O'REILLY

Overview of Algorithmic Trading With Python

By Deepak Kanungo
CEO, Hedged Capital LLC

Feb 2, 2023

Course Outline

1. Algorithmic trading systems
2. Analyzing trading data
3. Designing trading algorithms
4. Prototyping trading algorithms

Poll: Your trading expertise

- What is your level of expertise trading equities?
  - Advanced
  - Intermediate
  - Basic
  - None

Poll: Your Python expertise

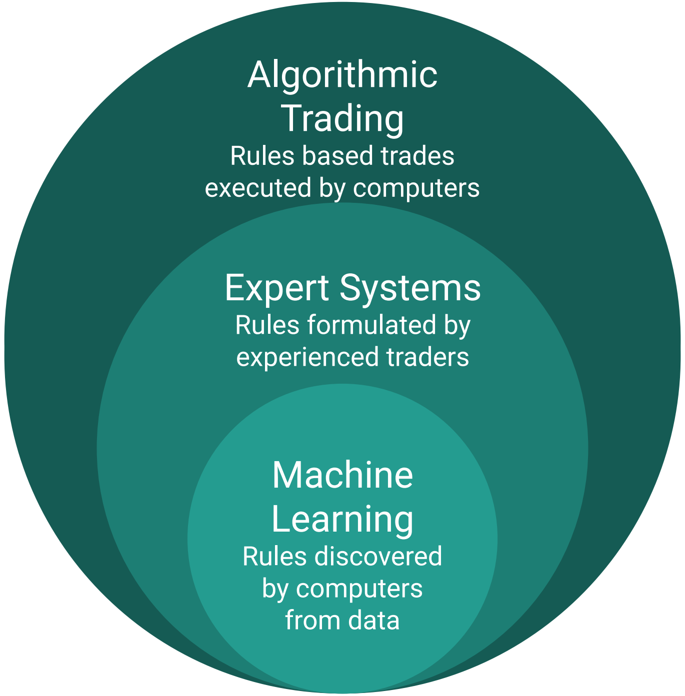
- What is your level of expertise in Python?
  - Advanced
  - Intermediate
  - Basic
  - None

Section 1

Algorithmic Trading Systems

- Trading strategies
- Trading models
- Machine learning

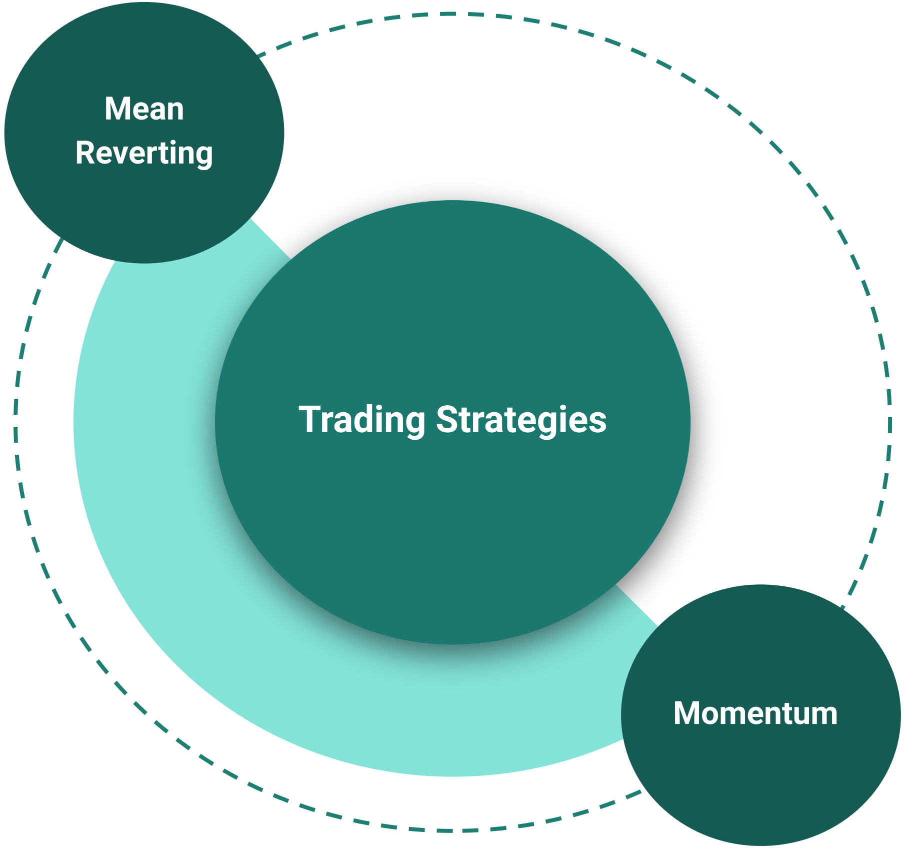
Algorithmic trading automates rules developed by humans and/or machines



Why trade algorithmically using AI/ML?

- Manual trading puts you at a massive disadvantage
  - Over 80% of all equity trading volume in the US is due to algorithmic trading
  - Human beings are too slow to react to price movements
  - Algorithms are not emotional
  - Quick decision-making based on predefined rules
  - Massive volumes of data are processed quickly
  - ML algorithms are very good at spotting patterns in high dimensional datasets
- Do other things with your time instead of staring at your screen!
  - Efficient division of labor between humans and machines
  - Keep your day job while you start a trading business
  - Spend more time with your family and friends

Basically there are only two types of trading strategies



Momentum trading strategies

- Encapsulates the adage, “The trend is your friend (until it ends)”
  - Strong empirical evidence
  - Supported by behavioral finance theory
- Many flavors of trending strategies:
  - Moving average crossover
  - 14-day momentum
  - Volatility breakout

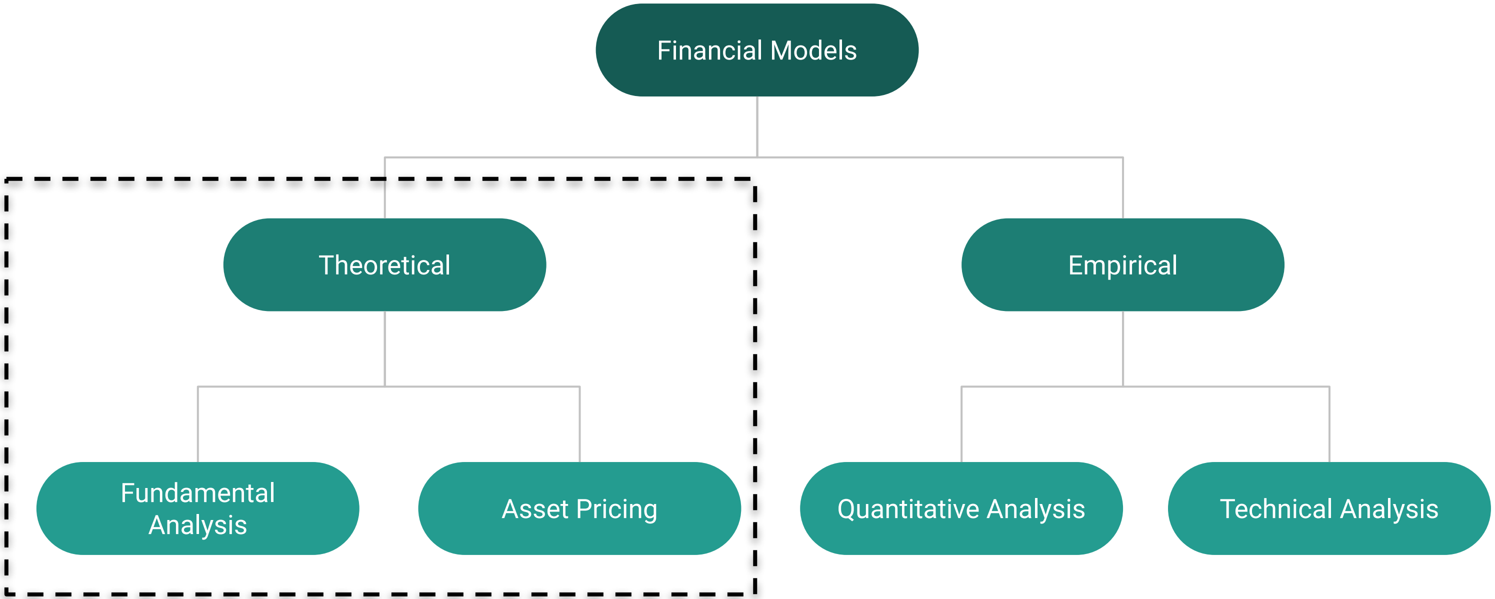
Dual moving average crossover strategy

Mean reverting trading strategies

- Encapsulates Buffet's advice, "Be greedy when others are fearful and fearful when others are greedy"
  - Law of large numbers ensures convergence to the mean
  - Explained by prospect theory of behavioral finance
- Different flavors of mean-reverting strategies:
  - Value based
  - Pairs trading
  - Volatility trading

