Financial Tick Data Analysis with pandas

Dr. Yves J. Hilpisch | Jason Ramchandani

PyConf Hyderabad, Online, 05. December 2020

THE PYTHON QUANTS

The AI Machine

Introduction

THE PYTHON QUANTS

SERVICES
for financial institutions globally

EVENTS
for Python quants &
algorithmic traders

CERTIFICATION
in cooperation with university

TRAINING
about Python for finance
& algorithmic trading

BOOKS
about Python and
finance

PLATFORM
for browser-based
data analytics

OPEN SOURCE
Python library
for financial analytics

http://tpq.io

The AI Machine

http://aimachine.io

http://certificate.tpq.io/tpq_top_algo_2019.pdf

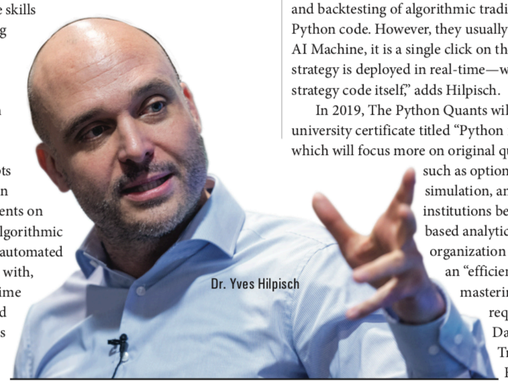
The Python Quants

First University Certificate in Python for Algorithmic Trading

Python programming has become a key skill in the financial industry. In areas such as financial data science, computational finance or algorithmic trading, Python has established itself as the primary technological platform. At the same time, the level of Python sophistication the industry is expecting from its employees and applicants is increasing steadily. The Python Quants Group is one of the leading providers of Python for Finance training programs.

Among others, The Python Quants have tailored a comprehensive online training program leading to the first University Certificate in Python for Algorithmic Trading. Be it an ambitious student with intrigue for algorithmic trading, or a major financial institution, The Python Quants, through this systematic training program, is equipping delegates with requisite skills and tools to formulate, backtest and deploy algorithmic trading strategies based on Python.

The topics covered in the training programs offered by The Python Quants are generally not found in the typical curriculum of financial engineering or quantitative finance Master programs. Dr. Yves Hilpisch, the firm's founder and managing partner, explains, "There are courses out there that show students how to apply machine learning for the formulation and backtesting of algorithmic trading strategies. However, none of them explains the difficulties or the skills required in deploying such algorithmic trading strategies in the real world. Besides providing an introductory course that teaches Python and financial concepts from scratch, we train our delegates and clients on how best to deploy algorithmic trading strategies in automated fashion in the cloud, with, among others, real-time risk management and monitoring," explains Hilpisch, an author of three books on the topic, with "Python for Finance" (2nd ed., O'Reilly) being the standard reference in the field.

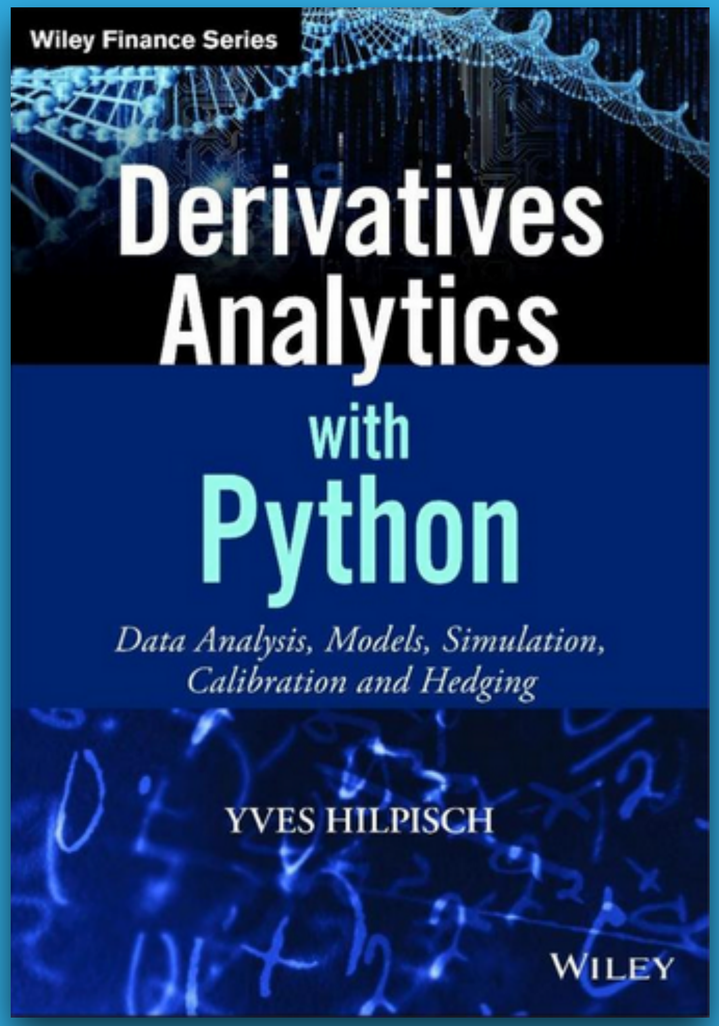

Dr. Yves Hilpisch

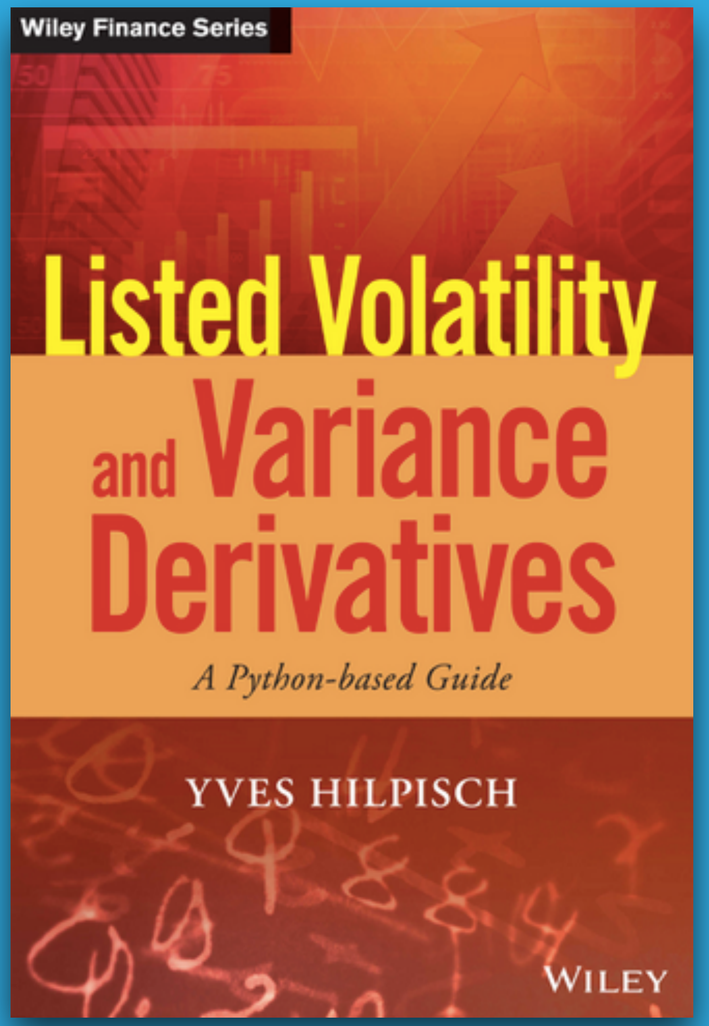
The organization's "Python for Algorithmic Trading University Certificate" consists of 200 hours of instruction, 1,200 pages of documentation and 1,000s of lines of Python code. In addition to offering both online and offline Python training, Hilpisch and his team also organize bespoke training events for financial institutions, hedge funds, banks, and asset management companies. "Most of the training is online since we have students and delegates from about 65 different countries in general. Most recently, we noticed that it's not just financial firms and students who want to deepen their algorithmic trading knowledge, but even professors of finance who want to get more involved in this popular topic," says Hilpisch.

