#PHXDATA20
VIRTUAL SUMMIT

Aerial Imagery to Predict the Wheat Futures Market

Avi Thaker
Senior AI/ML Engineer
GlaxoSmithKline

Lecture Highlights

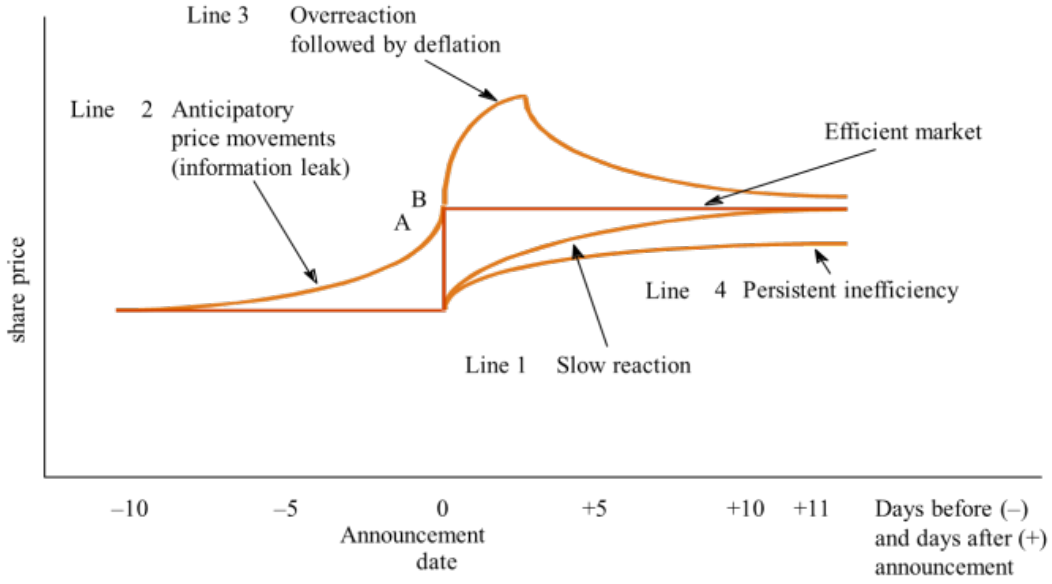
- Who I am
- Efficient market hypothesis and trading
- Background on wheat futures
- Demo via a Jupyter Notebook
- Appendix

Who I Am

- Senior Data Scientist - GSK
  - Build better drugs using AI
- Data Scientist - Microsoft
  - Deep learning for customers (computer vision)
  - Implemented a semantic knowledge graph for Ads
- Trader
  - Semi/Fully automated trading systems
  - Research Driven Process
  - Equity and Crypto Markets

Trading and The Efficient Market Hypothesis

- Impossible to “beat the market”
  - Market efficiency causes prices to incorporate and reflect all relevant information
  - Prices adjust quickly to new information
  - Prices should reflect all available information
  - Securities always trade at their fair value
  - The only way to obtain higher returns is to purchase riskier investments
  - Cannot predict trends, history does not indicate future
- The market is becoming increasingly more efficient
  - Warren Buffett? Citadel? Virtu?
  - When filing for its IPO in March 2014, it was disclosed that for five years Virtu Financial made profit 1,277 out of 1,278 days, losing money just one day.
- Quantitative and Systematic trading the efficient market hypothesis is not true



