COMP226: Slides 22

Biases

Rahul Savani

rahul.savani@liverpool.ac.uk

Overview

- Recap on **Look-ahead Bias**
- Discussion of
    - **Survivorship Bias**
    - **Data-snooping Bias**
    - **Time period Bias**

Biases

There are a number of **common biases that can arise when using historic data** for trading strategy development. They usually bias results upwards

We discuss briefly several such biases. We already mentioned the first, **look-ahead bias**, where the backtest uses information that would not be available when a trading decision is made. E.g.:

- moving average calculation uses the close price on Monday when it make a trading decision at the open on Monday: available when it is not.

- strategy uses annual earnings data to place trades in January, even though the earnings data is not available until March

Survivorship bias

- Survivorship bias **occurs when backtests are conducted on datasets that excludes data for companies/funds etc. that no longer exist** (e.g. due to bankruptcy)

- Such backtests are likely to have upwardly biased results, since the surviving entities will tend to have performed better than those that no longer exist

- **Example 1**: mutual fund databases only contain funds that are currently in existence. Generally, funds that have ceased to exist have lower returns relative to surviving funds, and thus the average mutual fund return will be overestimated

- **Example 2**: data for stocks listed on an exchange will be subject to survivorship bias as delisted companies often have poor performance

Data-snooping bias

- Data-snooping bias can **occur when data is used more than once to develop a model (by trial and error)**
- It is particularly problematic for trading strategy development due to **widespread use of the same datasets** - is difficult to avoid in trading strategy design because **historical data is, by definition, limited**
- It is not surprising that models that appear to do well historically can be found by **exhaustive/extensive search**
- **Cross-validation** (discussed earlier) can help to avoid data-snooping bias.
- Having strategies that **make intuitive sense** can also protect against data-snooping bias and overfitting

Time period bias

- Time period bias **occurs when a test design is based on a time period that may make the results time-period specific**

- For example, a strategy may be very biased to go long, and then would have excellent performance during a bull market (period of rising prices), but would perform poorly during a bear market (period of falling prices)

- Thus it is desirable that the test period contains **different types of market conditions**. However, if the time period is too long, the fundamental market structure may have changed during the time frame, resulting in two data sets that reflect different relationships