While the Quant Platform is the most popular choice, especially for users in the financial sector who don't have access to a full-fledged, interactive, financial analytics environment, the team at The Python Quants is currently developing The AI Machine—a new platform which leverages artificial intelligence to formulate and deploy algorithmic trading strategies in a standardized manner. Hilpisch explains that it's relatively easy to write Python code for an algorithmic trading strategy, but the same can't be said about the deployment of such a strategy. "There are a few platforms out there that allow the formulation and backtesting of algorithmic trading strategies by the use of Python code. However, they usually stop exactly there. With The AI Machine, it is a single click on the 'GO LIVE' button and the strategy is deployed in real-time—without any changes to the strategy code itself," adds Hilpisch.

In 2019, The Python Quants will be introducing a new university certificate titled "Python for Computational Finance," which will focus more on original quantitative finance topics, such as option pricing, Monte Carlo simulation, and hedging. As financial institutions begin to perceive Python-based analytics as a prerequisite skill, the organization will continue to provide an "efficient and structured way of mastering all the tools and skills required in Python for Financial Data Science, Algorithmic Trading, and Computational Finance."CM

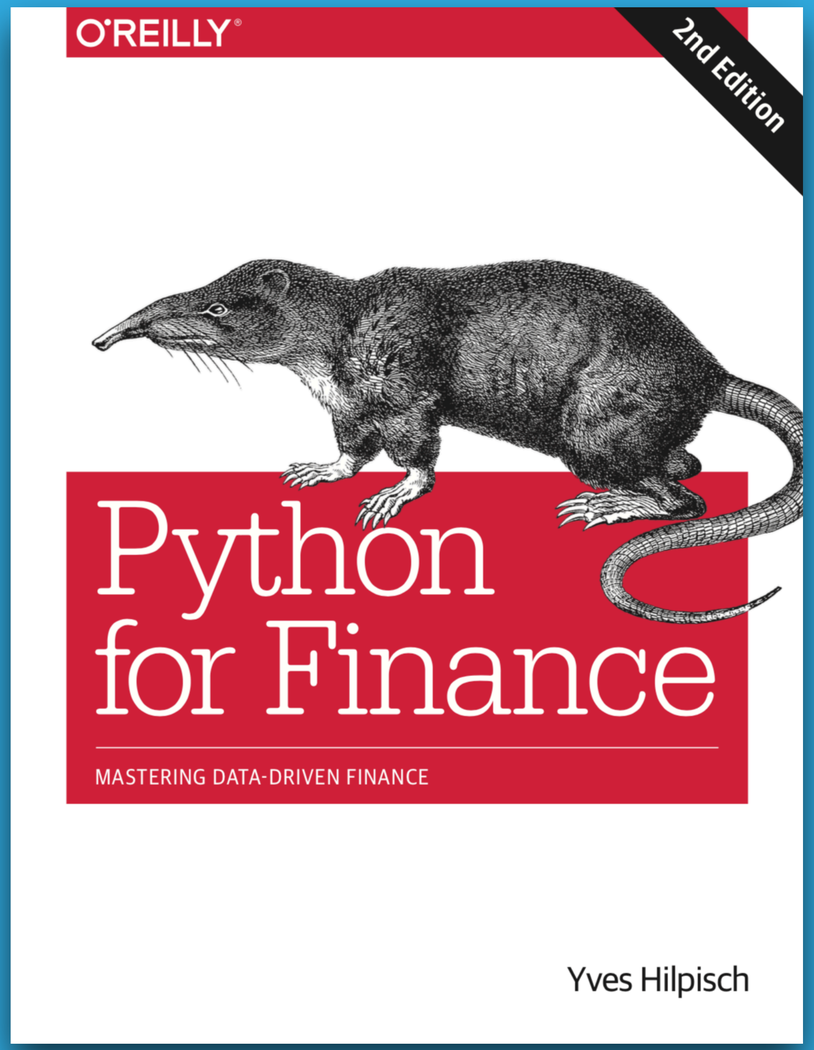
Quant Finance with Python

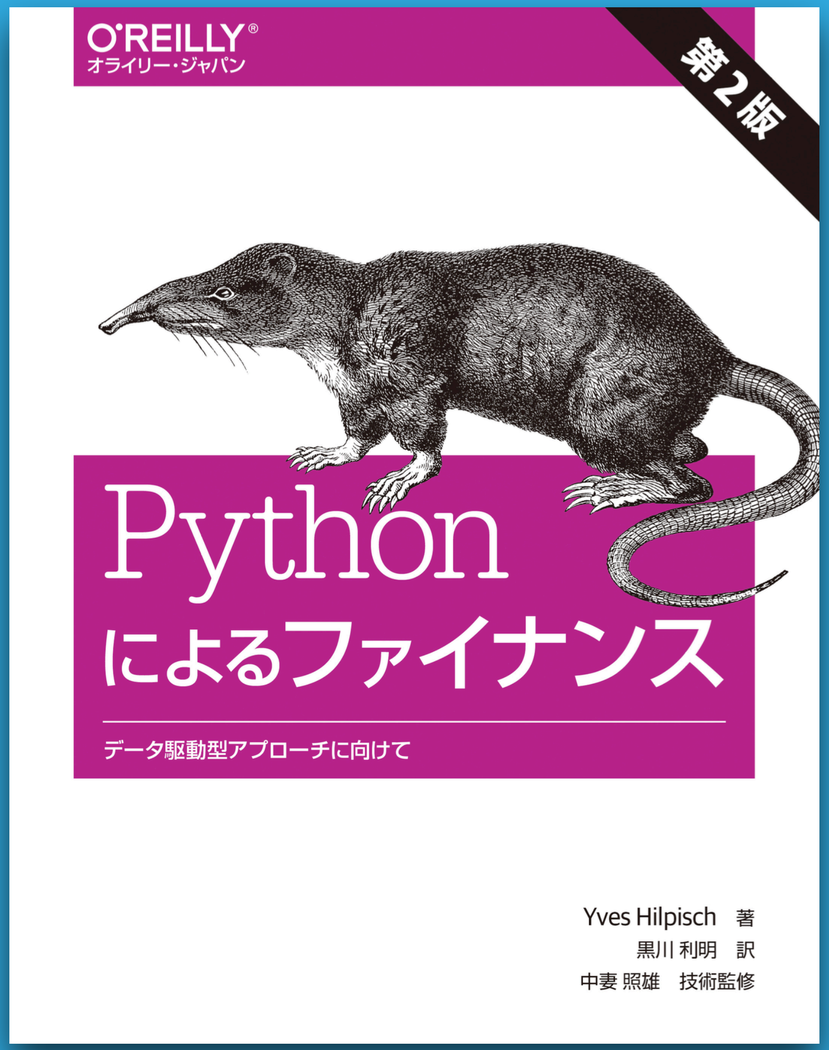




http://books.tpq.io

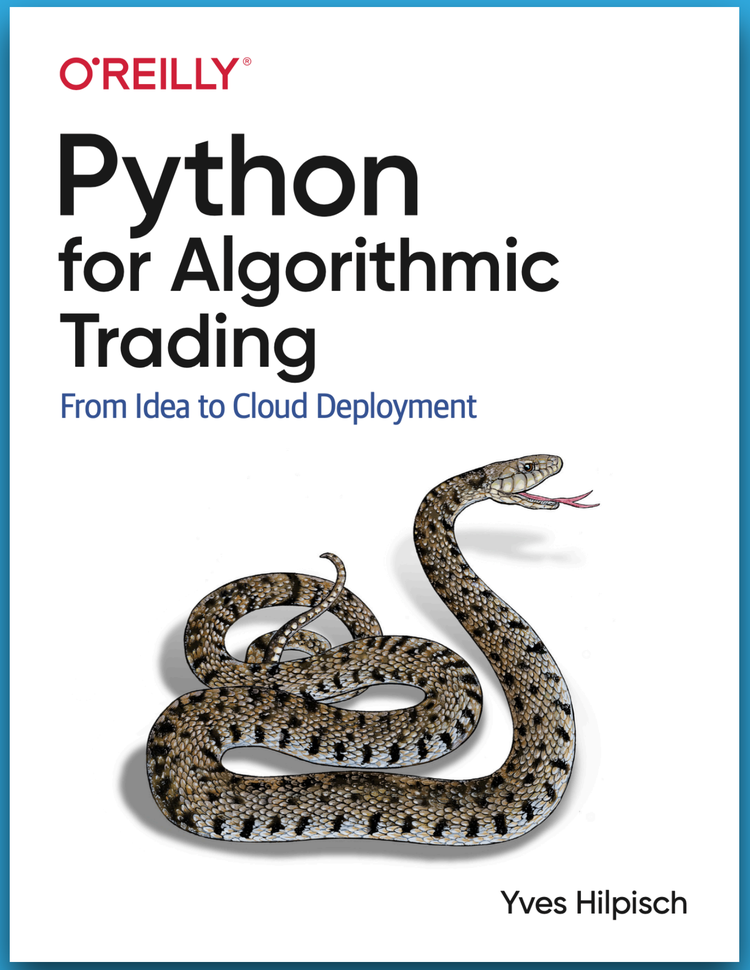
Python for Finance

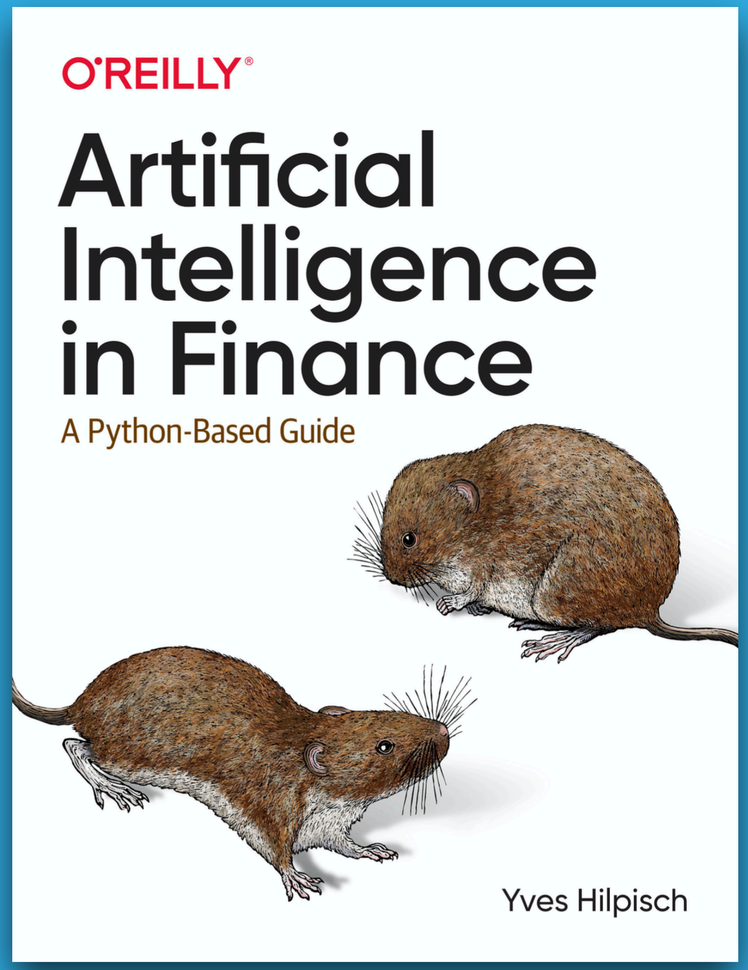




http://books.tpq.io

Python & AI for Finance & Trading





http://books.tpq.io

AGENDA

1. Data-Driven Finance
2. The Financial Tick Data
3. Data Preprocessing
4. Efficient Markets
5. AI-First Algorithmic Trading
6. AI-Powered Market Prediction
7. Conclusions

Data-Driven Finance

FINANCIAL TIMES
TUESDAY 3 OCTOBER 2017
WORLD BUSINESS NEWSPAPER
MIDDLE EAST
Mohamed El-Erian
Reasons to worry about the Fed's 'beautiful normalisation' — PAGE 20
Torturing Theresa
