Optimizing Trading Strategies Without Overfitting

Ernie Chan, Ph.D. and Ray Ng, Ph.D.
QTS Capital Management, LLC.

Typical Backtest Workflow

Prices $(x_1, x_2, \ldots, x_n)$

⇩

Trading signals $(B_1, \_, \_, S_4, \_ \ldots, S_n)$

⇩

Positions $(L_{1,} L_{2,} L_{3,} 0_{2,} \ldots, 0_n)$

⇩

PLs ($1, $1.2, -$0.6, ..., $0.3)

Optimizing Trading Signals

- Optimize trading strategy ≈ Optimize sum(PLs) by tweaking trading signals.
- Number(trading signals) << Number(prices) typically.
  - Easy to cherry-pick trading signals for optimization.
  - Overfitting/data Snooping Bias.
  - No predictive power on unseen/out-of-sample data!

Remedies for Overfitting

- Increase length of historical backtest period.
  - Subject to data availability
  - Regime changes $\Rightarrow$ old prices may be irrelevant.
- Create mathematical model of historical prices, then analytically find optimal trading signals
  - Effectively infinite backtest period.
  - Historical price models tend to be oversimplified.
  - Only analytically solvable for Trading Signals and performance objective linearly related to prices.

Remedies for Overfitting

- Simulate historical prices with similar statistics as actual historical prices.
  - As large number of price series as practical.
  - Can capture as many quirks of actual historical prices as necessary.
    - E.g. serial correlation, volatility clustering, tail events, …
  - Can be used to optimize nonlinear trading signals and performance objectives.

