10 Financial Applications of Machine Learning

Seeing Beyond the Hype

- Financial ML offers the opportunity to gain insight from data:
  - Modelling non-linear relationships in a high-dimensional space
  - Analysing unstructured data (asynchronous, categorical)
  - Learning patterns with complex interactions (hierarchical, non-parametric)
  - Focusing on predictability over parametric adjudication
  - Controlling for overfitting (early-stopping, cross-validation)

- At the same time, **Finance is not a plug-and-play subject** as it relates to machine learning
  - Modelling financial series is harder than driving cars or recognising faces
  - **An ML algorithm will always find a pattern, even if there is none!**

What is Machine Learning?

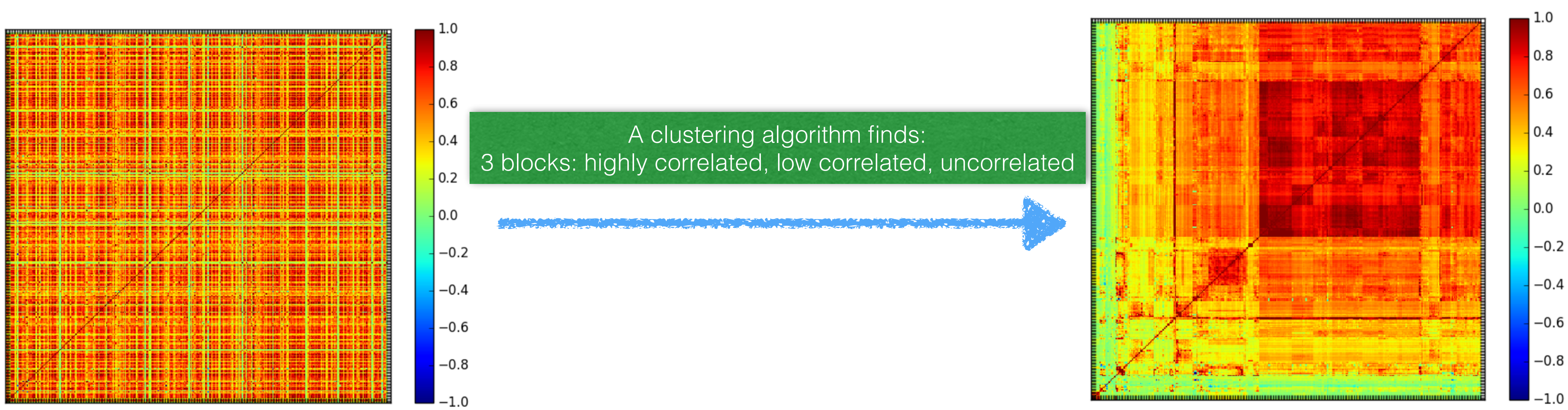
"Machine Learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead."

