Trading Algorithms

BACKTESTING

Lecturer: Reza Entezari-Maleki

entezari@iust.ac.ir

School of Computer Engineering

Iran University of Science and Technology

Outlines

- Definition and importance
- Performance measures
- Backtesting biases and challenges
- When Not to Backtest a Strategy
- Will a Backtest be Predictive of Future Returns?

Definition and importance

- **Backtesting is the general method for seeing how well a strategy or model would have done ex-post**.
- It involves applying a strategy or predictive model to historical data to determine its **accuracy**.
- Backtesting assesses the viability of a trading strategy by discovering how it would play out using historical data.
- If Backtesting works, traders and analysts **may have the confidence** to employ it going forward.
- The **underlying theory** is that any strategy that worked well in the past is likely to work well in the future, and conversely, any strategy that performed poorly in the past is likely to perform poorly in the future.

Definition and importance …

- When testing an idea on historical data, it is beneficial to **reserve a time period** of historical data for testing purposes. If it is successful, testing it on alternate time periods or **out-of-sample data** can help confirm its potential viability.
- Backtesting can be used to test and compare different trading strategies without the need to risk capital.
- Ideally, our Backtesting program can be transformed into an **automated execution program** by the push of a button.
- A Backtest should consider **all trading costs**, however insignificant, as these can add up over the course of the Backtesting period and **drastically affect the appearance of a strategy's profitability**.

Definition and importance …

- If you have developed a **strategy from scratch**, you would certainly want to know how it has performed.
- Even if you read about **a strategy from a publication**, and you trust that the author did not lie about its stated performance, it is still imperative that you independently Backtest the strategy.
- Backtesting a published strategy allows you to conduct true out-of-sample testing in the period following publication. If that out-of-sample performance proves poor, then one has to be concerned that the strategy may have worked only on a limited dataset.
- Many authors will claim in their articles that the Backtest results were verified with out-of-sample data. But, actually, if the out-of-sample testing results were poor, the authors could have **just changed some parameters**, or they could have tweaked the model substantially so that the results look good with the out-of-sample data.

Definition and importance …

- Often, the **profitability of a strategy depends sensitively on the details of implementation**.
  - Are the stock orders supposed to be sent as market-on-open orders or as market orders just after the open?
  - Are we supposed to use the bid or ask price to trigger a trade, or are we supposed to use the last price?
  - Have we taken into account the fact that some stocks were hard to borrow and cannot easily be shorted at any reasonable size?
  - In Backtesting an intermarket pair-trading strategy in futures, have we made sure that the closing prices of the two markets occur at the same time?

Definition and importance …

- Once we have implemented every detail of a strategy as a Backtest program, we can then **put them under the microscope** and **look for pitfalls in the Backtesting process or in the strategy itself**.
- We often can find ways to **refine** and **improve the strategy** to make it more profitable or less risky.
- If the results of the Backtest are not good enough, we can **modify our hypothesis and repeat the process**.
- A Backtest is usually coded by a programmer **running a simulation** on the trading strategy.
- It is also essential that the model is tested across many different market conditions to assess performance objectively. Variables within the model are then tweaked for optimization against several different Backtesting measures.

Performance measures

- There are some of the starting metrics that will help an investor measure the performance of her/his own strategy.
- In general the metrics cover **two important aspects of a strategy**: how your portfolio value changed and the risk in achieving those gains/losses. Also there is a **third category** that combines both of above-mentioned aspects.
- By understanding these two areas, it will help the investor identify weaknesses and strengths of a strategy.
  - **Metrics focused on P/L**
    - The metrics in this category all tell you something about how much money you made (or lost) with a particular strategy.
  - **Metrics focused on risk**
    - Equally important to seeing massive returns, is understanding the probability (risk) of the strategy losing money in the future.
  - **Combined risk and reward metrics**
    - There are then a third type of metrics which give you a combined view of the returns and risk in one metric.

Performance measures …

➢ **Return on Investment (ROI),** also known as **Cumulative Return**

- The annual return is the return that an investment provides over a period of time, expressed as a time-weighted annual percentage. Sources of returns can include dividends, returns of capital and capital appreciation.

$$\mathrm{ROI} = \frac{\text{Net Return on Investment}}{\text{Cost of Investment}} \times 100\%$$

$$\mathrm{ROI} = \frac{\mathrm{FVI} - \mathrm{IVI}}{\text{Cost of Investment}} \times 100\%$$

where:

$\mathrm{FVI} = \text{Final value of investment}$

$\mathrm{IVI} = \text{Initial value of investment}$

$$\mathrm{ROI} = \text{Capital Gains\%} - \text{Commission\%} + \text{Dividend Yield}$$

Performance measures …

- **Example (ROI):** An investor purchases property A, which is valued at $500,000. One year later, the investor sells the property for $1,000,000. Compute the ROI metric for this investment.
  - **Solution:** ROI = (1,000,000 – 500,000) / (500,000) = 1 or 100%

- **Example (ROI):** Assume an investor bought 1,000 shares of the hypothetical company A at $10 per share. One year later, the investor sold the shares for $12.50. The investor earned dividends of $500 over the one-year holding period. The investor spent a total of $125 on trading commissions in order to buy and sell the shares. Compute the ROI metric for this investment.
  - **Solution:**

$$\begin{aligned}\mathrm{ROI} &= \frac{(\$12.50 - \$10) \times 1000 + \$500 - \$125}{\$10 \times 1000} \times 100 \\ &= 28.75\%\end{aligned}$$

