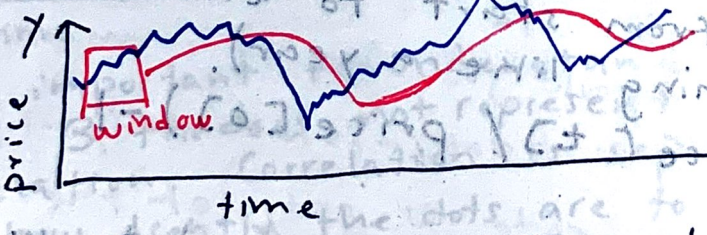
You could expect data like the number of employees, the date/time, the company name, the price of the stock, the company's hometown and more inside a CSV stock data file.

Numpy is sort of a wrapper for underlying C and Fortran code... Pandas may act as a wrapper for Numpy, and thus we may use Numpy arrays and Pandas dataframes very closely and similarly.

Rolling statistics

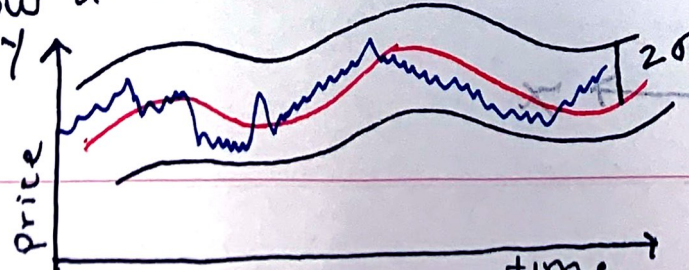
As oppose to global statistics, rolling statistics work based on "windows" or periods of "time".

- Rolling mean
  computes the mean for a window, then we compute the mean for the window + 1 day... and so on. example:



Rolling statistics often later serve as basis for technical indicators

- Bollinger Bands®
  Consists of looking at the recent volatility of the stock and determining that if it has been volatile, then we may discard crossings below and above the mean, and vice versa.



{set limits at ±2σ (std. dev.)}

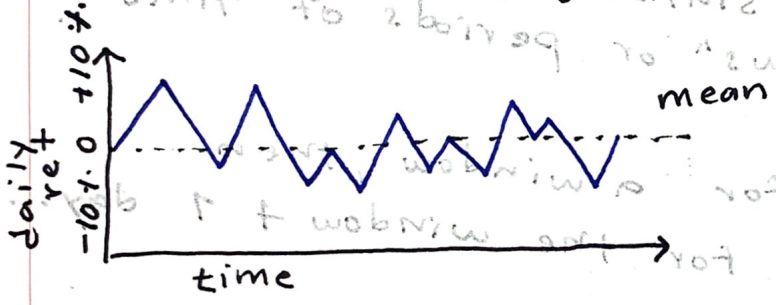
~~It is important to note that~~
The hypothesis with Bollinger bands is that if the price crosses any of the two limits, then it may be a signal to buy/sell.

Daily returns

How much do prices go up or down during a given day?

daily_ret[t] = (price[t]/price[t-1]) - 1

We would produce a graph that would something like the following:



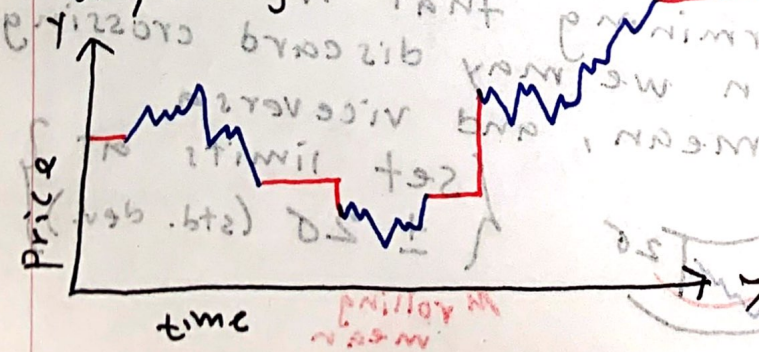
Cumulative returns

Percentage change from start to end of a whole period (something like a year).

cumret[t] = (price[t]/price[0]) - 1

Incomplete data

When we have periods with missing data we consider that the value (say, the price) kept constant (representing we do not know anything about the past/future).