Wikipedia

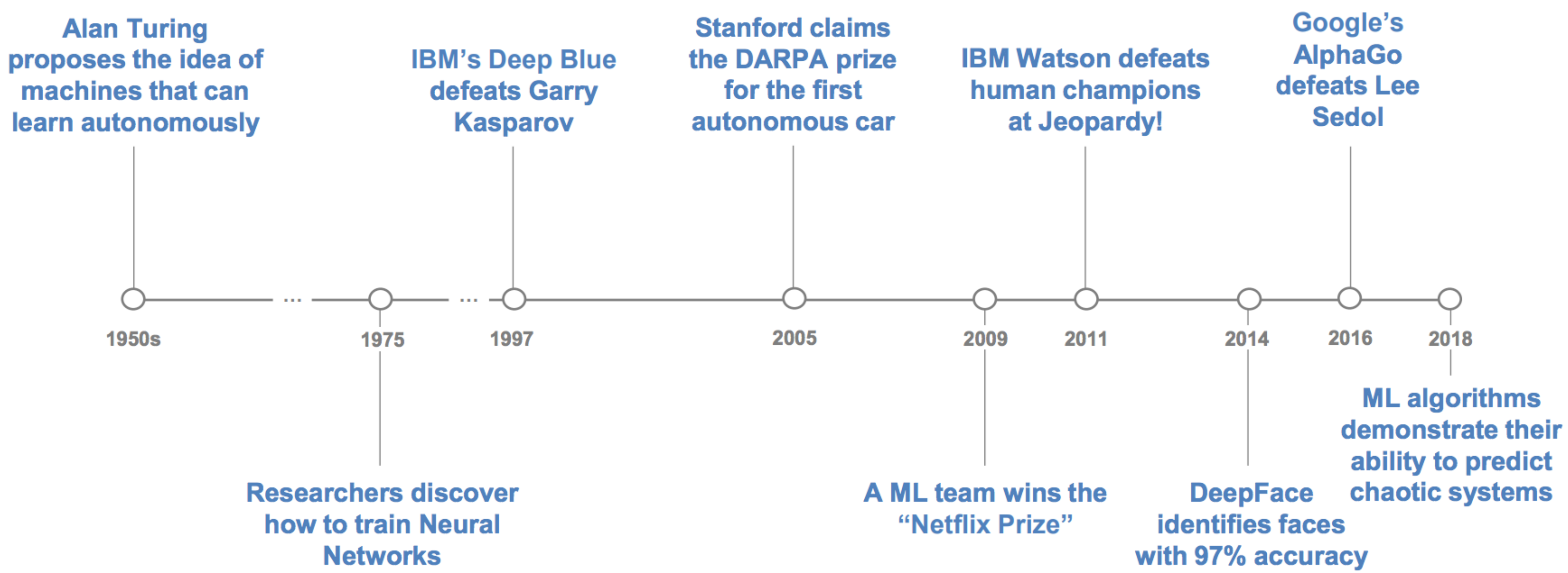
"An ML algorithm learns complex patterns in a high-dimensional space without being specifically directed."