The Futures Market

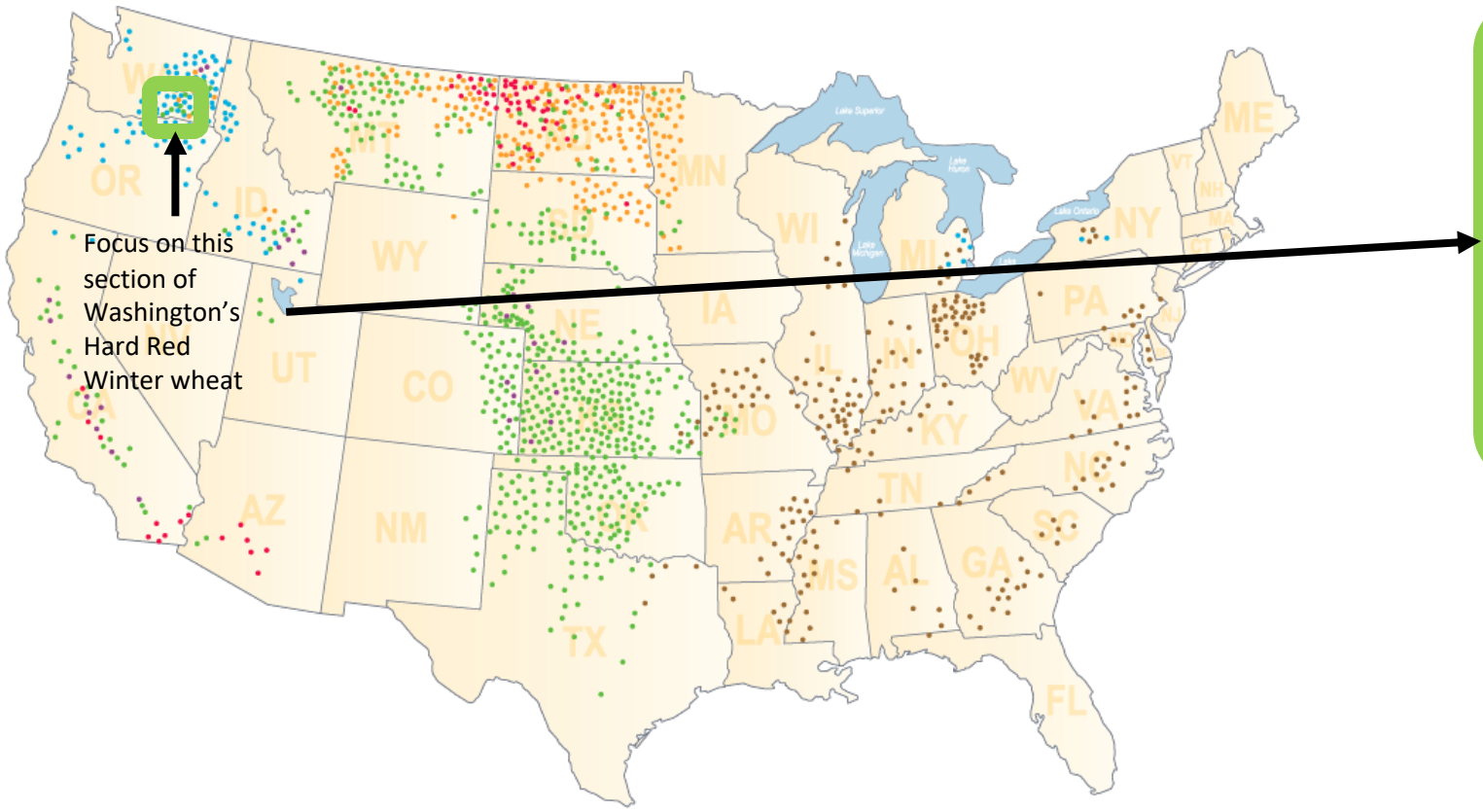
- Auction market in which participants can buy and sell commodity and futures contracts for delivery on a specified future date.
- Futures are exchange traded contracts that lock in future delivery of a commodity or security at a price set today
  - Often used to hedge – lock in a price for a given time
  - Speculators use it to bet on the direction
- Example
  - Spot price of oil is $35 USD a barrel (buy now)
  - Oil producer plans to create 1 million barrels of oil the next year ready to be delivered 3 months from now (March 2021)
  - If they expect the oil price to fall, they could lock in a price at $35.43/barrel and guarantee the ability to sell this price. A speculator could take the other side of the position to speculate if the price will rise
- Contracts are standardized to time
- We will act as speculators to guess what the price of wheat will be in some future time
- We will use a continuous contract so simplify this (no settlement date)

