Algorithmic Trading – Ch. 1

E. P. Chan
Ch. 1: Table of Contents

• Back-testing
  • Pitfalls
    • Data-snooping
    • Look-ahead Bias
    • Dividend/Stock Adjustments
    • Survivorship Bias
    • Short Constraints
    • Primary vs. Secondary Markets & Venue Dependency (FX)
    • Futures: Traded vs Settlement prices, Continuous contracts
  • Hypothesis Testing
    • Gaussian (Normal) Distribution
    • Monte Carlo Simulation (of Data)
      • Lo, Mamaysky, & Wang Simulation of Comparable Trades

• Automated Execution
Data-Snooping

• “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.”

• Linearity imposes simplicity and minimizes data snooping

• Discussion Questions:
  • Can we test the significance of additional trading parameters with an F-test on the significance of dummy variables assigned to different trading conditions?
  • Can artificially “test-forward” by testing our strategy on randomized historical start & end dates?
  • How can we use a back-tested linear model to inform live-trading models?
Monte Carlo Simulation

• Simulate data for which the strategy is applied to with the same first four moments as the historical data – (mean, deviation, skewness, kurtosis)

• Any serial relationship would be lost in the simulated data if the trading strategy captures this, then the simulated returns should be worse than the observed

• Discussion Questions:
  • Does this only work if the trading strategy captures serial correlation – i.e. would it be meaningful to use this significance test for a buy and hold strategy
Lo, Mamaysky, & Wang – Simulated Trades

- Simulate different trades (holding the # of long/shorts constant) for the observed historical data
- How to conceptualize what this is testing:
  - Creating hypothetical benchmarks?
  - Rejecting the null hypothesis that the trading strategy is no better than a random trading strategy?

Discussion Questions
- Given existence of data-snooping, is this test meaningful?
- P39, in example 1.1, “There is not a single sample out of 100,000 where the average strategy return is greater than or equal to the observed return. Clearly, the third test is much weaker for this strategy.”
- In which case(s) is the Monte Carlo / Simulated Trades a better test of significance?
Ch. 1: Table of Contents

• Back-testing
  • Pitfalls
    • Data-snooping
    • Look-ahead Bias
    • Dividend/Stock Adjustments
    • Survivorship Bias
    • Short Constraints
    • Primary vs. Secondary Markets & Venue Dependency (FX)
    • Futures: Traded vs Settlement prices, Continuous contracts

• Hypothesis Testing
  • Gaussian (Normal) Distribution
  • Monte Carlo Simulation (of Data)
    • Lo, Mamaysky, & Wang Simulation of Comparable Trades

• Automated Execution