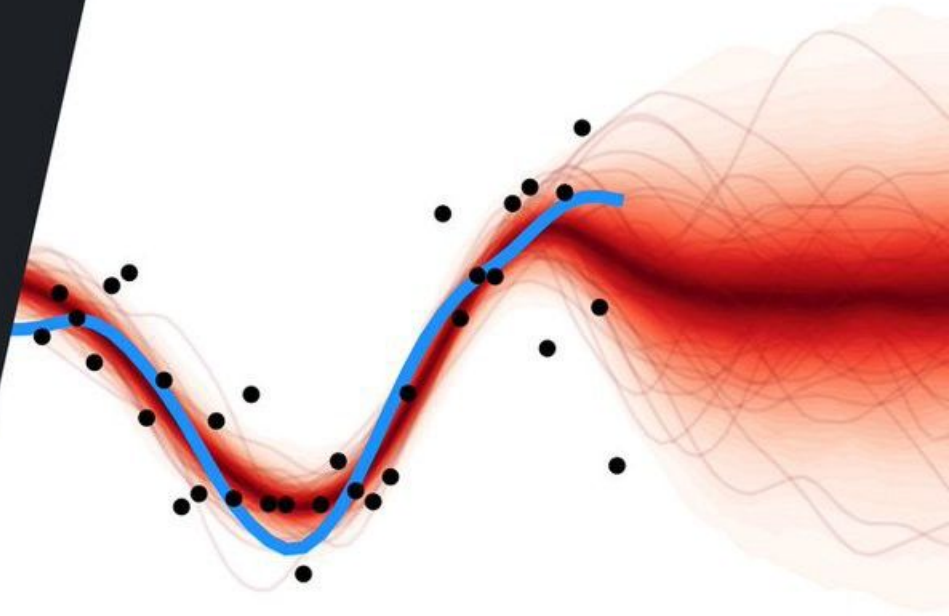
QUANTCON 2018

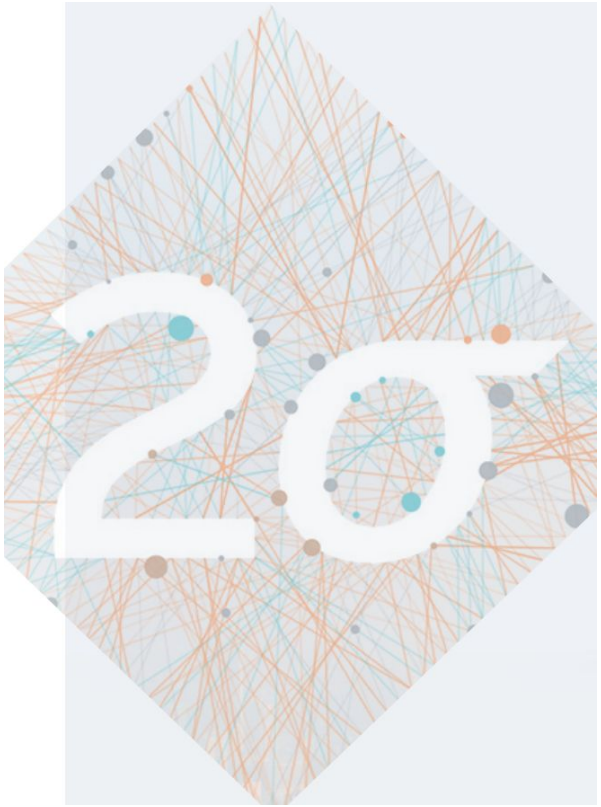
April 27-28 / NYC / #QUANTCON

QUANTCON.COM



Hosted by Quantopian

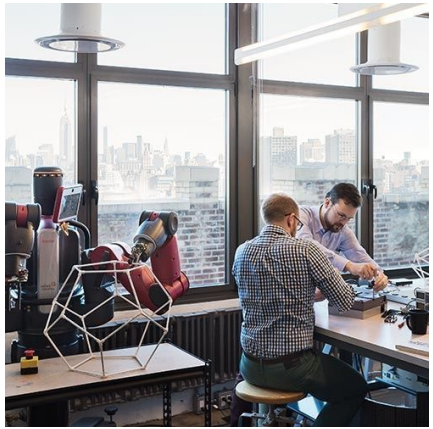
What is a Hedge Fund



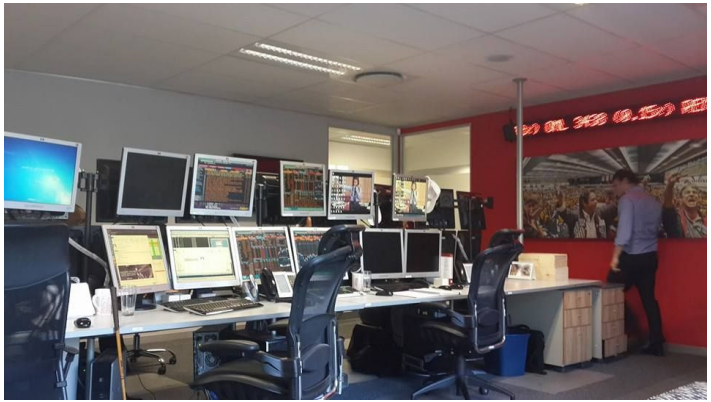
The Buy Side

Day Traders

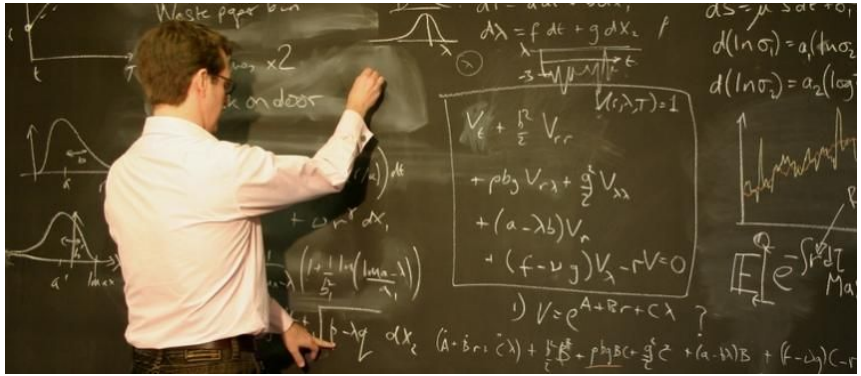
Quant PM



Discretionary PM



Asset Pricing



Machine Learning Topics

1. An Introduction to Deep Reinforcement Learning Applied to Trading
2. Reinforcement Learning for Trading - Practical Examples
3. Automation of Equity Markets, the Evolution of High Frequency Trading and the Applicability of Deep Learning
4. Using Neural Networks for Time Series Prediction
5. Turning the Internet into the Semantic Web using Machine Learning
6. Bayesian Optimization to Simultaneously Tune Multiple Metrics
7. Big Data and Machine Readable News to Trade Markets
8. Statistical Algorithm Selection: A Data Science Approach to Managing Systematic Trading Strategies Developed by 'The Crowd’
9. The 7 Reasons Most Machine Learning Funds Fail

Marcos López de Prado

Research fellow at Lawrence Berkeley National Laboratory

One of the top-10 most read authors in finance

Published dozens of scientific articles on ML and supercomputing in the leading academic journals

Multiple international patent applications on algorithmic trading

Financial Machine Learning

ML is only beginning to transform Finance:

– 2016: Studies show that ML methods (like Hierarchical Risk Parity) deliver portfolios that systematically outperform Markowitz optimization out-of-sample.

– 2016: The GIS-Liquid Strategies group manages $13 billion with 12 people.

– 2017: Four funds of Man/AHL manage $12.3 billion using AI.