Mean reverting trading strategy

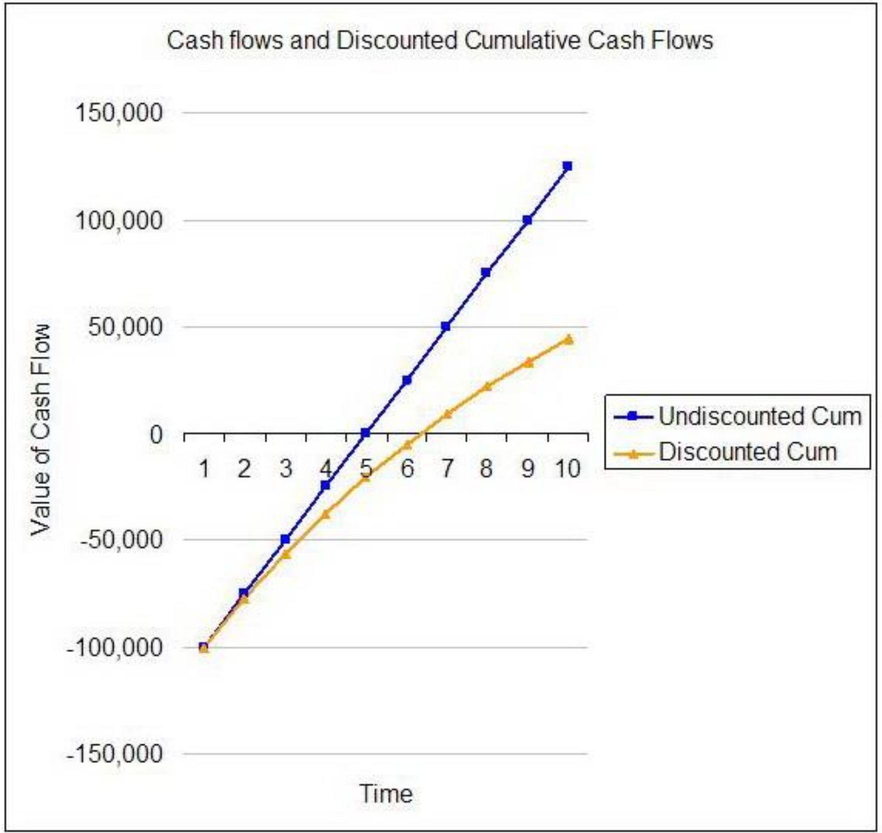
Two main types of financial models used



Theoretical models: Discounted Cash Flow

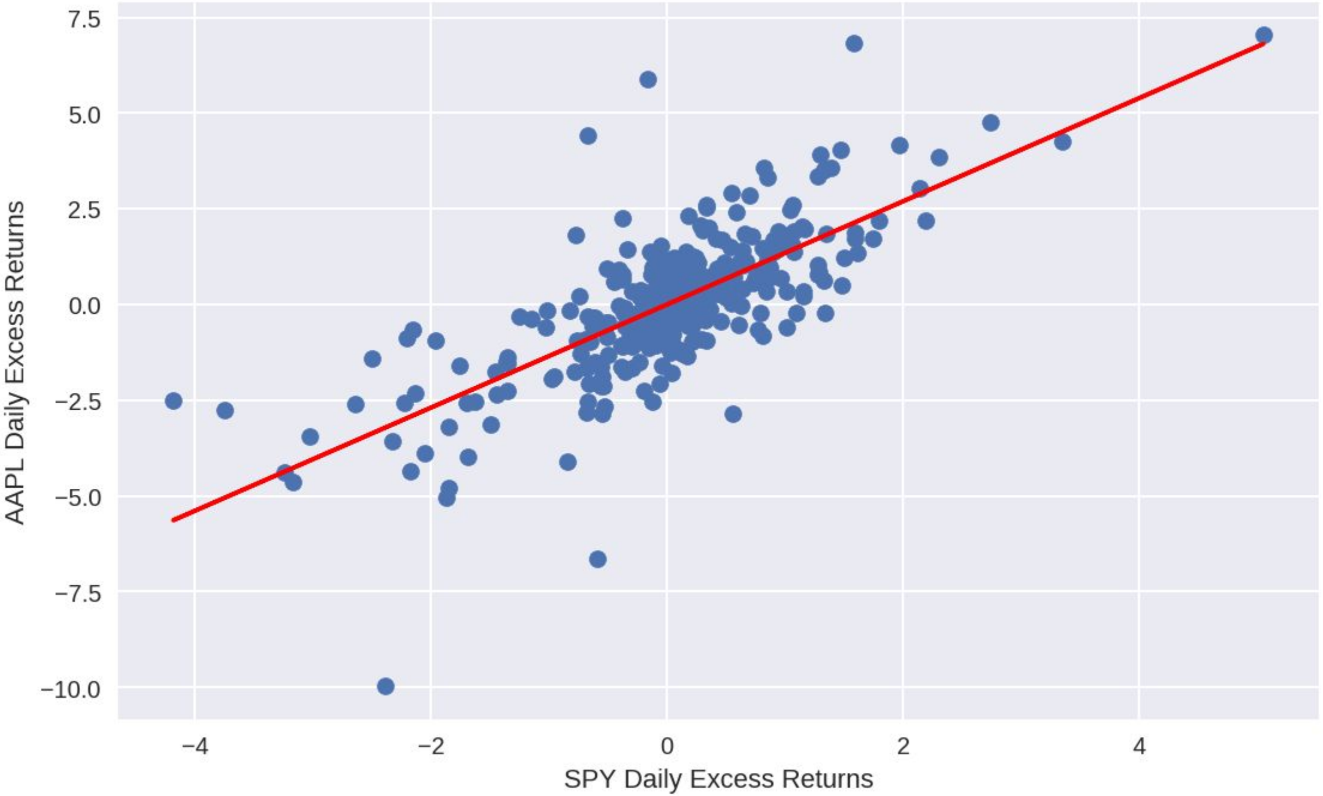
- Discounted Cash Flow (DCF) models are logically sound but hard to do
  - Forecast free cash flows (FCF) over 3-5 years
    - FCF = cash from operations - capital expenditures
  - Forecast terminal value of the company into perpetuity
  - Estimate a discount rate(s) to account for risk
  - Forecast macroeconomic variables like tax rates, inflation rates, GDP, exchange rates
- Net Present Value (NPV) formula estimates the company's value
- Estimate of value is based on many forecasts and assumptions being valid
  - Very sensitive to minor changes in discount and growth rates

Discounted cash flow: cash tomorrow is worth less than the same amount today



Assumes interest rates are positive!

Key statistical concepts in financial models: linear correlations



Theoretical models: Capital Asset Pricing

- Market equilibrium theory about risk, return and diversification
  - Derived from Markowitz’s Modern Portfolio Theory (MPT)
  - Capital markets are completely efficient
- Return on equity depends only on a stock’s sensitivity to market risk
  - No reward for idiosyncratic risk of a company since it can be diversified away
  - Only systematic equity risk exposure to the market is rewarded
- Formula uses a single variable linear regression
  - Beta coefficient captures sensitivity of a stock to the movements of the market
  - If market goes up/(down) by 1%, equity return goes up/(down) by beta% on average

Capital Asset Pricing Model

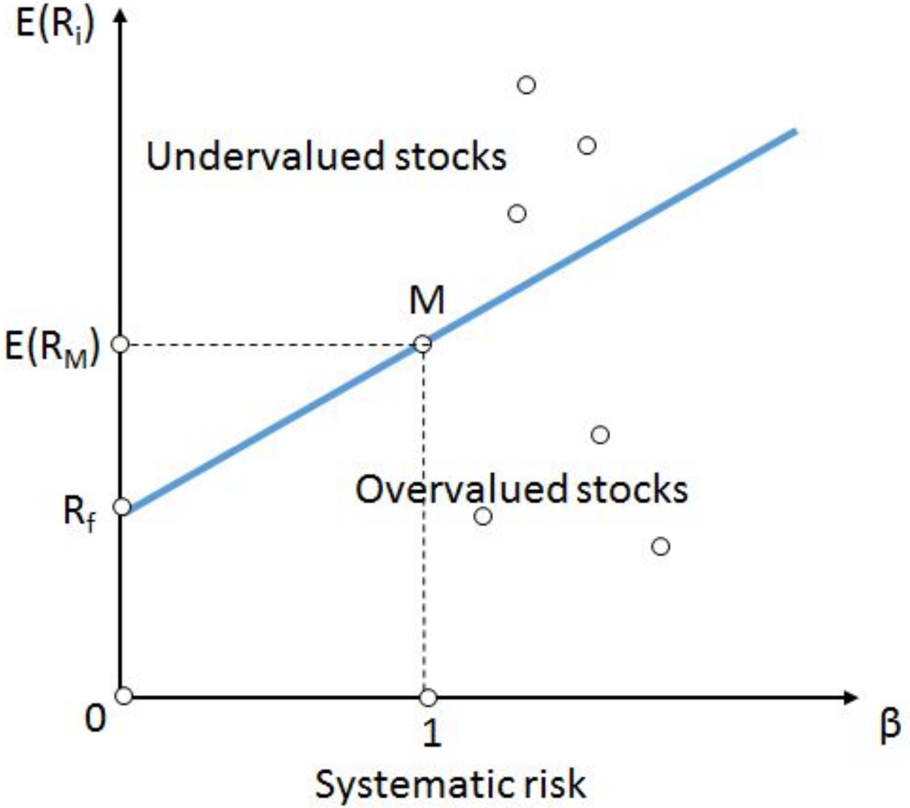


Image copyright: Wikipedia

Theoretical models: Arbitrage Pricing

- A market equilibrium theory based on the law of one price
  - Does not assume markets are efficient but well functioning
  - Mispricings do not persist because market participants arbitrage it away
- Allows more than one macroeconomic risk factor to explain price returns
  - Risk factors may include changes in GDP, inflation rate, yield curve, junk bond premiums
  - Generalizes the single factor of the Capital Asset Pricing Model (CAPM)
- Formula uses multiple linear regression with each factor having its own beta coefficient
  - Does not prescribe the type or number of factors to be used in the model
  - Allows practitioners to create different models without violating the tenets of MPT

Efficient Market Hypothesis: you can’t beat the markets consistently

- Efficient market hypothesis: asset prices reflect all available information. A hierarchy of three hypotheses with cumulative information sets:
    - Weak: asset prices reflect all publicly available historical data
        - Technical trading strategies do not work!
    - Semistrong: asset prices reflect all publicly available historical and fundamental data
        - Technical and fundamental trading strategies do not work!
    - Strong: Asset prices reflect all private and public information
        - Technical, fundamental and insider trading strategies do not work!
- Grossman-Stiglitz paradox: If market prices reflect all available information, what is the incentive for any participant to acquire the information on which prices are based?

What kind of efficiency is this? Investors confuse ZOOM and ZM? For months?

Reference: Market Watch article published online on April 18, 2019

Warren Buffett thinks academic theories of finance are ridiculous!

Some of Buffett's quotes over the years:

- *"I'd be a bum on the street with a tin cup if the markets were always efficient"*
- *"In my opinion, the continuous 63-year arbitrage experience of Graham-Newman Corp, Buffett Partnership and Berkshire illustrate just how foolish EMT [Efficient Market Theory] is"*
- *"You can occasionally find markets that are ridiculously inefficient – or at least you can find them anywhere except at the finance departments of some leading business schools."*
- *"Naturally the disservice done to students and gullible investment professionals who have swallowed Efficient Market Hypothesis has been an extraordinary service to us. In any sort of a contest – financial, mental or physical – it's an enormous advantage to have opponents who have been taught that it's useless to even try"*