Analytical Optimization

- Example: a mean-reverting log price series x.
- Ornstein-Uhlenbeck equation

$$dx(t) = \kappa\big(\theta - x(t)\big)dt + \sigma dW(t)$$

κ: rate of mean reversion

θ: mean log price level

σ: conditional volatility of x

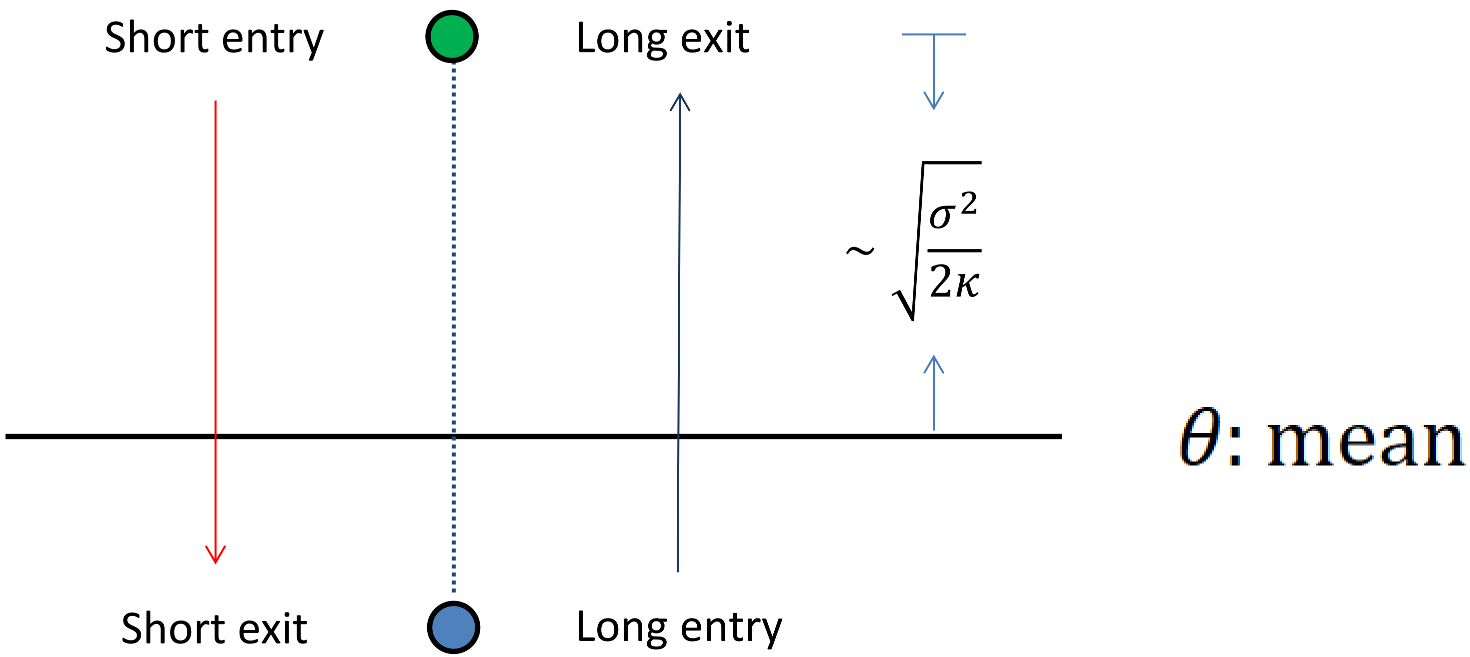
W: random walk

- What are optimal entry/exit levels?
  - Optimal ≡ maximum expected (discounted) profit *for single round-trip trade*.
  - Similar to optimal Bollinger bands.

Solving HJB

- *Cartea, 2015* demonstrated solution using Hamilton-Jacobi-Bellman equation (a PDE), familiar from stochastic control theory.
- Numerical solution to equation shows
  - Entry and exit levels are ***asymmetric*** w.r.t. mean, due to discount factor.
  - Entry level closer to mean level than exit level.
  - Distance of entry / exit levels to mean increases with decreasing κ.
  - Distance of entry / exit levels to mean increases with increasing σ.
  - (Last 2 points expected because unconditional volatility is $\sqrt{\frac{\sigma^2}{2\kappa}} \Leftrightarrow$ width of Bollinger bands.)
  - Long exit = short entry, vice versa.
  - Position is path-dependent.
  - Always in either long or short position.

Optimal Entry and Exit

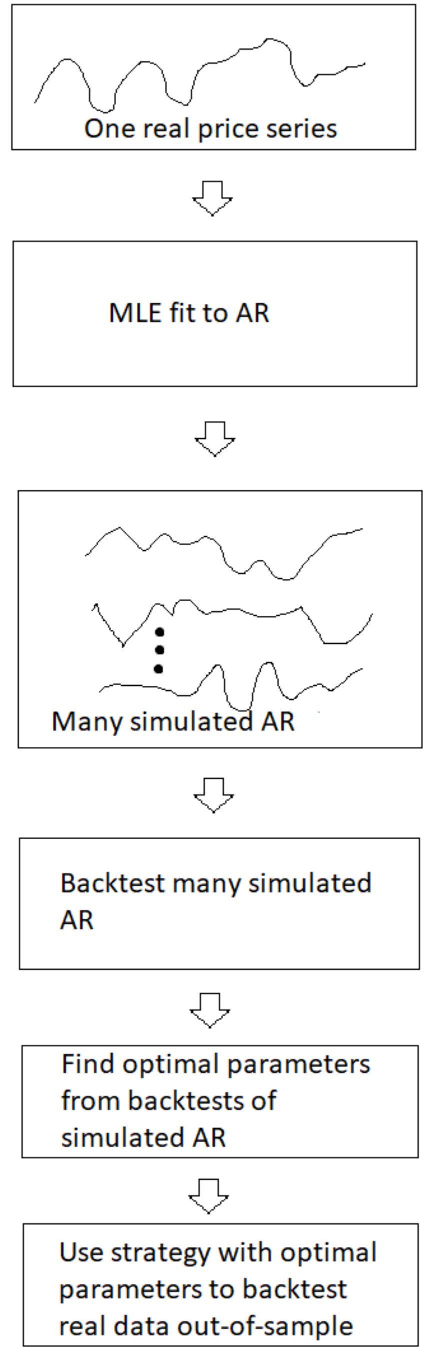


Analytical Optimization

- What if underlying price prices are not described by simple SDE like OU process?
  - Jumps, volatility clustering, long range correlations, etc.
- What if objective function is not discounted profit but a nonlinear function of PL?
  - **Sharpe ratio**, Calmar ratio, etc.
- What if objective function is **total PL**, not PL per trade?
- Even setting up HJB equation is too difficult.

Simulation for optimization

- We can simulate as many copies of price series as we like.
  - All follow the same time series model, e.g. AR(p).
- Find trading parameters that **maximizes the average** Sharpe ratio over all simulated price series.
  - Similar to solving HJB equation.
- Alternatively, find trading parameters that **most often maximizes** Sharpe ratio of a simulated price series.
  - Similar to maximum likelihood estimation.