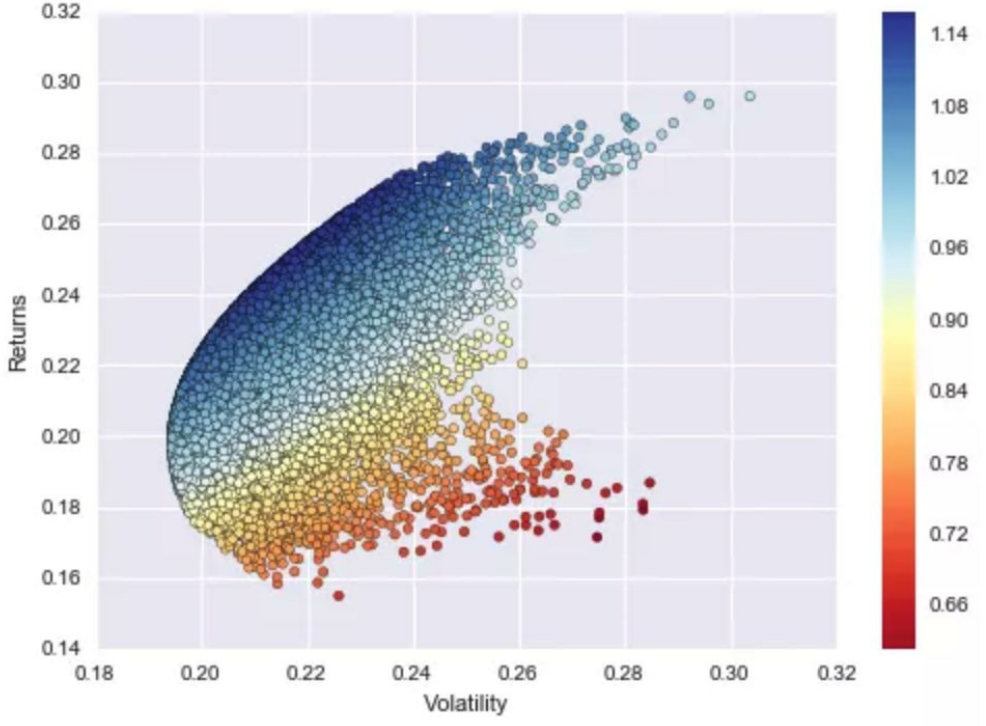
– 2018: KPMG's report argues that hedge funds must embrace technology or face 'treadmill to oblivion'.

– 2018: First graduate-level textbook on ML, specifically applied to Finance (Advances in Financial Machine Learning)

The dismal state of 21st century financial research

– Story-telling prevails over objective data analysis
– The curse of Econometrics and other 18th century mathematical tools
– Linear theoretical models: CAPM, APT, Risk Premia, etc.
– Multiple testing, selection bias, backtest overfitting
– Factor investing: A few (3?) factors are well understood, however studies show that “most claimed research findings are likely false.”

Mean Variance Portfolio Optimization



Searching for the efficient frontier (Python for Finance, 2017)

The Arbitrage Pricing Theory (APT)

- Its author is **Stephen Ross**
- Unlike the CAPM, this theory doesn't ask which portfolios are efficient
- **Key Hypotesis**: each stock's return depends partly on pervasive macroeconomi factors and partly on noise events that are unique features of firm
- The principal relationship is this:

$$\text{Return} = a + b_1(r_{factor\ 1}) + b_2(r_{factor\ 2}) + b_3(r_{factor\ 3}) + \ldots. + \text{noise}$$

- **Example factors**: oil price, interest-rate, return on market portfolio
- **Two sources of risk**: first the risk that come from the pervasive macroeconomic factors cannot be eliminates by diversification. Second the risk arising from possible events that are specific to the firm and that diversification eliminates specific risk. The investor can ignore it when deciding whether to buy or sell a stock
- **The risk premium on a stock is affected by factor or macroeconomic risk and not specific risk**

Pitfall #1: The Sisyphean Quants

- Discretionary portfolio managers (PMs) make investment decisions that do not follow a particular theory or rigorous rationale.
- Because nobody fully understands the logic behind their bets, they can hardly work as a team and develop deeper insights beyond the initial intuition.
- If 50 PMs tried to work together, they would influence each other until eventually 49 would follow the lead of 1.

For this reason, investment firms ask discretionary PMs to work in silos.

Silos prevent one PM from influencing the rest, hence protecting diversification.

The silo approach fails with quant portfolio managers

The boardroom mentality is, let us do with quants what has worked with discretionary PMs.

- Hire 50 PhDs, demand each produce an investment strategy within 6 months.
- Typically backfires, because each eventually settle for:

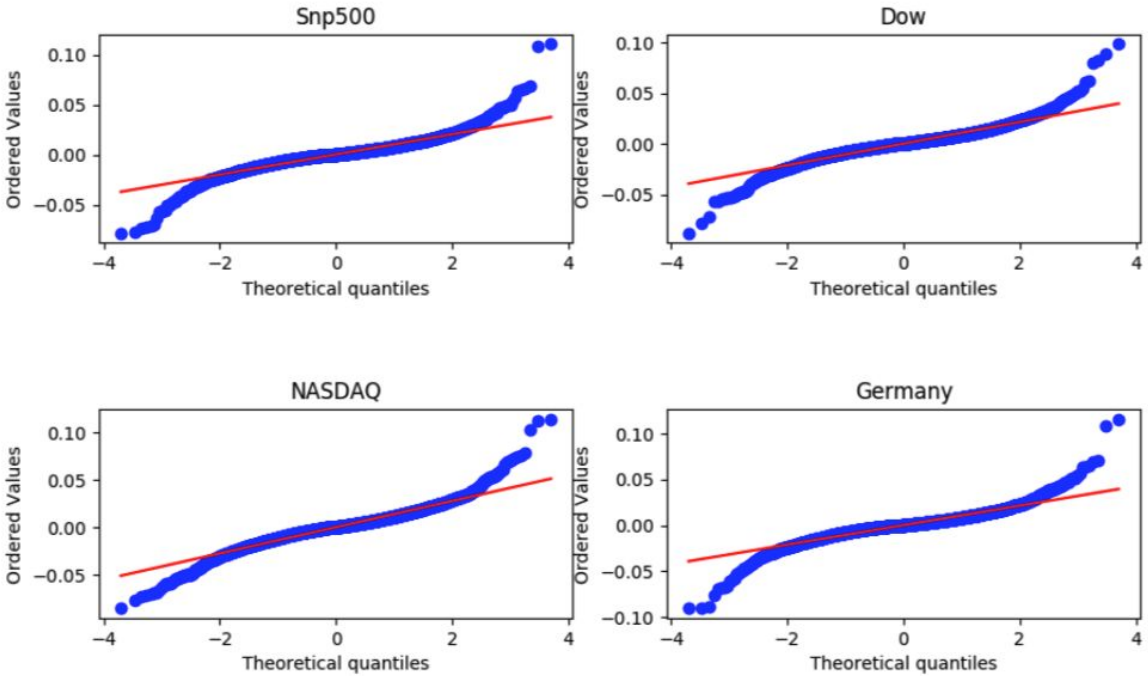
  – A false positive that looks great in an overfit backtest; or
  – A standard factor model, which is an overcrowded strategy with low Sharpe ratio, but at least has academic support.

- Disappoint the investment board, and the project will be cancelled.
- Even if 5 of those 50 PhDs found something, they would quit.

The Meta-Strategy Paradigm

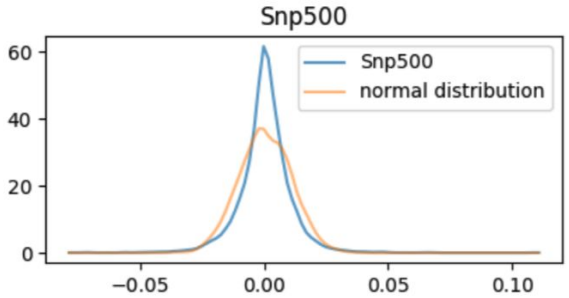
1. Identifying new strategies requires specialized teams working together.
2. It takes almost as much effort to produce one true investment strategy as to produce a hundred
3. Your firm must set up a research factory
    a. Where tasks of the assembly line are clearly divided into subtasks.
    b. Where quality is independently measured and monitored for each subtask.
    c. where the role of each quant is to specialize in a particular subtask, to become the best there is at it, while having a holistic view of the entire process.

