Predictive pricing algo for equities

* goal is to make money and learn

Use large data set with available tick info for stocks, apply ML algo to make money

Sharpe = return/risk

Alpha = edge you have that others probably don’t or isn’t common sense

Edge = expected money you made on your trade

Edge retained = look at PNL over 15, 60, 300, 900 seconds

Goal is to back test data gathered online to “fit” a strategy that predicts future prices

Volatility of a stock is how you assess its riskiness – a stock with a volatility of 16% means a 1% daily move, and it scales linearly.

Lets assume S&P has a volatility of 8%. We therefore expect 0.5% daily move, so if you trade spy (which is worth about $270), you expect a $1.35 daily move. Delving further, lets assume a trading day consists of 10 hours. 1 minute, 5 minute, 15 minute fraction of day = 1/6000, 5/6000, 15/6000. Sqrt of those numbers are 0.013, 0.029, 0.05.

With these ratios, we can expect an average move over 1, 5, 15 minutes to be $0.02, $0.04, $0.06.

So, this is a massive cash flow intensive problem and relies on speed and extremely accurate execution.

So project I propose to Vidal-

Find the biggest 200 stocks in the S&P 500. Use “risk” (vol) of S&P as baseline to keep things clean and almost everything will have a higher volatility in the actual index.

For these 200 stocks, find the top 25 correlated products, and include S&P 500.

To simplify, use end of day trading price to find daily moves (4:00PM EST price). Decide whether you want to look at 24 hour move of stocks or 9:30AM-4:00PM EST move (one trading day or 24 hours).

Process of doing this-

* Load last 2 years of data for each stock (opening and closing prices).
* Calculate daily move (decide if you want overnight or trading day, I recommend full 24 hours)
* Figure out a way to find out when the stocks had special events and take those days out of your data (earnings, dividends, foreseeable big price impacts like new chairman vote or anticipated USA elections).

So the way this would look would be M-T-W-Th-F are the trading days, and if apple had earnings on Thursday then you would simply delete that day from data set.

\*note, make sure you know what time of day the events take place, since it could effect the day before or day after (10am event vs 5pm event).

Now the model is built-

* we can recreate a basket of stocks that we expect to mimic apple’s move in price (all 20 stocks contribute a ratio of the total)
* If the basket moves up relative to the stock, we expect the stock to be a good buy
* Similarly, if it moves down relative, we expect it to be a good sell.
* Ideally, we can sell the basket and buy the stock, or visa versa, but because you cannot short sell (sell something without owning it), we will just need to take the long side of the trades

It will be important to note how your trades performed, because if the market goes down, any long strategy will lose money. However, if you prove a good model where you can mimic stocks through a basket, and it is accurate, we can eventually move to a place where we can short sell and simply capture the profits of the mispricing due to our correlation models.

Apple- 15 stocks with high correlations

Figure out what combination of the stocks (needs to add to 100% of apple) makes the highest correlation