Boris Johnson's bid to dictate May's Brexit strategy — JANAN GANESH, PAGE 11
Confined in a circle
The myths that hold back women in Indian society — AMY KAZMIN, PAGE 10
Las Vegas reels from worst US mass shooting
Briefing
Catalan president urges Brussels to mediate in independence clash
Region seeks to avoid 'traumatic split' from Spain
EU says dispute is 'internal matter'
Datawatch
Knowledge.
Integrity.
Discretion.
Of all our fine properties, these are the most valuable.
A Nikkei Company
Financial Times
Secure
https://www.ft.com
FTSE 100 -0.08%
S&P 500 +0.37%
Euro/Dollar -0.17%
Pound/Dollar -0.20%
Brent Crude Oil +0.33%
Visit Markets Data
FINANCIAL TIMES
myFT
HOME
WORLD
US
COMPANIES
MARKETS
OPINION
WORK & CAREERS
LIFE & ARTS
Portfolio
Account Settings
fastFT
Fast Europe Open: Italy, eurozone industrial production
Tencent-backed carmaker NIO prices shares at bottom of range
Oil prices climb as hurricane threatens US
Hong Kong bear market worsens as Asia stocks fall
US banks
Jamie Dimon hands more responsibilities to lieutenants
Wall Street's longest-serving chief executive says he is 'more like the coach now'
AN HOUR AGO
Lehman/US bank capital: loss cause
After the crisis, the banks are safer but debt is a danger
Financial crisis 2008: A reporter's memories from the front lines
FT Series Financial crisis: Are we safer now?
JPMorgan: defying attempts to end 'too big to fail'
AN HOUR AGO
Feedback
TODAY IN PERSONAL JOURNAL
A Celebrity Selfie Machine
PLUS Secondhand Stress at Work
THE WALL STREET JOURNAL.
WEDNESDAY, DECEMBER 11, 2013 - VOL. CCLXII NO. 138
WSJ.com
$2.00
What's News
Business & Finance
World-Wide
Deal Brings Stability to U.S. Budget
Congressional Negotiators Avert January Shutdown and Soften Sequester Cuts; Airline Fees to Climb
Thousands Honor Memory of South Africa's Liberator
RIP NELSON MANDELA
Longtime Insider Is GM's First Female CEO
PARTY DISCIPLINE
China Spins New Lesson From Soviet Fall
Ukrainian Forces Confront Protesters
Bank Rule Challenges Wall Street
Here's Your Holiday Bonus, Now Start Running
Workers Win All-They-Can-Grab Sprees From Companies; 'Supermarket Sweep'
Every investor is unique. That's why there's iShares Core.
iShares by BLACKROCK

