US Stock Returns: Forecastibility vs. Random Walk Theory

Outline

- Introduction
- Literature review
- Hypothesis
  - Research questions
- Data
  - Define variables
  - Sample characteristics
- Methodology
  - Models
- Results
- Conclusion

Introduction

- What I'm doing, why, my expectations Motivation- a universally accepted answer has yet to be found
  - open-ended question, payoff, it’s an intellectual challenge

- EMH
  1. Strong form
  2. Semi-strong
  3. Weak form
  4. “random walk” represents random departures
- What anomalies exist-warren buffet, value line ranking system

Literature Review

- Selecting the model- what's been tried before?
    1. Monte carlo
    2. Martingale
    3. Non-linear time series models (MGTE)
    4. Kalmans Filter
    5. Variable order Markov Tree
    6. Market timing

- Predictors and Data criteria
    1. News sentiment
    2. Increasing the forecast horizon
    3. Long term growth expectation
    4. Seeking correlation between returns
        1. Local and global trend
    5. Crude oil (market data or and as a predictor)
    6. Sign direction of returns

- Results – inconclusive as a whole

Hypothesis

Hypothesis:ho≠ h1

h1

Research -The benefits would be huge- large profits

the academic insight into our (market) system and investor behavior

Data- description of variables 1

- Returns defined- levels or first differences
- Dependent variables Yi,t+1 for each security. There are 4,000 equations?
- I adjust the data for splits (preventing false outliers)
- How
- -throw a graph in

Data- sample selection

Full historical data set:[2007-2017]

In-sample composed of ½ these data

Out-of-sample –Forecast the other ½ of data

Monthly returns (Δfirst differences and in levels for robust check)

Data sources: ActiveTic, Capital IQ

Data- remove probability of bias 4

Mean reverting

Complex Event Processing

Back testing:

Data-snooping bias

Walk-forward testing

Methodology- Batch vs. Recursion 1

- Linear regression
  - Gather all the data then estimate (batch processing)

- Algorithms
  - Recursive-estimates in real time as each successive datum arrives
- Neural networks- machine learning

Methodology 2

- Kalman Filter-optimal estimator
- Minimizes the mean square error of the estimated parameters

if all noise is Gaussian, if not then given only the mean and standard deviation of noise, than this is the best *linear* estimator.

- “Filters out” the noise
- Projects onto the state estimate
- vector difference equation with additive white noise models unpredictable disturbances

Methodology- the models 4

Kalman Filter- requires only its current observed value and previously calculated state

State-space models

Uses two equations (recursively)

Measurement: $\boldsymbol{y}_{t+1} = \boldsymbol{X}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_{t+1}$ $N \sim (0, R)$

Transition: $\boldsymbol{\beta}_{t+1} = \boldsymbol{A} + \boldsymbol{B}\boldsymbol{\beta}_{\boldsymbol{t}} + \boldsymbol{\epsilon}_{t+1}$ $N \sim (0, \Omega)$

Methodology 3

- Variable Order Markov Tree
- Algorithms-properties of