Markets make a mockery of academic models that pervade theory and practice

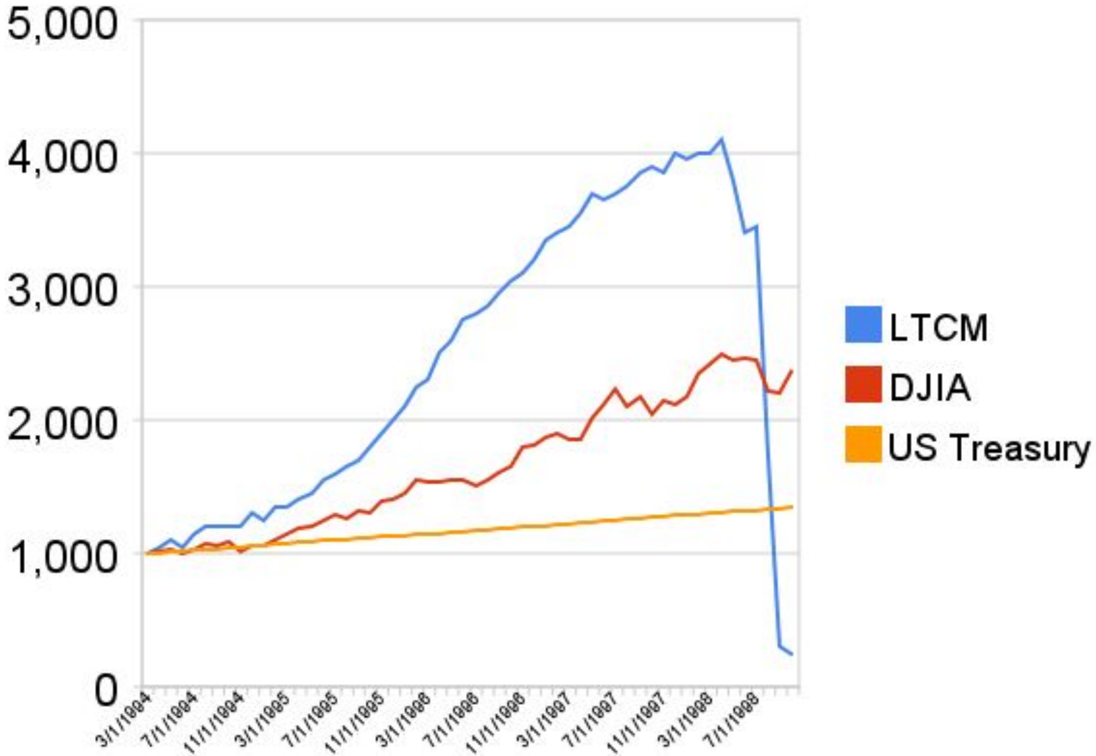
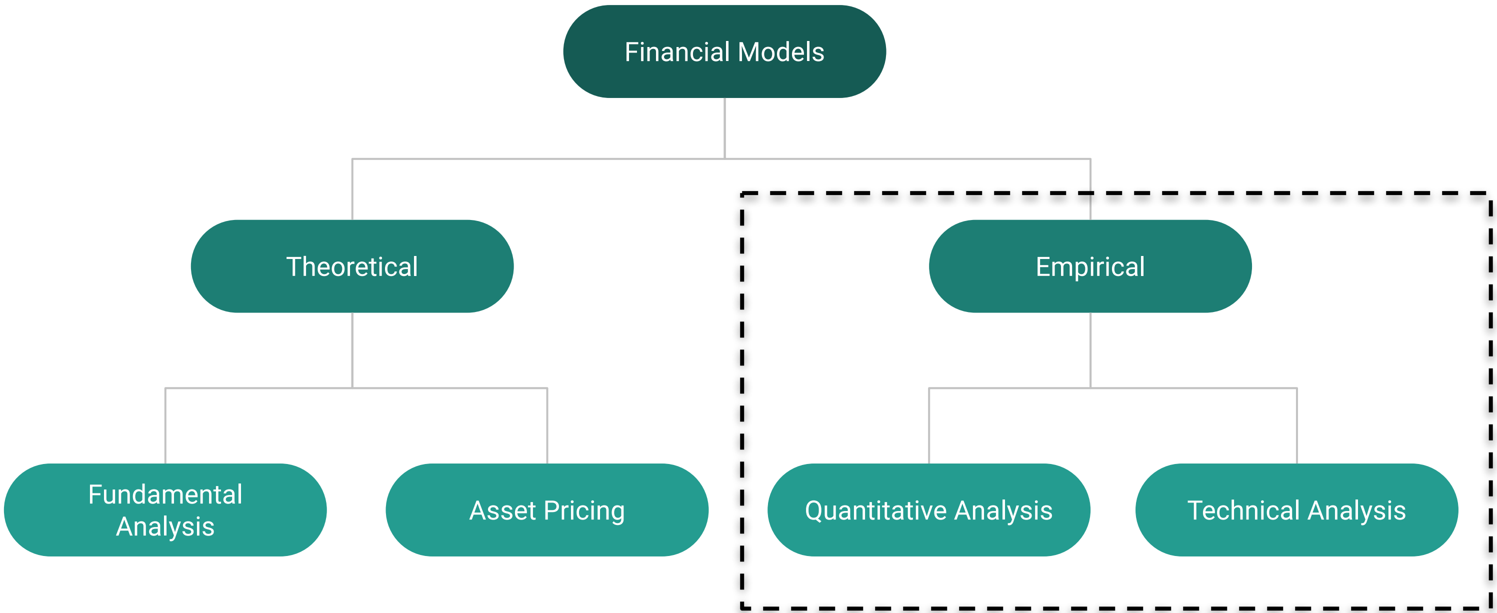


Image source: Creative Commons

Empirically based financial models



Exercise: Importing financial data

- Use built-in Colab to import data from various sources
- Click on link here
- Read chapter 6 on Data Loading, Storage, and File Formats for additional technical details

Freely available data sources

- FRED: Economic and market data
  - A vast repository of economic and financial data
- Alpha Vantage: Equities, currencies, cryptocurrencies, sectors, indicators
  - Realtime and historical API
- Yahoo Finance: free one stop shop for various financial securities
  - Unstable endpoints and data reader needs a fix
- Quandl: One stop shop for all kinds of data
  - Free and premium APIs
- IEX: Equity, ETF market and fundamental data
  - Free and premium APIs
- EDGAR: Company fundamental data
- Hedged Capital: ML generated market analytics
  - Insights about which factors moved the markets in the prior week

Empirical models: technical analysis

- Technical analysis is based on the theory of supply and demand
  - All information is in the price and volume of the security
  - Equity prices follow trends: uptrend, downtrend or sideways
  - Historical patterns will repeat in the future
- Accepts many hypotheses of behavioral finance
  - People are not the rational economic agents of modern financial theory
  - Supply and demand based on fear and greed drives prices and volumes
- Some models use complex mathematics and “theories”
  - Signals can be quite subjective and seem arbitrary

Technical analysis is about recognizing past patterns that will repeat in the future

Image copyright: Wikipedia

Empirical models: quantitative analysis

- Statistical models that are based on historical data and experimentation
  - No economic rationale is required
  - Prone to discovering spurious relationships
- Multi-Factor models are generally multiple linear regression models
  - Any number of factors e.g. macroeconomic, fundamental, technical, sentiment, other
  - Note that APT is a multi-factor model based on an economic theory
- Machine learning algorithms are used in N-factor models
  - Factors are called risk factors, independent variables, predictors or features
  - Feature engineering is the difficult part
  - Overfitting noisy financial data is a massive problem

A single factor market model

‘The Trinity Of Errors’ in all financial models

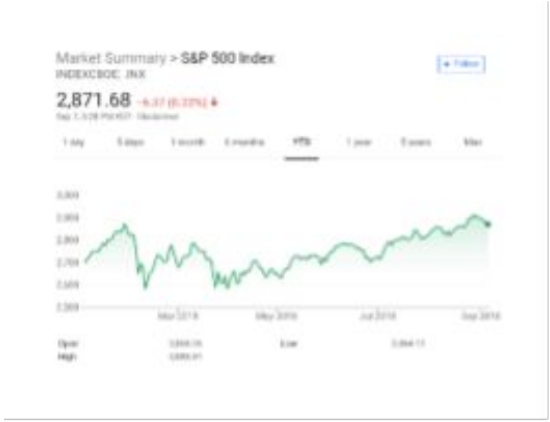
“All models are wrong, but some are useful.” - George Box, Statistician

- All models are affected by the Trinity of Errors:
  - Errors in model specification
  - Errors in model parameter estimates
  - Errors from the failure of a model to adapt to structural changes
- Imperative need to quantify the uncertainty in all estimates
  - The veneer of a physical science lulls adherents into a false sense of certainty
  - Disastrous economic consequences are too many to list but LTCM is instructive

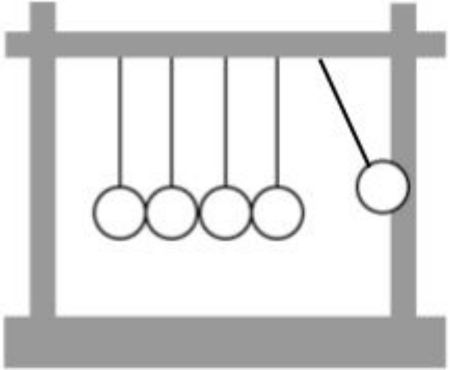
For a detailed discussion, please see chapter 1 of my book, Probabilistic Machine Learning [here](#)

Finance is not physics. Not even close.

"I can calculate the movement of the stars, but not the madness of men." - Isaac Newton, after losing a fortune on his investment in the South Sea Company



≠



Social systems are extremely complex

- People are complex, emotional beings
  - Endowed with free will to flout rules
  - Inconsistent/irrational behavior
  - Continually react to the actions of others
- Latent cognitive biases used in decision-making
- Market participants profit by beating or gaming the system
- Asymmetry of information and processing abilities
- Capitalism and technological innovations lead to dynamic social systems