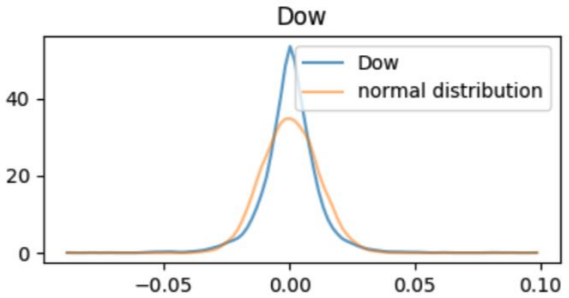
Pitfall #2: Backstory

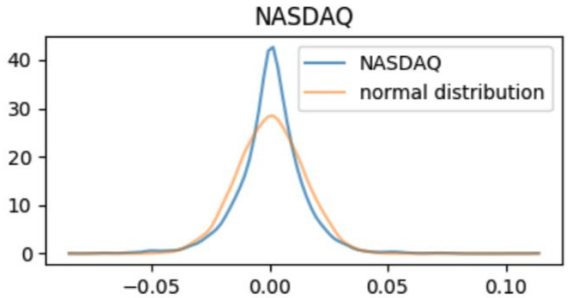


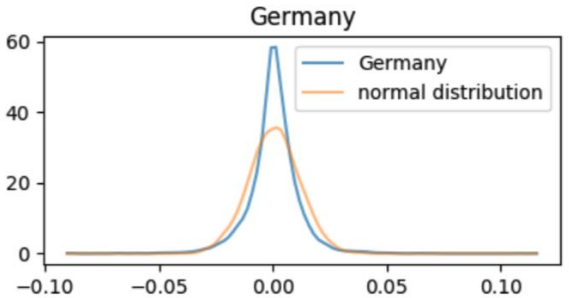
Assumption that Returns follow Normal Distribution

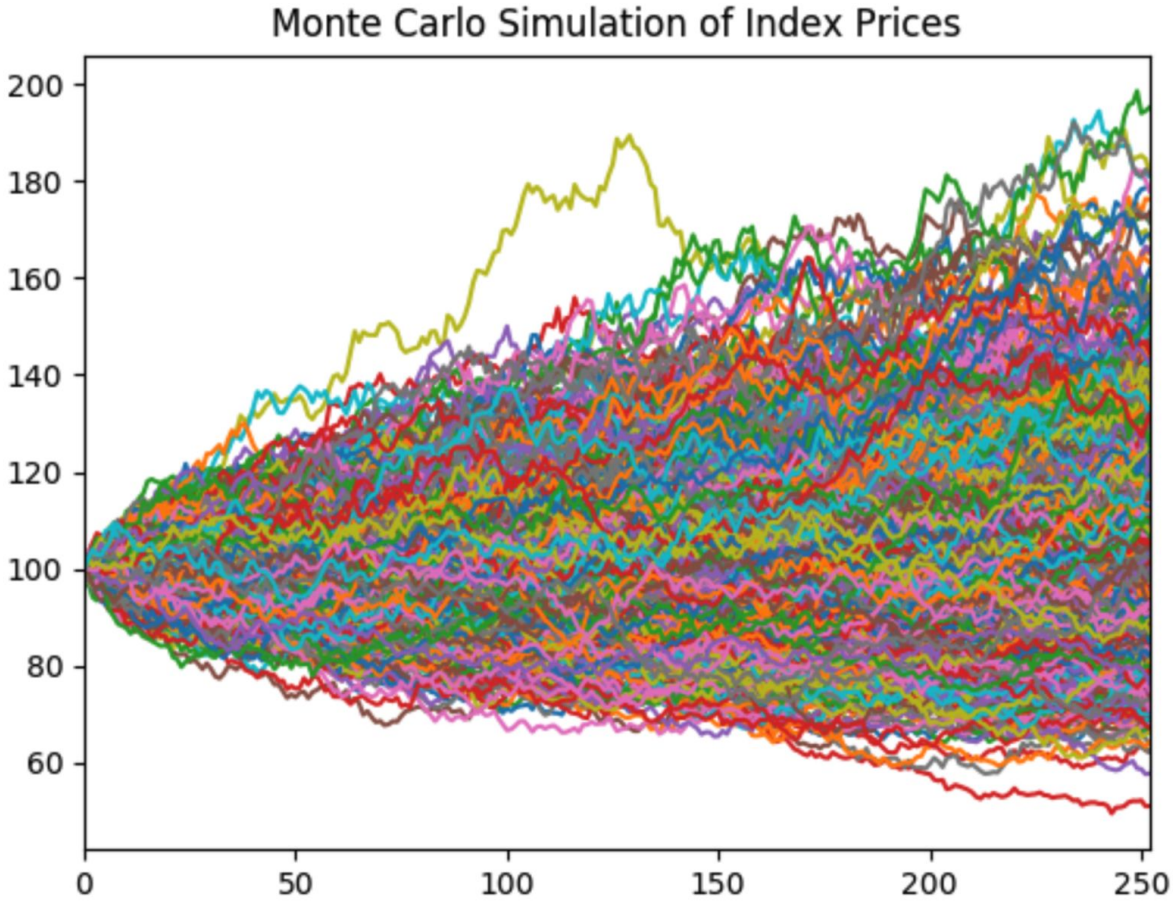
Distribution Comparison










Monte Carlo Simulation of Index Prices
200
180
160
140
120
100
80
60
0
50
100
150
200
250

