

#PHXDATA20
VIRTUAL SUMMIT

Aerial Imagery to Predict the Wheat Futures Market

Avi Thaker

*Senior AI/ML Engineer
GlaxoSmithKline*



Agenda

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

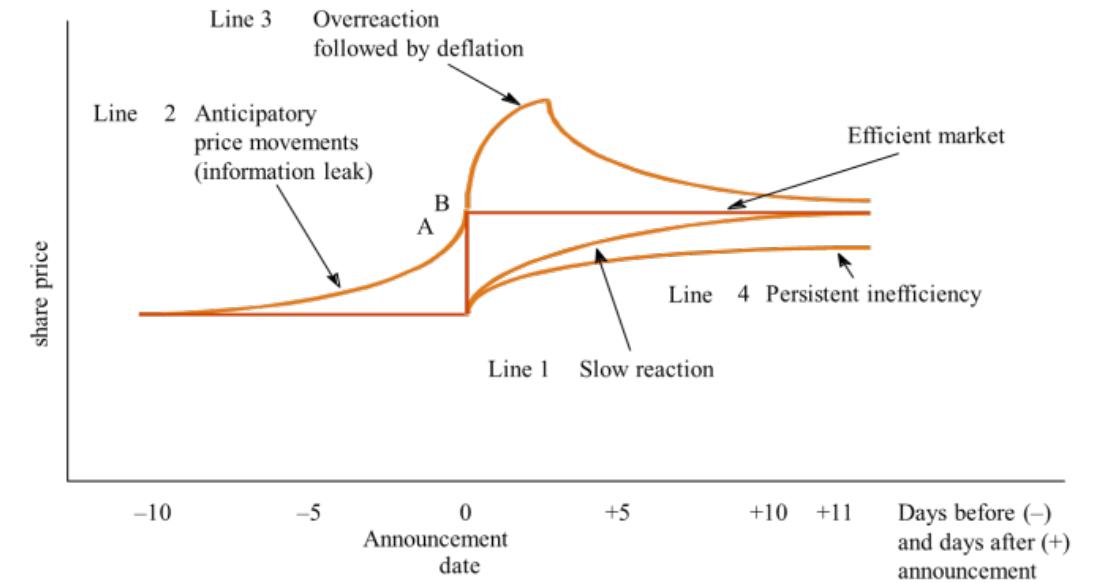
Who I Am

- Senior AI/ML Engineer – GSK
 - Help build better drugs using AI
- Senior Data Scientist – Zume
 - Make Pizza and farming efficient
- 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 (formerly) crypto markets



Trading and The Efficient Market Hypothesis

- The market is becoming **increasingly more efficient**
 - Market efficiency causes prices to incorporate and reflect all relevant information
 - Prices adjust quickly, and reflect available information to new information
 - Securities always trade at their fair value
 - Risk is needed to obtain higher returns
 - Cannot predict trends, history does not indicate future
- Impossible to “**beat the market**”
 - 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 suggest the **efficient market hypothesis is not (completely) true**