Section 2

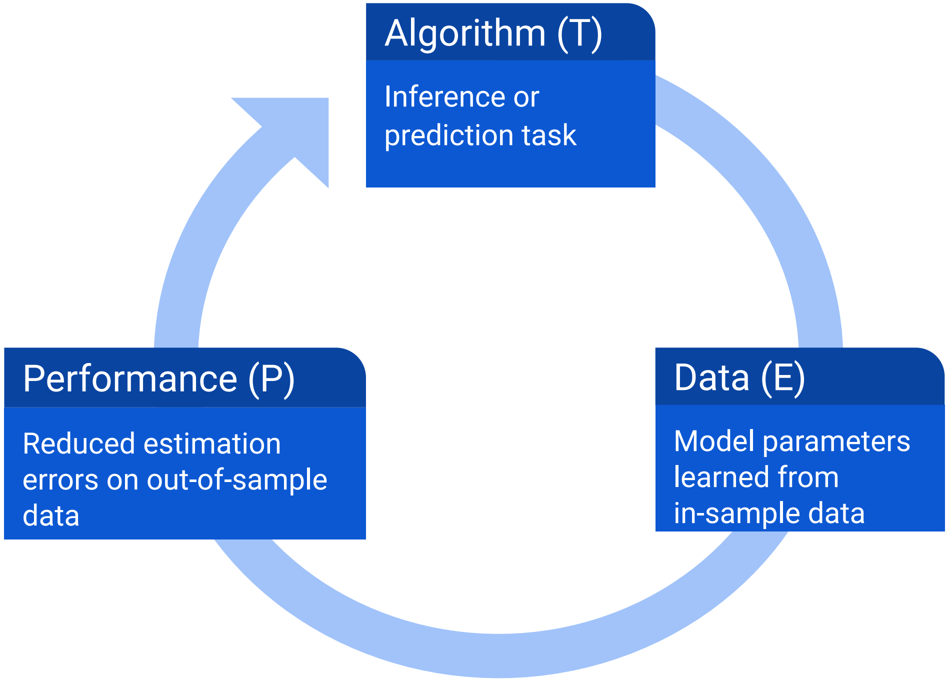
Analyzing Trading Data

- Machine learning systems
- Development process
- Performance metrics

How machine learning systems are different from expert systems of the past

A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E

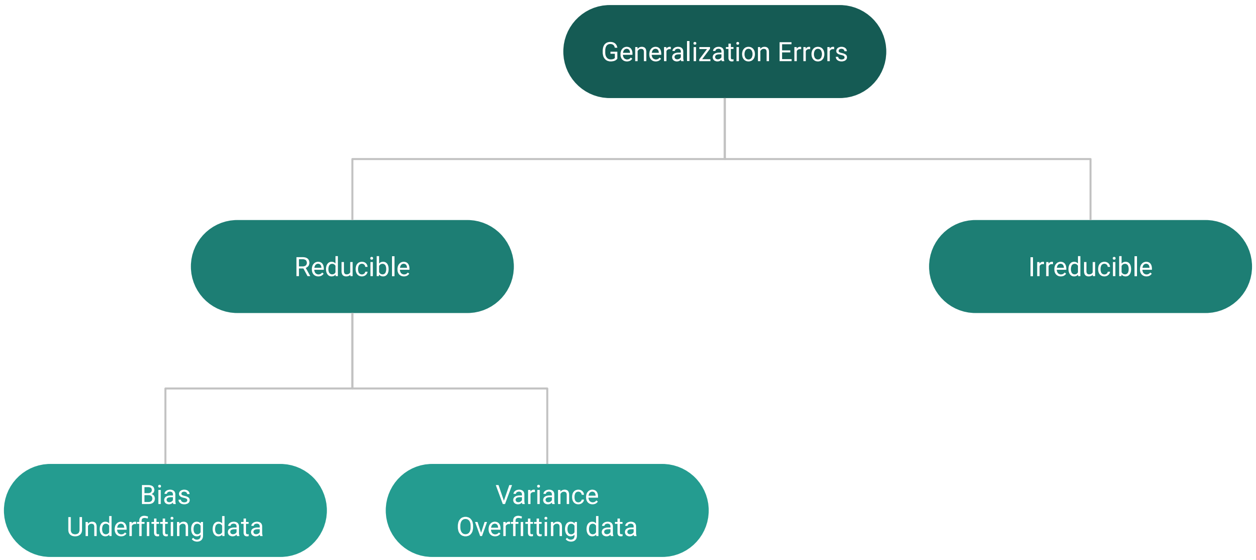
- *Tom Mitchell, 1997*



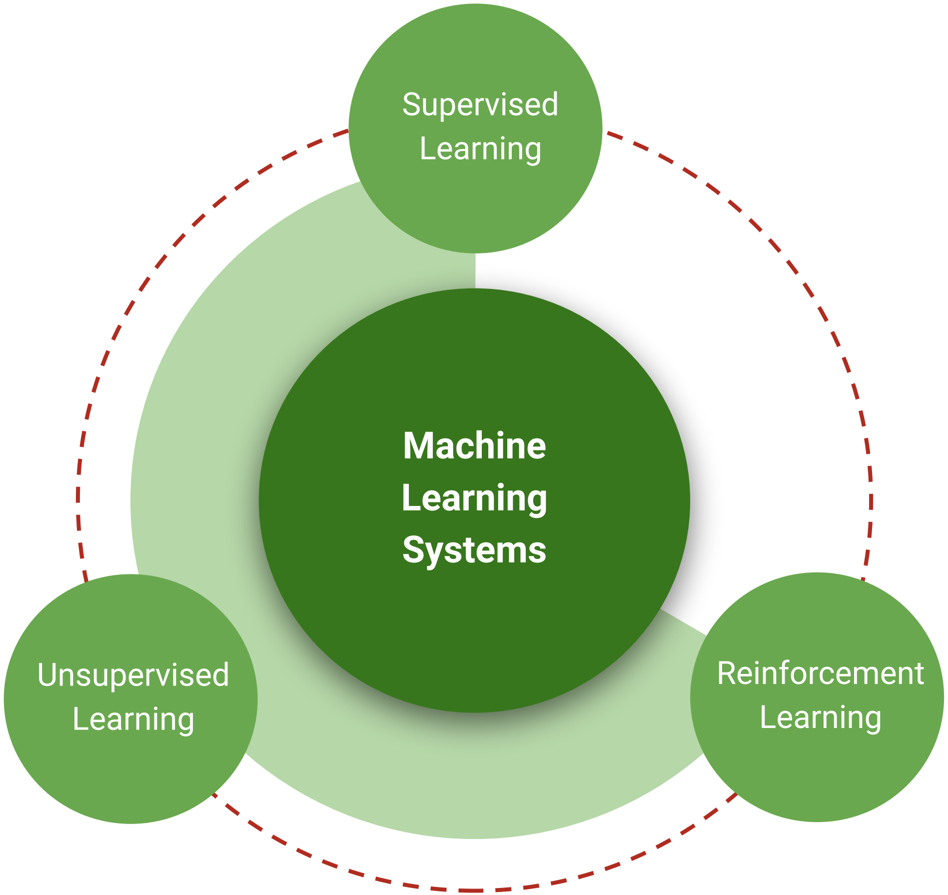
Why use machine learning (ML) in trading?

- Machine learning helps traders build more robust and insightful models given the woeful inadequacy of financial theories
- No need to hardcode rules into models. Useful when there are too many rules or they are difficult to formulate
- Models are able to detect patterns in very high dimensional datasets
- Able to mitigate the curse of high dimensions
- Engineer better, more realistic, financial models

Generalization errors of ML models



Three Types of Machine Learning Systems



Three Types of Machine Learning Systems

- Supervised Learning: data provided to the learning algorithm includes pairs of inputs and desired outputs
  - Regression versus classification algorithms
- Unsupervised Learning: data provided to the learning algorithm only includes inputs
  - Data exploration versus data transformation algorithms
- Reinforcement Learning: algorithm continually updates a strategy based on feedback from its environment to maximize a cumulative reward
  - Model-based versus model-free algorithms

“No free lunch” theorems by David Wolpert

- Age-old philosophical problem of induction
  - Consistent out-of-sample predictions require the underlying phenomena to have structural unity so the future can resemble the past
- Every model is a simplified version of the reality it imitates
  - Assumptions are made to focus on important details
  - Assumptions are data/problem specific
- If you don’t make any assumptions about the data, the performance of all algorithms will be equivalent when averaged across all possible problems
- There is no bias-free optimal machine learning algorithm in ML
  - Optimal algorithm depends on the nature of the data

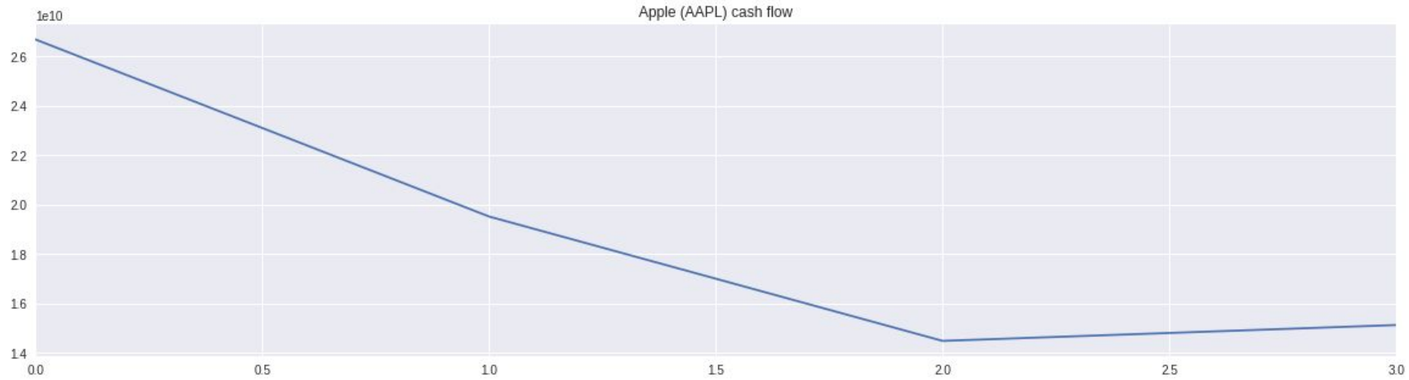
Challenges in machine learning

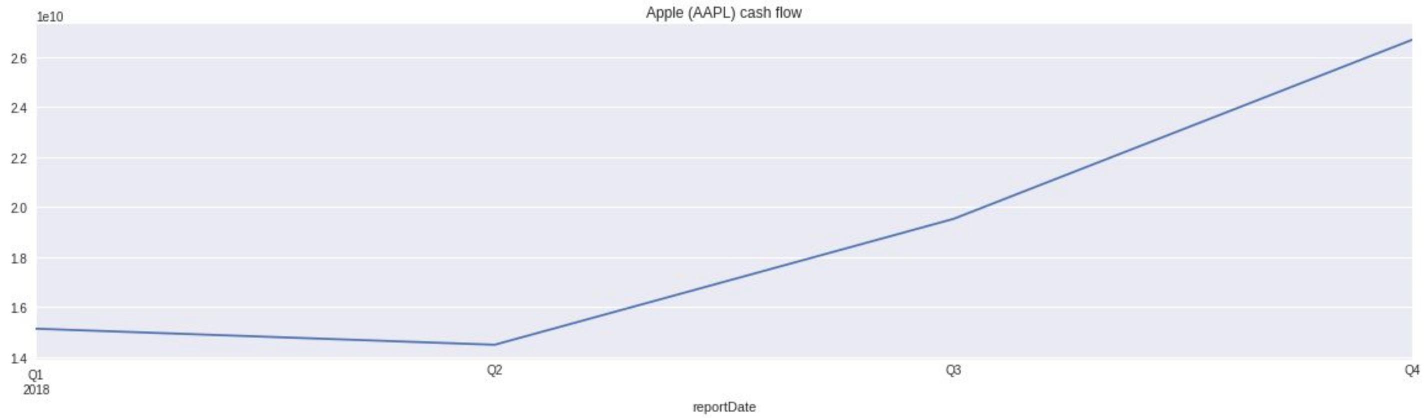
- Data-related
  - Poor quality datasets
  - Small datasets for training and validation
  - Sample dataset on which model is trained is not representative of the population
- Model-related
  - Poor feature selection or omission of key features or correlation among features
  - Algorithm is too simple to capture complexity of the phenomenon (bias/underfitting)
  - Algorithm is too complex and learns the noise in the data (variance/overfitting)

Why financial AI/ML is so hard

- Patterns in the market data change continually, often abruptly
  - Previous datasets no longer represent the new data generating stochastic process
    - Data are insufficient when market changes discontinuously
  - Some previous features become irrelevant or misleading
    - New features may be need to be engineered or discovered
  - Learning algorithm may need to be changed or may not be robust to handle disruption
    - May need to be retrained based on new dataset if possible