Inefficient Sampling

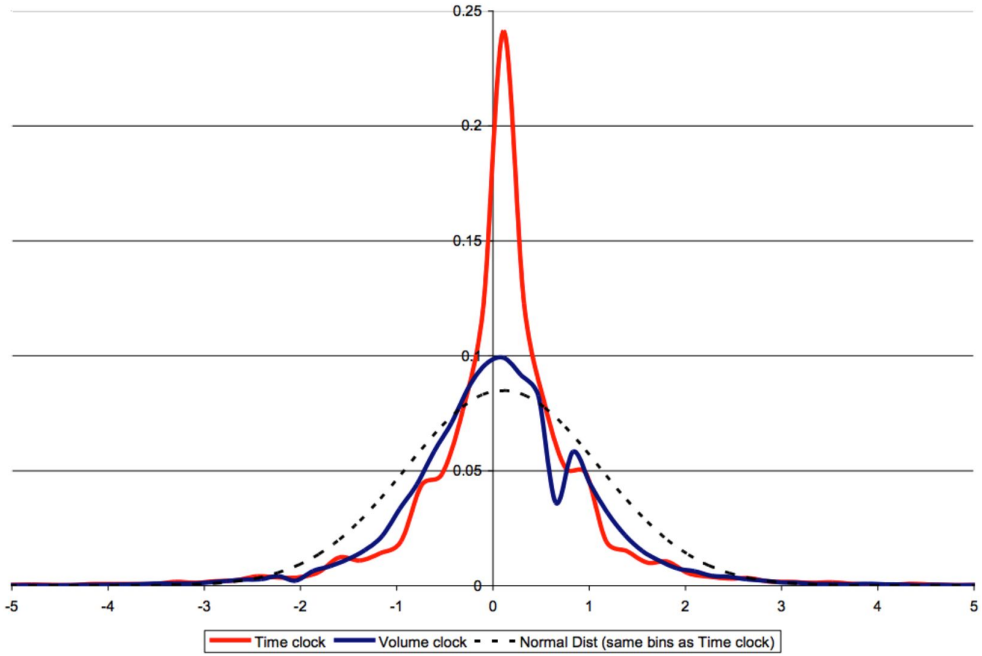
- Information does not arrive to the market at a constant rate.
- Sampling data in chronological intervals means that the informational content of the individual observations is far from constant.
- A better approach is to sample observations as a subordinated process of the amount of information exchanged:
  – Trade bars.
  – Volume bars.
  – Dollar bars.
  – Volatility or runs bars.
  – Order imbalance bars.
  – Entropy bars.

Mandlebrot and Taylor [1967] were among the first to realize that sampling as a function of the number of transactions exhibited desirable statistical properties. Multiple studies have confirmed that sampling as a function of trading activity allows us to achieve returns closer to IID Normal Ane and Geman [2000]

This is important because many statistical methods rely on the assumption that observations are drawn from an IID Gaussian process.

Yes - at this moment we are all in awe.

Exhibit 2 – Partial recovery of Normality through a price sampling process subordinated to a volume clock



Search or jump to... / Pull requests Issues Marketplace Explore

Jackal08 / financial-data-structures

Unwatch 3 Star 1 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

This program is to help users create structured financial data in the form of time, tick, volume, and dollar bars from unstructured tick data. The user passes tick data to the create_bars(data, units=1000, type='tick') function and it returns the desired structured data. Everything can be found in the main.py file. I left lots of comments in the...

