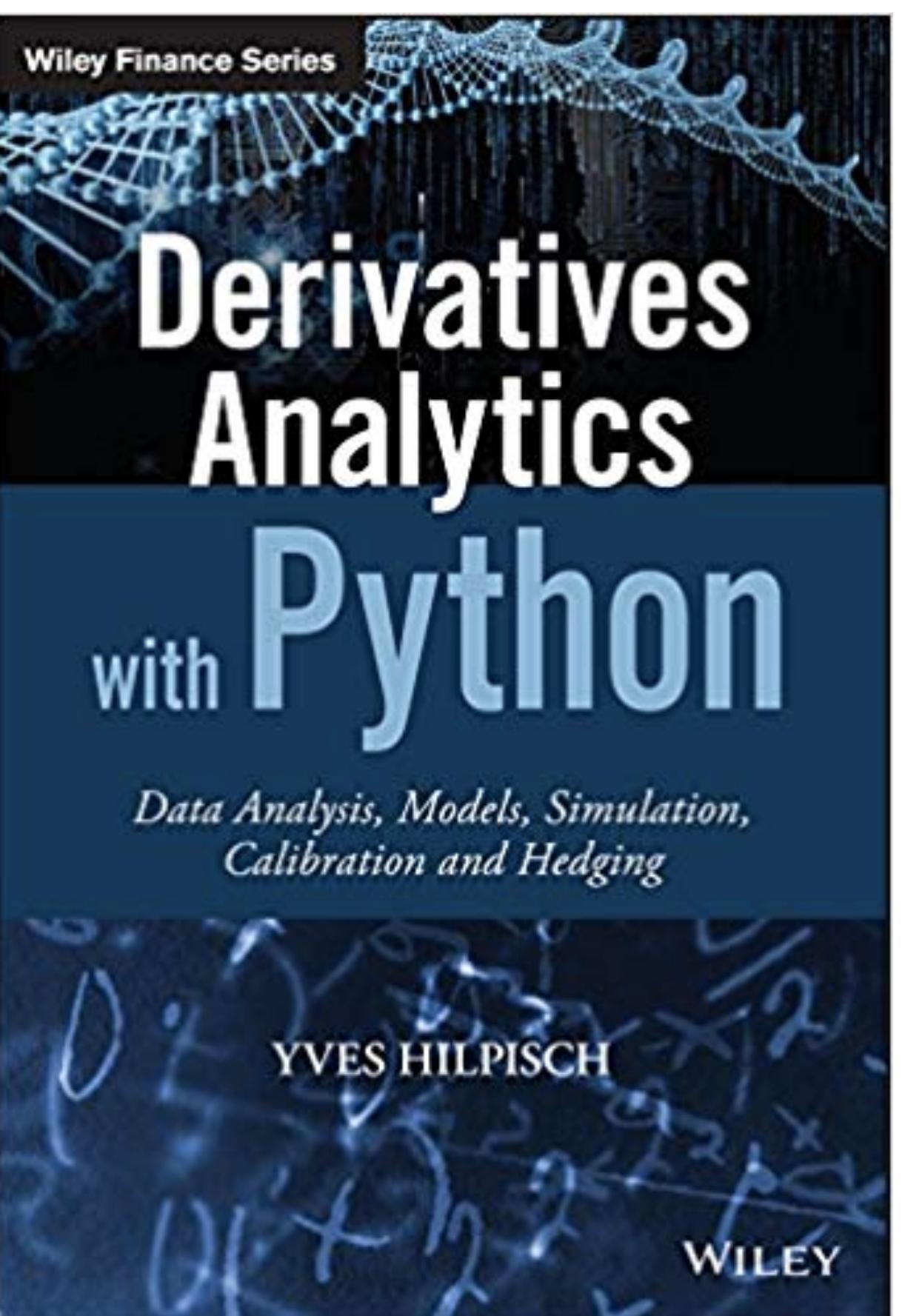