Need to examine and clean data imported from various sources





Issues with ingesting data

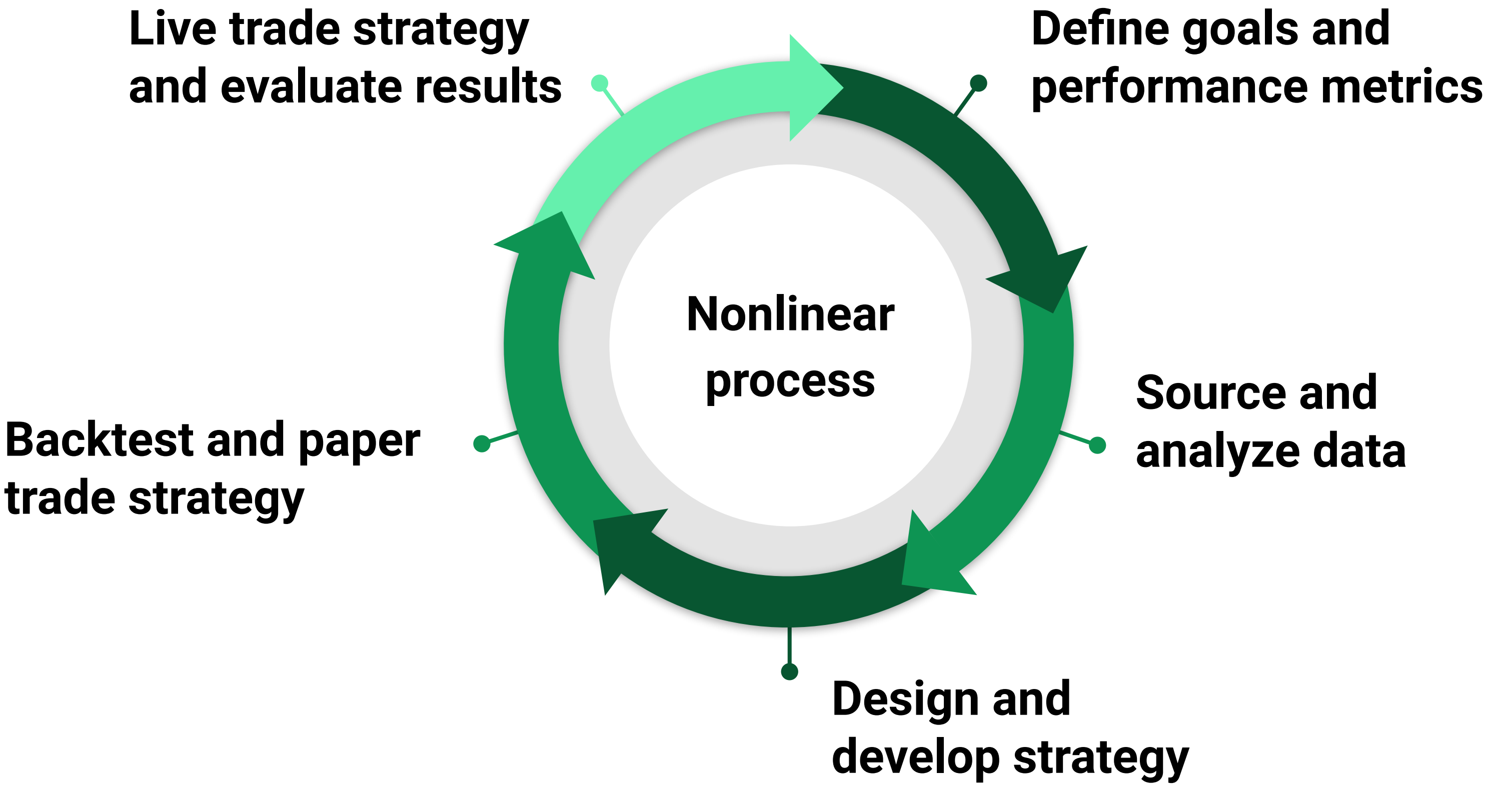
- Data always need to be cleaned and prepared
  - Not a number (NAN)
  - Converting JSON texts to float
  - Data may be inaccurate
  - Data may have different frequencies
  - Some prices may be stale
- Read chapter 7 of Data Cleaning and Preparation for more technical details

Sources should have point-in-time (PIT) data without look-ahead bias

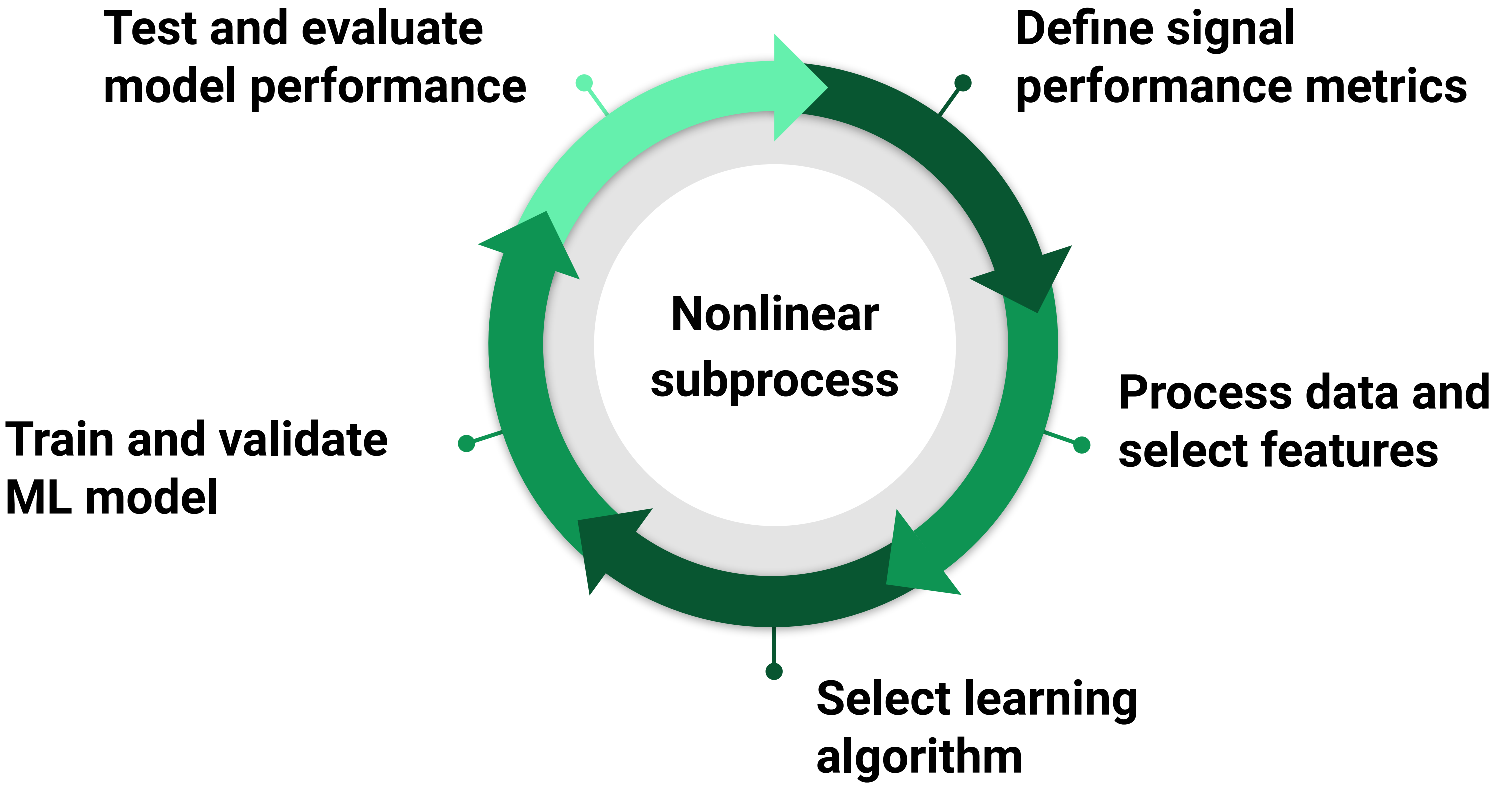
Exercise: Analyze and visualize ingested data

- Use built-in Colab to import data from various sources
- Click on link here
- Use the matplotlib and pandas to visualize and analyze data
  - Various built-in methods
- Generally leads to better understanding of the problem
  - Sometimes, key insights are gained that lead to quality signals

Algorithmic trading system development process



Subprocess for designing a machine learning model for generating trading signals



Commonly used risk-adjusted performance metrics

Two types of risk-adjusted performance metrics:

- Theory based: Jensen’s alpha
  - The difference between the realized return, calculated using its MM, and the expected return predicted by CAPM, is called Jensen's Alpha
- Ratio based: Sharpe, Treynor, Sortino, Sterling