Month	OPTIONS	CHARTS	LAST	CHANGE	PRIOR SETTLE	OPEN	HIGH	LOW	VOLUME	UPDATED
DEC 2020	OPT		33.99	-1.80	35.79	35.24	35.28	33.64	39,798	18:01:34 CT 01 Nov 2020
JAN 2021	OPT		34.44	-1.71	36.15	35.50	35.58	34.04	20,537	18:01:23 CT 01 Nov 2020
FEB 2021	OPT		34.92	-1.65	36.57	36.07	36.07	34.50	9,424	18:01:12 CT 01 Nov 2020
MAR 2021	OPT		35.43	-1.57	37.00	36.37	36.45	35.00	6,203	18:01:04 CT 01 Nov 2020
APR 2021	OPT		35.87	-1.53	37.40	36.76	36.76	35.53	4,890	18:00:49 CT 01 Nov 2020
MAY 2021	OPT		36.24	-1.52	37.76	36.91	36.96	36.00	2,270	18:01:12 CT 01 Nov 2020

Aerial Imagery and Futures

- **Thesis**:
  - Crop health can be an influence the spot price of a tradeable good
- **Execution**:
  - Take aerial imagery of Hard Red Winter wheat farms, use that as a proxy for health
  - Google Earth gives us free historical imagery over time
  - Bet on the direction that the wheat prices will go over time
- **Machine Learning**:
  - ML models can ”learn” the influence of crop health from market prices
  - Use pretrained models that have been trained on large datasets

Looking at Wheat



https://www.uswheat.org/working-with-buyers/wheat-classes/

HARD RED WINTER | HARD RED SPRING | SOFT RED WINTER | SOFT WHITE | HARD WHITE | DURUM

A look at the six classes of wheat grown in the U.S. and the food products made from them.

Hard Red Winter

Versatile, with excellent milling and baking characteristics for pan bread, HRW is also a choice wheat for Asian noodles, hard rolls, flat breads, general purpose flour and cereal.

Hard Red Spring

The aristocrat of wheat when it comes to "designer" wheat foods like hearth breads, rolls, croissants, bagels and pizza crust, HRS is also a valued improver in flour blends.

Soft Red Winter

Versatile weak-gluten wheat with excellent milling and baking characteristics for cookies, crackers, pretzels, pastries and flat breads.

Soft White

A low moisture wheat with high extraction rates, providing a whiter product for exquisite cakes, pastries and Asian-style noodles, SW is also ideally suited to Middle Eastern flat breads.

Hard White

The newest class of U.S. wheat, HW receives enthusiastic reviews when used for Asian noodles, whole wheat or high extraction applications, pan breads and flat breads.

Durum

The hardest of all wheats, durum has a rich amber color and high gluten content, ideal for pasta, couscous and some Mediterranean breads.