Advances in Financial Machine Learning

Suppose you have a 1000x1000 correlation matrix…



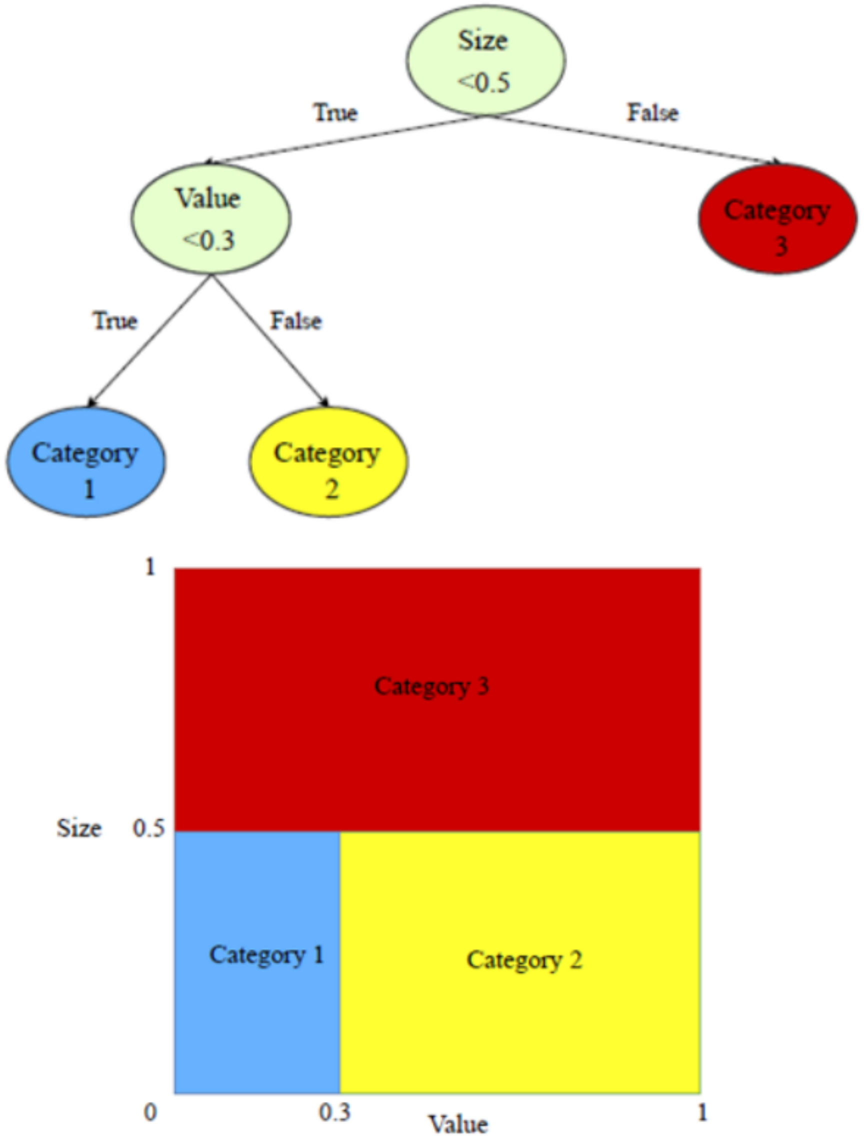
Timeline


Alan Turing proposes the idea of machines that can learn autonomously
IBM's Deep Blue defeats Garry Kasparov
Stanford claims the DARPA prize for the first autonomous car
IBM Watson defeats human champions at Jeopardy!
Google's AlphaGo defeats Lee Sedol
1950s
...
1975
...
1997
2005
2009
2011
2014
2016
2018
Researchers discover how to train Neural Networks
A ML team wins the "Netflix Prize"
DeepFace identifies faces with 97% accuracy
ML algorithms demonstrate their ability to predict chaotic systems

Applications of Financial ML

1. Price Prediction

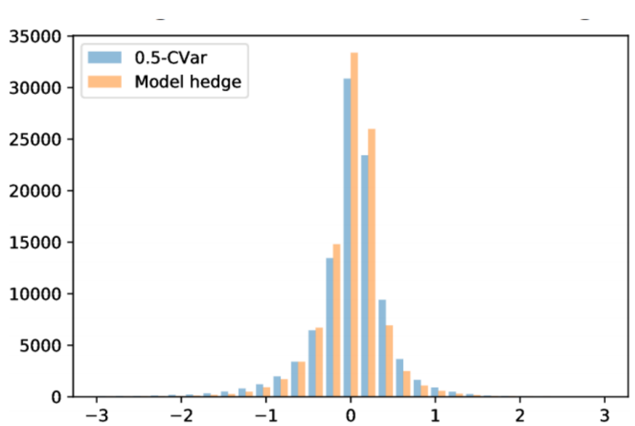
- ML methods allow the modelling of complex relations among the 6 or 7 widely accepted economic factors, including
  - Non-linear relations
  - Threshold relations
  - Hierarchical relations
  - Categorical variable
  - Unknown specification
  - Interaction effects
  - Control variables
- Econometric methods fail to recognise complex relationship, hence leading to inferior results