Key statistical concepts in financial models: mean and standard deviation

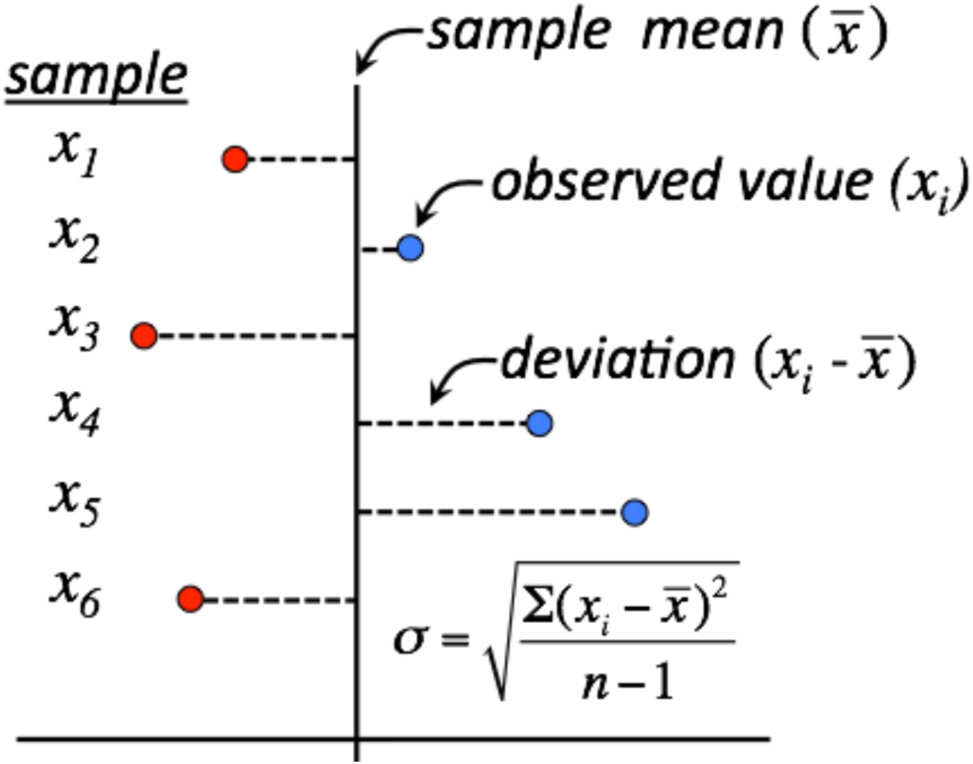


Image source: Wikimedia Commons

Key statistical concept in financial models: Gaussian or Normal distribution

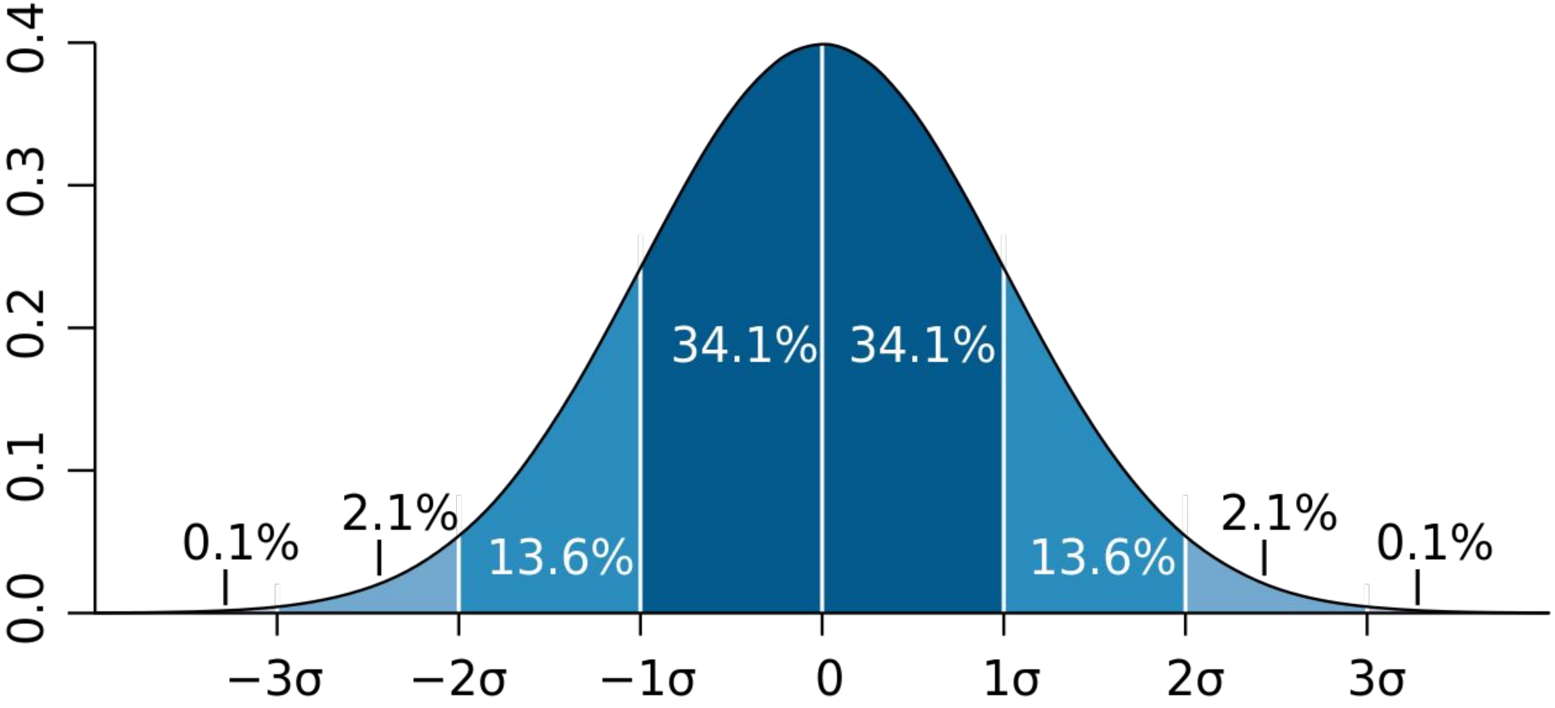


Image source: Wikimedia Commons

Sharpe Ratio: volatility adjusted excess returns

- Measures excess returns per unit of volatility
  - Sharpe ratio = (Expected rate of return - riskless rate)/standard deviation of returns
  - Expected return is estimated using the mean of past returns
  - Expected volatility is estimated using the standard deviation of past returns
  - Unaffected by leverage used to produce returns
  - Sensitive to unit of time. Need to annualize for comparison purposes
- Used on for measuring stand-alone risk only
- Penalizes upside returns that deviate from the benchmark
- Assumes a normal distribution and ignores risks from outlier events

Treynor Ratio: beta adjusted excess returns

- Measures excess returns per unit of systematic risk
  - Treynor ratio = (Expected rate of return - riskless rate)/beta of returns
  - Expected return is estimated using the mean of past returns
  - Expected volatility is estimated using the standard deviation of past returns
  - Sensitive to unit of time. Need to annualize for comparison purposes
- Used to compare investments to be added to a well-diversified portfolio
- Penalizes upside returns that deviate from the benchmark
- Assumes a normal distribution and ignores risks from outlier events

Sortino Ratio: downside volatility adjusted excess returns

- Measures excess returns per unit of downside volatility. Target rate is whatever benchmark return the investor sets it to be
  - Sortino ratio = (Expected rate of return - target rate)/semi-standard deviation of returns
  - Expected return is estimated using the mean of past returns
  - Expected downside volatility is estimated using semi-standard deviation of past returns
- Used on a stand-alone basis and not as part of a diversified portfolio
- Does not penalizes upside returns that deviate from the target return
- Does not assume a normal distribution
- Analytically not so tractable as standard deviation

Sterling Ratio: common sense return and risk

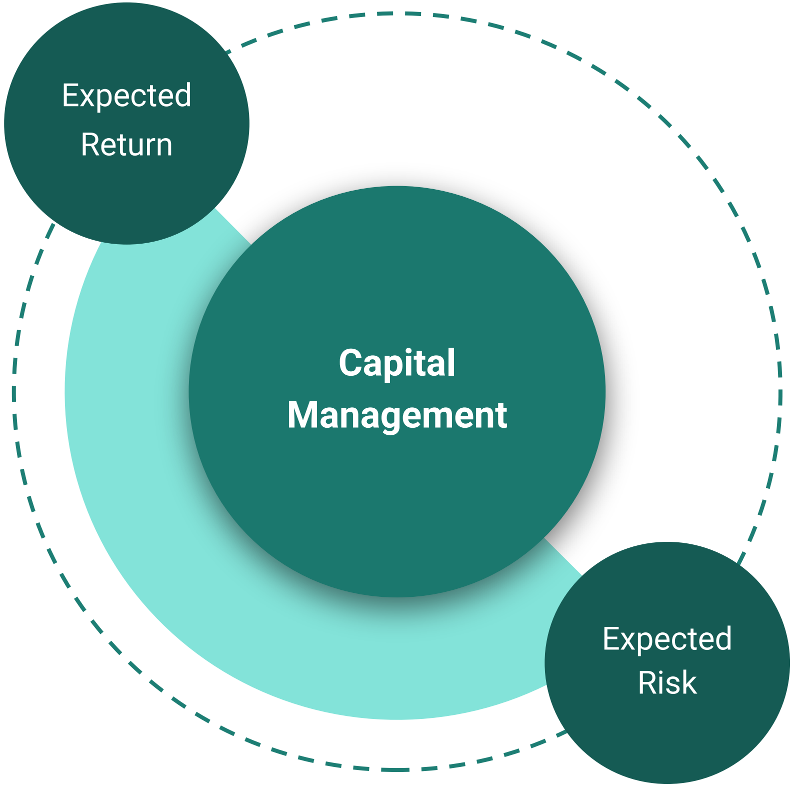
- Sterling ratio is based on common sense understanding of risk:
  - Sterling ratio = (Average return - risk free rate)/average of the most significant drawdowns
  - Drawdown is the difference between a peak and a trough of a curve over a period of time
  - Used by commodity trading advisors
  - Several variations on the ratio used

Section 3

Designing Trading Algorithms

- Expected return
- Expected risk
- Capital management

The three key elements in designing a trading strategy



Key statistical concept: the law of large numbers

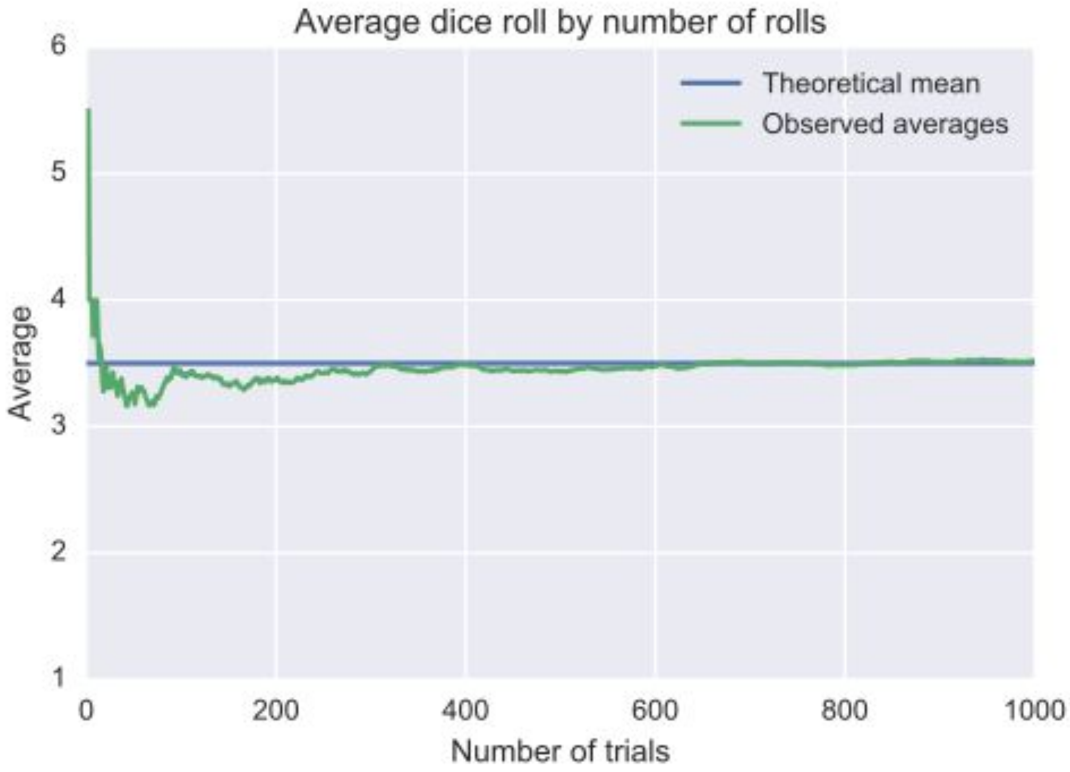


Image source: Wikimedia Commons

Expected return: an arithmetic mean

- Conventional return measurements are incomplete and misleading
    - Reward/Risk: Buy a lottery ticket for the best ratio
    - Win ratio: Many small winning trades can be wiped out by one losing trade
- Expected return: probability-weighted average of event payoffs/outcomes
    - Expected return = (prob event $_1$ * event $_1$ payoff) + …… + (prob event $_N$ * event $_N$ payoff)
    - Expected return may be fictional e.g. expected value of a fair die = 21/6 = 3.5
- Probabilities are difficult to estimate except in simple games of chance
    - Even in games of chance, probability of an event is a measure of its long run frequency
    - Probabilities keep changing depending on new information and events
    - Probabilities of unique or uncommon events are hard to estimate

Poll Question: Maximizing the expected return of trades

Say a market maker uses a weighted coin that has a 80% probability of turning up heads. She offers you the chance of a lifetime:

On heads, she pays you 5 times the value of your net worth. If tails, you lose everything you own except the clothes you are wearing.

- *Expected value of net worth = (5* net worth*0.8) + (0 *0.2) = 4 * net worth*
- *Do you take the wager?*
    a. *Yes*
    b. *No*

Poll Answer: Maximizing the expected return of trades

No! Do not make any trade where you risk blowing up your account!