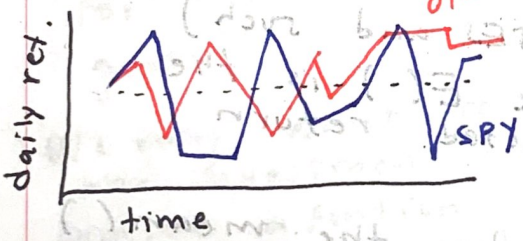
ML for trading

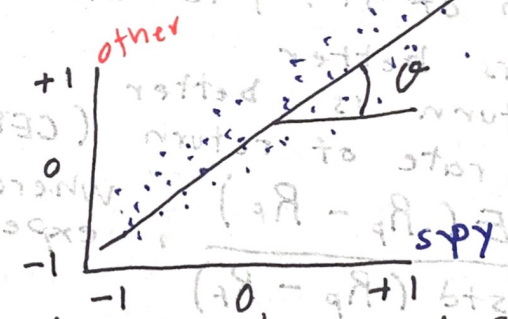
Analyze daily returns

We will have a histogram that represents the daily returns of a stock. We can analyze things like the mean, the std. deviation and kurtosis.

- Kurtosis
  It tells us how different our distribution is from a normal Gaussian distribution.
  If positive, it mens we have "fat tails", where more cases deviate more strongly from the mean, and viceversa also applies. (negative kurtosis, "skinny tails").

We may also analyze daily returns as a scatterplot of two stocks:

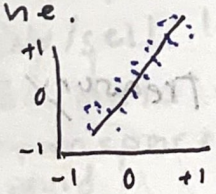




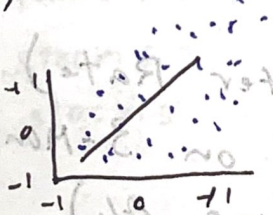
β (slope): tells us how reactive the stock is to the market.

α (distance to the origin): tells us if the stock on average is better/worse performing than the market.

It is important to understand that β does not represent correlation. Correlation is represented by how tightly the dots are to the line.



high-corr.



low-corr.

Portfolio statistics

To calculate how the value of the portfolio changes over time we may do the following (daily):

```
start_val = 1000000
start_date = 2009-1-1
end_date = 2011-12-31
symbols = ['SPY', 'XOM', 'GOOG', 'GLD']
allocs = [0.4, 0.4, 0.1, 0.]
```

```
normed = prices / prices[0]
alloced = normed * allocs
pos_vals = pos_vals.
    sum(axis=1)
```

We may analyze things like the cumulative return, the average return, the std. deviation of the returns and the sharpe ratio.

Sharpe ratio

- Sharpe ratio
  the idea of this ratio is to compare the return of the investment, ~~compared~~ in relation to the risk of it. The SR considers:
  - lower risk is better
  - higher return is better
  - risk free rate of return (CETES and such)

$$SR = \frac{E(R_p - R_f)}{std(R_p - R_f)}$$, where $E(\ )$ is the expected return

In Python, the $E(\ )$ would just be the mean() of the past $(R_p - R_f)$ or (daily_rets - daily_rf).

Traditionally, risk free rate is only of the following:

- LIBOR (London Interbank offer Rate)
- 3mo T-Bill (Interest Rate on 3-Month Treasury bill)
- 0%. (In 2015 the return was 0%.)
- CETES (in Mexico) (Mexican companies...)

A shortcut to find the RF rate is:

$$daily\_rf = \sqrt[252]{1.0 + 0.1} - 1$$

ML for trading

Note: SR can vary widely depending on how frequently you sample.

- SR is an annual measure
- $SR_{annualized} = K \cdot SR$

$K = \sqrt{\#\text{Samples per year}}$

- daily: $K = \sqrt{252}$
- weekly: $K = \sqrt{52}$
- monthly: $K = \sqrt{12}$

Then:

$$SR = \sqrt{252} \cdot \frac{\text{mean}(\text{daily\_ret} - \text{daily\_rf})}{\text{std}(\text{daily\_ret})}$$

sharpe ratio

{ bps = basis point, which is 100th of a percent }

Optimizers

Essentially, fit polynomials to data points (usually containing noise and randomness) to line, whatever the degree may be. We optimize for the derivative of the polynomial (initial guess).