Performance measures …

- **Compound Annual Growth Rate (CAGR),** also known a **Compound Annual Return**
  - A compound annual growth rate (CAGR) measures the rate of return for an investment — such as a mutual fund or bond — over an investment period, such as 5 or 10 years.

$$\mathrm{CAGR} = \left(\frac{V_{\mathrm{final}}}{V_{\mathrm{begin}}}\right)^{1/t} - 1$$

CAGR = compound annual growth rate
V_{begin} = beginning value
V_{final} = final value
t = time in years

Performance measures …

- **Example (CAGR):** Imagine you invested $10,000 in a portfolio with the returns outlined below:
  - From Jan. 1, 2018, to Jan. 1, 2019, your portfolio grew to $13,000 (or 30% in year one).
  - On Jan. 1, 2020, the portfolio was $14,000 (or 7.69% from January 2019 to January 2020).
  - On Jan. 1, 2021, the portfolio ended with $19,000 (or 35.71% from January 2020 to January 2021).
- **Solution:**

CAGR = (($19,000/$10,000)^(1/3) – 1) * 100 = 23.86%

Performance measures …

- **Profitability**
  - Two metrics Gross Profit and Net Profit can be defined to show the profitability of a strategy. Also, we can compute their corresponding ratios named Gross Profit Margin Ratio and Net Profit Margin Ration.

Gross profit = Total revenue – Cost of goods sold

Net profit = Gross profit – Expenses

Gross Profit Margin Ratio = (Gross Profit/Number of Sales) * 100

Net Profit Margin Ration = (Net Income/Number of Sales) * 100

Performance measures …

Maximum Drawdown

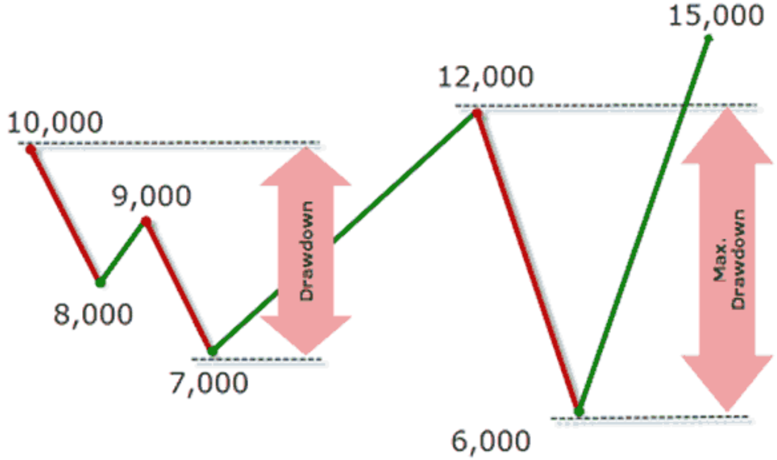
- The maximum drawdown is a financial indicator that shows how much the value of an investment has lost from its last peak or maximum value. It is expressed as a percentage.

MD = (LP - PV) / PV × 100%

MD: Maximum drawdown, in percent;

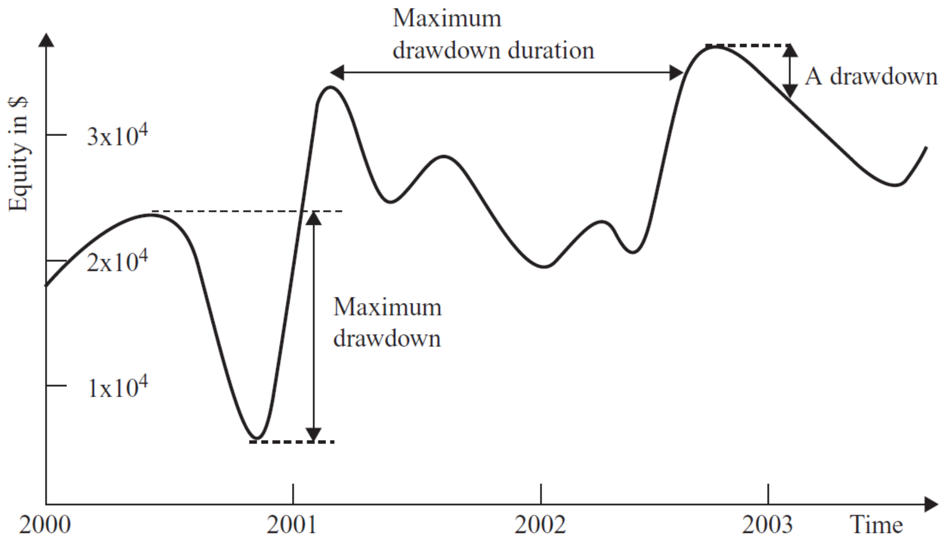
LP: Lowest value after peak value; and

PV: Peak value.



