3 Big Mistakes of Backtesting

Here are the 3 mistakes

1. Overfitting
2. Look-ahead bias
3. P-hacking

Overfitting

What is overfitting in trading?

Overfitting in trading is the process of designing a trading system that adapts so closely to historical data that it becomes ineffective in the future.

Is overfitting bad?

Yes.

The past does not predict the future perfectly, especially in financial markets.

Adapting strategies too closely to past data will result in an inflexibility to adapt to the future. Hence, it leads to poor performance in the future.
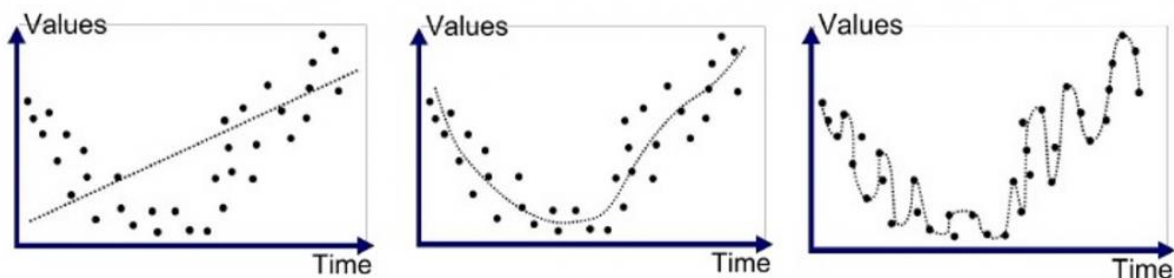
Overfitting in math

How do we overfit?

We overfit by adapting our strategies to noise in the markets instead of signals.

Signals are useful fundamental information. Noise are distractions that don't offer useful qualities.

Example:

The strategy is to pair trade ETF A and ETF B. The performance is not great in the last 3 years. They are not cointegrated well.
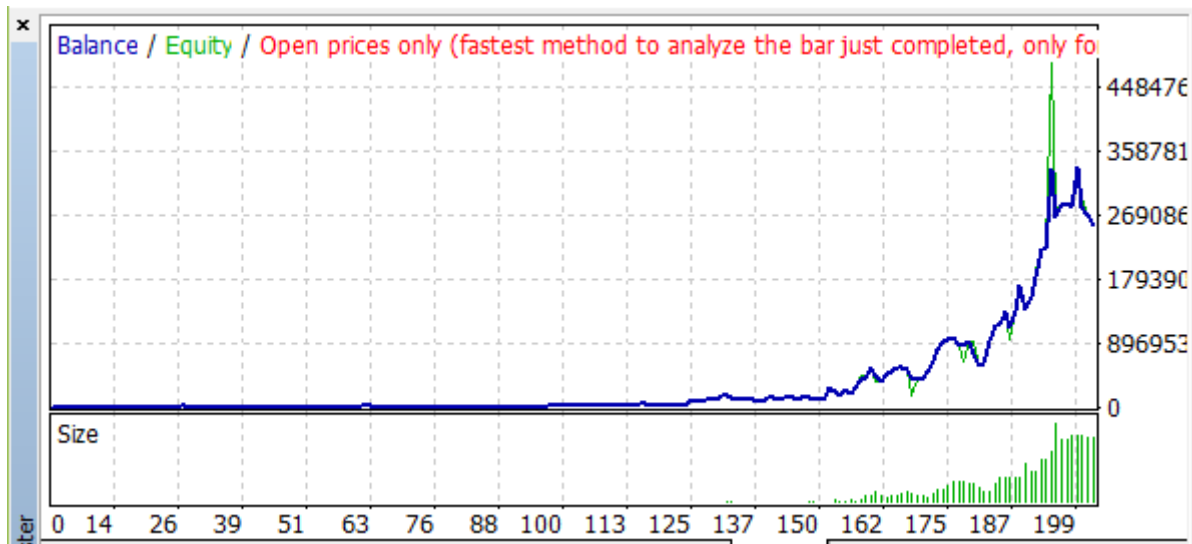
But you tinker with the trading size, take-profit and stop losses and what time of the day you can trade etc.