Edit

financial machine-learning features Manage topics

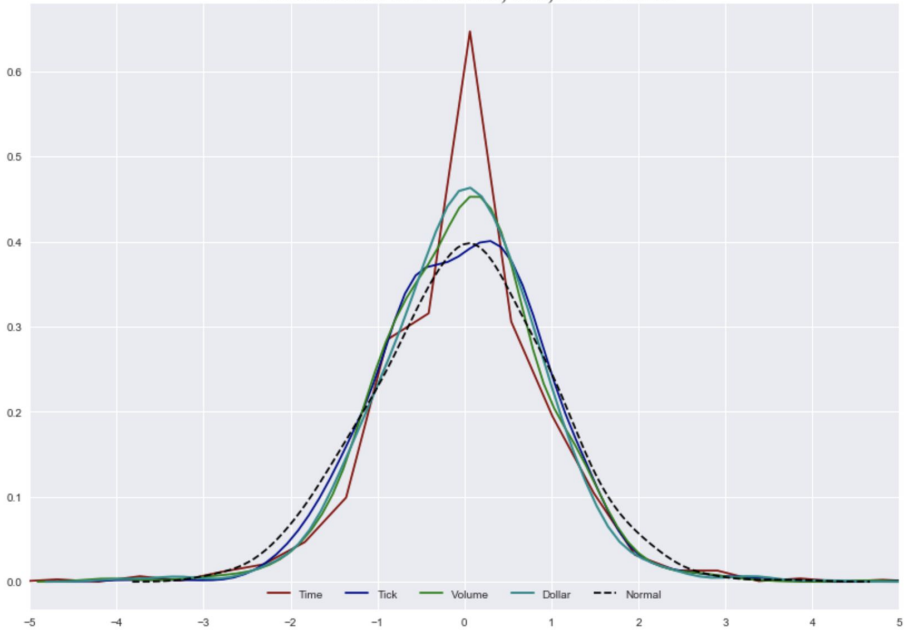
4 commits 1 branch 0 releases 1 contributor MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

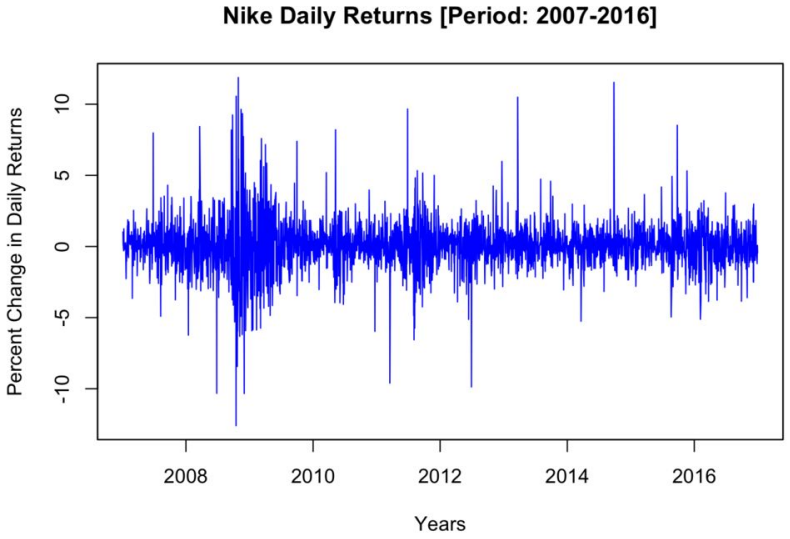
Jacques Joubert Added Jupyter Notebook & html files (Deeper analysis) ... Latest commit f392367 7 days ago

images	Added Jupyter Notebook & html files (Deeper analysis)	7 days ago
jupyter_notebooks	Added Jupyter Notebook & html files (Deeper analysis)	7 days ago
raw_tick_data	initial commit	17 days ago
saved_data	initial commit	17 days ago
LICENSE	Initial commit	17 days ago
ReadMe.md	Added Jupyter Notebook & html files (Deeper analysis)	7 days ago
cython_loops.c	initial commit	17 days ago