One real price series
MLE fit to AR
Many simulated AR
Backtest many simulated AR
Find optimal parameters from backtests of simulated AR
Use strategy with optimal parameters to backtest real data out-of-sample

Example: AUDCAD

- ADF test indicates hourly AUDCAD prices are stationary with p-Value better than 1%.
- Assume AR(1) model on daily log prices x.

$$x(t) = a_1 x(t-1) + a_0 + \sigma_0 \epsilon(t)$$

$$\epsilon \sim \mathcal{N}(0,1)$$

  - For illustrative purpose only.
  - Train (a_0, a_1, σ_0) on first half of data using MLE.

Optimal trading of AUDCAD

- Simulate 10,000 log price series based on fitted AR(1).
  - Each series is about 3.7 years (~10 x halflife).
- On each series, backtest a simple strategy:

  Buy if expected log return > $k\sigma_0$

  Sell if expected log return < $-k\sigma_0$

  Flatten otherwise.
- Apply 1.8 bps per side transaction cost.

Simulation Results

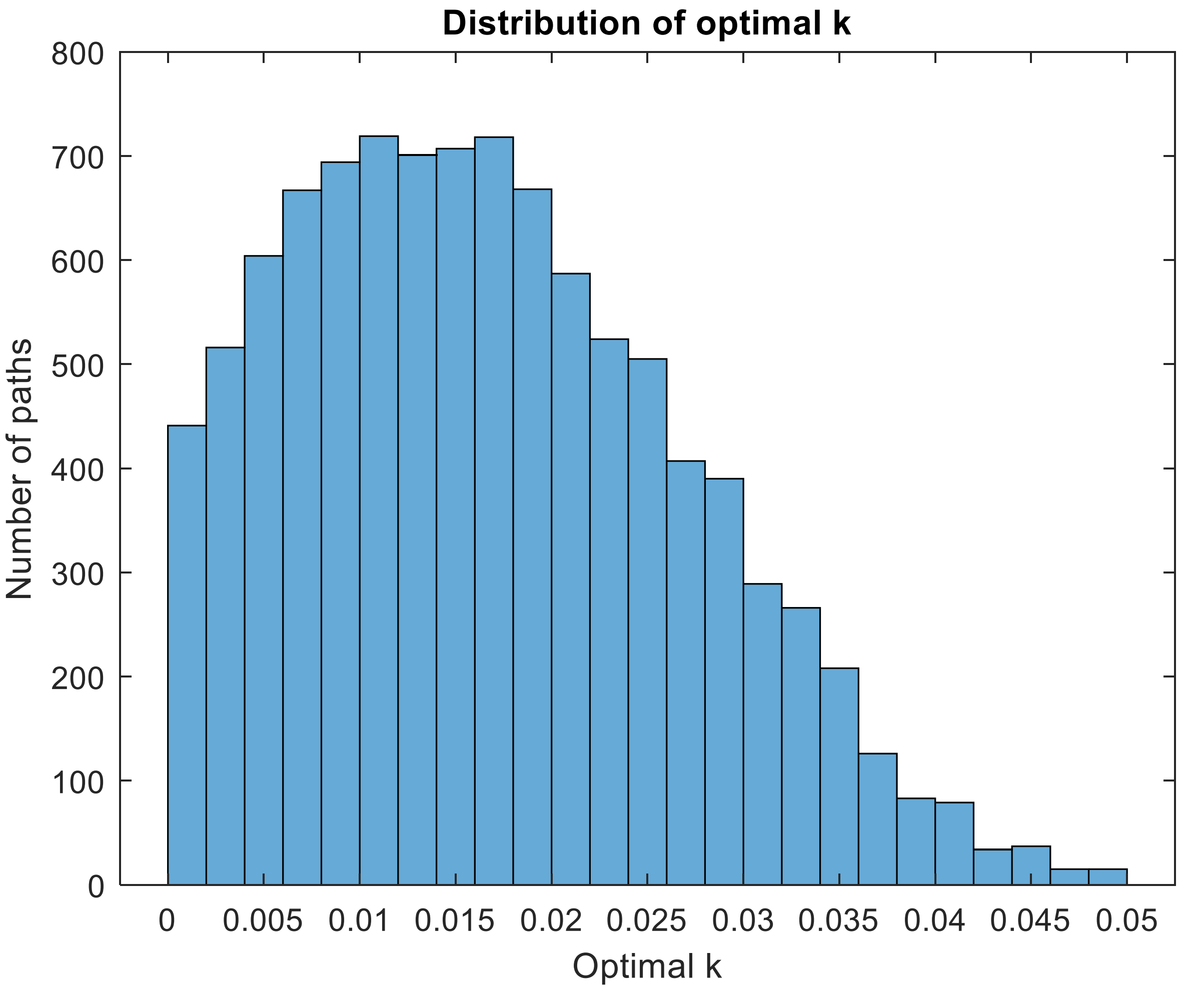
- Maximizing the average Sharpe ratio gives optimal k=0.0088±0.0002.