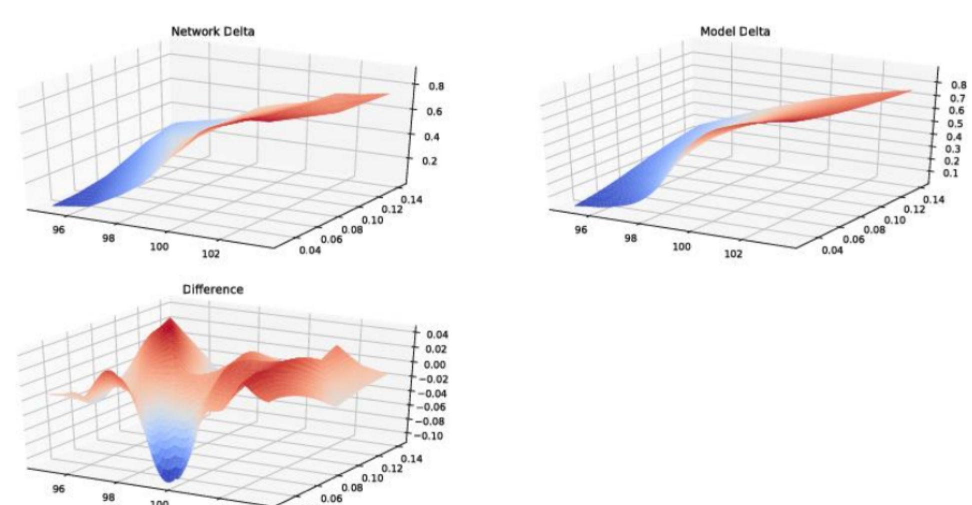
2. Hedging

- Analytical hedging is problematic in presence of market frictions, such as transition costs, market impact, liquidity constraints, risk limits, etc.

- Reinforcement learning approaches are Greek-free and model free. They are purely empirical, with very few theoretical assumptions
  - These models consider many more variables and data points when making hedging decisions, and can generate more accurate hedges at greater speeds



Comparison of model hedge and deep hedge associated to 50%-expected shortfall criterion

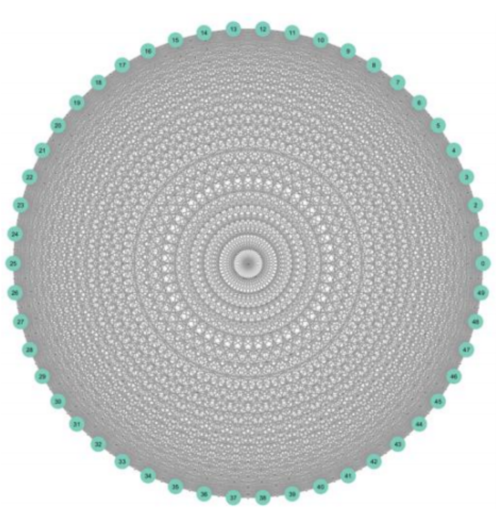


3. Portfolio Construction / Risk Analysis

- Most firms continue to allocate trillions of dollars using mean-variance portfolio optimization (MVO). "*The most expensive piece of beautiful math in history.*"

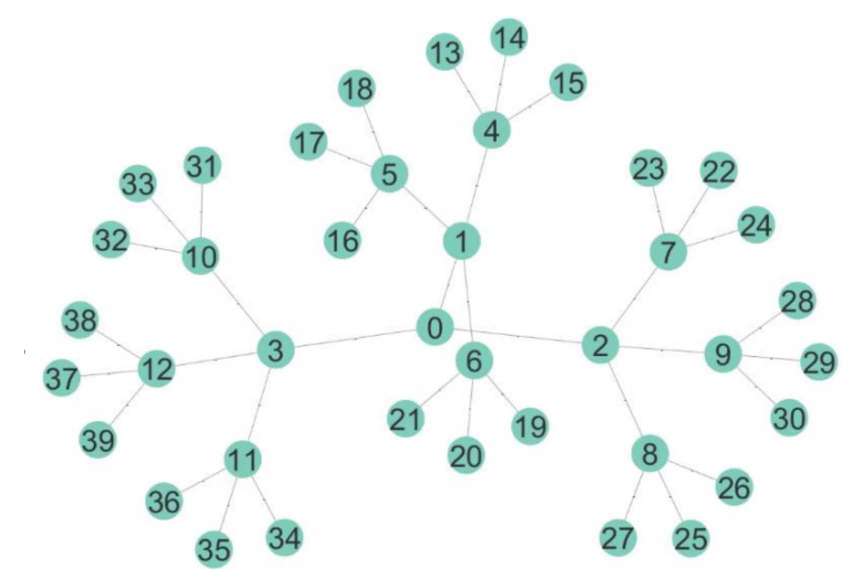
- It is widely known that MVO underperforms the naïve allocation out-of-sample (De Miguel et al. [2009]).

- In contrast, ML solutions outperform MVO (and 1/N) out-of-sample, with **gains in Sharpe ratio that exceed 31%** (López de Prado [2016]).



Covariance-based models require the independent estimation of $N(N+1)/2$ variables.

ML models need only $N-1$ *hierarchical* estimates, making them more robust and reliable.