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

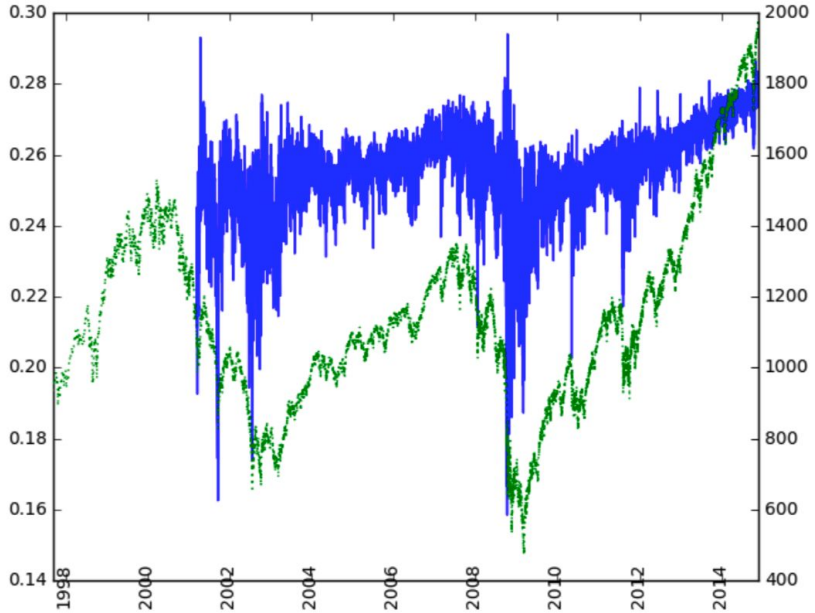


Pitfall #3 Integer Differentiation



The Stationarity vs. Memory Dilemma

- In order to perform inferential analyses, researchers need to work with invariant processes, such as:
  - returns on prices (or changes in log-prices)
  - changes in yield
  - changes in volatility
- These operations make the series stationary, at the expense of removing all memory from the original series.
- Memory is the basis for the model's predictive power.
  - For example, equilibrium (stationary) models need some memory to assess how far the price process has drifted away from the long-term expected value in order to generate a forecast.
- The dilemma is:
  - returns are stationary however memory-less; and
  - prices have memory however they are non-stationary.



- Green line: E-mini S&P 500 futures trade bars of size 1E4
- Blue line: Fractionally differentiated ($d = .4$)
- Over a short time span, it resembles returns
- Over a longer time span, it resembles price levels

Pitfall #4: Wrong Labeling

- Virtually all ML papers in finance label observations using the fixed-time horizon method.
- Consider a set of features $\{X_i\}_{i=1,\ldots,I}$, drawn from some bars with index $t = 1, \ldots, T$, where $I \leq T$. An observation X_i is assigned a label

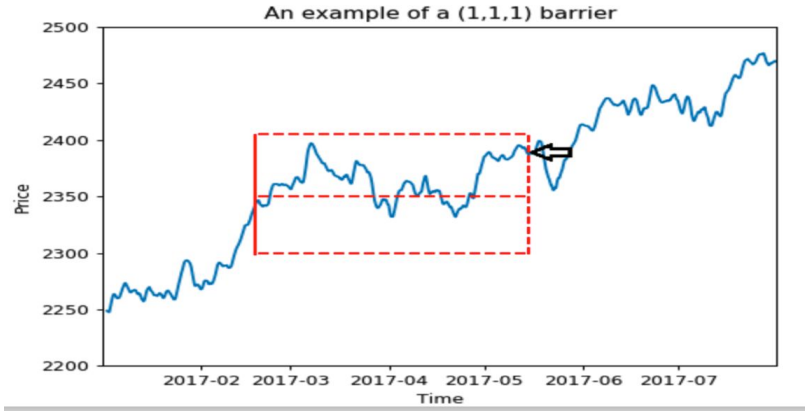
$$y_i \in \{-1,0,1\},\ y_i = \begin{cases} -1 & \text{if } r_{t_{i,0},t_{i,0}+h} < -\tau \\ 0 & \text{if } \left|r_{t_{i,0},t_{i,0}+h}\right| \leq \tau \\ 1 & \text{if } r_{t_{i,0},t_{i,0}+h} > \tau \end{cases}$$

where τ is a pre-defined constant threshold, $t_{i,0}$ is the index of the bar immediately after X_i takes place, $t_{i,0} + h$ is the index of *h* bars after $t_{i,0}$, and $r_{t_{i,0},t_{i,0}+h}$ is the price return over a bar horizon *h.*

- Because the literature almost always works with time bars, *h* implies a fixed-time horizon.

Triple Barrier Method

- The Triple Barrier Method labels an observation according to the first barrier touched out of three barriers.
  - Two horizontal barriers are defined by profit-taking and stop-loss limits, which are a dynamic function of estimated volatility (whether realized or implied).
  - A third, vertical barrier, is defined in terms of number of bars elapsed since the position was taken (an expiration limit).
- The barrier that is touched first by the price path determines the label:
  - Upper horizontal barrier: Label 1.
  - Lower horizontal barrier: Label -1.
  - Vertical barrier: Label 0.




NYSE
WALL ST


www.WeedWorldCandies.global
WARNING!
Do Not Attempt to
SMOKE
THIS
VEHICLE
404.496.3602


Make
Google
do it.
Make Google do it.