Performance measures …

- **Maximum Drawdown Duration,** also known as **Drawdown Period**
  - The drawdown duration is the length of any peak to peak period, or the time between new equity highs. The max drawdown duration is the worst (the maximum/longest) amount of time for an account to raise back to its peak level after a loss.



Performance measures …

➢ Annualized Actual Volatility (AAV)

- Annualized Actual Volatility is measured as annualized standard deviation of the continuously compounded daily returns of the asset. It is also often referred to as **realized**, **historical**, or **actual**, **asset volatility.**

R_i = Daily portfolio return on the i^{th} day

R_{av} = Mean return in n days

Variance = $\sum (R_{av} - R_i)^2 / n$

Daily volatility = standard deviation = $\sqrt{(\sum (R_{av} - R_i)^2 / n)}$

Annualized volatility = = $\sqrt{252} * \sqrt{(\sum (R_{av} - R_i)^2 / n)}$

Performance measures …

Beta (β)

- Annualized Beta (β) is a measure of the volatility—or systematic risk—of a security or portfolio compared to the market as a whole (usually the S&P 500).

- Beta data about an individual stock can only provide an investor with an approximation of how much risk the stock will add to a **diversified portfolio**.

- For beta to be meaningful, the stock should be related to the benchmark that is used in the calculation. The S&P500 has a beta of 1.0.

- **Beta vs Volatility**
  - Beta compares the change in a stock's price with the market, while implied volatility forecasts the future performance of a stock price.

Performance measures …

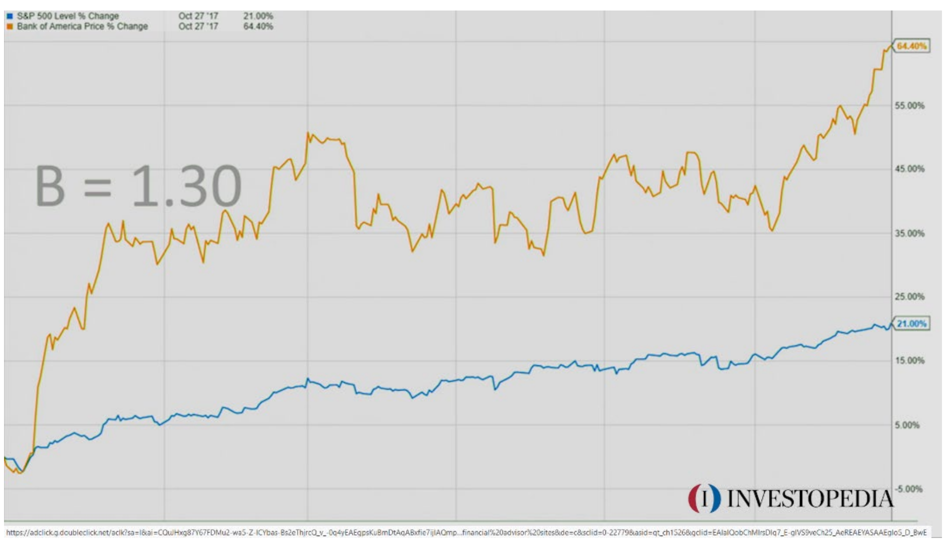
- **Beta (β)**

$$\text{Beta coefficient}(\beta) = \frac{\text{Covariance}(R_e, R_m)}{\text{Variance}(R_m)}$$

where:

R_e = the return on an individual stock

R_m = the return on the overall market



Performance measures …

➢ Value at Risk (VaR)

- The maximum loss in a given holding period to a certain confidence level.
- It is a measure of the risk of loss for investments. It estimates **how much a set of investments might lose (with a given probability)**, *given normal market conditions*, in a set **time period such as a day**.
- VaR modeling determines the potential for loss in the entity being assessed and the probability that the defined loss will occur.
- This metric can be computed in <u>several ways</u>, including the historical, variance-covariance, and Monte Carlo methods.
- Investment banks commonly apply VaR modeling to firm-wide risk due to the potential for independent trading desks to unintentionally expose the firm to highly correlated assets.

Performance measures …

- **Conditional Value at Risk (CVaR),** also known as **Expected Shortfall**
  - It is a risk assessment measure that quantifies the amount of tail risk an investment portfolio has.
  - CVaR is derived by taking a **weighted average** of the extreme losses in the tail of the distribution of possible returns, beyond the Value at Risk (VaR) cutoff point.
  - Conditional value at risk is used in portfolio optimization for effective risk management.
  - The use of CVaR as opposed to just VaR tends to lead to a more conservative approach in terms of risk exposure.
  - The choice between VaR and CVaR is not always clear, but **volatile and engineered investments can benefit from CVaR** as a check to the assumptions imposed by VaR.