$$Argmax_k\{E_{path}[Sharpe(path)|k]\}$$

- In contrast, k=0.01±0.006 maximizes the likelihood that a path has highest Sharpe ratio

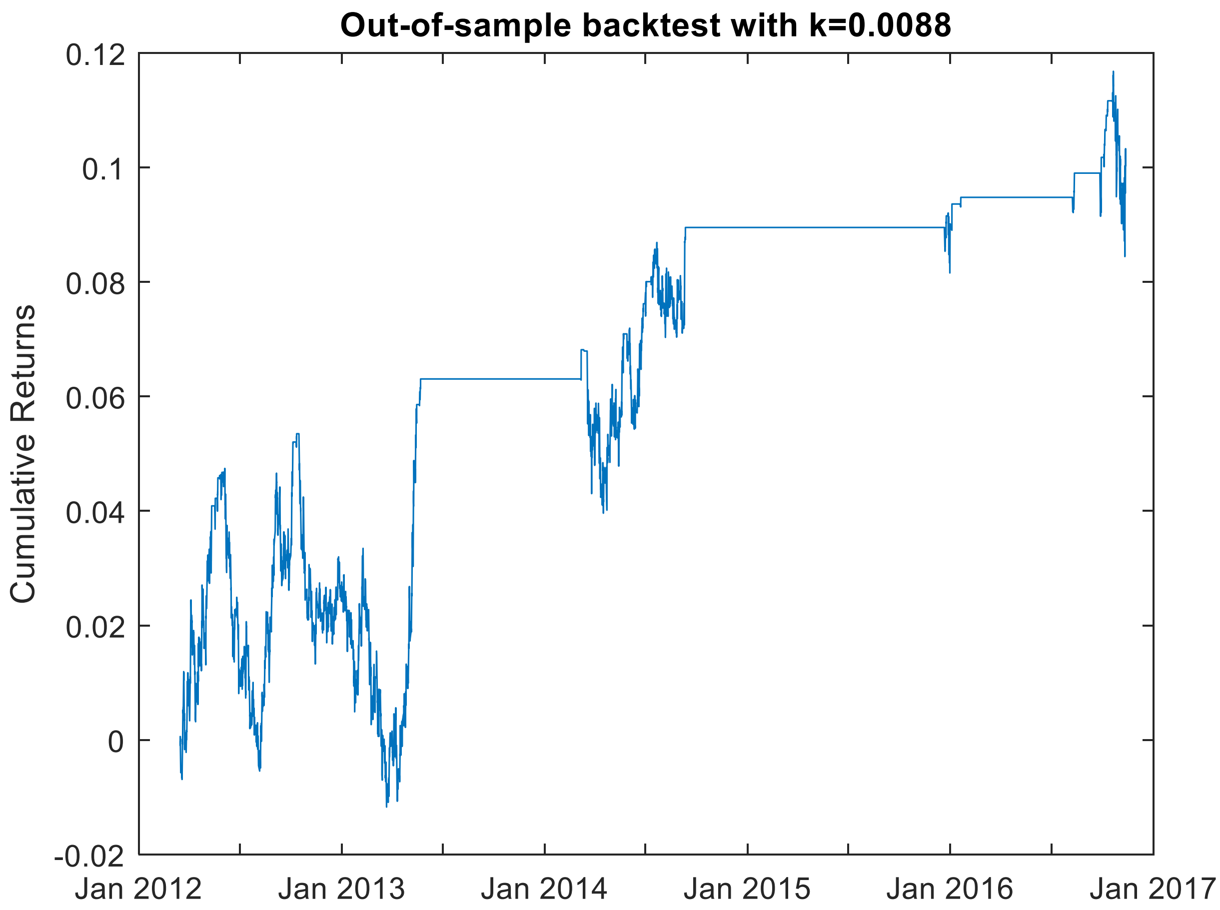
$$Argmax_k\{P_{path}[Argmax_{\hat{k}}[Sharpe(\hat{k}, \text{path})]]\}$$

- In general, the first method is more accurate since all paths are used to determine E_{path}.