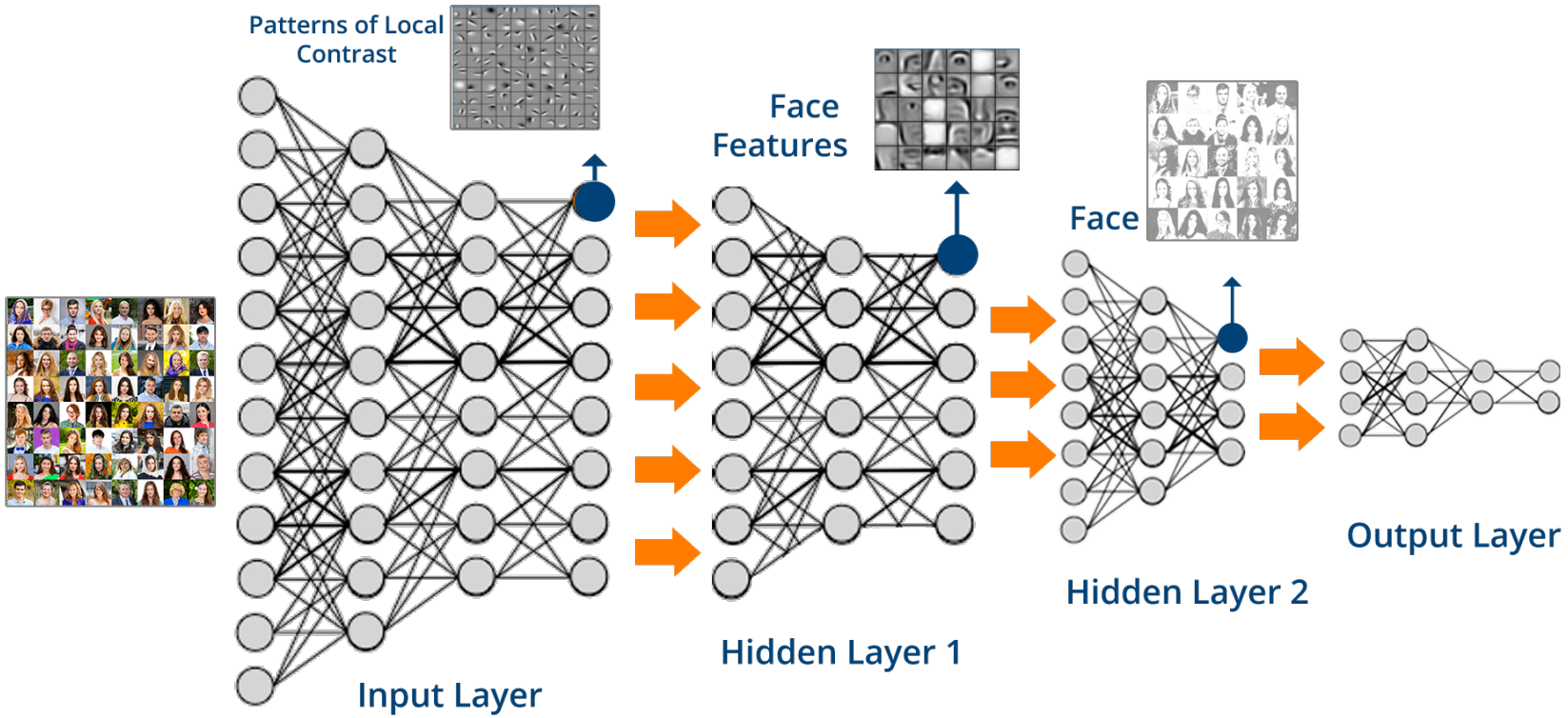
National Association of Wheat Growers
www.wheatworld.org

U.S. WHEAT ASSOCIATES
www.uswheat.org

Wheat Foods Council
www.wheatfoods.org

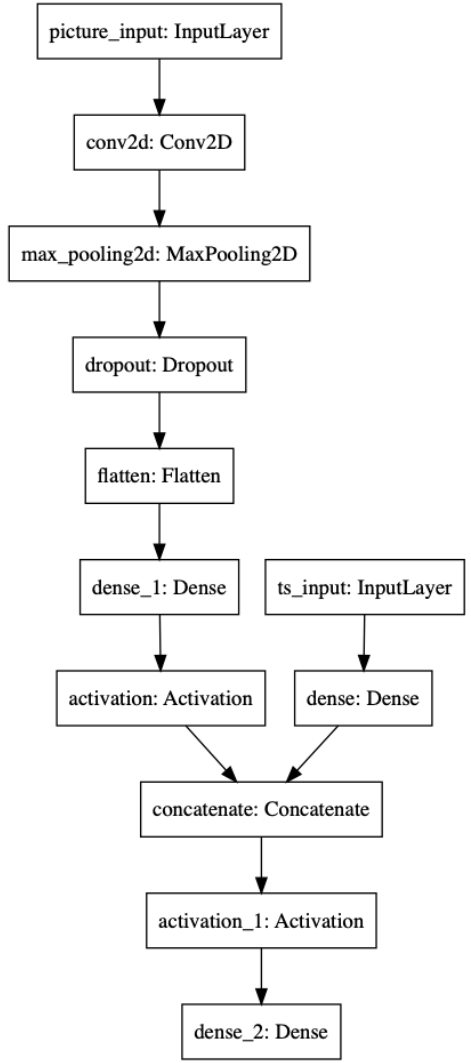
Machine Learning: What Does a Neural Network Learn?