Performance measures …

- **VaR vs. CVaR**
  - if VaR(95) = 3%
    - **5% chance to lose 3% or more on a given day**
  - if CVaR(95) = 4.5%
    - 5**% chance to lose <u>in average</u> 4.5% on a given day**
  - CVaR gives us an average expected loss, while VaR gives us a range of potential losses, Loss accurate and **lower approximation of risk**

Performance measures …

➢ Information Ratio

- Information ratio is the measure to use when you want to assess a long-only strategy.
- The benchmark is usually the market index to which the securities you are trading belong.
- For example, if you trade only small-cap stocks, the market index should be the Standard & Poor's small-cap (S&P 600) index rather than the S&P 500.
- **If you are trading just gold futures, then the market index should be gold spot price, rather than a stock index.**

$$Information\ Ratio = \frac{\text{Average of Excess Returns}}{\text{Standard Deviation of Excess Returns}}$$

$$\text{Excess Returns} = \text{Portfolio Returns} - \text{Benchmark Returns}$$

Performance measures …

- **Sharpe Ratio**
  - The Sharpe ratio is one of the most famous ratios out there because it captures two things: (1) The amount of **reward** you are getting, i.e. your returns, and (2) how much **risk** you are taking on.
  - The Sharpe ratio is actually a special case of the information ratio, so that the benchmark to use is always the risk-free rate.
  - It then takes this and makes it into a metric that can be verbalized as **for every unit of reward, how much risk am I taking on?**

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio

R_f = risk-free rate

σ_p = standard deviation of the portfolio's excess return

The Sharpe ratio is a ratio of reward vs risk, so anything greater than one is phenomenal. This means that for every unit of risk, you are getting more reward.

Performance measures …

- The ratios use something very different from CAGR. Instead of using account values (your total asset value at specific times), it uses returns to ultimately calculate risk (the standard deviation in returns) and reward (the average in returns).

- **Standard deviation** is a measure of **how spread out your data is**. In the case of returns, how spread out your returns is, means how volatile your account value is.

- **The more spread out your returns are, the more likely you are to run into losses, and therefore more emotion and therefore more volatility.**

- Our Risk Free Rate is our benchmark for how much we would make if we chose a "**risk free investment**," as typically this is the treasury yield, or can also be the estimated percentage that you would make if you did a buy and hold, or whatever you define as "risk free".

Performance measures …

- The various Sharpe ratios are going to be different depending on the sample time and sample size.

- Every minute, a stock might move 0.01% but every day, stocks might move 1 to 2%. Well, clearly the Sharpe ratios are going to be very different.

- That's why we have to **annualize** them all so that they are consistent and easily comparable. This makes it possible for us to compare strategies that are running on minute data, and strategies that run every hour or every day (it really doesn't matter).

- For example, when calculating the Sharpe Ratio using monthly data, the Sharpe Ratio **is annualized by multiplying the entire result by the square root of 12**.

Performance measures …

Sortino Ratio

- The Sortino ratio is almost the exact same as the Sharpe ratio, except one difference: instead of analyzing ALL the standard deviations of returns, we only care about the negative returns.

- Why do we do this? Well I don't care if my model necessarily makes 25% and then 2% the next day, they're all wins in my book, what I care more about is that I don't want to be losing 25% one day and then 2% the next day, I'd much rather only lose 2% maximum.

$$\text{Sortino Ratio} = \frac{R_p - r_f}{\sigma_d}$$

where:

R_p = Actual or expected portfolio return

r_f = Risk-free rate

σ_d = Standard deviation of the downside

Performance measures …

When would I use Sharpe vs Sortino?

- No one ratio or metric tells the **whole story**.

- By comparison, the Sharpe ratio treats upside and downside risks in the same way. It means that even those investments that produce gains are penalized, which should not be the case.

- Therefore, the Sortino ratio should be used to assess the performance of high volatility assets, such as shares. In comparison, the Sharpe ratio is more suitable for analyzing low volatility assets, such as bonds.

- If you care about overall volatility, use Sharpe, if you care about making sure that your model reducing negative downside, then use Sortino.

Performance measures …

Alpha (α)

- Alpha is a measure of the performance of an investment as compared to a suitable benchmark index, such as the S&P 500.

- It is actually the excess return on an investment after adjusting for market-related volatility and random fluctuations.

- An alpha of one (the baseline value is zero) shows that the return on the investment during a specified time frame outperformed the overall market average by 1%.

- A negative alpha number reflects an investment that is underperforming as compared to the market average.

- Active portfolio managers seek to generate alpha in **diversified portfolios**, with diversification intended to eliminate unsystematic risk.

Performance measures …

- **Alpha (α)**

$$\alpha = R - R_f - \beta * (R_m - R_f)$$

- Where:
  - **R** represents the portfolio return
  - **R_f** represents the risk-free rate of return
  - **β** represents the systematic risk of a portfolio
  - **R_m** represents the market return, per a benchmark

- **Example (Alpha):** Assume that the actual return of the fund is 30%, the risk-free rate is 8%, beta is 1.1, and the benchmark index return is 20%. Now compute the alpha metric.

- **Solution:**

Alpha = (0.30 - 0.08) – 1.1 * (0.20 - 0.08)
= 0.088 or 8.8%

Backtesting biases and challenges

- Although Backtesting almost every strategy allows for unique opportunities in committing errors in Backtesting, there are a number of common themes, some generally applicable to all markets, others pertain to specific ones.
- **Backtesting seems easy** given that the trades were made using a computer algorithm in our case, **but there are numerous ways in which it can go wrong**.
- Usually, an erroneous Backtest would produce a historical performance that is better than what we would have obtained in actual trading.
- If one blithely goes ahead and Backtests a strategy without taking care to avoid these pitfalls, the Backtesting will be useless, or worse, it will be misleading and may cause significant financial losses.