We may want to optimize the allocation of the investment in a portfolio. We would need to provide a funtion that 'describes' the market, provide the intial guess, and optimize it.

Types of funds

- ETF (Electronically traded f.)
  - Buy/sell like stocks
  - Baskets of stocks
  - Transparent
  - Liquid
- Hedge Funds
  - Buy/sell by agreement
  - No disclosure
  - Not transparent
- Mutual Funds
  - Buy/sell at the end of the day
  - Quarterly disclosure
  - Less transparent

ML for trading

How are the different funds compensated?
- ETFs → Expense Ratio (0.01% – 1.00%)
- Mutual Funds → Expense Ratio (0.5% – 3.00%)
- Hedge Funds → "Two and Twenty" (2% of the AUM, plus 20% of the profits)

{AUM: Assets Under Management}

Hedge Fund management

– How funds attract investors —

Who?
- Individual
- Institutions
- Funds of funds

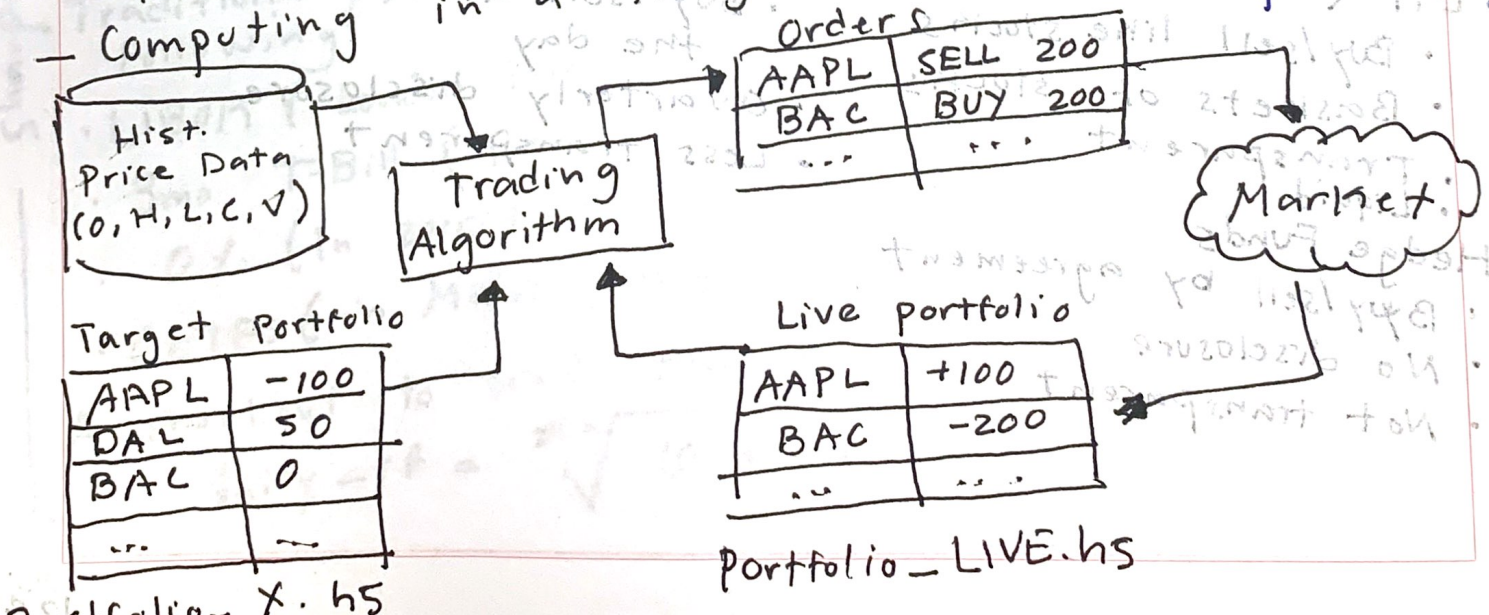
Why?
- Track record
- Simulation + story
- Good portfolio fit