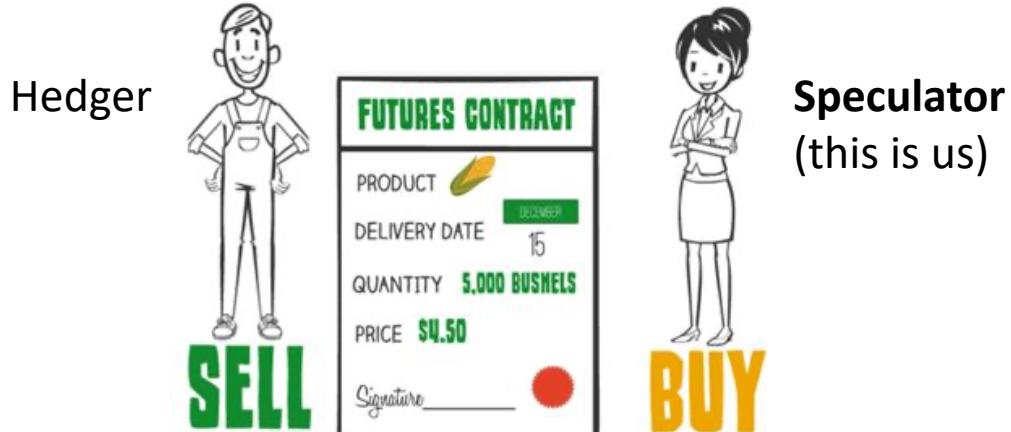


<http://seadnafinnegan.blogspot.com/2014/10/efficient-market-hypothesis.html>

Contracts are standardized to time

The Futures Market

- An auction market
- Spot Price
 - Current market price (immediate payment and delivery)
- Futures contract
 - Buyer is taking the obligation to buy and receive underlying asset when contract expires
 - Seller is taking the obligation to provide and deliver the underlying asset at the expiration date
 - Trade around the spot price



<https://grainphd.com/education/how-do-i-use-the-futures-market/>

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

<https://www.cmegroup.com>

We will use a continuous contract:

- No delivery date
- Price can be thought of as continuous

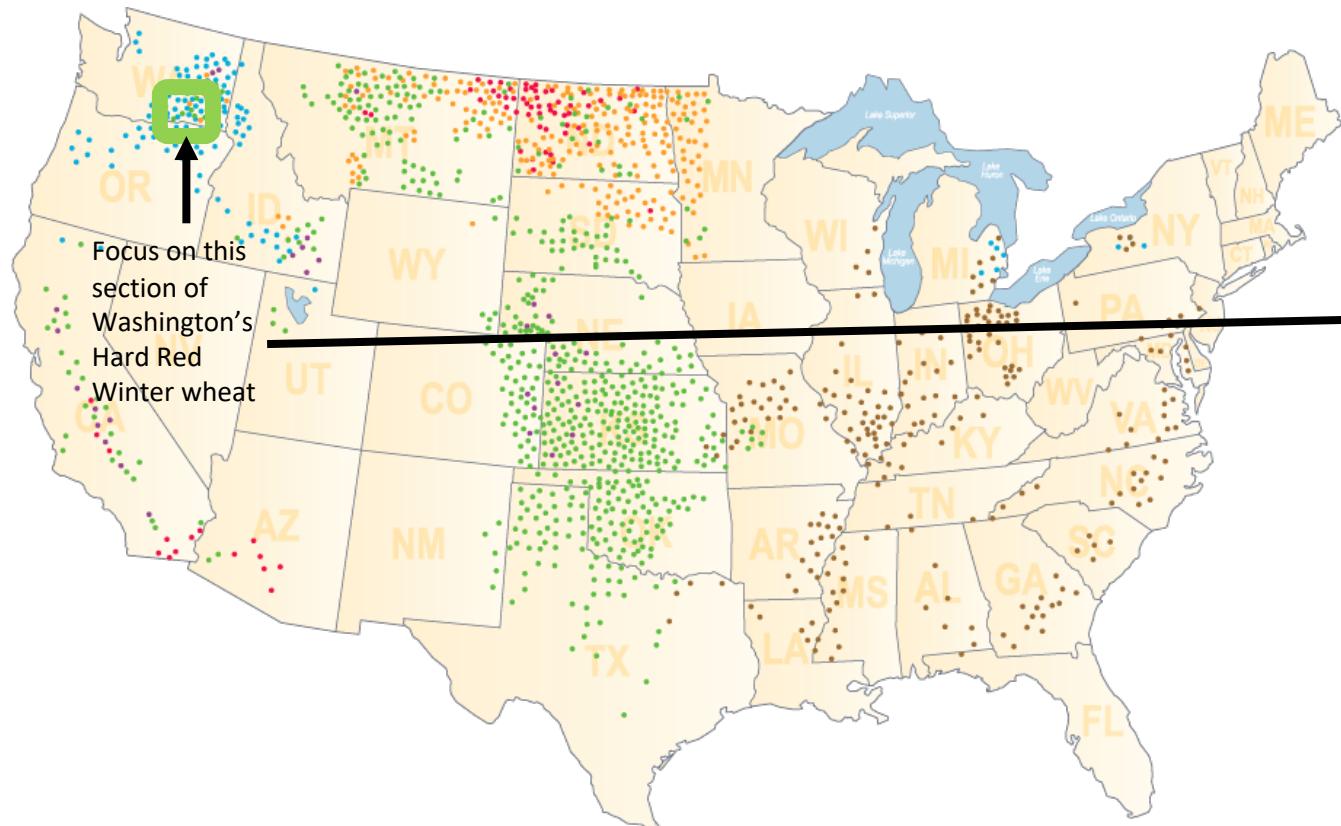
Aerial Imagery and Futures

- **Thesis:**
 - Crop images can suggest future price changes
- **Execution:**
 - Take aerial imagery of Hard Red Winter wheat farms
 - Google Earth gives us free historical imagery over time
 - **Speculate** on the direction that the wheat prices will go over time
- **Machine Learning:**
 - ML models can "learn" the influence of imagery to predict market prices
 - Health and possibly supply can be latent variables
 - Deep learning, especially convolutional neural networks are good at understanding imagery



<https://www.drschaer.com/us/institute/a/nutritionwheat-free-diet>

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.



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.

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.

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

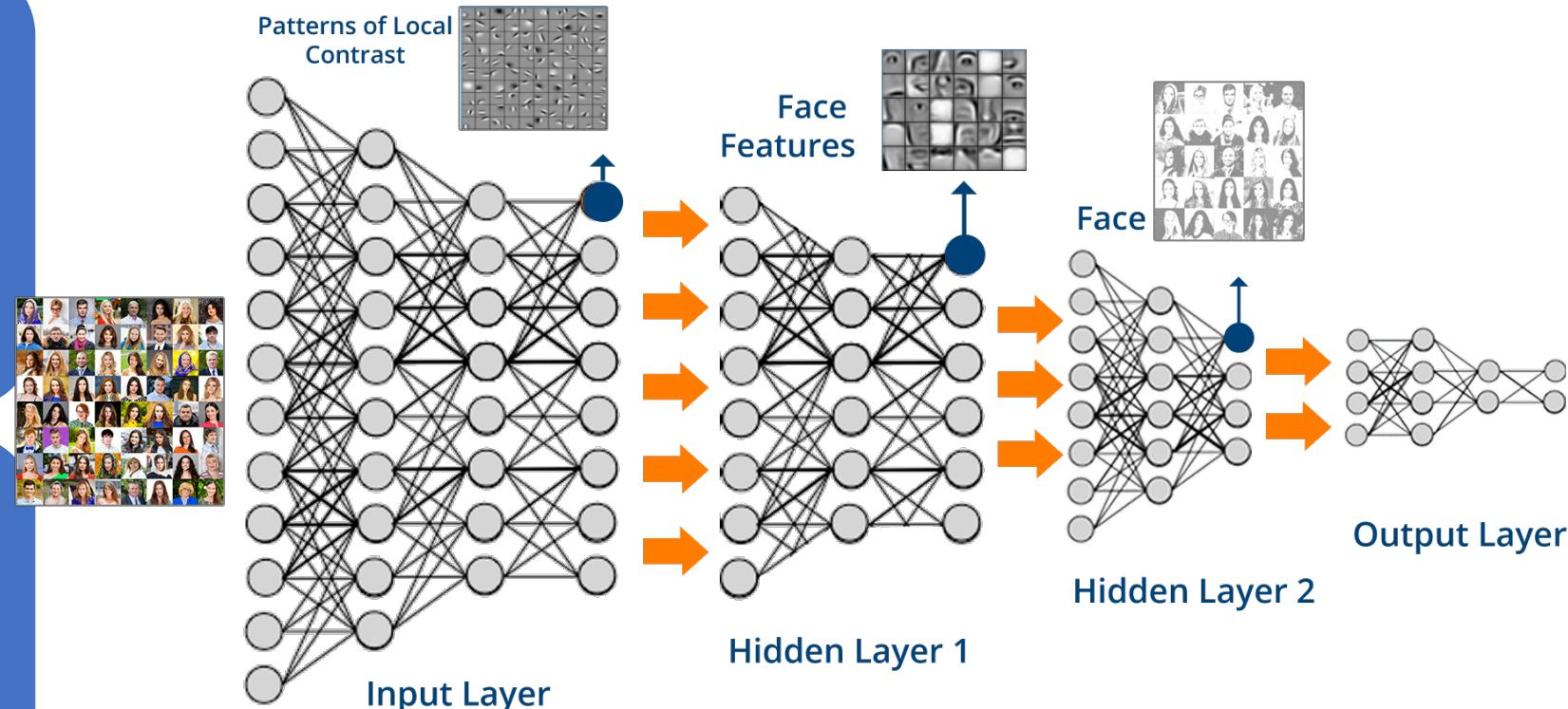
U.S. WHEAT
ASSOCIATES
www.uswheat.org

Wheat Foods
COUNCIL
www.wheatfoods.org

Machine Learning: What Does a Neural Network Learn?