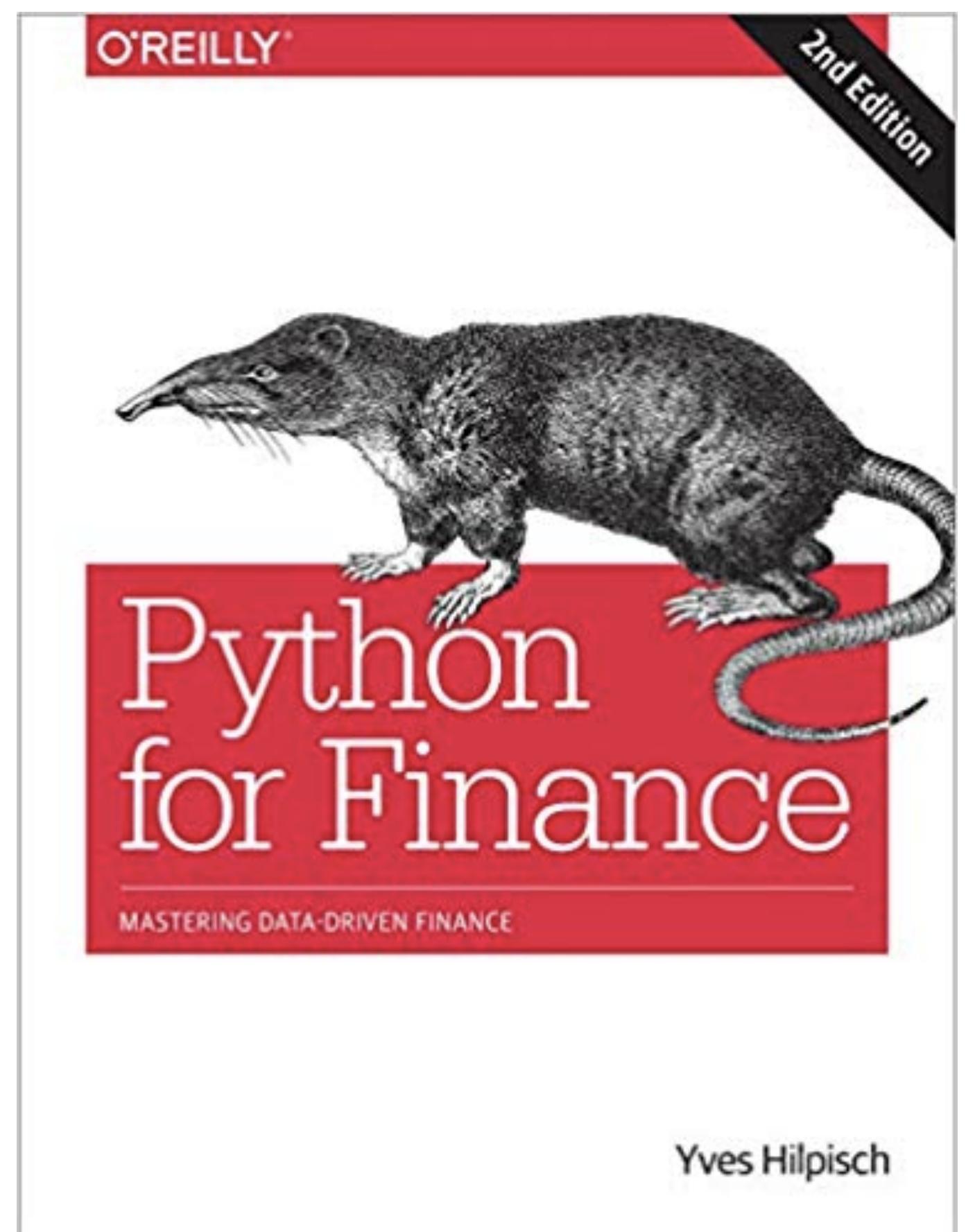
 QuantStart