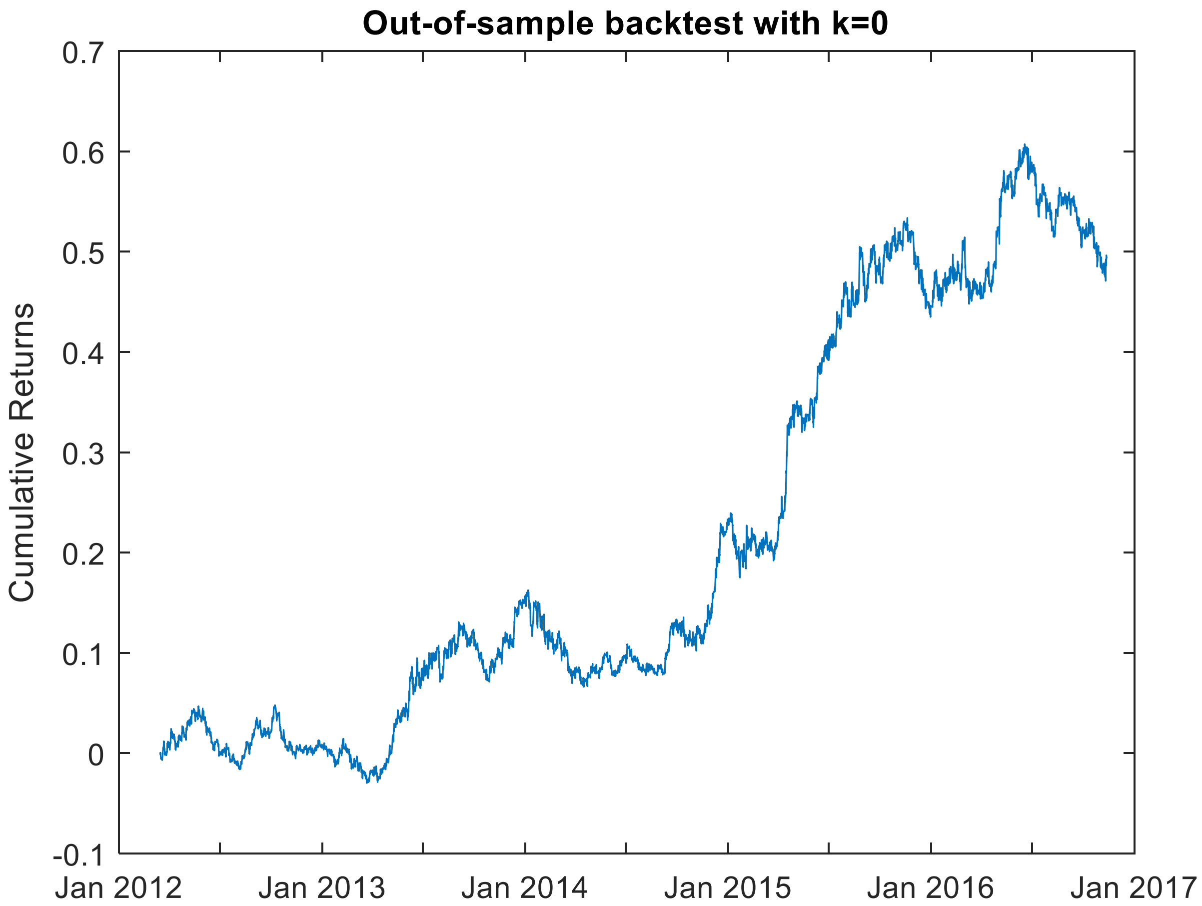


Distribution of optimal k
800
700
600
500
400
300
200
100
0
Number of paths
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05
Optimal k

OOS Backtest Optimal Parameter



OOS Backtest **Sub**optimal Parameter



Suboptimal > optimal?

- Backtest of "optimal" parameter underperforms that of "suboptimal" parameter out-of-sample.
- AR(1) model may need refitting periodically.
- **Nobody promises that for a particular realized path, our optimal k will maximize Sharpe!**
  - It is worth trading a range of k in the vicinity of the optimal for diversification.
- See similar work by Carr and Lopez de Prado, 2014.

Further work

- Can easily optimize other nonlinear functions of prices instead
  - Calmar ratio.
  - CVaR .
- Can easily extend this to more complicated time series models
  - AR+GARCH
  - Nonlinear generative models: e.g. LSTM (recurrent neural network)
- Can easily extend this to more complicated trading strategies, with multiple parameters.

Conclusion

- Optimizing trading strategy parameters on historical data invites overfitting.
- More robust to fit time series (not trading) models on historical data instead.
- Fitted time series model can be used to simulate arbitrary number of time series.
- Can find optimal trading parameters on simulated time series to arbitrary precision.

Thank you for your time!

www.qtscm.com

Twitter: @chanep

Blog: epchan.blogspot.com