– Goals and metrics —

Goals
- Beat a benchmark (ex. beat SP500)
- Absolute return (provide positive return Long/short)

Metrics
- Cumulative return
- Volatility
- Risk/reward

– Computing in a Hedge Fund — complete



ML for trading

• Target Portfolio

N-day forecast

AAPL	-0.01
BAC	+0.05
...	...

forecast_X.h5

Current Portfolio

AAPL	-100
...	...

Portfolio Optimizer

Hist. Price Data

Risk Constraints

Target Portfolio

AAPL	-100
BAC	0
...	...

• N-day forecast

Information Feed

Hist Price Data

Forecasting Algorithm

(ML model...)

N-day Forecast

AAPL	-0.01
BAC	+0.05
BOA	+0.00
...	...

forecast_X.h5

Market Mechanics

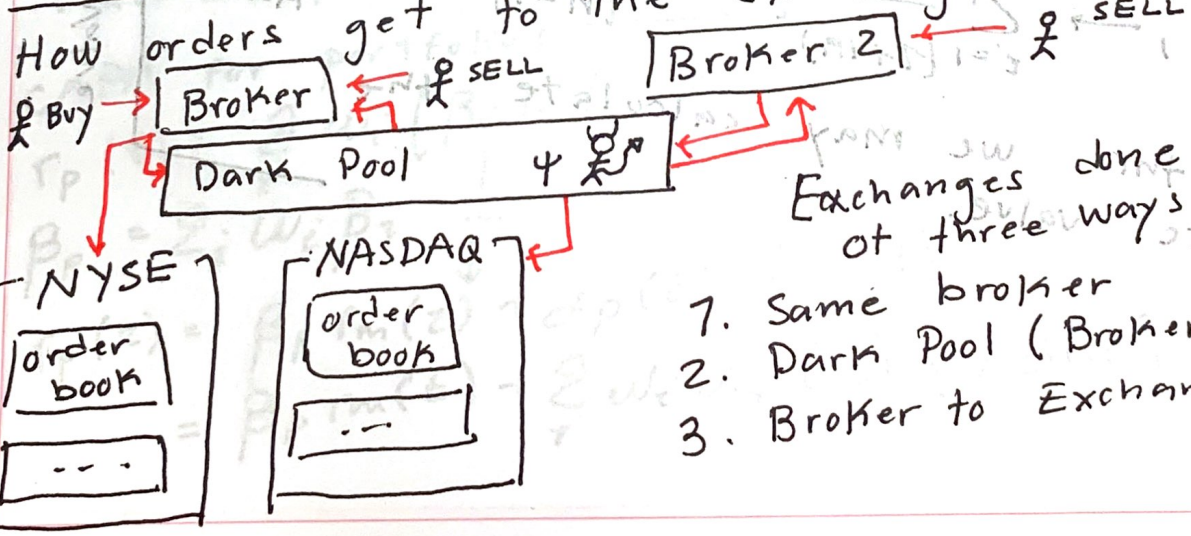
Order
- Buy or Sell
- Symbol
- # shares
- Limit or market
- Price

The Order Book

ex.

BID/ASK	Price	Size
ASK	100.10	100
ASK	100.05	500
BID	99.95	1000
...	...	...

How orders get to the exchange:



Exchanges done in one of three ways.

1. Same broker
2. Dark Pool (Broker to Broker2)
3. Broker to Exchange

ML for trading

Two ways hedge funds exploit market mechanics
A. Order book exploit" (FlashBoys ..)
B. Geografic exploit arbitrage

— Additional order types

Exchanges
- Buy or Sell
- Market, limit

Broker
- stop loss
- stop gain
- trailing stop
- Sell short

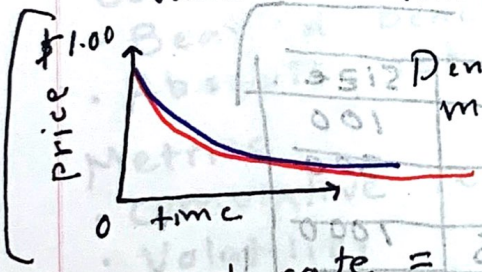
— Value of a company (acumulated dividends)