Backtesting biases and challenges …

- But even if a Backtest is done correctly without pitfalls and with high statistical significance, it doesn’t necessarily mean that it is predictive of future returns.
- In trading model, traders must avoid **bias**. **Backtesting biases refer to how the results of a trading strategy Backtest can be misleading**.
- The strategy must be tested on several different time periods with an unbiased and representative sample of stocks.
- If a trader were to pick and choose the stocks and time period in which their strategy is Backtested against, the model would be fundamentally flawed.
- While the test may yield positive results, this would only be because the model was created to fit this data perfectly. Therefore, it is essential that different datasets are used throughout the process.

Backtesting biases and challenges …

Look-Ahead Bias

- As its name implies, look-ahead bias means that your Backtest program is using tomorrow’s prices to determine today’s trading signals.
- Or, more generally, it is using future information to make a **prediction at the current time.**
- A common example of look-ahead bias is to use a day’s high or low price to determine the entry signal during the same day during Backtesting.
- **Before the close of a trading day, we can’t know what the high and low price of the day are.**
- Look-ahead bias is essentially a programming error and can infect only a Backtest program but not a live trading program because there is no way a live trading program can obtain future information.

Backtesting biases and challenges …

Look-Ahead Bias …

- How do we avoid look-ahead bias?
- Use **lagged historical data** for calculating signals at every opportunity.
- Lagging a series of data means that you calculate all the quantities like moving averages, highs and lows, or even volume, based on data up to the close of the previous trading period only. (Of course, you needn't lag the data if your strategy enters only at the close of the period.)
- Even with all the care and caution that goes into creating a Backtest program without look-ahead bias, sometimes we may still let some of it slip in.
- **Some look-ahead bias is quite subtle in nature and not easy to avoid**.

Backtesting biases and challenges …

Look-Ahead Bias …

- It is best to do a final checkup of your Backtest program using this method:
- Run the program using all your historical data; generate and save the resulting position to file A (position file is the file that contains all the recommended positions generated by the program on each day).
- Now truncate your historical data so that the most recent portion (say N days) is removed.
- So if the last day in the original data is T, then the last day in the truncated data should be T-N. N could be 10 days to 100 days.
- Now run the Backtest program again using the truncated data and save the resulting positions into a new file B.
- Truncate the most recent N rows of the positions file A, so that both A and B have the same number of rows (days) in them, and the last day in both file A and B should be T-N.
- Finally, check if A and B are identical in their positions. If not, you have a look-ahead bias in your Backtest program that you must find and correct, because the discrepancies in positions mean that you are inadvertently using the truncated part of the historical data (the part that lies ahead of day T-N) in determining the positions in file A.

Backtesting biases and challenges …

- **Insufficient Sample Bias**
  - Many researchers observe a single relationship and draw a conclusion from it. This is statistically unacceptable, of course.
  - In order to form a conclusion about an observation, quite a large number of observations need to be present.

Backtesting biases and challenges …

Data-Snooping Bias (or over-fitting bias) and the Beauty of Linearity

- Data-snooping bias is caused by having too many free parameters that are fitted to random ethereal market patterns in the past to make historical performance look good.
- **These random market patterns are unlikely to recur in the future**, so a model fitted to these patterns is unlikely to have much predictive power.
- The way to detect data-snooping bias is well known: We should test the model on out-of-sample data and reject a model that doesn't pass the **out-of-sample test**.
- But this is easier said than done. Are we really willing to give up on possibly weeks of work and toss out the model completely?
- Few of us are blessed with such decisiveness. Many of us will instead tweak the model this way or that so that it finally performs reasonably well on both the in-sample and the out-of-sample result.
- But, by doing this **we have just turned the out-of-sample data into in-sample data!**

Backtesting biases and challenges …

Data-Snooping Bias (or over-fitting bias) and the Beauty of Linearity …

- One way to overcome this is using the idea of cross-validation, that is, you should select a number of different subsets of the data for training and tweaking your model and, more important, making sure that the model performs well on these different subsets.
- One reason why we prefer models with a high Sharpe ratio and short maximum drawdown duration is that this almost **automatically ensures that the model will pass the cross-validation test**: the *only subsets where the model will fail the test are those rare drawdown periods*.
- There is a general approach to trading strategy construction that can **minimize data-snooping bias**: **make the model as simple as possible, with as few parameters as possible.**
- It should be noted that a model with few parameters but lots of complicated trading rules are just as susceptible to data-snooping bias.

Backtesting biases and challenges …

- **Data-Snooping Bias (or over-fitting bias) and the Beauty of Linearity …**

  - The conclusion is that nonlinear models are more susceptible to data-snooping bias than linear models because nonlinear models not only are more complicated but they usually have more free parameters than linear models.

  - The most basic safeguard against data-snooping bias is to ensure that you have a sufficient amount of Backtest data relative to the number of free parameters you want to optimize.

  - **As a rule of thumb, let's assume that the number of data points needed for optimizing your parameters is equal to 252 times the number of free parameters your model has.**

  - This assumption of proportionality is not based on any survey of the vast statistical literature, but purely **on experience**.