AAPL.O VOV

APPLIB

MARKETS | COMPANY | NEWS | CHARTING | PORTFOLIOS & MONITORS | SEARCH TOOLS | DEVELOPMENT AND ADMIN TOOLS | MY LAYOUTS

AAPL.O APPLE INC

123.08 c USD Vol 13,358 CAM 72 CCR 96

Updated: 03-Dec-2020 11:00:20
Valuation VOV

United States | NASDAQ Global Select Consolidated | Phones & Handheld Devices

Overview | News & Research | Price & Charts | Estimates | Financials | ESG | Event | Ownership | Debt & Credit | Peers & Valuation | Derivatives | Filings

PRICE PERFORMANCE

Open	
Prev. Close	123.08
Bid / Ask	123.24 / 123.35
VWAP	
Turnover	--
Volume	13,358
Short Interest	0.500%
YTD	67.66%
Beta (5Y Monthly)	1.272
Mkt Cap - Default	USD 2.093T
PE (LTM)	37.676
Div Yield	0.666%
DR	BRL AAPL34.SA (1:0.1)
DR Type	--
DR Bank	--

Today | 5D | 3M | 6M | 1Y | 5Y | YTD

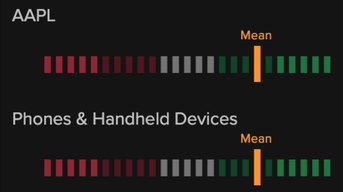
No Benchmark

52Wk: 53.153 (23-Mar) — Last 123.13 (Daily) — 137.980 (02-Sep)

Next Earn Report: 26-Jan-2021

Free Float	16.99B	Asset Ty...	Ordinary Share	5 yr CDS	28.35 bps
Outstanding	17.00B	Share Class	--	Δ Today	1.00%
IPO Date	12-Dec-1980	Lot Size	100	Δ 1 Week	2.72

RECOMMENDATIONS



Analysts Per Level	
Strong Buy	11
Buy	19
Hold	8
Sell	1
Strong Sell	1

Price Target (Mean) USD 125.92 Upside 2%

STARMINE MODELS

Highlights | All Models

Bullish		Bearish	
Price Momentum	97	Intrinsic Valuation	19
Analyst Revisions	79	Relative Valuation	23
Smart Holdings	77	Insider	27
Short Interest	95		
Earnings Quality	95		
Credit Risk - Combined	96		
Credit Risk - Smart Ratios	92		
Credit Risk - Structural	88		
Credit Risk - Text Mining	85		
Combined Alpha Model	72		

RELATIVE VALUATION

	Global Rank	Trailing 12 Months AAPL	Trailing 12 Months Industry Median	Next 12 Months AAPL	Next 12 Months Industry Median
PE	32	36.284	8.368	30.782	13.425
EV/EBITDA	24	26.610	8.037	22.303	8.094
Div Yield	31	0.653%	1.808%	0.666%	3.242%
EV/Sales	18	7.514	1.149	6.363	1.013
P/CF	31	26.393	5.920	24.353	11.821
P/B	2	33.482	1.505	47.447	2.774

ESTIMATES

Guidance | Summary

	QTR Dec-2020 EPS	QTR Dec-2020 Rev	FY Sep-2021 EPS	FY Sep-2021 Rev
Mean Estimate	1.39	102.00B	3.96	315.25B
Smart Estimate	1.39	101.94B	3.95	314.43B
Predicted Surprise	0.00%	-0.06%	-0.32%	-0.26%
Mean Chg %	0.62%	0.29%	0.95%	0.33%

NEWS

News | Investor Briefs | Global Press

03-Dec-2020

10:15:43	BUSINESS Apple sued over old iPhones in Europe Tech giant i...	LATIME
10:13:01	《外企动向》苹果因旧款手机电池问题在欧洲多国面临诉讼	HKETNN
10:12:11	《外企動向》蘋果因舊款手機電池問題在歐洲多國面臨訴訟	HKETNN
09:49:33	USPTO ISSUES TRADEMARK: WORKS WITH APPLE IBEACON	USFEDE
09:16:24	Analysts' Weekly Ratings Updates for Apple (AAPL)	TICREP
09:11:02	Analysts' Weekly Ratings Updates for Apple (AAPL)	AMEBAN
09:00:25	《A股焦点》欧菲光：被苹果移出相机模块供应链消息不属实	HKETNN
08:59:39	《A股焦點》歐菲光：被蘋果移出相機模塊供應鏈消息不屬實	HKETNN
07:13:26	Dow Jones Selected Stocks 1710 - December 03	AAP
06:47:44	RPT-Hong Kong-Le magnat Jimmy Lai détenu pour des accusa...	RTRS

RESEARCH

Contributor | Internal

03-Dec-2020	Equiti Global Arabic Smart Report	Equiti
02-Dec-2020	Trefis Report: Apple - $99.92 Trefis Price Estimate	Trefis
02-Dec-2020	Equiti Global Arabic Smart Report	Equiti
01-Dec-2020	Comprehensive Technical and Fundamental Analysis...	Stock Trade...
01-Dec-2020	Equiti Global Arabic Smart Report	Equiti
30-Nov-2020	Equiti Global Arabic Smart Report	Equiti
29-Nov-2020	"The Economy Matters" Report for AAPL: the econo...	MacroRisk ...

EVENTS

Upcoming | Past

26-Jan-2021

NTS	Q1 2021 Apple Inc Earnings Release

25-Feb-2021

18:00:00	Apple Inc Annual Shareholders Meeting

PEERS

Company Name	Mkt Cap	% Chan...	Reven...	EBITD...	EV/ ...
Apple Inc	2.09T		274.52B	28.17%	23.39
HP Inc	31.34B	1.97%	56.64B	8.40%	6.63

[4]:
```
%%time
data = ek.get_timeseries('AAPL.O', fields='*',
                         start_date='2020-12-02 16:00:00',
                         end_date='2020-12-02 16:30:00',
                         interval='tick')
```

```
CPU times: user 115 ms, sys: 6.49 ms, total: 122 ms
Wall time: 4.65 s
```