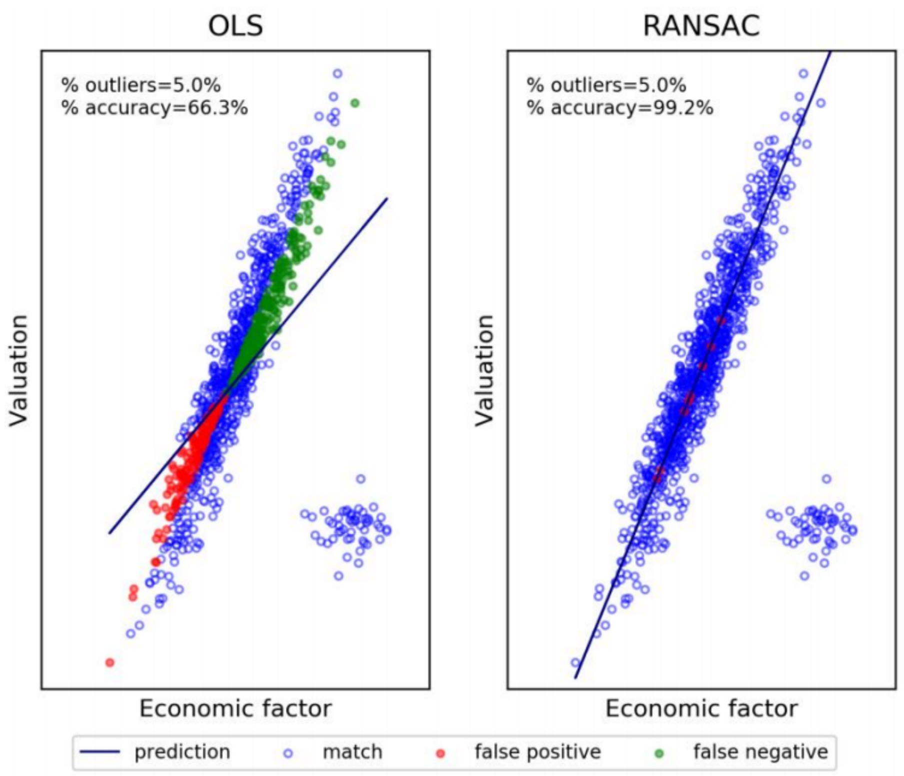


4. Structural Breaks / Outlier Detection

Cross-sectional studies are particularly sensitive to the presence of outliers. Even a small percentage of outliers can cause a very large percentage of wrong signals: Buys that should be sells (false positives), and sells that should be buys (false negatives).

In this plot we run a regression on a cross-section of securities, where a very small percentage (only 5%) are outliers:

- The **red dots** are securities that are expensive, but the regression wrongly classified as cheap.
- The **green dots** are securities that are cheap, but the regression wrongly classified as expensive.

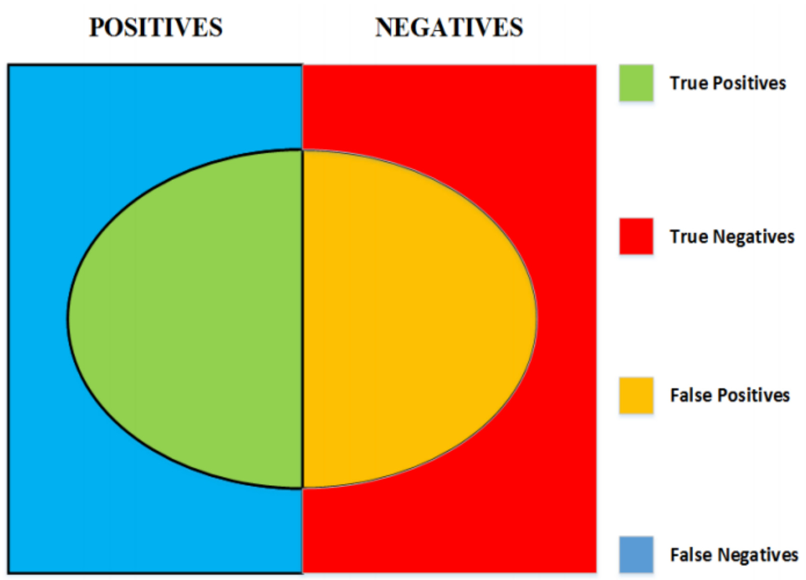


With only 5% of outliers, the cross-sectional regression produced a 34% classification error. In contrast, RANSAC's classification error was 1%, involving borderline cases.