Backtesting biases and challenges …

Data-Snooping Bias (or over-fitting bias) and the Beauty of Linearity …

- A more rigorous method (albeit **more computationally intensive**) to overcome data-snooping bias is to use moving optimization of the parameters.
- In this case, the parameters themselves are constantly adapting to the changing historical data, and data-snooping bias with respect to parameters is eliminated. These kind of models are named parameterless trading models.
- A parameterless model, or a model with no free parameters, does not mean that the model does not contain, for example, any lookback period for calculating trends, or thresholds for entry or exit, because it would be impossible.
- It just means that all such parameters are dynamically optimized in a moving lookback window.
- The advantage of a parameterless trading model is that it minimizes the **danger of overfitting the model** to multiple input parameters (the so-called data-snooping bias). So the Backtest performance should be much closer to the actual forward performance.

Backtesting biases and challenges …

Survivorship Bias

- A historical database of stock prices that does not include stocks that have disappeared due to bankruptcies, delistings, mergers, or acquisitions suffer from the so-called survivorship bias, because **only survivors of those often unpleasant events remain in the database**.
- Backtesting a strategy using data with survivorship bias can be dangerous because it may inflate the historical performance of the strategy.
- Imagine an extreme case: *suppose your model asks you to just buy the one stock that dropped the most in the previous day and hold it forever*.
- In actuality, this strategy will most certainly perform poorly because in many cases the company whose stock dropped the most in the previous day will go on to bankruptcy, resulting in 100 percent loss of the stock position.
- But if your historical data do not include delisted stocks—that is, they contain only stocks that survive until today—then the Backtest result may look excellent.

Backtesting biases and challenges …

Survivorship Bias …

- Survivorship bias is more dangerous to mean-reverting long-only stock strategies than to mean-reverting long-short or short-only strategies.
- This is because, this bias tends to inflate the Backtest performance of a long-only strategy that first buys low and then sells high, whereas it will deflate the Backtest performance of a short-only strategy that first sells high and then buys low.
- Those stocks that went to zero would have done very well with a short-only strategy, but they would not be present in Backtest data with survivorship bias.
- The profitable short momentum trade will tend to be omitted in data with survivorship bias, and thus the Backtest return will be deflated.

Backtesting biases and challenges …

- **An Example of How Survivorship Bias Can Artificially Inflate a Strategy's Performance**

  - Let's say from a universe of the 1,000 largest stocks (based on market capitalization), we pick 10 that have the lowest closing prices at the beginning of the year and hold them (with equal initial capital) for one year.
  - Let's look at what we would have picked if we had **a good, survivorship-bias-free database**.
  - The NaNs indicate those with nonexistent closing prices on 1/2/2002.
  - The Terminal Price column indicates the last prices at which the stocks were traded on or before 1/2/2002.

SYMBOL	Closing Price on 1/2/2001	Closing Price on 1/2/2002	Terminal Price
ETYS	0.2188	NaN	0.125
MDM	0.3125	0.49	0.49
INTW	0.4063	NaN	0.11
FDHG	0.5	NaN	0.33
OGNC	0.6875	NaN	0.2
MPLX	0.7188	NaN	0.8
GTS	0.75	NaN	0.35
BUYX	0.75	NaN	0.17
PSIX	0.75	NaN	0.2188

Backtesting biases and challenges …

➢ **An Example of How Survivorship Bias Can Artificially Inflate a Strategy's Performance …**

- All but MDM were delisted sometime between 1/2/2001 and 1/2/2002.
- The total return on this portfolio in that year was **–42 percent**.
- Now, let's look at what we would have picked if our database had survivorship bias and actually **missed all those stocks that were delisted that year**.
- We would then have picked the following list instead:

SYMBOL	Closing Price on 1/2/2001	Closing Price on 1/2/2002
MDM	0.3125	0.49
ENGA	0.8438	0.44
NEOF	0.875	27.9
ENP	0.875	0.05
MVL	0.9583	2.5
URBN	1.0156	3.0688
FNV	1.0625	0.81
APT	1.125	0.88
FLIR	1.2813	9.475
RAZF	1.3438	0.25