[5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 32239 entries, 2020-12-02 16:00:00.010000 to 2020-12-02 16:29:59.914000
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   VALUE   32191 non-null  float64
 1   VOLUME  32239 non-null  Int64
dtypes: Int64(1), float64(1)
memory usage: 787.1 KB
```

[6]:
```
data.head()
```

[6]:

	AAPL.O	VALUE	VOLUME
Date			
2020-12-02 16:00:00.010		122.00	37
2020-12-02 16:00:00.010		122.00	2
2020-12-02 16:00:00.010		122.01	600
2020-12-02 16:00:00.010		122.01	300
2020-12-02 16:00:00.010		122.00	20

[7]:
```
news = ek.get_news_headlines('R:AAPL.O',
                             date_from='2020-10-13',
                             date_to='2020-10-14',
                             count=5
                            )
```

[8]:
```
news
```

[8]:

	versionCreated	text	storyId	sourceCode
2020-10-13 21:40:03.235	2020-10-13 23:55:17.584000+00:00	Refinitiv Newscasts - Post-Market Wrap: Octobe...	urn:newsml:reuters.com:20201013:nRTV99mSrx:4	NS:CNBC
2020-10-13 21:07:10.816	2020-10-13 23:17:28.661000+00:00	Refinitiv Newscasts - Apple enters 5G race wit...	urn:newsml:reuters.com:20201013:nRTV8h4JYx:4	NS:RTRS
2020-10-13 21:05:12.773	2020-10-13 23:14:26.185000+00:00	Refinitiv Newscasts - T-Mobile CEO on 5G servi...	urn:newsml:reuters.com:20201013:nRTV3ghlDN:2	NS:CNBC
2020-10-13 23:12:27.333	2020-10-13 23:12:27.333000+00:00	Prime Day 2020: The best deals from Apple, Mic...	urn:newsml:reuters.com:20201013:nNRAd6qnfu:1	NS:INDEPE
2020-10-13 20:58:43.544	2020-10-13 23:10:15.457000+00:00	Refinitiv Newscasts - I feel more bullish afte...	urn:newsml:reuters.com:20201013:nRTV6vSQ8N:5	NS:CNBC

[9]:
```
from IPython.display import HTML
```

[10]:
```
HTML(ek.get_news_story(news['storyId'].iloc[1]))
```

[10]:
Click the following link to watch video: https://share.insider.thomsonreuters.com/link?entryId=1_eg1wg2n1&referenceId=tag:reuters.com,2020:newsml_OVCZYVZPN_K15&pageId=ReutersNews
Source: Thomson Reuters

Description: Apple on Tuesday launched four versions of its new flagship iPhone 12 with faster 5G connectivity in hopes of sparking an upgrade cycle that will keep sales booming through the end of the year. Conway G.Gittens has more on what's coming.

EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"[1] examines why so much of physics can be neatly explained with simple mathematical formulas such as $f = ma$ or $e = mc^2$. Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematics. Economists suffer from physics envy over their inability to neatly model human behavior. An informal, incomplete grammar of the English language runs over 1,700 pages.[2] Perhaps when it comes to natural language processing and related fields, we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.

One of us, as an undergraduate at Brown University, remembers the excitement of having access to the Brown Corpus, containing one million English words.[3] Since then, our field has seen several notable corpora that are about 100 times larger, and in 2006, Google released a trillion-word corpus with frequency counts for all sequences up to five words long.[4] In some ways this corpus is a step backwards from the Brown Corpus: it's taken from unfiltered Web pages and thus contains incomplete sentences, spelling errors, grammatical errors, and all sorts of other errors. It's not annotated with carefully hand-corrected part-of-speech tags. But the fact that it's a million times larger than the Brown Corpus outweighs these drawbacks. A trillion-word corpus—along with other Web-derived corpora of millions, billions, or trillions of links, videos, images, tables, and user interactions—captures even very rare aspects of human behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data.

Learning from Text at Web Scale

The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The reason for these successes is not that these tasks are easier than other tasks; they are in fact much harder than tasks such as document classification that extract just a few bits of information from each document. The reason is that translation is a natural task routinely done every day for a real human need (think of the operations of the European Union or of news agencies). The same is true of speech transcription (think of closed-caption broadcasts). In other words, a large training set of the input-output behavior that we seek to automate is available to us *in the wild*. In contrast, traditional natural language processing problems such as document classification, part-of-speech tagging, named-entity recognition, or parsing are not routine tasks, so they have no large corpus available in the wild. Instead, a corpus for these tasks requires skilled human annotation. Such annotation is not only slow and expensive to acquire but also difficult for experts to agree on, being bedeviled by many of the difficulties we discuss later in relation to the Semantic Web. The first lesson of Web-scale learning is to use available large-scale data rather than hoping for annotated data that isn't available. For instance, we find that useful semantic relationships can be automatically learned from the statistics of search queries and the corresponding results[5] or from the accumulated evidence of Web-based text patterns and formatted tables,[6] in both cases without needing any manually annotated data.

1541-1672/09/$25.00 © 2009 IEEE
Published by the IEEE Computer Society

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" examines why so much of physics can be neatly explained with simple mathematical formulas such as $f = ma$ or $e = mc^2$. Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematics. Economists suffer from physics envy over their inability to neatly [and successfully] model human behavior. An informal, incomplete grammar of the English language runs over 1,700 pages. Perhaps when it comes to natural language processing and related fields, we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.

IP[y]:
IPython

NumPy

matplotlib

pandas

$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$

The Financial Tick Data

[7]:
```
raw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10108751 entries, 0 to 10108750
Data columns (total 10 columns):
 #   Column     Dtype
---  ------     -----
 0   #RIC       object
 1   Domain     object
 2   Date-Time  object
 3   Type       object
 4   Price      float64
 5   Volume     float64
 6   Bid Price  float64
 7   Bid Size   float64
 8   Ask Price  float64
 9   Ask Size   float64
dtypes: float64(6), object(4)
memory usage: 771.2+ MB
```

[8]:
```
raw.head()
```

[8]:

	#RIC	Domain	Date-Time	Type	Price	Volume	Bid Price	Bid Size	Ask Price	Ask Size
0	NIFc1	Market Price	2019-08-01T09:15:01.197913842+0530	Trade	11065.95	9975.0	NaN	NaN	NaN	NaN
1	NIFc1	Market Price	2019-08-01T09:15:01.197913842+0530	Quote	NaN	NaN	11063.45	150.0	11065.9	150.0
2	NIFc1	Market Price	2019-08-01T09:15:02.266641886+0530	Trade	11069.40	11775.0	NaN	NaN	NaN	NaN
3	NIFc1	Market Price	2019-08-01T09:15:02.266641886+0530	Quote	NaN	NaN	11065.90	525.0	11066.4	525.0
4	NIFc1	Market Price	2019-08-01T09:15:03.301959220+0530	Trade	11066.55	11550.0	NaN	NaN	NaN	NaN

[9]:
```
raw.tail()
```

[9]:

	#RIC	Domain	Date-Time	Type	Price	Volume	Bid Price	Bid Size	Ask Price	Ask Size
10108746	NIFc1	Market Price	2020-08-31T15:29:58.141881872+0530	Quote	NaN	NaN	11357.00	75.0	11359.8	750.0
10108747	NIFc1	Market Price	2020-08-31T15:29:59.149444032+0530	Trade	11359.8	3675.0	NaN	NaN	NaN	NaN
10108748	NIFc1	Market Price	2020-08-31T15:29:59.149444032+0530	Quote	NaN	NaN	11357.05	150.0	11359.8	1050.0
10108749	NIFc1	Market Price	2020-08-31T15:30:05.141130971+0530	Trade	11359.8	3000.0	NaN	NaN	NaN	NaN
10108750	NIFc1	Market Price	2020-08-31T15:30:05.141130971+0530	Quote	NaN	NaN	11359.75	75.0	11359.8	1275.0

Data Preprocessing

Notebooks and Resources under http://bit.ly/pyconf_hyd_2020

Downloading the Data

[2]:
```
url = 'https://certificate.tpq.io/nif_refinitiv_tick_data.csv.gz'
```

[3]:
```
%time c = urllib.request.urlretrieve(url)
```