- Early layers learn small local features
- Later layers learn higher level features



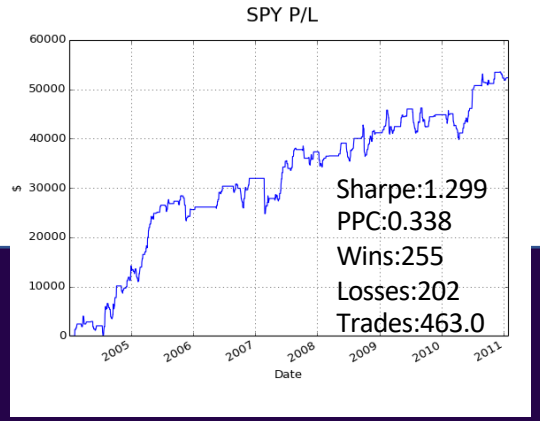
Demo

https://github.com/athaker/econ_136/tree/master/phxdata20



Backtesting Risk

- Does the strategy work across many assets
- How many years does it work for?
- Does it escape the bid-ask bounce?
- Risk Tolerance?
  - Maximum Drawdown?
- Fees? Trading frequency?
- Provide evidence of profitability
  - Curve fitting/ optimization bias
  - In-sample vs out-of-sample
  - Forward looking bias

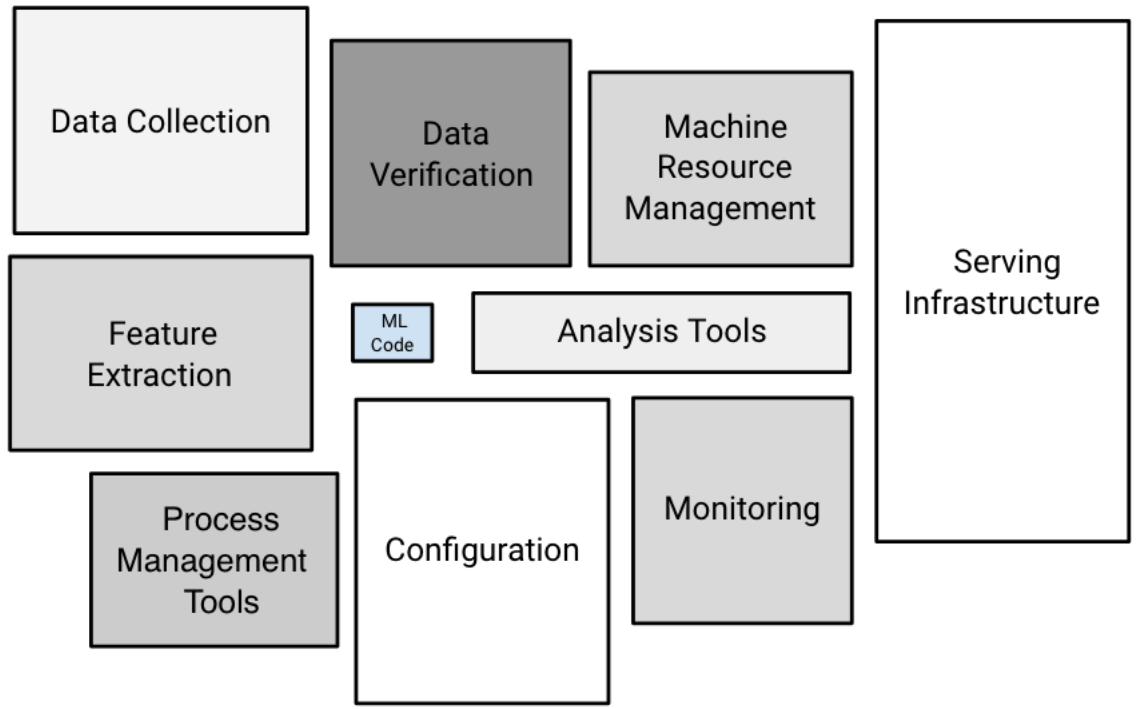


In Sample: SPY 2004-2010
Out of Sample: Assets Randomly Selected:
ADBE XLNX BBBY CFN EMC ADP AFL DE T SPLS DG ADS ALL MET CL PX WYN

ML/DL is Only a Fraction of The Work

For the wheat example to trade it in real-time you need:

- Satellite imagery or drones to take pictures
- Data storage and cleaning
- Programmatic interfaces to the exchanges with the price feeds
- Tools to help people build better models (tensorflow, etc.)
- Monitoring to ensure that the process is going correctly
- Much much more



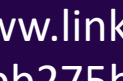
Often the ML problem is the easy part

https://developers.google.com/machine-learning/crash-course

#PHXDATA20
VIRTUAL SUMMIT

www.phxdataconference.com

Let's Connect!