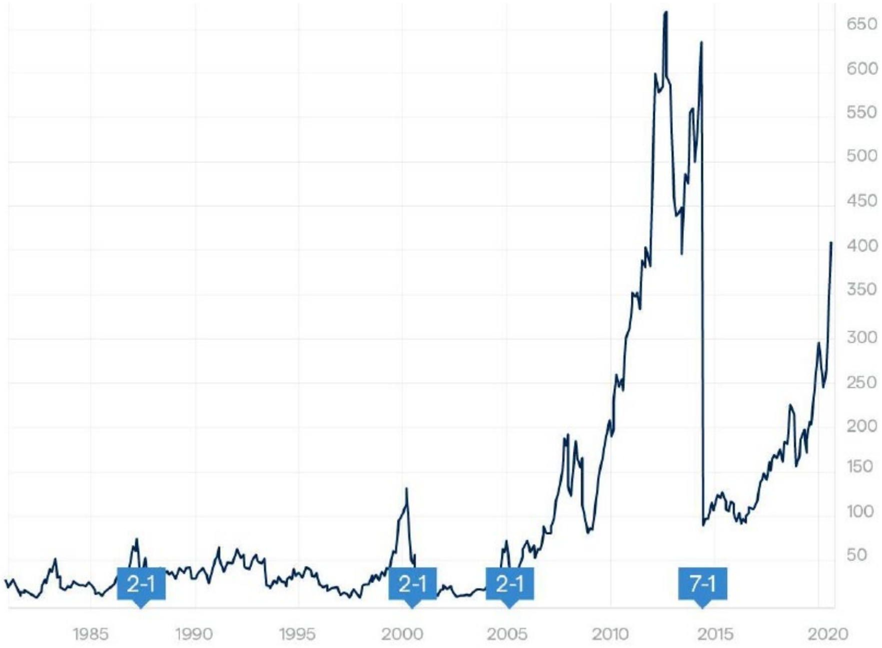
Backtesting biases and challenges …

- **An Example of How Survivorship Bias Can Artificially Inflate a Strategy’s Performance …**
  - Notice that since we select only those stocks that “survived” until at least 1/2/2002, they all have closing prices on that day.
  - The total return on this portfolio was 388 percent!
  - In this example, –42 percent was the actual return a trader would experience following this strategy, whereas 388 percent is a fictitious return that was due to survivorship bias in our database.

Backtesting biases and challenges …

Stock Splits and Dividend Adjustments

- Whenever a company’s stock has an N-to-1 split, the stock price will be divided by N times.
- However, if you own a number of shares of that company’s stock before the split, you will own N times as many shares after the split, so there is in fact no change in the total market value.
- But **in a Backtest**, we typically are **looking at just the price series** to determine our trading signals, not the market-value series of some hypothetical account.
- So unless we back-adjust the prices before the ex-date of the split by dividing them by N, we will see a **sudden drop in price** on the ex-date, and that might **trigger some erroneous trading signals**.



Backtesting biases and challenges …

Stock Splits and Dividend Adjustments …

- If it is a reverse 1-to-N split, we would have to multiply the historical prices before the ex-date by N.
- Similarly, when a company pays a cash (or stock) dividend of $d per share, the stock price will also go down by $d (absent other market movements).
- That is because if you own that stock before the dividend ex-date, you will get **cash (or stock) distributions in your brokerage account**, so again there should be no change in the total market value.
- If you do not back-adjust the historical price series prior to the ex-date, the sudden drop in price may also trigger an erroneous trading signal.
- This adjustment, too, should be applied to any historical data used in the live trading model just before the market opens on an ex-date.

Backtesting biases and challenges …

- **Stock Splits and Dividend Adjustments …**
  - When a company had its stocks **split N to 1** (N is usually 2, but can be a fraction like 0.5 as well. When N is smaller than 1, it is called a *reverse split*) with an ex-date of T, all the prices before T need to be multiplied by 1/N.
  - Similarly, when a company issued a dividend $d per share with an ex-date of T, all the prices before T need to be multiplied by the number (Close(T − 1) − d)/Close(T − 1), where Close(T − 1) is the closing price of the trading day before T.
  - Notice that we adjust the historical prices by a **multiplier instead of subtracting $d** so that the historical daily returns will remain the same pre- and postadjustment.
  - **This is the way Yahoo! Finance adjusts its historical data, and is the most common way.**
  - If you adjust by subtracting $d instead, the historical daily changes in prices will be the same pre- and postadjustment, but not the daily returns.

Backtesting biases and challenges …

Stock Splits and Dividend Adjustments …

- We look at IGE, an ETF that has had both splits and dividends in its history (our reference date is November 2007), so we only consider the prices before this date.
- It had a 2:1 split on June 9, 2005 (the ex-date).
- Let's look at the **unadjusted prices** around that date.

Date	Open	High	Low	Close
6/10/2005	73.98	74.08	73.31	74
6/9/2005	72.45	73.74	72.23	73.74
6/8/2005	144.13	146.44	143.75	144.48
6/7/2005	145	146.07	144.11	144.11

Backtesting biases and challenges …

- **Stock Splits and Dividend Adjustments …**
  - We need to adjust the prices prior to 6/9/2005 due to this split.
  - This is easy: N = 2 here, and all we need to do is to multiply those prices by 1/2.
  - The following table shows the adjusted prices:

Date	Open	High	Low	Close
6/10/2005	73.98	74.08	73.31	74
6/9/2005	72.45	73.74	72.23	73.74
6/8/2005	72.065	73.22	71.875	72.24
6/7/2005	72.5	73.035	72.055	72.055

  - Now, we see that the **adjusted close prices here do not match the adjusted close prices displayed in the Yahoo! Finance table**.
  - The reason for this is that there have been dividends distributed after 6/9/2005, so the Yahoo! prices have been adjusted for all those as well.

Backtesting biases and challenges …

Stock Splits and Dividend Adjustments …

- Since each adjustment is a multiplier, the aggregate adjustment is just the product of all the individual multipliers.
- Here are the dividends from 6/9/2005 to November 2007, together with the unadjusted closing prices of the previous trading days and the resulting individual multipliers
- So the aggregate multiplier for the dividends is simply $0.998618 \times 0.997488 \times \cdots \times 0.997214 = 0.976773$.
- This multiplier should be applied to all the unadjusted prices on or after 6/9/2005.