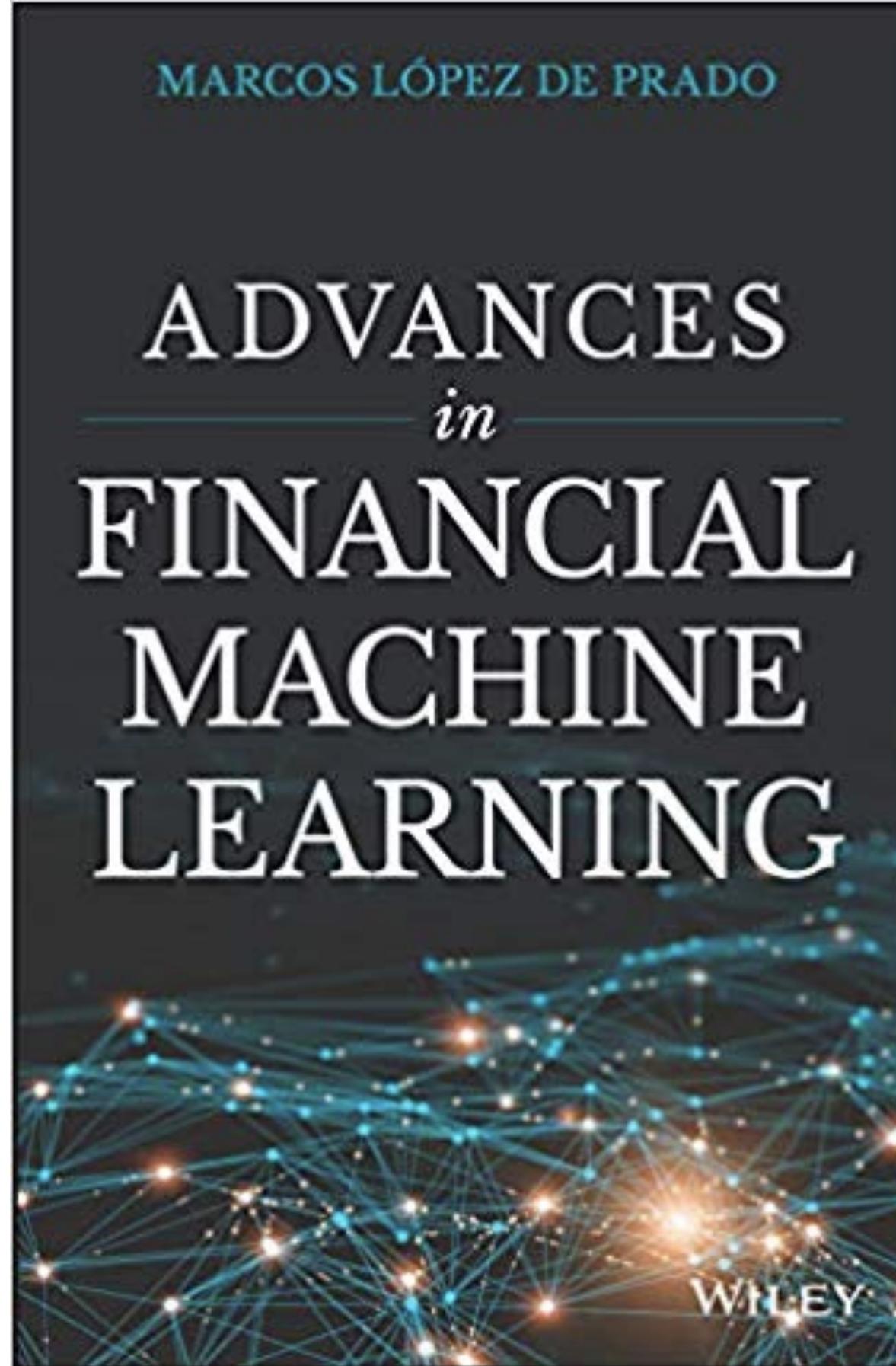
Advanced Algorithmic Trading

Bayesian statistics, time series analysis and machine learning for profitable systematic trading strategies



Michael L. Halls-Moore, PhD.





Browse > Data Science > Machine Learning

Machine Learning and Reinforcement Learning in Finance Specialization

Reinforce Your Career: Machine Learning in Finance. Extend your expertise of algorithms and tools needed to predict financial markets.

★★★★★ 3.8 (822 ratings)



Igor Halperin

Enroll

Starts Feb 29

Premium Access: €34/mo

Financial aid available

9,092 already enrolled

id	era	feature1	...	feature310	target
n2b2e3dd163cb422	era1	0.75	...	0.00	0.25
n177021a571c94c8	era1	1.00	...	0.25	0.75
n7830fa4c0cd8466	era1	0.25	...	1.00	0.00
nc584a184cee941b	era1	0.25	...	0.00	1.00
nc5ab8667901946a	era1	0.75	...	0.25	0.25
n84e624e4714a7ca	era1	0.00	...	0.75	1.00

<http://numer.ai/>