- Instrinsic value
- Books value
- Market cap

Find "true" value

$$\text{Present Value} = \text{Future Value} / (1 + IR)^i$$

IR = Interest Rate
i = period to be delivered (years)



Pending on the IR, the money will "depreciate" more during any given period

Discount rate = how much an investment should offer based on how risky it is.

$$\sum_{i=1}^{\infty} \frac{1}{n^i} \rightarrow \sum_{i=1}^{\infty} \frac{F.V._i}{(1+IR)^i} = \frac{F.V.}{(n-1)} = \frac{F.V.}{\text{Discount Rate}}$$

And by this we may calculate the Intrinsic value.

ML for trading

"Total assets minus intangible assets and liabilities." [Book Value]

[market cap = #shares · price per share] Market cap.

CAPM (The Capital Asset Pricing Model)

where w_i (weights in the portfolio), and $\sum_i abs(w_i) = 1.0$, then $r_p(t) = \sum_i w_i r_i(t)$ (portfolio return).

- What is the market portfolio? — where $w_i = \frac{mkt\ cap_i}{\sum_j mkt\ cap_j}$

Markets
- US: SP500
- UK: FTA
- Japan: TOPIX

Tend to be cap. weighted

The CAPM equation

$$r_i(t) = \beta_i r_m(t) + \alpha_i(t)$$

r_i = return of given stock
r_m = return of the market
β_i = how much the stock responds to the market (slope)
α_i = "residual"

"The market strongly affects the individual stocks"
CAPM also says that the expected α_i is zero! and random

- CAPM vs Active Management —
  - passive: buy index and hold
  - active: picks stocks
  - > is α random?

- CAPM for portfolios —

$$r_p(t) = \sum_i w_i (\beta_i r_m(t) + \alpha_i(t))$$

$$\beta_p = \sum_i w_i \beta_i$$

$$r_p(t) = \beta_p r_m(t) + \alpha_p(t) \quad \{CAPM\}$$

$$= \beta_p r_m(t) + \sum_i w_i \alpha_i(t) \quad \{Active\}$$

ML for trading

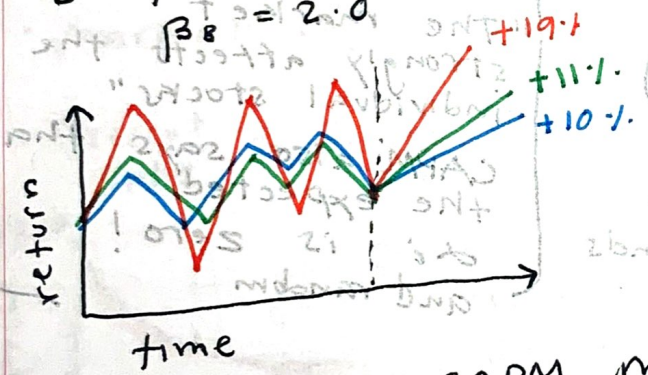
Taking into account the CAPM formula, the only way to beat the market would be to predict β. The Efficient Markets Hypothesis (EMH) says you can't predict the market.

- Arbitrage Pricing Theory (APT)
  - Developed by Stephen Ross in 1976
  - Decompose β into smaller β's for each of the sectors of the market

- Two stock scenario (example)

A: predict +1% over mkt. → long + $50
$\beta_A = 1.0$

B: predict -1% below mkt. → short - $50
$\beta_B = 2.0$



	%	$
r_A	11 %	5.50
r_B	-19 %	-9.50
Total	-4 %	-4.00

Two stock CAPM math

$$r_p = \sum_i w_i (\beta_i r_m + \alpha_i)$$
$$= (w_A \beta_A + w_B \beta_B) r_m + w_A \alpha_A + w_B \alpha_B$$

* From here we could calculate the percentual return of the portfolio.

Continues in all-digital PDF...