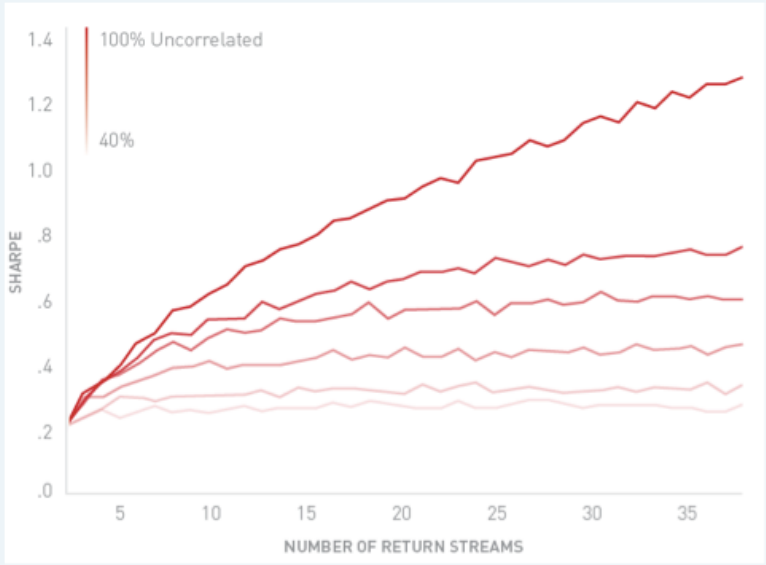


https://www.linkedin.com/in/avi-thaker-64bb275b/

https://github.com/athaker

Appendix: Correlation and Risk

Achieving High Portfolio Sharpe Ratio from holding Low Sharpe Ratio, but uncorrelated, individual algos



Simulation Setup:

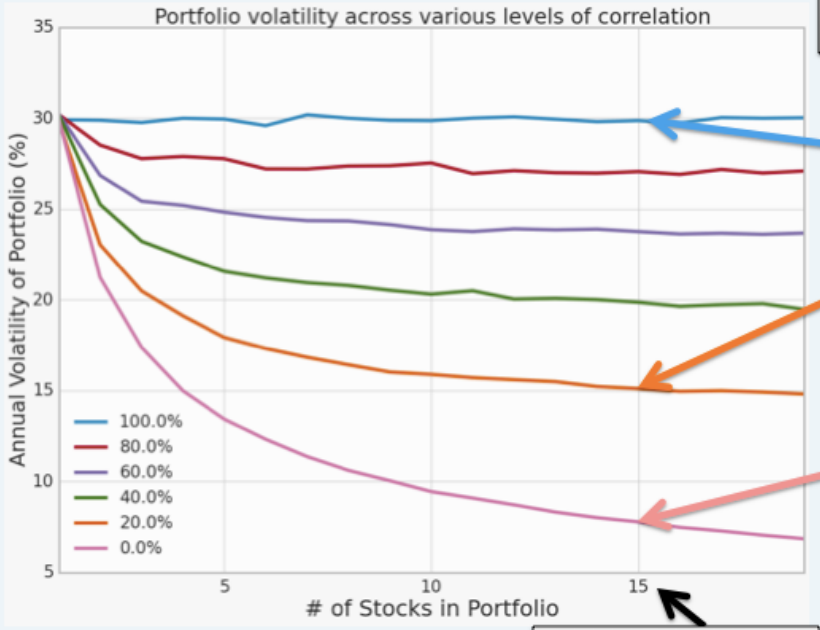
- Each algorithm individually has a Sharpe Ratio of only 0.2
- Hold X strategies across varying correlation assumptions

Results

Sharpe Ratio increases dramatically as you add more uncorrelated algorithms to your portfolio

Quantopian

Investing in uncorrelated algorithms can reduce overall portfolio risk by 50% - 75%