```
CPU times: user 753 ms, sys: 746 ms, total: 1.5 s
Wall time: 26 s
```

[4]:
```
c
```

[4]:
```
('/var/folders/31/s079lk6j3111vjcp2qlvczy00000gn/T/tmpf7lk1jru',
 <http.client.HTTPMessage at 0x1165fb820>)
```

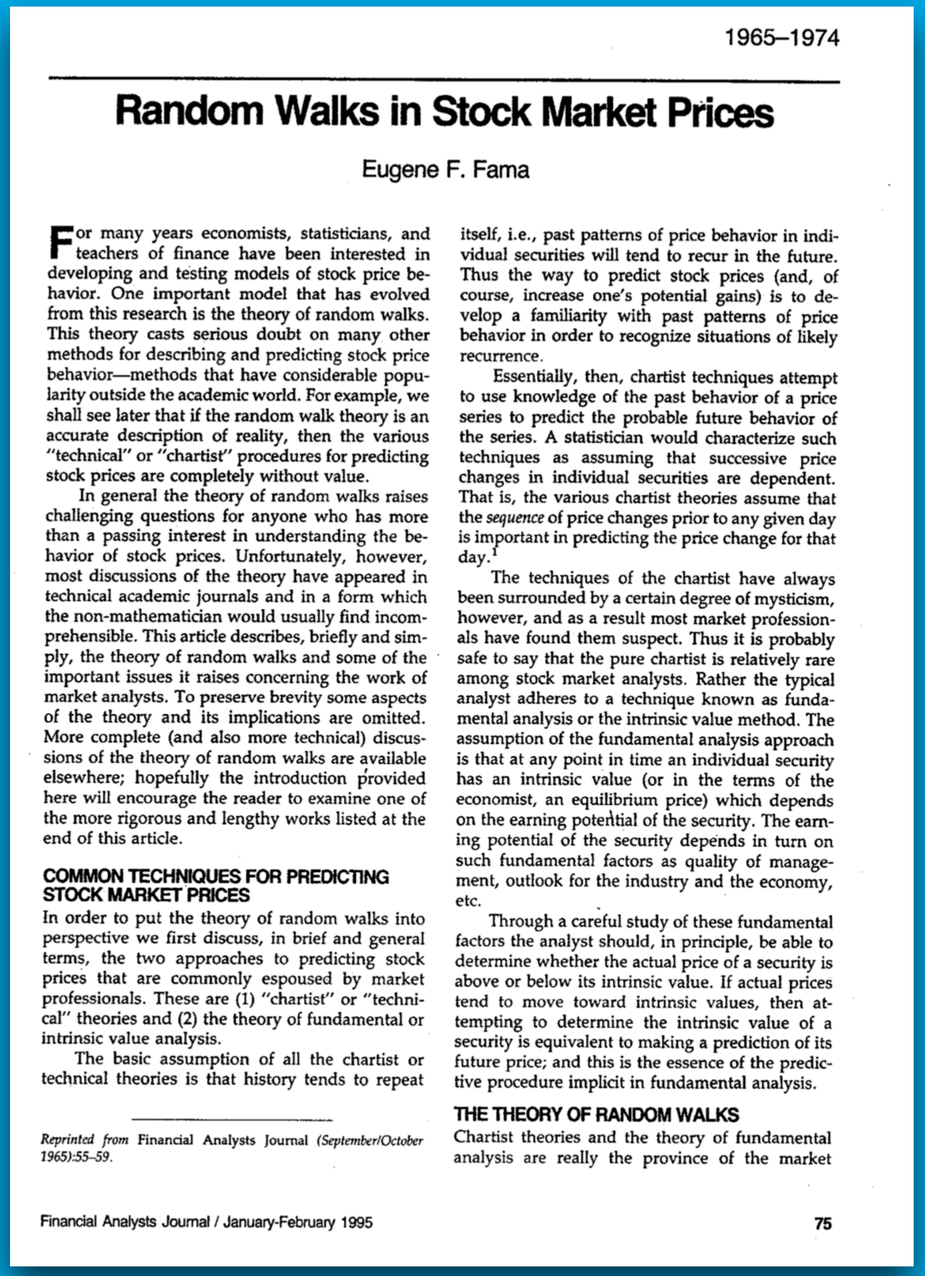
Reading the Data

[5]:
```
%time raw = pd.read_csv(c[0], compression='gzip')
```