- Expected return relies on the law of large numbers to be realized
- Need to stay solvent to realize long run positive expectation of trades
- “Bulls make money, bears make money but pigs get slaughtered”

Image by Nick Youngson CC BY-SA 3.0

You will almost surely go broke if you try to maximize positive expected return

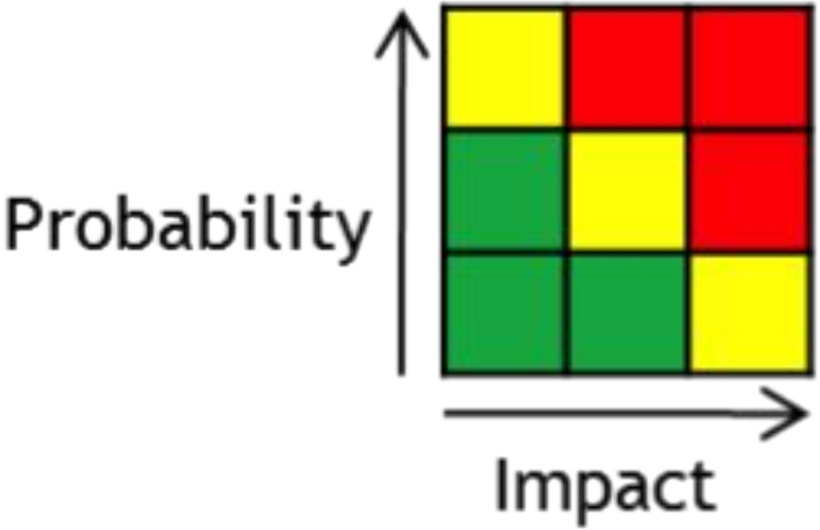
- Assume probability of winning a trade is p where $0.5 < p < 1$
- Probability of winning N trades in succession is p^N
- Therefore probability of losing at least one trade is $1 - p^N$
- If we bet everything we have on each successive trade, we will almost surely go broke:

Probability of bankruptcy = $1 - p^N \rightarrow 1$ since $p^N \rightarrow 0$ as N becomes large

Maximizing the positive expectation of returns risks ruin in the long run

- It is necessary for trades/investments to have positive expected returns
  - Take the opposite side of negative expectation trades if possible or sit in cash
  - Finding trades with positive expectation is necessary but not sufficient
- But how much capital do we allocate to trades with positive expectation?
  - Stake too much, you risk making huge losses
  - Stake too little, you waste a favorable opportunity

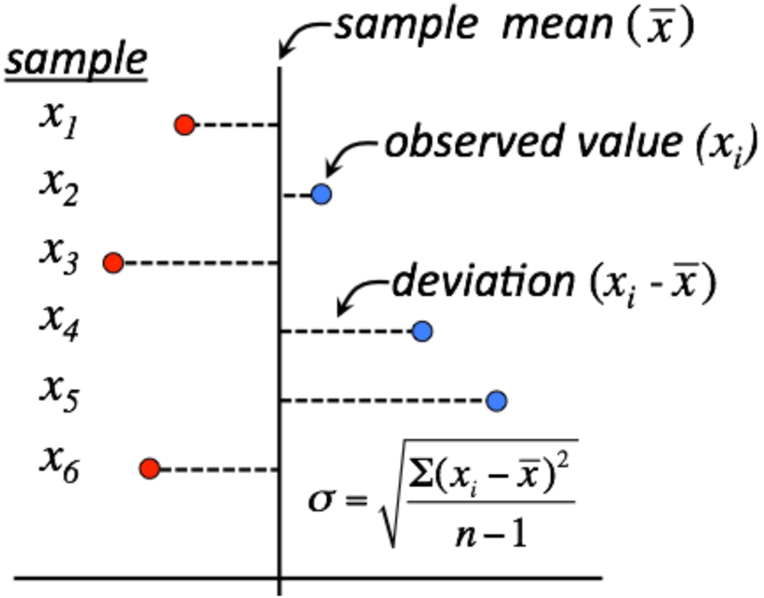
Common sense risk: chance of losing your capital



Risk is the probability of losing some or all of your capital over a period of time

Image source: Wikimedia Commons

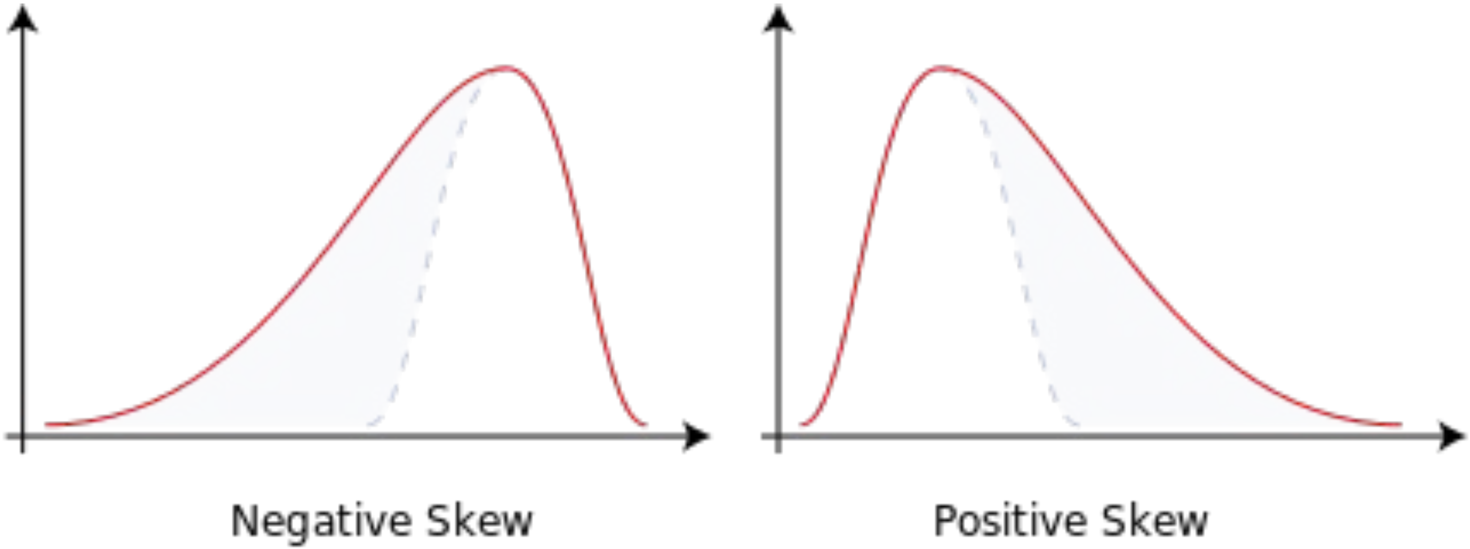
Volatility and beta are nonsensical measures of risk



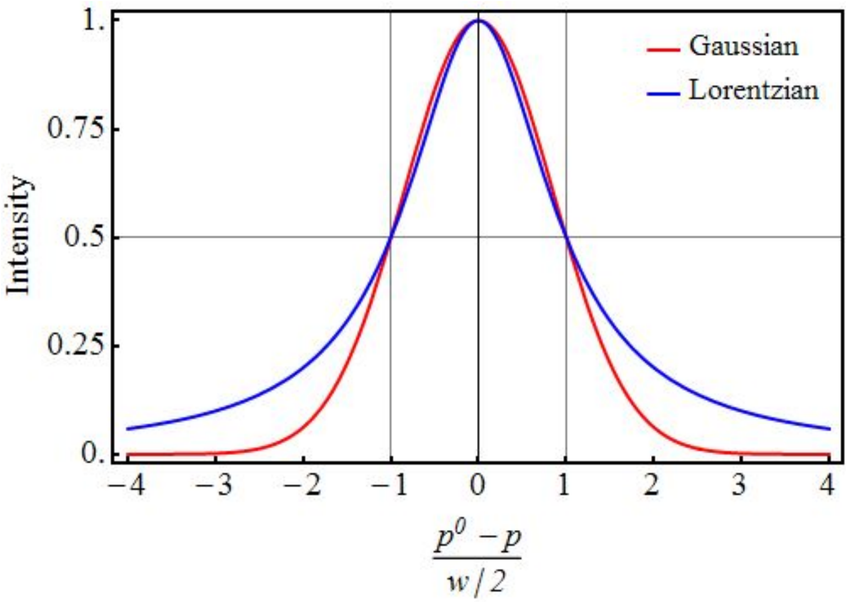
- Volatility treats profits that exceed expectations as risk, which is ridiculous
  - If positive deviations are a risk for you, please transfer them to me - for free!
- Steady losses can have a much lower volatility than irregular profits
  - Constant losses have no volatility!

Volatility is an absurd measure of asymmetric risk

- Volatility ignores the skewness of risks, which is absurd
  - Volatility will be the same for both distributions above
- By ignoring skewness, volatility miscalculates asymmetric risks
  - All things being equal, investors prefer positively skewed returns



Volatility is a disastrous measure of tail risks



- Unless returns are approximately normal, volatility does not capture risks posed by outlier or tail events
  - Known fact that all market return distributions have fat-tails (kurtosis)
  - Many fat-tailed distributions do not have finite variances

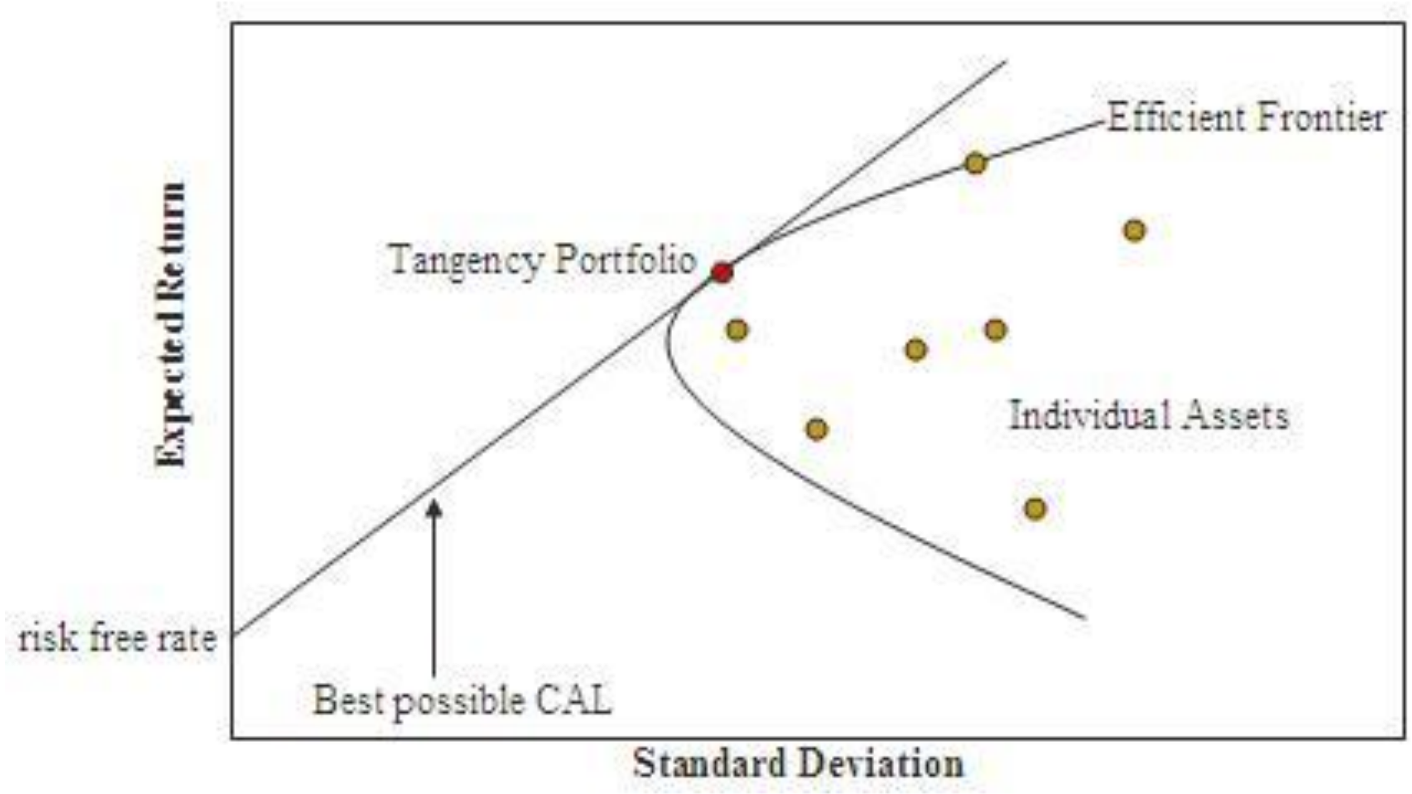
Volatility and beta are bad measures of dynamic risk