Ex-Date	Dividend	Prev Close	Multiplier
9/26/2007	0.177	128.08	0.998618
6/29/2007	0.3	119.44	0.997488
12/21/2006	0.322	102.61	0.996862
9/27/2006	0.258	91.53	0.997181
6/23/2006	0.32	92.2	0.996529
3/27/2006	0.253	94.79	0.997331
12/23/2005	0.236	89.87	0.997374
9/26/2005	0.184	89	0.997933
6/21/2005	0.217	77.9	0.997214

Backtesting biases and challenges …

- **Stock Splits and Dividend Adjustments …**
  - You can see that the adjusted closing prices from this calculations and from Yahoo! are the same.
  - Of course, there are other dividend distributions and further splits on IGE from November 2007 on, **so todays Yahoo! table will not look like the table below**.
  - It is a good exercise to check that you can make further adjustments based on those dividends and splits that result in the same adjusted prices as your current Yahoo! table.

Date	Open	High	Low	Close	Volume	Adj Close
6/10/2005	73.98	74.08	73.31	74	179300	72.28
6/9/2005	72.45	73.74	72.23	73.74	853200	72.03
6/8/2005	144.13	146.44	143.75	144.48	109600	70.56
6/7/2005	145	146.07	144.11	144.11	58000	70.38

Backtesting biases and challenges …

- **There are other types of biases that should be taken into consideration whenever a strategy is Backtested.**
  - Venue dependence of currency quotes
  - Short-sale constraints
  - Primary versus consolidated stock prices
  - Optimal period bias
  - Timelessness and temporal bias
  - Holding and transaction costs
  - Issues with market capitalization and Liquidity

Python programming example

First simple strategy with look-ahead bias

When Not to Backtest a Strategy

- We have spent much effort earlier to show the importance of Backtesting every strategy before trading it.
- The fact behind recommending against Backtesting some strategies is that there are some published strategies that are so obviously flawed; it would be a waste of time to even consider them.
- Given what you know now about **common pitfalls of Backtesting**, you are in a good position to judge whether you would want to Backtest a strategy without even knowing the details.
- We will look at a few examples here.

When Not to Backtest a Strategy …

- **Example 1:** A strategy that has a Backtest annualized return of 30 percent and a Sharpe ratio of 0.3, and a maximum drawdown duration of two years.
  - The **low Sharpe ratio** coupled with the **long drawdown duration** indicates that the **strategy is not consistent**.
  - The high average return may be just a fluke, and it is not likely to repeat itself when we start to trade the strategy live.
  - Another way to say this is that the high return is likely the result of data-snooping bias, and the long drawdown duration will make it unlikely that the strategy will pass a cross-validation test.
  - **Do not bother to Backtest high return but low Sharpe ratio strategies.**
  - Also, do not bother to Backtest strategies with a maximum drawdown duration longer than what you or your investors can possibly endure.

When Not to Backtest a Strategy …

- **Example 2:** A long-only crude oil futures strategy returned 20 percent in 2007, with a Sharpe ratio of 1.5.
    - A quick check of the total return of holding the front-month crude oil futures in 2007 reveals that it was 47 percent, with a Sharpe ratio of 1.7.
    - Hence, this trading strategy is not in any way superior to a **simple buy-and-hold** strategy!
    - Moral of the story: We must always choose the **appropriate benchmark** to measure a trading strategy against.
    - **The appropriate benchmark of a long-only strategy is the return of a buy-and-hold position—the information ratio rather than the Sharpe ratio.**

When Not to Backtest a Strategy …

- **Example 3:** A non-linear trading strategy has about 100 free parameters and it generates a Backtest Sharpe ratio of 6.
  - Since this model has a large number of free parameters it is prone to data snooping bias.
  - With at least 100 parameters, we can certainly fit the model to any time series we want and obtain a fantastic Sharpe ratio.
  - Needless to say, it will have **little** or **no predictive power** going forward due to data-snooping bias.

When Not to Backtest a Strategy …

- **Example 4:** A high-frequency E-mini S&P500 futures trading strategy has a Backtest annual average return of 200 percent and a Sharpe ratio of 6. Its average holding period is 50 seconds.
  - Can we really Backtest a high-frequency trading strategy?
  - **The performance of a high-frequency trading strategy depends on the order types used and the execution method in general.**
  - Furthermore, it depends crucially on the **network connection** and **market microstructure**.
  - Since the act of placing or executing an order might alter the behavior of the other market participants, a trader should be very skeptical of a so-called Backtest of a high-frequency strategy.
  - **Life is too short to Backtest every single strategy that we read about, so we hope awareness of the common pitfalls of Backtesting will help you select what strategies to Backtest.**

Will a Backtest be Predictive of Future Returns?

- Even if we manage to avoid all the common pitfalls outlined earlier and there are enough trades to ensure statistical significance of the Backtest, the **predictive power of any Backtest rests on the central assumption that the statistical properties of the price series are unchanging**, so that the trading rules that were profitable in the past will be profitable in the future.

- **This assumption is, of course, invalidated often in varying degrees: A country's economic prospect changes, a company's management changes, and a financial market's structure changes.**

- In the past decades, we have witnessed numerous instances of the last category of changes such as decimalization, 2008 financial crisis, 2010 Uptick Rule, and 2020 Covid crisis.

Python programming example

Second simple buy-sell strategy