Assume each algo in a portfolio has 30% volatility

If they are 100% correlated, then entire portfolio also has 30% volatility…

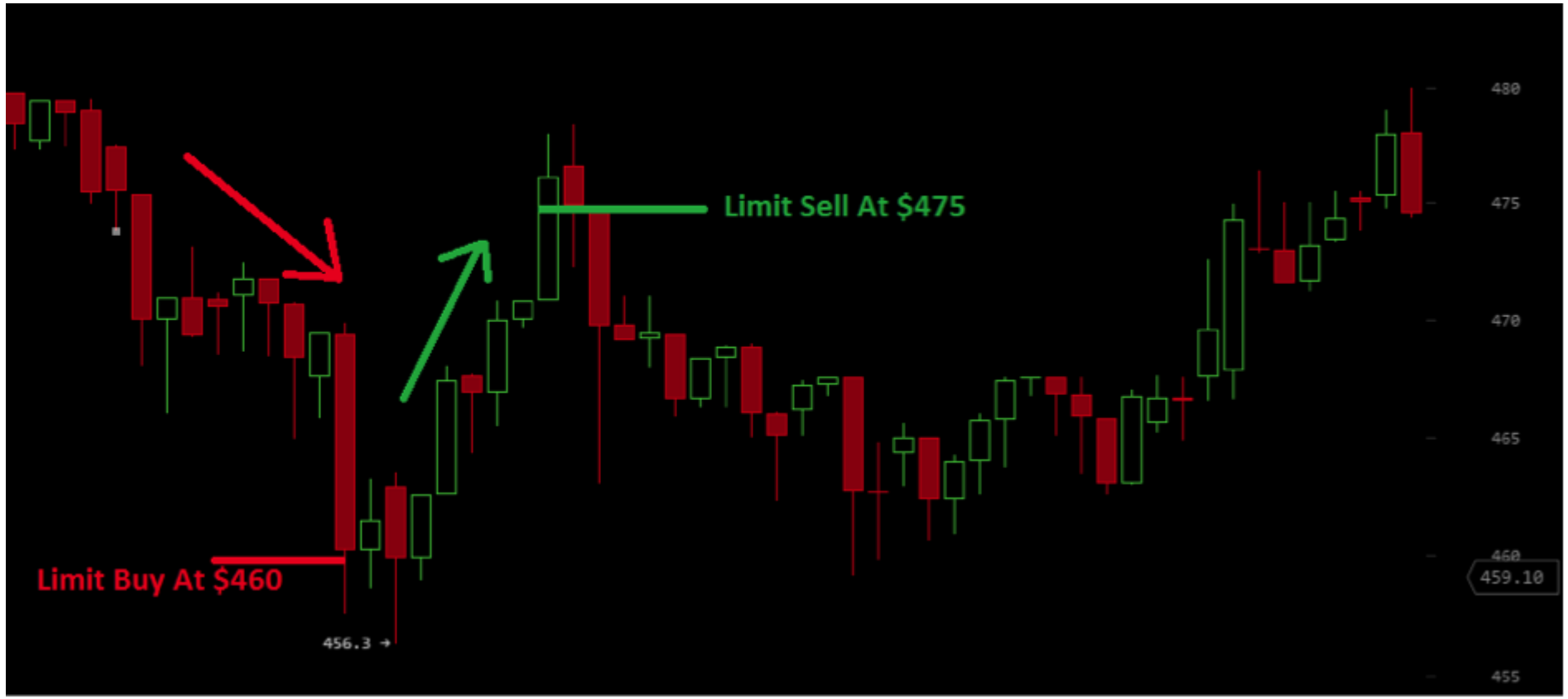
…But, if they are only 20% correlated, overall portfolio volatility is reduced by half to 15%

…And…perfect *uncorrelation* reduces annual volatility to *only 8% !*

Only need to hold 15 algos in the portfolio

Quantopian

Appendix: Limit Order Execution



For an order to be executed, a trade must cross below your buy, or a trade happens at your price, when you have been filled in the queue