```
CPU times: user 7.18 s, sys: 721 ms, total: 7.9 s
Wall time: 9.14 s
```

[6]:
```
urllib.request.urlcleanup()
```

Efficient Markets



1965–1974

Random Walks in Stock Market Prices

Eugene F. Fama

For many years economists, statisticians, and teachers of finance have been interested in developing and testing models of stock price behavior. One important model that has evolved from this research is the theory of random walks. This theory casts serious doubt on many other methods for describing and predicting stock price behavior—methods that have considerable popularity outside the academic world. For example, we shall see later that if the random walk theory is an accurate description of reality, then the various "technical" or "chartist" procedures for predicting stock prices are completely without value.

In general the theory of random walks raises challenging questions for anyone who has more than a passing interest in understanding the behavior of stock prices. Unfortunately, however, most discussions of the theory have appeared in technical academic journals and in a form which the non-mathematician would usually find incomprehensible. This article describes, briefly and simply, the theory of random walks and some of the important issues it raises concerning the work of market analysts. To preserve brevity some aspects of the theory and its implications are omitted. More complete (and also more technical) discussions of the theory of random walks are available elsewhere; hopefully the introduction provided here will encourage the reader to examine one of the more rigorous and lengthy works listed at the end of this article.

COMMON TECHNIQUES FOR PREDICTING STOCK MARKET PRICES

In order to put the theory of random walks into perspective we first discuss, in brief and general terms, the two approaches to predicting stock prices that are commonly espoused by market professionals. These are (1) "chartist" or "technical" theories and (2) the theory of fundamental or intrinsic value analysis.

The basic assumption of all the chartist or technical theories is that history tends to repeat itself, i.e., past patterns of price behavior in individual securities will tend to recur in the future. Thus the way to predict stock prices (and, of course, increase one's potential gains) is to develop a familiarity with past patterns of price behavior in order to recognize situations of likely recurrence.

Essentially, then, chartist techniques attempt to use knowledge of the past behavior of a price series to predict the probable future behavior of the series. A statistician would characterize such techniques as assuming that successive price changes in individual securities are dependent. That is, the various chartist theories assume that the *sequence* of price changes prior to any given day is important in predicting the price change for that day.[1]

The techniques of the chartist have always been surrounded by a certain degree of mysticism, however, and as a result most market professionals have found them suspect. Thus it is probably safe to say that the pure chartist is relatively rare among stock market analysts. Rather the typical analyst adheres to a technique known as fundamental analysis or the intrinsic value method. The assumption of the fundamental analysis approach is that at any point in time an individual security has an intrinsic value (or in the terms of the economist, an equilibrium price) which depends on the earning potential of the security. The earning potential of the security depends in turn on such fundamental factors as quality of management, outlook for the industry and the economy, etc.

Through a careful study of these fundamental factors the analyst should, in principle, be able to determine whether the actual price of a security is above or below its intrinsic value. If actual prices tend to move toward intrinsic values, then attempting to determine the intrinsic value of a security is equivalent to making a prediction of its future price; and this is the essence of the predictive procedure implicit in fundamental analysis.

THE THEORY OF RANDOM WALKS

Chartist theories and the theory of fundamental analysis are really the province of the market

Reprinted from Financial Analysts Journal *(September/October 1965):55–59.*

Financial Analysts Journal / January-February 1995 75

Eugene F. Fama (1965):

"For many years, economists, statisticians, and teachers of finance have been interested in developing and testing models of stock price behavior. One important model that has evolved from this research is the theory of random walks. This theory casts serious doubt on many other methods for describing and predicting stock price behavior—methods that have considerable popularity outside the academic world. For example, we shall see later that, if the random-walk theory is an accurate description of reality, then the various "technical" or "chartist" procedures for predicting stock prices are completely without value."—Eugene F. Fama (1965): "Random Walks in Stock Market Prices"