Early layers
learn small
local features

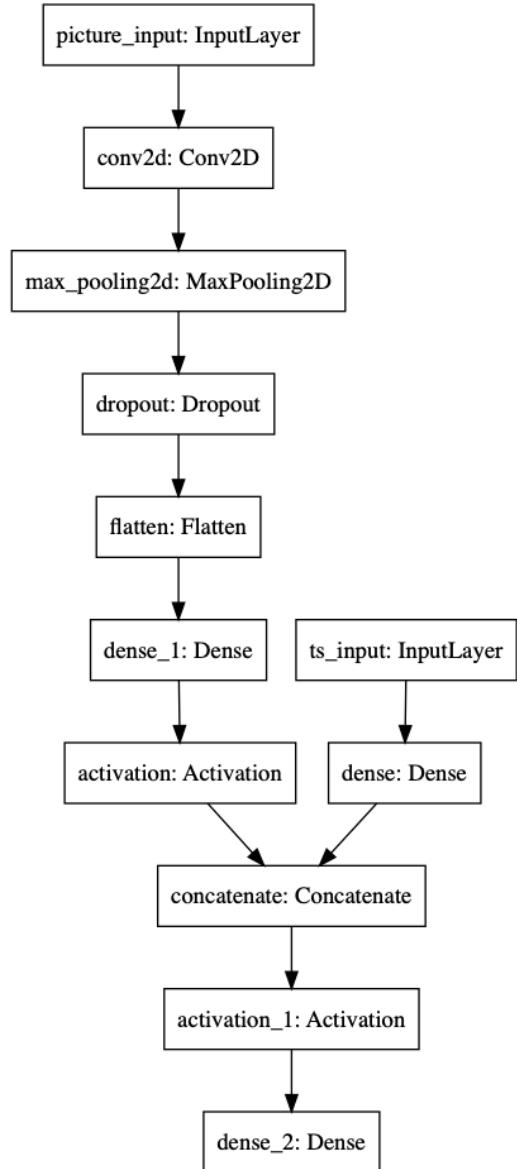
Later layers
learn higher
level features



<https://techxplore.com/news/2019-05-convolutional-neural-network-facial-recognition.html>

Demo

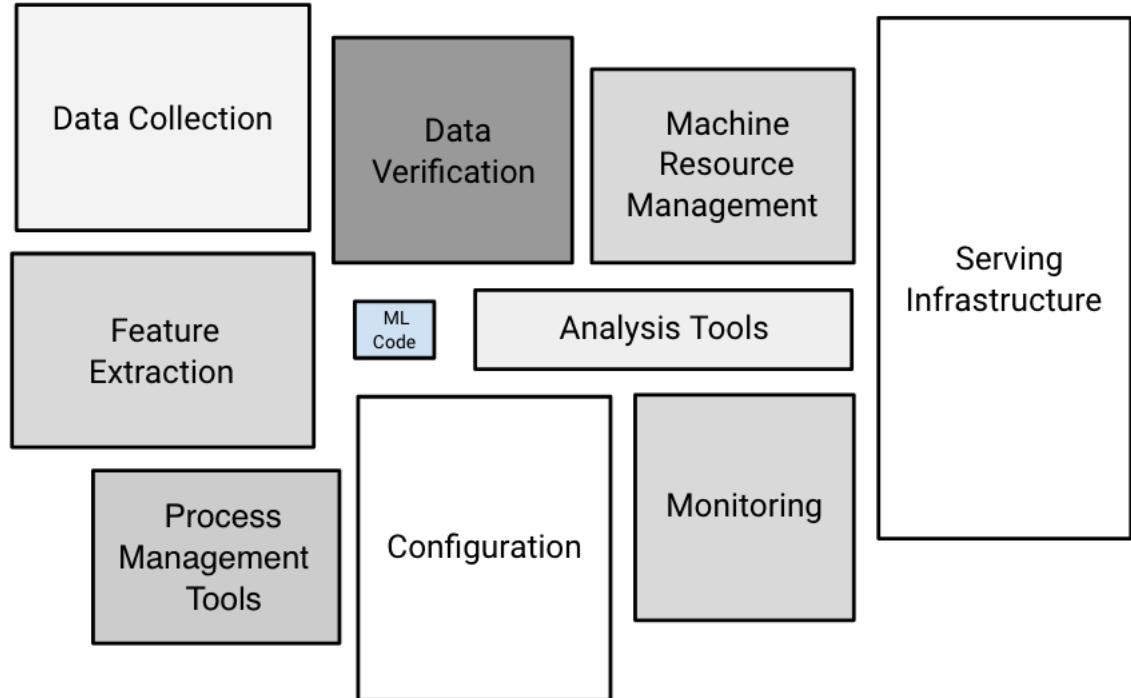
[https://github.com/athaker/econ_136
/tree/master/phxdata20](https://github.com/athaker/econ_136/tree/master/phxdata20)