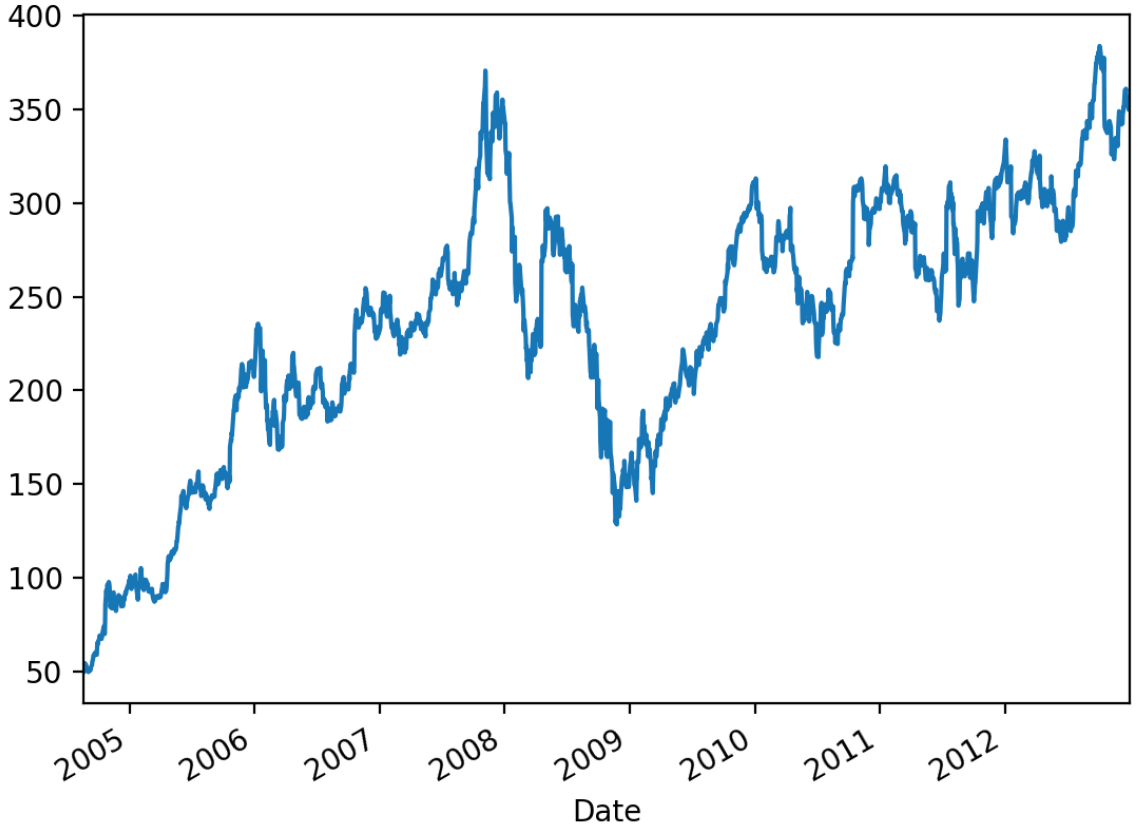
Section Title

- The measure of the largest drop from peak to bottom (in percentage)
  - It is a pain index measure
- Extremely important to measure the duration of the drawdown
  - Do you want to be losing money for years?

$$D(T) = \max_{t \in (0,T)} \{X(t) - X(T)\}$$

$$\text{MDD}(T) = \max_{t \in (0,T)} [\max_{t \in (0,\tau)} \{X(t) - X(\tau)\}]$$

- Where $X = (X(t), t \geq 0)$ is a random process
- Simply put maximum drawdown is:
  - (Peak value before largest drop – lowest value before new high) / Peak value before drop



Appendix: Sharpe Ratio

$$Sharpe = \frac{r_p - r_f}{\sigma_p}$$

$$r_p = portfolio\ return$$

$$r_f = risk\ free\ rate$$

$$\sigma_p = standard\ deviation\ of\ return$$

Measures risk adjusted performance

- Risk vs. Reward

Higher is usually better

Risk free rate sometimes assumed to be 0

Usually annualized and volatility taken as standard deviation

- Monthly: Volatility sampled monthly * sqrt(12)
- Daily: Volatility sampled daily * sqrt(252)
- Minutely: Volatility sampled minutely * sqrt(390*252)

Appendix: Profit Per Contract (PPC)

$$\frac{r_a}{c * t_s}$$

$$r_a = average\ return$$

$$c = number\ of\ contracts\ traded$$

$$t_s = tick\ size$$

- A measure of profitability, measured in ticks
- A highly liquid stock usually has a tick size of a penny
- If your strategy has more than 2 ticks, it is considered profitable (can escape the bid/ask bounce), if testing on bar data without limit order execution on bar closes
  - You can submit market orders and still make money
    - Assumes liquidity!!!!!

Section Title