Whenever you suspect the presence of outliers in your data, consider applying RANSAC or similar ML methods.

5. Bet Sizing / Alpha Capture

- Suppose that you have a model for making a buy-or-sell decision:
  - You just need to learn the size of that bet, which includes the possibility of no bet at all (zero size).
  - This is a situation that practitioners face regularly. We often know whether we want to buy or sell a product, and the only remaining question is how much money we should risk in such bet.
  - **Meta-labeling: Label the outcomes of the primary model as 1 (gain) or 0 (loss)**



- Meta-labeling builds a secondary ML model that learns how to use a primary exogenous model.
- The secondary model does not learn the *side*. It only learns the *size*.
- We can maximize the F1-score:

$$F1 = 2\frac{precision \cdot recall}{precision + recall}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

6. Feature Importance

- ML algorithms identify patterns in a high dimensional space.
- These patterns associate features with outcomes.
- The nature of the relationship can be extremely complex, however we can always study what features are more important.
  - E.g., even if a ML algorithm may not derive an analytical formula for Newton's Gravitational Law, it will tell us that mass and distance are the key features.

In traditional statistical analysis, key features are often missed as a result of the model's misspecification.

In ML analysis, we give up closed-form specifications in exchange for identifying what variables are important for forecasting.

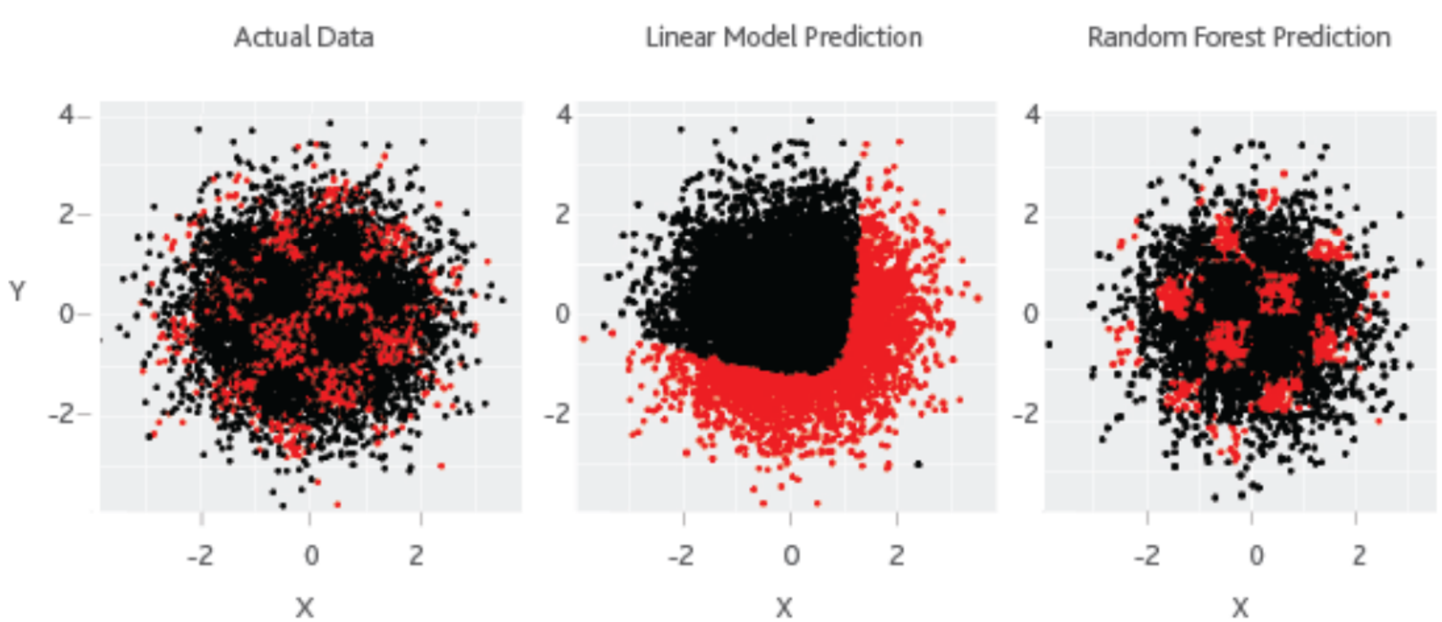
Once we know what are the factors at play, we can develop a theory of how.