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

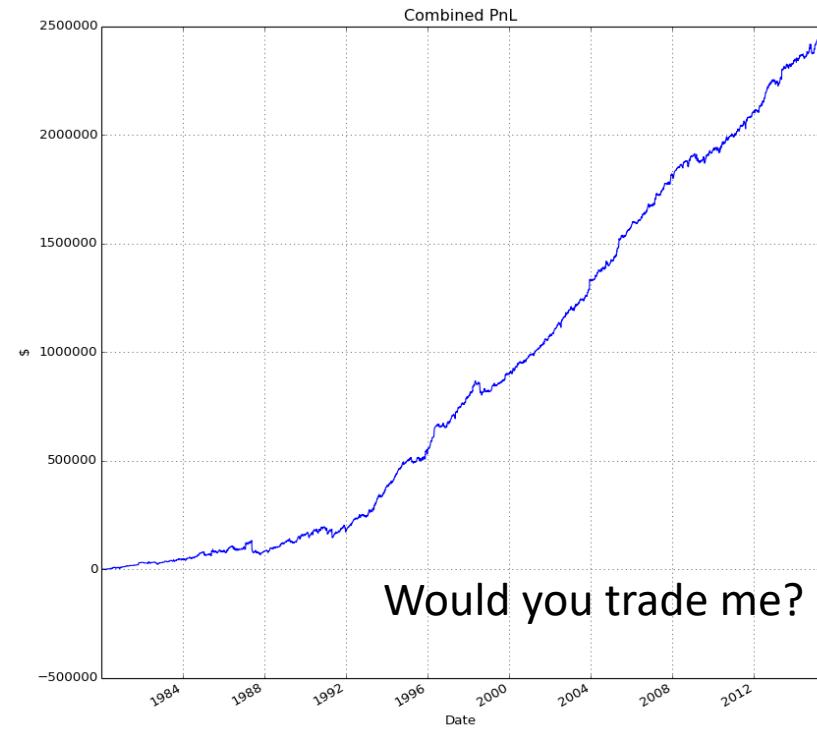
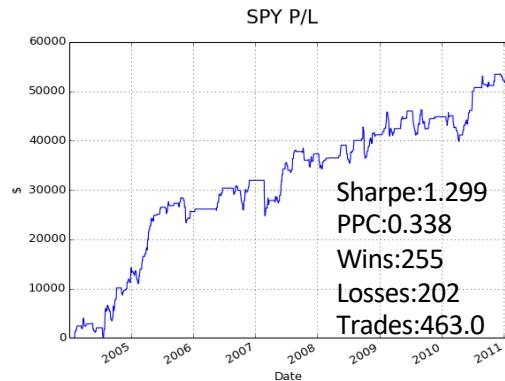


Often the ML problem is the easy part

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

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

#PHXDATA20 VIRTUAL SUMMIT

www.phxdataconference.com

Let's Connect!