After the 20th try, you find a combination that makes the last 3 years' performance fantastic.

This is overfitting. This combination is so specific for the last 3 years that will very likely not work for the future.

Here is an example of me overfitting a strategy from AT101:

I kept tinkering with the strategy inputs and had this performance.

And just like that, we have just turned $10K into $2.5 million in 9 months!

If it were so easy, everyone would be a billionaire.

Thought experiment

Here is a thought experiment to reiterate the idea of overfitting

Let's say that you have 5 variables:

- Historical Price Returns
- Volume
- Orderbook data
- Google Trends data
- Reddit r/wallstreetbets data

You run a regression with these variables and found some relationship between these 5 variables and returns.

You think "I found a model that works! Time to be rich!"

Is it really time to be rich? Maybe. But probably not. You've probably overfitted.

Here is the thought experiment, replace these variables with something ridiculous.

E.g. Replace...

- Historical Price returns with the change in the number of people playing Fortnite
- Volume with the number of cars in Europe
- Orderbook data with the temperature in USA
- Google Trends data with number of penguins in Antarctica
- Reddit r/wallstreetbets data with number of people wearing blue shirts in Singapore

Pick a time period and I've sure you can massage the data in a way to find some sort of relationship.

This model is most probably not profitable moving forward.

Look-ahead bias

What is look-ahead bias in trading?

Look-ahead bias occurs when your strategy is trading based on information that it is not supposed to know (i.e. from the "future").

How does look-ahead bias occur?

Let's illustrate this with an example.

Here is Amazon's stock performance from 2013 to 2020.

Wow, it is trending up rather smoothly. I'll design a trading model that invests in Amazon as it trends up.

I then test my trading model on this same dataset (2013 to 2020).

To my non-surprise, the model performs brilliantly and I make a lot of hypothetical monies. You don't say!

To illustrate, when the backtesting is running trades in 2013, the strategy took into account what Amazon's 2014 to 2020 stock behaviour would be like.

This is because we analysed 2013 to 2020 data when designing the trading model (that's how we determine the strategy will be a trending one).

The model is said to have "looked ahead" into the future.

Thus, there is look-ahead bias in our model. We built a model based on data we were not supposed to know.

Is look-ahead bias bad?

Yes.

Backtests with look-ahead bias will hold little meaning.

P-Hacking

P-hacking is also known as Data dredging, data fishing, data snooping, and data butchery.

What is p-hacking?

P-hacking is the process of finding misleading patterns in the data.

It is done by testing many pattern variations and only focusing on the positive finds while ignoring the negative ones.

P-hacking is quite similar to overfitting. Compared to overfitting, p-hacking usually involves mixing and matching data from different assets, market info, economic data, alternative data etc. Overfitting usually refers to adapting to past data.

How does p-hacking bias occur?

I know the above definition is not very clear so let's look at an example.

Example:

You want to run a pairs trading strategy.

1. You have price data on 5000 stocks from 2017 to 2019.

2. With 5000 stocks, you can create 12497500 possible pairs of stocks[1].

3. You use Python to scan them all and find that 50 out of those 12497500 pairs cointegrate well from 2017 to 2019.

---

[1] https://www.hackmath.net/en/calculator/n-choose-k?n=5000&k=2&order=0&repeat=0

4. You then claim that you found 50 working strategies.

This is p-hacking.

The mistake here is that those 50 pairs probably didn't cointegrate well before 2017 and won't cointegrate well after 2019.

It is like a coincidence that those pairs did well from 2017 to 2019.

If you change the date period to any other period, you will find other sets of pairs that do well.

The next time you see some strategies published in academic papers or online. Do think if those strategies consist of these 3 mistakes!

Most beginners commit these 3 mistakes and get over-enthusiastic about their backtests. They then trade their strategy live and lose money.