Michael Jensen (1978): “Some Anomalous Evidence Regarding Market Efficiency”:

“A market is efficient with respect to an information set S if it is impossible to make economic profits by trading on the basis of information set S.”

If a stock price follows a (simple) random walk (no drift & normally distributed returns), then it rises and falls with the same probability of 50% (“toss of a coin”).

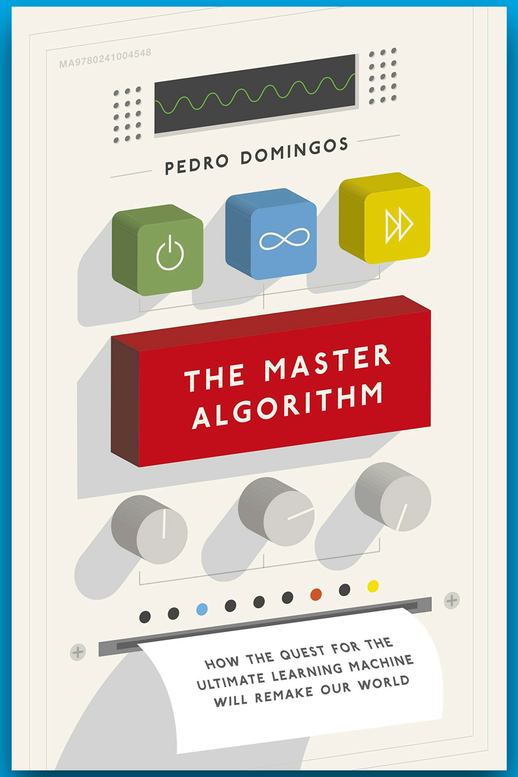
In such a case, the best predictor of tomorrow’s stock price — in a least-squares sense — is today’s stock price.

AI-First Algorithmic Trading

scientific method

noun

a method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.
"criticism is the backbone of **the scientific method**"



"The grand aim of science is to cover the greatest number of experimental facts by logical deduction from the smallest number of hypotheses or axioms."
— Albert Einstein

"Machine learning is the scientific method on steroids. It follows the same process of generating, testing, and discarding or refining hypotheses. But while a scientist may spend his or her whole life coming up with and testing a few hundred hypotheses, a machine-learning system can do the same in a second. Machine learning automates discovery. It's no surprise, then that it's revolutionizing science as much as it's revolutionizing business."

Bullish patterns (going up)

Flag

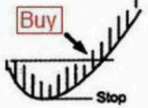
Pennant

Cup with handle

Ascending triangle

Symetrical triangle

Measured move up



Ascending scallop

3 rising valleys

Bearish patterns (going down)

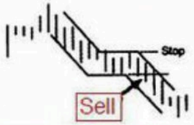
Flag

Pennant

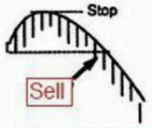
Invertede Cup with handle

Descending triangle

Symetrical triangle



Measured move down



Descending scallop

3 descending peaks

Reversal patterns

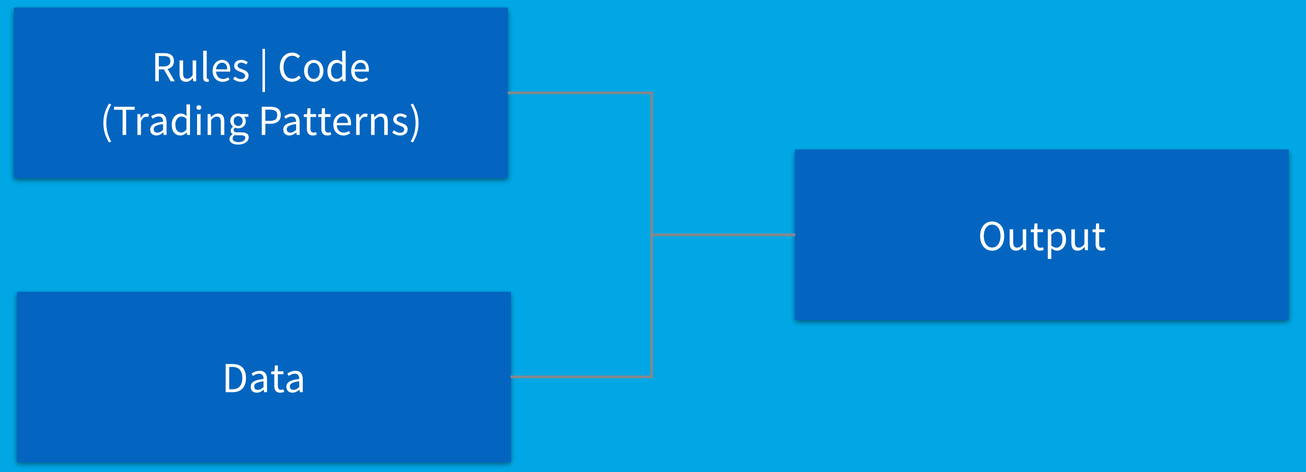
Double bottoms

Diamond bottoms

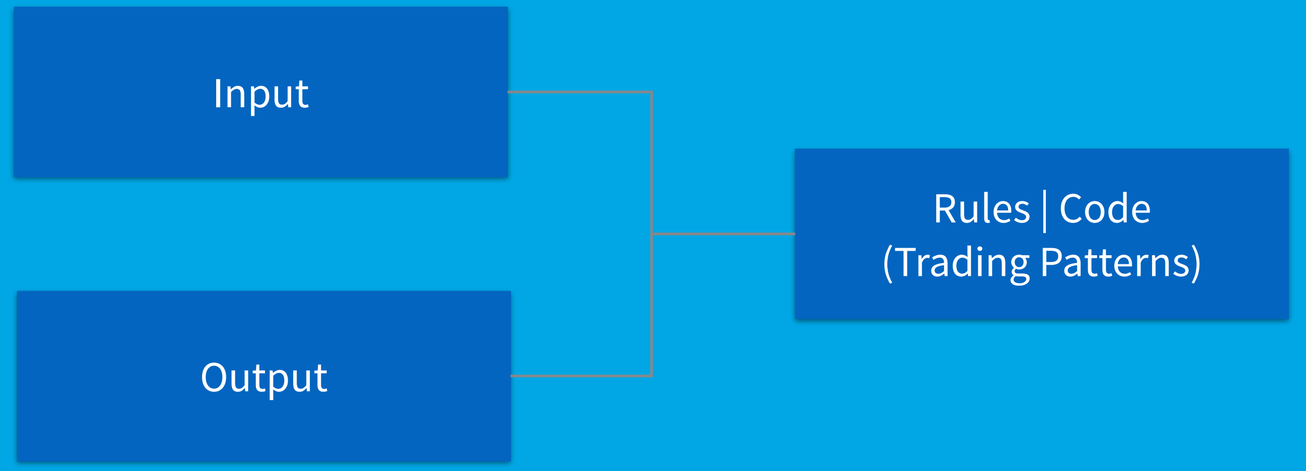
Tops rectangle

Head and shoulders top

Programming.



Machine Learning.



Financial Markets

Finance History

AI in Finance = finaince

“normative economics = assumptions, axioms, etc.”

“positive economics = data, relationships, etc.”

x

(too) “simple and elegant theories”

“general, parametrizable, trainable algorithms”

f(•)

m(•, a, b)

“hardly any supporting empirical evidence”

“might show good performance, but black box”

y

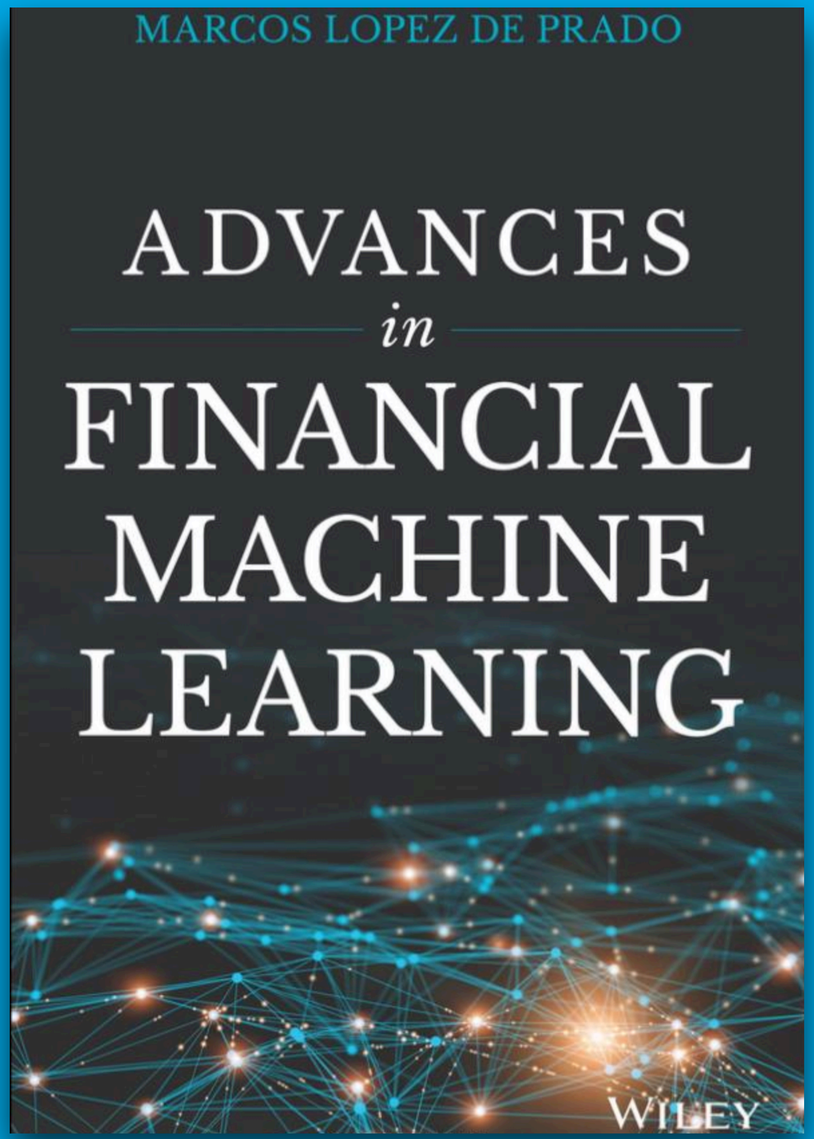
f(x) ≠ y

m(x, a*, b*) ≈ y

“non-linear, complex, changing”

“brain-driven & beauty myth”

“data-driven & AI-first”



"The essential tool of econometrics is multivariate linear regression, an 18th-century technology that was already mastered by Gauss before 1794 … It is hard to believe that something as complex as 21st-century finance could be grasped by something as simple as inverting a covariance matrix."

"… what if economists finally started to consider non-linear functions?"

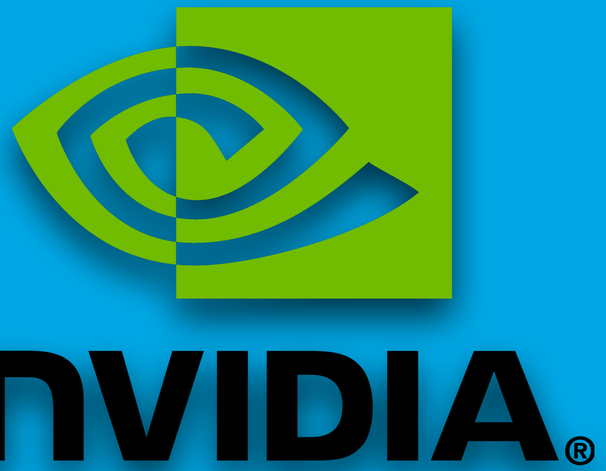
"An ML algorithm can spot patterns in a 100-dimensional world as easily as in our familiar 3-dimensional one."

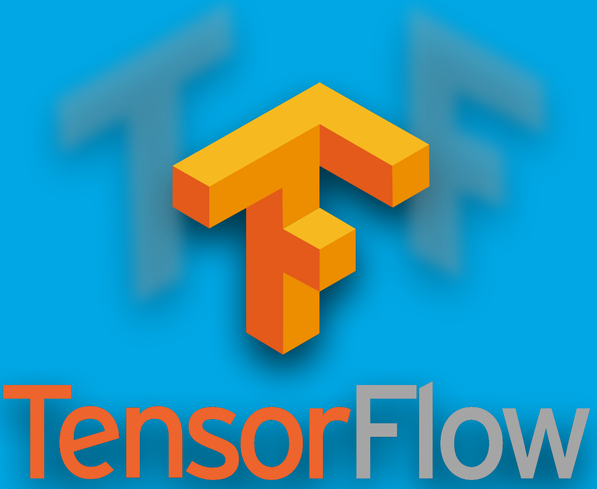
"Econometrics might be good enough to succeed in financial academia (for now), but succeeding in practice requires ML."

Marcos López de Prado (2018)


K
Keras


scikit
learn


NVIDIA


TensorFlow

AI-Powered Market Prediction

Notebooks and Resources under http://bit.ly/pyconf_hyd_2020

Preparing Features & Labels Data

[5]:
```
data['RET'] = np.log(data['Price'] / data['Price'].shift(1))
```

[6]:
```
window = 20
```

[7]:
```
data['SMA'] = data['Price'].rolling(window).mean()
```

[8]:
```
data['MOM'] = data['RET'].rolling(window).mean()
```

[9]:
```
data['VOL'] = data['RET'].rolling(window).std()
```

[10]:
```
data.dropna(inplace=True)
```

[11]:
```
data['D'] = np.sign(data['RET'])  # labels
data['D'] = data['D'].astype(int)
```

[12]:
```
data.head()
```

[12]:

Date-Time	Price	Volume	RET	SMA	MOM	VOL	D
2019-08-05 16:00:00+05:30	10896.0	675.0	0.001561	10947.5925	-0.000801	0.005633	1
2019-08-06 10:00:00+05:30	10942.2	75.0	0.004231	10943.6025	-0.000363	0.005669	1
2019-08-06 11:00:00+05:30	10935.0	600.0	-0.000658	10940.0500	-0.000324	0.005663	-1
2019-08-06 12:00:00+05:30	10930.1	75.0	-0.000448	10936.0625	-0.000363	0.005661	-1
2019-08-06 13:00:00+05:30	10978.0	225.0	0.004373	10936.2150	0.000014	0.005715	1

Conclusions

1. The availability of **big financial data** (historical—streaming, structured—unstructured) gives rise to ***data-driven finance***.
2. It might be assumed that the **“unreasonable effectiveness of big data”** holds true in the financial domain as well.
3. Due to the availability of big data (e.g. billions of hours of virtual car driving, billions of self-played games), **Artificial Intelligence** (AI) is changing almost every area of our lives.
4. It is to be assumed that in the same way the **combination of *data-driven and AI-first finance*** will influence and change finance and algorithmic trading for good.

The Python Quants GmbH

Dr. Yves J. Hilpisch
+49 3212 112 9194
tpq.io | ai@tpq.io | @dyjh