<https://www.linkedin.com/in/avi-thaker>

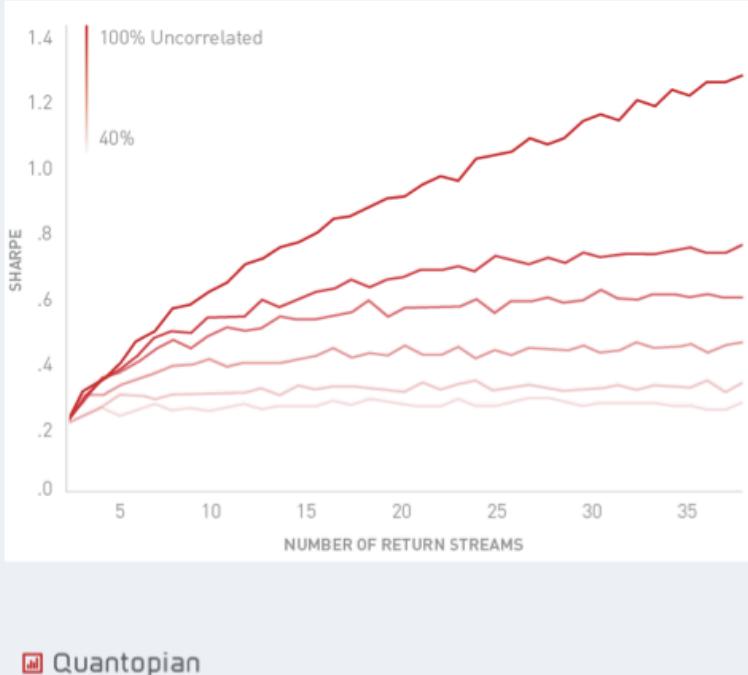
<https://github.com/athaker>

Questions?

Jobs: <https://jobs.gsk.com/en-gb/jobs/258163?lang=en-us>

Appendix: Correlation and Risk

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



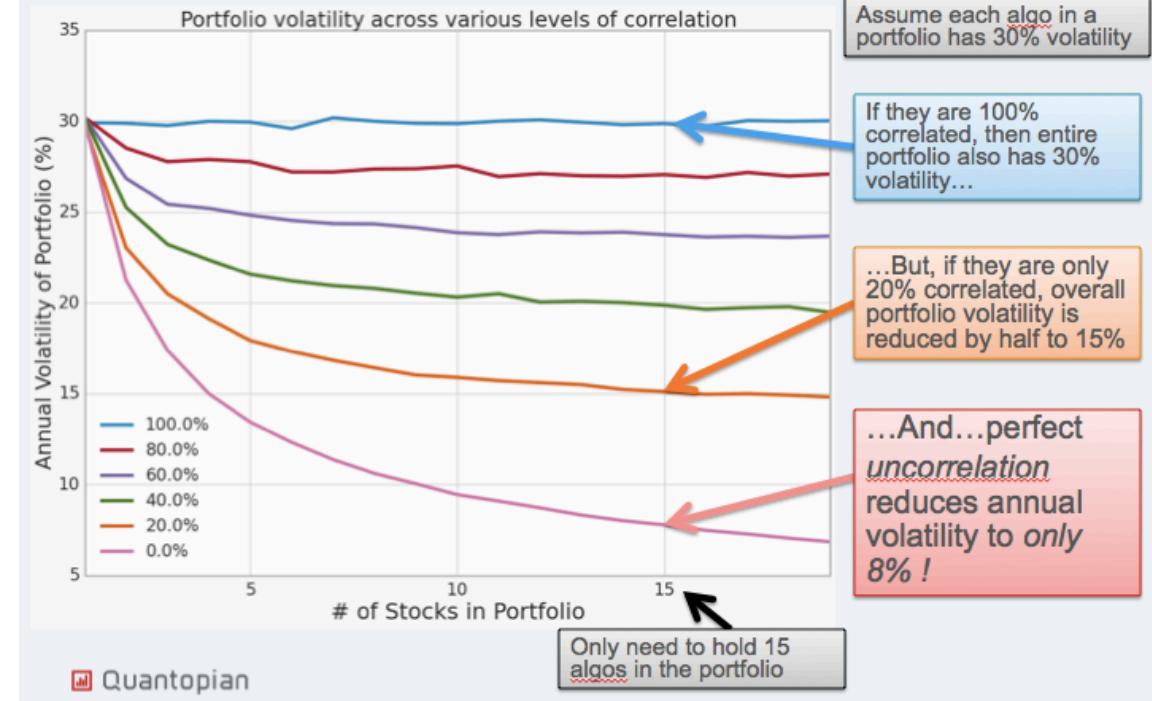
Simulation Setup:

- 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

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



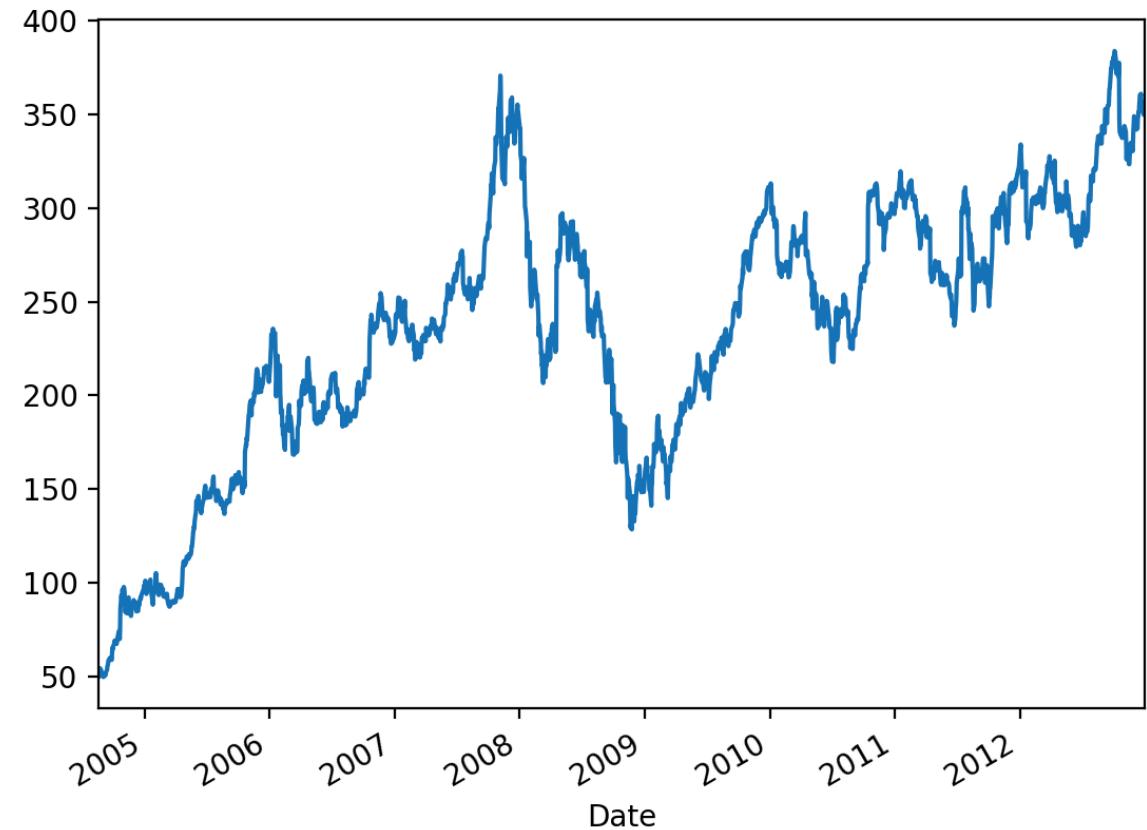
Appendix: Limit Order Execution



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

Maximum Drawdown

- 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?
- $$D(T) = \max_{t \in (0, T)} \{X(t) - X(T)\}$$
- $$\text{MDD}(T) = \max_{t \in (0, T)} [\max_{\tau \in (0, t)} \{X(t) - X(\tau)\}]$$
- Where $X = (X(t), t \geq 0)$ is a random process
 - Simply put maximum drawdown is:
 - $(\text{Peak value before largest drop} - \text{lowest value before new high}) / \text{Peak value before drop}$



Appendix: Sharpe Ratio

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

r_p = portfolio return

r_f = risk free rate

σ_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!!!!

End

Thank You!

Questions?