Backtesting and its Pitfalls

Ernest Chan, Ph.D.
QTS Capital Management, LLC.

About Me

- As a quant, I have been developing and trading algorithmic strategies at Morgan Stanley, Credit Suisse, and various hedge funds since 1997.
- My book *Quantitative Trading* (Wiley 2009) has dealt extensively with backtesting.
- My forthcoming book Algorithmic Trading: Strategies and Pitfalls (2013) will discuss more backtesting nuances.
- At QTS Capital Management, we managed our fund with only backtested strategies.

What is backtesting?

- Feed historical data instead of live market data into a computerized trading program → get historical performance instead of suffering live performance.
- Backtesting is what distinguishes algorithmic from discretionary trading.
- Differs from “simulation”:
  - Historical data is real, not simulated.
  - Objective is *not* stress-testing the program under extreme market conditions.

The importance of backtesting

- If you developed a strategy from scratch, obviously you would want to know if it works, without risking real capital.
- If you obtained a strategy from someone else, you would want to find out if you can:
  a) replicate every detail of the strategy;
  b) ensure the stated performance is not obtained due to some backtesting pitfalls ... which I will discuss;
  c) improve its performance by small changes in the original strategy.

How to decide whether to backtest a strategy

- Backtesting involves a lot of work, and there are so many prospective strategies. Is there a shortcut?
  - Answer: Yes!
- In a series of examples, I will
  - give an overview of some of the pitfalls of backtesting; and
  - show how to judge whether a strategy is worth backtesting.

How to decide whether to backtest a strategy

- **Example 1**: You read about a strategy that has annualized returns of 30%, a Sharpe ratio of 0.7, and a maximum drawdown duration of 2 years, and a maximum drawdown of 15%.
  - *Would you backtest it?*

How to decide whether to backtest a strategy

- My answer: No!
  - Very few ~~investor~~ trader has the stomach for a strategy that kept losing for 2 years.
  - Low Sharpe ratio (<1) and long drawdown duration indicates performance is not consistent: high average returns may be due to a fluke (overfitting).
- Moral: don’t bother to backtest high returns but low Sharpe ratio strategies.

How to decide whether to backtest a strategy

- **Example 2**: An article describes the backtest of a "high frequency" E-mini strategy that has an annual average return of 20% and a Sharpe ratio of 2. Its average holding period is 48 minutes.
  - *Would you backtest it?*
  - *(Do you have enough information?)*

How to decide whether to backtest a strategy

- My answer: Maybe ... but
  - Has the author included transaction costs?
  - can the strategy be readily implemented using limit orders instead of market orders?
- High frequency strategies are very sensitive to transaction costs assumptions.
- Transaction costs are highly dependent on method of execution (limit vs. market orders).
- Backtesting tells only a small part of the story for HF strategies, though it is still a useful filter.
- Moral: be skeptical of backtest results of HF strategies in general.

How to decide whether to backtest a strategy

- **Example 3**: A simple “buy-low-sell-high” strategy advises you to pick 10 lowest-priced stocks in an index in the beginning of the year, and holds them for a year. The reported return in 2001 is 388%.
  - *Would you backtest this?*
  - *(Do you have enough information?)*

How to decide whether to backtest a strategy

- Answer: Probably not, unless the backtest was done on the index components *as they existed at the beginning of 2001*.
- Moral: “Survivorship bias” is a common cause of inflated backtest performance.

3 common biases

- Look-ahead bias
  - Using tomorrow prices to trade today.
- Survivorship bias
  - Backtesting with current stock universe, not with historical stock universe.
- Data-snooping bias
  - Overfitting to historical data with large number of model parameters or trading rules.
- All 3 biases tend to inflate backtest performances and over-estimate live returns.

Look-ahead bias

- Very common programming bug: it happens to the best of us.
  - E.g. 1) using a day's high or low prices as inputs to trigger a trading signal at the market open of that day.
  - E.g. 2) using a hedge ratio determined by regression over a time period to determine trading signals over the same period.
  - E.g. 3) optimizing parameters over some time period and measuring performance on same period.
- Easy to commit, but also easy to detect.

Look-ahead bias

- Common detection method: see whether your backtest program can generate positions at the last bar.
- Look-ahead bias will also be detected when you build an automated trading system for the strategy.

Look-ahead bias

- How to avoid look-ahead bias?
  - Depends on backtesting software.
- Ratings based on ability to avoid look-ahead bias:

Best	**Deltix, Progress Apama, Quanthouse, etc.**
Good	Excel
Poor	C++, Java, MATLAB, etc.

Survivorship bias

- Using a universe of stocks that have *survived* till today to backtest, ignoring the stocks that have disappeared due to bankruptcies, acquisitions, etc.

Survivorship bias

- E.g. “Buy low-price stocks” strategy
  - Pick 10 stocks with the lowest prices from a 1000-stock universe at beginning of year.
  - Sell them at end of year.
  - First, look at the picks in 2001 using a survivorship-bias-free database.

Survivorship-bias-free table

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
RTHM	0.8125	NaN	0.3000

Survivorship bias

- All but MDM were delisted sometime between Jan 2, 2001 and Jan 2, 2002.
- Terminal Price indicates last prices stocks were traded.
- Total portfolio return: -42%.
- Next, look at picks if our database *has* survivorship bias.

With survivorship bias

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

Survivorship bias

- (The prices were dividend and split adjusted – hence the 4 decimal places.)
- All these stocks survived till 1/2/2002, so all have closing prices on that day.
- Total portfolio return: 388%
- Highly unrealistic estimate of our would-be actual performance!

Survivorship bias

- What's the cure?
  - Buy expensive survivorship-bias-free database.
    - E.g. data from tickdata.com, crsp.com, kibot.com, csidata.com
  - Use only recent data for testing (no more than 3 years.)
  - Collect your own data day-by-day for future testing.

Data-snooping bias

- Model complexity > data complexity

  $\Rightarrow$ data-snooping bias.
- Model is picking up noise and non-recurring patterns in the past.
- Models with data-snooping bias has
  - great backtest performance
  - will suffer degraded performance in real trading.
- Complexity can result from:
  - Too many parameters.
  - Too many trading rules.

Data-snooping bias

- What's the cure?
  - Build models that are based on some well-researched financial/economic principles or phenomena: e.g. cointegration, PEAD (Post-Earnings-Announcement-Drift), etc.
  - Try not to use a pure data-driven/data-mining approach.
  - Rule-of-thumb: every additional parameter to be optimized requires an extra year of daily data.
  - Out-of-sample testing.

Data-snooping bias

- Optimizing parameters with data in moving lookback window.
- Parameterless trading models: no entry and exit parameters!

Conclusions

- There are many ways where backtesting can go wrong, and they usually inflate the expected performance.
- To avoid these pitfalls:
  - Pick the right backtesting platform.
  - Choose the right dataset.
  - Choose a reasonable model.

Keep in touch!

- Through email: ernest@epchan.com
- Through my blog: epchan.blogspot.com
- Through my consulting practice: www.epchan.com
- I run workshops in quantitative trading in London (UK), Hong Kong and Singapore, as well as through online learning.