- Estimates of volatility and beta are based on historical data
  - Misleading since volatility and beta change over time, sometimes abruptly
  - Changes to beta may be nonlinear
- Portfolio diversification has severe limitations
  - In the 2008 market crash, all equities were positively correlated and approached 1
  - In times of market stress, hedging with options and futures are better

Modern portfolio theory (MPT)

- Developed by Harry Markowitz in the 1950s. Key theme is quantification of the benefits of diversification using correlation of returns
  - Trade-off between return (mean) and risk (variance). Ignores other moments
  - Return distributions are normal and relationships are linear
  - Assumes markets are efficient, investors are rational and risk averse
- Student, William Sharpe, used MPT to derive CAPM discussed earlier

Markowitz’s capital allocation strategy



Issues with MPT

- Unrealistic assumptions about normality and linear relationships
  - Ignores skewness and kurtosis of asset price returns
  - Diversification is reduced or eliminated in extreme events
- Portfolio weights can be extremely sensitive to estimates of returns, variances and covariances
  - Small changes in return estimates can completely change the composition of the optimal portfolio
  - Fat tailed distributions can introduce large errors in covariances among securities
- MPT portfolio is riskier with suboptimal returns
  - Diversification leads to “diworsification”
  - Buffet has called it a lot of nonsense

What does race track betting and information theory have in common?

Image copyright: Wikimedia Commons

The Kelly Criterion: maximize E log W

- Trader's challenge:
  - If we have a trading strategy with positive expectation (edge), how do we maximize our future wealth without the risk of losing all or most of our capital?
- Kelly's solution: maximize E log W
  - Maximize the expected compound growth rate of capital (expected logarithm of wealth)

Kelly formula: edge over odds

- Kelly position size = (expected value)/(fractional odds)
  - Fractional odds = average profit/average loss per unit staked
  - Simple version of formula for binary outcomes. Helps with risk management
- Mathematically proven to outperform any other capital allocation strategy
  - Used by great investors and traders like Thorpe, Buffet and Simons
  - It is a travesty that it is not taught in academia or professional schools

Poll Question: How much capital to allocate to trades to maximize wealth?

The same market maker gives you another weighted coin that has a 55% probability of turning up heads. She offers you an infinite series of trades:

- *On heads, you get two times your stake. On tails, you lose your entire stake. How much capital do you allocate to maximize your capital in the long term?*
  a. *55%*
  b. *20%*
  c. *10%*
  d. *5%*

Poll Answer: Position size that maximizes capital growth and terminal wealth

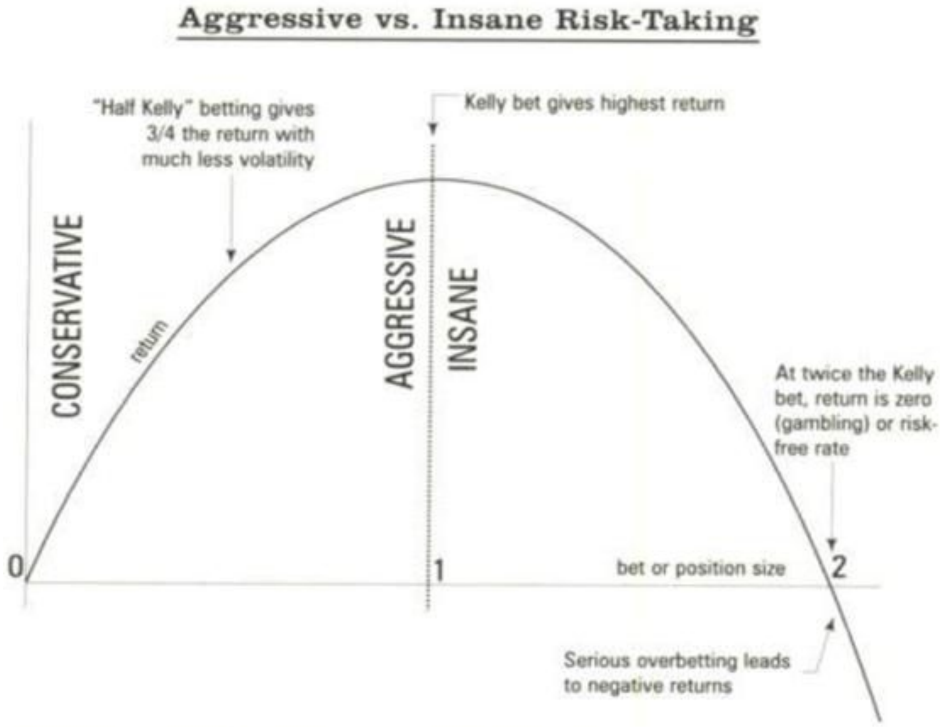
Let's solve for the Kelly position size:

- Fractional odds = average profit/average loss = (2-1)/1 = 1
- Expected return (edge) = (1*0.55 - 1*0.45) = 0.10
- Kelly position size = 0.10/1 = 0.10 or 10% of your capital

Exercise: Position sizing: the sufficient condition

- Mathematically indisputable that the Kelly Criterion maximizes terminal wealth in the shortest amount of time without the risk of going broke
    - Exponential growth since profits are reinvested
    - Multi-period, myopic trading strategy
- Kelly position sizing has built-in risk management
    - Kelly position size is a fraction of your risk capital
    - Position size becomes small as losses accumulate
- Click here for Colab

Risk is over-allocating capital to positive expectation trades



Source: "Fortune's Formula," William Poundstone

Challenges of Kelly position sizing

- Unfortunately, capital allocation is not that simple
  - Difficult to estimate odds and one’s edge consistently
  - Edge and odds are not static
  - Finite trading time horizon
- Fractional Kelly sizing helps avoid overbetting
  - Hedge against overestimating one’s edge
  - Hedge against mis-estimating event odds
  - Hedge against changing edge and odds

Kelly strategy may not be suitable for all

- Kelly strategy can have highly volatile returns
  - Maximum capital growth comes at a price - higher volatility of returns
- Even positive expectation trades can have a string of losses
  - May impair capital base if position sizes are large
- Other capital allocation strategies may be appropriate for people with different goals or time horizons
  - However, other capital allocation strategies are sub-optimal for growing terminal wealth

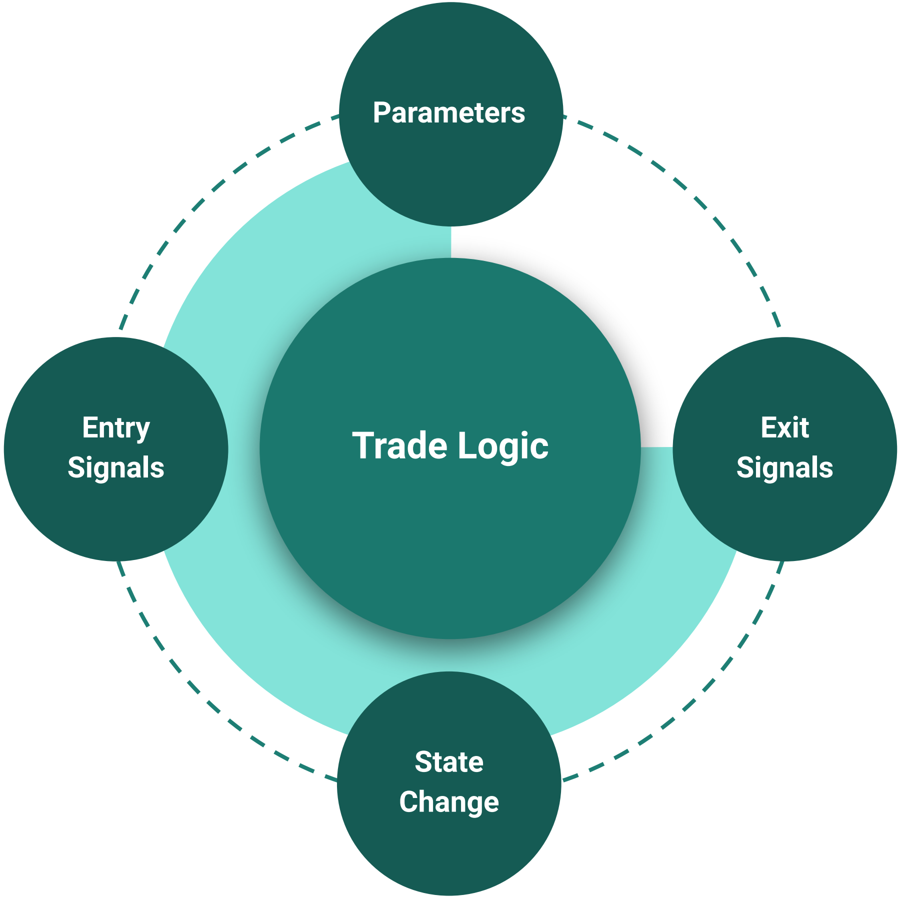
Section 4

Prototyping Trading Algorithms

- Algorithm framework
- Backtesting algorithms
- Backtesting pitfalls

Recall our dual moving average crossover strategy

Architecture of trading algorithms



Exercise: Develop dual moving average algorithm

- Click here for colab document
- Read section on Simple Moving Averages to learn more about the vectorized backtesting of trading strategies

Major pitfalls of backtesting

- Overfitting
  - Creating a overly complex trading strategy to fit historical patterns. Biggest problem with machine learning algorithms. Predictive power of strategy declines dramatically in paper and real-time trading
- Data dredging
  - Mining data exhaustively to find correlations using multiple factors. Correlations will be spurious - based on random noise in historical data which do not persist into the future

Major pitfalls of backtesting

- Hindsight bias
  - Looking at past events with the knowledge of its actual outcomes. Leads you to oversimplify the causal relationships which may not improve your ability to predict future outcomes based on that causal relationship
- Sample selection bias
  - Choosing a non-random historical data sample that does not represent the population distribution or trading environment of interest. Leads to erroneous conclusions.

Avoiding backtesting pitfalls

- Avoid the pitfalls and biases of backtesting by increasing your awareness
- Have results evaluated by impartial colleagues or third-parties
- Count the number of successful and unsuccessful tests performed
- Paper-trading/forward testing

Algorithmic trading: implementation issues

- Many things can go wrong in algorithmic trading:
  - Software bugs
  - Hardware failure
  - Internet connection failure
  - Power failure
  - Other
- Systems need human oversight and a disaster recovery plan
  - My primary risks trading algorithmically are mainly technology/system related
  - Redundant systems are a solution
  - Need to manage costs of redundant systems

Knight Capital Group lost $460 million in less than an hour because of software glitch

Image copyright: Stan Honda

Image copyright: Bloomberg Media