7. Credit Ratings, Analyst Recommendations

- Stock analysts apply a number of models and heuristics to produce credit and investment ratings.

- These decisions are not entirely arbitrary, and correspond to **a complex logic that cannot be represented with a simple set of formulas or a well-defined procedure.**

- Machine learning algorithms have been successful at replicating a large percentage of recommendations produced by bank analysts and credit rating agencies.



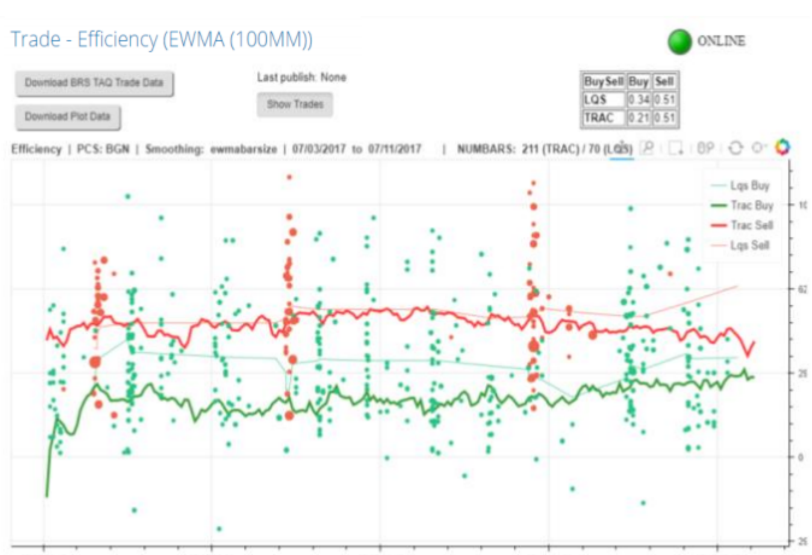
In this example by Moody's, the left figure shows a scatter plot of bonds as a function of two features (X,Y), where defaults are colored in red. The middle plot shows that traditional econometric methods fail at modelling this complex, non-linear relationship. The right plot shows that a very simple ML algorithm performs well.

8. Sentiment Analysis / Recommender Systems

- In the plot below, an algorithm has identified news articles containing information relevant to Tesla (TSLA US Equity).
    - **Blue bars**: Daily count of the total number of articles. The average is 458 articles/day, with a maximum of ~5000.
    - **Green bars**: Daily count of articles expressing a positive sentiment.
    - **Red bars**: Daily count of articles expressing a negative sentiment.

9. Execution

- Credit instruments are traded over-the-counter.
- Many investment grade bonds are not traded for days and even weeks.
- Kernel-based methods identify “similar” trades based on their common features.
  - The set of common trades enables us to derive theoretical prices.
  - If we buy a bond at a price higher than subsequent “similar” bonds, we can bust the trade



This plot shows the trade efficient of buys (green) and sales (red):

