

Types of Alpha (Part 1): Data

We don't refer to just price (and maybe volume) data here. When we develop trading strategies using data, we don't add layers and layers of technical indicators that are all based on price and hope some mathematical formulas will output some sort of magical market alpha.

The idea behind trading using data is to find data with some sort of predictive value, analyse them to derive some sort of useful information. We are going to refer to these data as Quality Data.

Note that the term "data" here refers to more than numbers. It could refer to text, images, videos and sound etc.

Quality Data – Data with predictive value on an asset's future price."

First, let's first look at some examples.

Examples

- Collecting satellite images of corn farms. Predicting future harvests based on the colour of the crops.
- Analysing tweets to predict market sentiment.
- Sending someone to record the number of trucks leaving a timber processing factory every day. Thus, estimating supply and sales.

These examples are simplified but they should give you a rough understanding of the types of quality data and how they can be used to discover market inefficiencies.



Satellite imagery analysis of corn farms

Understanding how to use data to discover market inefficiencies

As you can tell by now, the success of a data-based trading strategy is to get quality data. Quality data lead to insights that allow us to predict certain market movement.

Not only does quality data need to have predictive value, it should also not be easily accessed by others. Thus, either the data is restricted to a select few, or that you obtain this data before others.

Once we have quality data, we analyse it using statistics, machine learning or simply using simple logic.

How a retail trader can apply this

Here is a rough guide on analysing and trading using data

1. Get quality data.

This is the hardest part. You need to have an understanding on what moves the market. Once you do, you need to collect the data. Data can be free, bought, or collected on your own.

Big hedge funds spend millions on quality data, retail traders can't do that. You need to be creative here, find quality data in areas that the big boys aren't looking at – frontier markets, exotic products, products with low liquidity etc. Big funds have certain restrictions that retail traders do not have¹. We should play this to our advantage.

To use a business example, Google will not touch a \$20,000 per month revenue opportunity – it is too small. However, this is big enough for a one-man tech company. Similarly, we should target these pockets of alpha in the market, until we grow big enough to play in the same playground as the big boys.

Note that you need at least 2 types of data. 1) Quality data that predicts the price behaviour of asset X and 2) The price of asset X.

2. Clean the data²

Once you get the data, you need to clean it to make sure it is not (too) erroneous. Cleaning data entails ensuring the data is accurate, that there is no missing or wrong data. Cleaning data is tough work.

3. Analyse the data, develop a strategy and backtest it

There are plenty of powerful statistical software/tools at your disposal today. Some of the more popular ones are Python, R and MATLAB. We can use these tools to conduct statistical analysis or build models with machine learning using these data.

If all else fails, use Excel. Excel can't manage big data but once we throw in VBA, it is still a decently powerful tool.

¹ <http://blog.algotrading101.com/design-theories/retail-traders-vs-institutional-traders-hedge-funds/>

² https://en.wikipedia.org/wiki/Data_cleansing

There are 2 broad steps here. First, get trading insights from these data and develop a strategy. Second, backtest this strategy. Make sure you're not overfitting³. Always do out-of-sample/forward tests⁴.

4. Run the strategy on the live market

Trade it! This can be done manually or using algorithms.

Machine learning vs Data?

It is not useful to compare machine learning⁵ with data-based market research. Machine learning is a method to learn and develop models based on data. It is a subset of data-based market research.

Note that data-based market research can be done without machine learning. You don't need to train a machine/model to tell you that if the corn crops are twice as yellow as in the last 10 years, we are getting a bumper harvest.

Big Data or small data?

There is quite a bit of hype about big data. Big data refers to large datasets. Data with millions of elements. Big data can generate some insights that smaller data sets can't. But that doesn't always mean one has more predictive value than the other.

We are indifferent to bigger or smaller data sets. As long as the data has predictive value on some asset's future price, it is quality data.

Useful/Interesting links

- <http://www.businesstimes.com.sg/banking-finance/big-data-is-a-big-mess-for-hedge-funds-hunting-trade-signals>
- <https://www.forbes.com/sites/tomcoughlin/2017/07/24/analysis-of-dark-data-provides-market-advantages/#78206105872b>
- <https://www.thestreet.com/story/13044694/1/how-traders-are-using-text-and-data-mining-to-beat-the-market.html>
- <https://www.ibtimes.co.uk/big-data-hoarding-hedge-funds-managing-twitter-firehose-1572455>
- <https://www.reuters.com/article/us-markets-saft/low-hanging-fruit-from-tweet-sentiment-investing-plucked-james-saft-idUSKCN1162VJ>
- <https://www.ibtimes.co.uk/londons-krzana-storms-into-big-data-league-energy-traders-hedge-fund-clients-1579299>
- <https://www.technologyreview.com/s/602096/crystal-ball-for-corn-crop-yields-will-revolutionize-commodity-trading/>
- <https://www.quora.com/How-is-big-data-analytics-used-for-stock-market-trading-quant-and-Non-quant-trading>

³ <http://blog.algotrading101.com/design-theories/what-is-curve-fitting-overfitting-in-trading/>

⁴ https://en.wikipedia.org/wiki/Walk_forward_optimization

⁵ https://en.wikipedia.org/wiki/Machine_learning