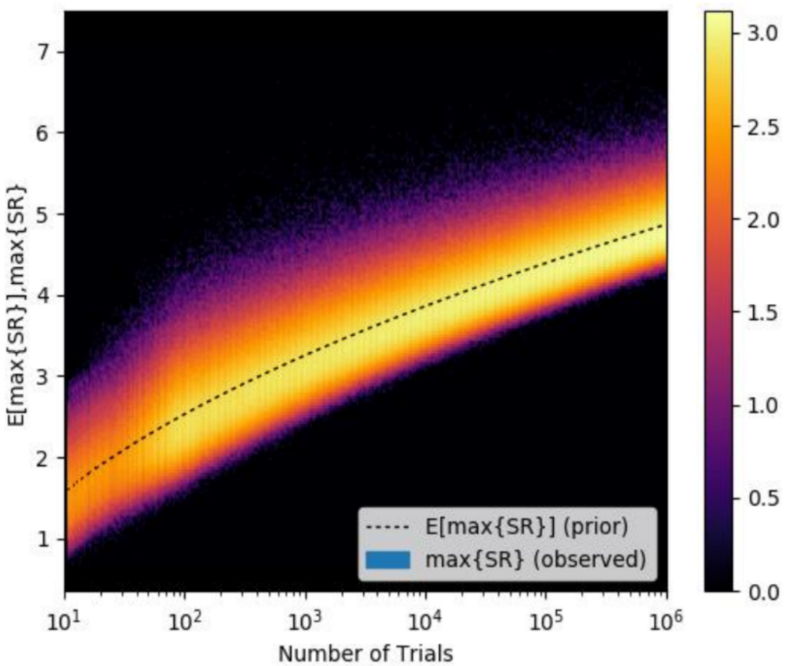
- A **buy** has efficiency 0 when it prints at the quoted offer, and it has efficiency 100 when it prints at the quoted bid.
- A **sale** has efficiency 0 when it prints at the quoted bid, and it has efficiency 100 when it prints at the quoted offer.
- Both have efficiency 50 at the mid.

In this example, the rebalancing of the portfolio has been profitable, as it has captured about 1/3 of the bid-ask spread (approx. 50 bps in price).

10. Detection of False Investment Strategies

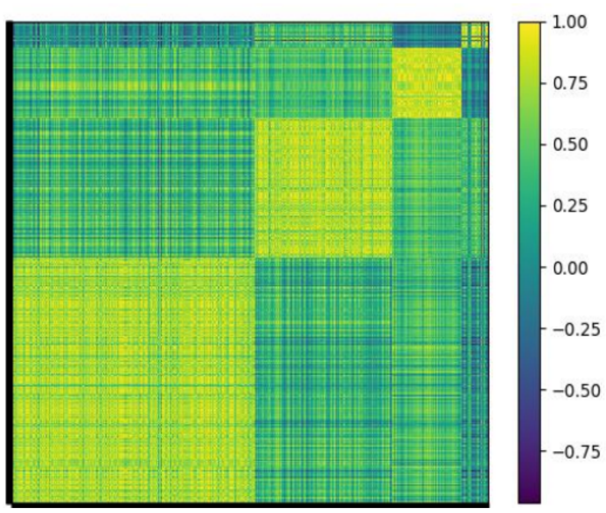
The y-axis displays the distribution of the maximum Sharpe ratios (max{SR}) for a given number of trials (x-axis). A lighter color indicates a higher probability of obtaining that result, and the dash-line indicates the expected value.

For example, after only 1,000 independent backtests, the expected maximum Sharpe ratio (E[max{SR}]) is 3.26, even if the true Sharpe ratio of the strategy is zero!



Most quantitative firms invest in false discoveries.

Solution: Deflate the Sharpe ratio by the number and variance of trials.



Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Strat Count	3265	1843	930	347
aSR	1.5733	1.4907	2.0275	1.0158
SR	0.0974	0.0923	0.1255	0.0629
Skew	-0.3333	-0.4520	-0.4194	0.8058
Kurt	11.2773	6.0953	7.4035	14.2807
T	2172	2168	2174	2172
StartDt	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-05-01	2018-04-25	2018-05-03	2018-05-01
Freq	261.0474	261.0821	261.1159	261.0474
sqrt(V[SR_k])	0.0257	0.0256	0.0256	0.0257
E[max SR_k]	0.0270	0.0270	0.0270	0.0270
DSR	0.9993	0.9985	1.0000	0.9558